DOCTOR OF PHILOSOPHY

A predictive model for assessing the reuse potential of structural elements at the end-of-life of a building

Rakhshanbabanari, Kambiz

Award date:
2021

Awarding institution:
Coventry University

A predictive model for assessing the reuse potential of structural elements at the end-of-life of a building

By

Kambiz Rakhshanbabanari

PhD

February 2021

A predictive model for assessing the reuse potential of structural elements at the end-of-life of a building

By
Kambiz Rakhshanbabanari

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

February 2021

Coventry University

Certificate of Ethical Approval

Applicant:

Kambiz Rakhshanbabanari

Project Title:

A predictive model for assessing the reuse potential of structural elements at the end-of-life of a building based on professional experience - Towards achieving a circular economy

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

13 April 2020

Project Reference Number:

P105548

Abstract

The reuse of building components can decrease the embodied energy and greenhouse gases of the construction activities and help get closer to a circular economy using fewer virgin materials. Part of the recent efforts to promote the reuse rates includes estimating the reusability of the load-bearing building components to assist the stakeholders in making sound judgements of the reuse potentials at the end-of-life of a building and alleviate the uncertainties and perceived risks. This study develops probabilistic models using advanced supervised machine learning techniques to predict the reuse potential of structural elements at the end-of-life of a building, from technical, economic, and social perspectives.

After performing a thorough literature search and identifying, analysing, and categorising the independent variables affecting the reusability of building structural elements, these factors were used to develop an online questionnaire. This questionnaire was then shared with a representative sample of practitioners in the construction industry, including managers, CEOs, architects, engineers, consultants, and deconstruction experts with previous experience in reusing recovered building structural components. The received questionnaires were reviewed, and the initial dataset was split into three separate datasets to address the technical, economic, and social aspects of the study. Then, the missing values were estimated, and the class imbalances were addressed using advanced techniques. In the next stage, and for each dataset, a total number of thirteen predictive models were developed in the R software using 13 advanced supervised machine learning methods. The performance and transparency of these models were compared to choose the best-practice Building Structural Elements Reusability Predictive Models (BSE-RPMs), which provide reliable predictions.

Random Forest (RF) models were selected as the best practice BSE-RPMs for all three datasets, with a considerable overall accuracy of 96%, 89%, and 94% for the technical, economic, and social models, respectively. Since RF models are known as black-box models, advanced supervised machine learning methods such as sensitivity analysis and visualisation techniques were employed to open the selected RF BSE-RPMs. Eventually, using advanced rule extraction methods, three easy-to-understand predictive models (learners) were developed for assessing the technical, economic, and social reusability of the load-bearing building components, with an overall accuracy of 85%, 82%, and 91%, respectively.

This research has contributed to promoting the reuse of building structural elements in two ways. First, using advanced supervised machine learning techniques such as the Boruta method and recursive feature elimination technique, this research identifies and ranks the main reusability factors based on the experience of the stakeholders with the recovered building structural elements in the building sector. Second, for the first time, it develops three sets of easy-to-understand learners (predictive rules) that can be used by practitioners to have an initial assessment of the technical, economic, and social reusability of the load-bearing components. The developed learners can be easily used by various stakeholders and have the potential to promote the reuse rate of the structural elements of the existing buildings, which were not designed for deconstruction. These sets of rules can also encourage more deconstruction projects since the developers would have a better judgment about the reusability of the structure of an existing building at its end-of-life, which, in turn, can accelerate the growth of reuse markets.

Dedication

To my wife, Tannaz

For all her love,

support, patience, and

encouragement

Acknowledgements

Firstly, I would like to thank my current director of studies, Dr Alireza Daneshkhah, for his consistent support and guidance during this project. His attention to different aspects of my research, his unwavering support for this project, and my career aspirations, along with his encouragement, helped me grow as a researcher. I would also like to thank my former director of studies, Professor Jean-Claude Morel, for the opportunity to start and complete this research. Professor Morel's dedication, support, and advice helped me throughout my PhD journey and provided me with the necessary infrastructure I required to perform this research successfully. Professor Morel's support as my academic advisor and mentor helped me to cope with unforeseen challenges in this project and my personal life. Likewise, I want to thank Dr Hafiz Alaka, my supervisor, for his help and support during the first stage of my PhD project. Eventually, I would like to thank Dr Messaoud Saidani, my supervisor, for his prompt support and enlightening advice throughout this project.

I would also like to thank reuse experts in the building sector that shared their expertise with this project through participating in the online questionnaire survey. Without their sincere answers to the cumbersome survey, this project could not have succeeded.

I also thank the Research Centre for the Built and Natural Environment and Research Centre for Computational Science and Mathematical Modelling, Coventry University, for their support and guidance throughout my PhD project. I am specifically thankful to Professor Eshmaiel Ganjian, Dr Azadeh Montazami, and Dr Ashish Shukla for their kind advice and support.

I would like to thank my dear wife, Tannaz Latifi, for her help and assistance in improving the visual quality of the figures of this work and my publications. Her love, support, and encouragement helped me to overcome all barriers throughout this project.

Furthermore, I would like to thank my mentor and friend, Dr Wilhelm Alexander Friess (University of Maine, USA), whose continuous encouragement and support inspired me to start and complete my PhD journey. While Dr Friess was not a part of the supervisory team, he offered his time to comment on my research difficulties and even helped proof-read some of my writings.

Eventually, I would like to thank Coventry University for funding this research project originally entitled "Recycling and Reuse of Construction and Building Materials in the Context of the Circular Economy".

List of publications

<u>Published articles:</u>

Rakhshan, K., Morel, J.-C., and Daneshkhah, A. (2021) 'Predicting the Technical Reusability of Load-Bearing Building Components: A Probabilistic Approach towards Developing a Circular Economy Framework'. Journal of Building Engineering 102791. available from https://doi.org/10.1016/j.jobe.2021.102791

- Based on Chapter 1 (Introduction), Chapter 4 (Quantitative study), Chapter 5 (Predictive models development), and Chapter 6 (Model selection (results) and discussion)

Rakhshan, K., Morel, J.-C., and Daneshkhah, A. (2021) 'A Probabilistic Predictive Model for Assessing the Economic Reusability of Load-Bearing Building Components: Developing a Circular Economy Framework'. *Sustainable Production and Consumption* 27, 630–642 available from https://doi.org/10.1016/j.spc.2021.01.031

- Based on Chapter 1 (Introduction), Chapter 4 (Quantitative study), Chapter 5 (Predictive models development), and Chapter 6 (Model selection (results) and discussion).

Rakhshan, K., Morel, J.-C., Alaka, H., and Charef, R. (2020) 'Components Reuse in the Building Sector – A Systematic Review'. *Waste Management & Research* [online] 38 (4), 347–370. available from https://doi.org/10.1177/0734242X20910463

- Forms Chapter 2 (Systematic review of factors affecting the reuse of load-bearing building components).

Table of Contents

List of Tables

List of Figures

List of Equations/Probability statements/Functional forms

Abbreviations

AAD	Average Absolute Deviation
AB	Adaptive Boosting
ACT	Action Research
AIC	Automation in Construction
ANN	Artificial Neural Networks
AR	Archival research
AUC	The area under the ROC curve
BCR	Load-bearing building components reuse
BE	Building and Environment
BIM	Building Information Modelling
BM	BART (Bayesian additive regression trees) Machine
BREEAM	Building Research Establishment Environmental Assessment Method
BRI	Building Research and Information
BS	British standard
BSE-RPM(s)	Building Structural Elements Reusability Predictive Model(s)
CA	Comparative analysis
CAW	Centre for Academic Writing
CD	Company documentation
CDW	Construction and demolition waste
CE	Conformité Européene
CEO	Chief executive officer
cF	cForest importance
CHIS	Chi-squared
C-Index	Co-occurrence index
CJCE	Canadian Journal of Civil Engineering
CME	Construction Management and Economics
CO_2	Carbon Dioxide
CS	Case study
DfD	design for deconstruction
DfMA	design for manufacture and assembly
DO	Direct observation
DR	Document review
DSA	Data-based sensitivity analysis

DT	Decision Trees
DV	Dependent variable
EC	European Commission
ECAM	Engineering, Construction and Architectural Management
ECO	Economic
ECO BSE-RPM(s)	Building Structural Elements Reusability Predictive Model(s) for the Economic dimension
ECOM	Economic models
ECO-RF BSE-RPM	Random Forest Building Structural Elements Reusability Predictive Model for the Economic dimension
EM	Expectation-Maximisation
ENV	Environmental
EX	Experiment
FGI	Focused-group interview
FN	False negative
FP	False positive
FS	Field study
GBP	British Pound sterling
GDP	Gross Domestic Product
GHG	Greenhouse gases
GI	Group Interview
GIS	Geographic Information System
GM	Group meetings
GP	Gaussian Processes
GPC	Gaussian processes for classification
GR	Gain ratio
I	Unspecified type Interviews
IG	Information gain
IV	Independent variables
JCCE	Journal of Computing in Civil Engineering
JCEM	Journal of Construction Engineering and Management
JME	Journal of Management in Engineering
kfCV	k-fold Cross-Validation
KNN	K-Nearest Neighbours

KT	Kruskal test
LCA	Life-cycle assessment
LDA	Linear Discriminant Analysis
LEED	Leadership in Energy and Environmental Design
LIR	Literature review
LOOCV	leave-one-out cross-validation
LR	Logistic Regression
MAR	Missing at random
MCAR	Missing completely at random
MI	Multiple imputation
MLE	Maximum likelihood estimation
MRMR	Minimum-redundancy-maximum-relevance
Mt	Megaton
MZSF	Maximum Z-statistic among the shadow features
NB	Naïve Bayes
NDT	Non-destructive testing
NMAR	Not missing at random
OBS	Observation
OECD	Organisation for Economic Co-operation and Development
oneR	One rule
ORG	Organisational
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRL	Propositional Rule Learner
Q	Questionnaire
QDA	Quadratic Discriminant Analysis
RBSE	Reuse of building structural elements
RCR	Resources, Conservation and Recycling
REG	Regulatory
RF	Random Forest(s)
RFE	Recursive feature elimination
ROC	Receiver operating characteristics curve
S	Survey (i.e. empirical survey, etc.)
SA	Sensitivity analysis
SD	System dynamics

SI	Structured interviews
SLR	Systematic literature review
SMOTE	Synthetic Minority Oversampling Technique
SMOTE-NC	Synthetic Minority Oversampling Technique for Nominal and Continuous
SOC	Social
SOC BSE-RPM(s)	Building Structural Elements Reusability Predictive Model(s) for the Social dimension
SOC-RF BSE-RPM	Random Forest Building Structural Elements Reusability Predictive Model for the Social dimension
SSI	Semi-structured interviews
SU	Symmetrical uncertainty
SVM	Support Vector Machine
T	Theoretical study
TEC	Technical
TEC BSE-RPM(s)	Building Structural Elements Reusability Predictive Model(s) for the Technical dimension
TEC-RF BSE-RPM	Random Forest Building Structural Elements Reusability Predictive Model for the Technical dimension
TF	Theoretical framework
TN	True negative
TP	True positive
UI	Unstructured interview
USD	United States Dollar
VEC	Variable effect characteristic curve

Chapter 1 – Introduction

1.1 Background

The construction industry is a leading economic sector that employs around 2.8% of the global workforce (around 97.5 million employees) (Statista 2019, The World Bank 2019) and accounts for 1.7% to 10.5% of the Gross Domestic Product (GDP) in European countries (UNECE 2020). In 2019, 2.7 million careers in the UK (equal to 7.8% of the total labour force based on The World Bank Data) were related to the construction industry (Statista 2021a, The World Bank 2019). Also, the construction activities in the UK (comprising of commercial and social, residential, and infrastructure subsectors (*Government Construction Strategy* 2011)) accounted for £118 billion, equal to 6.02% of the total GDP in 2020 (Statista 2021a, 2021b). The global value of construction activities in 2020 was around £9.1 trillion and is expected to reach £11.6 trillion by 2030[1] (Designing Buildings Wiki 2021). According to (Barbosa et al. 2017), the expected growth in the construction industry spending is approximately 3.6% per year, expecting to reach 14% of the global GDP by 2025. The International Finance Corporation (a member of the World Bank Group) forecasts that this growth takes place mostly in residential, non-residential, and infrastructure projects (International Finance Corporation 2018).

The construction industry is also a leader in the consumption of resources and the emission of greenhouse gases (GHG) (UNEP 2020, OECD 2019, International Finance Corporation 2018, World Economic Forum 2016). According to (OECD 2019), this sector is the largest consumer of raw materials, and construction-related activities account for 25% to 40% of the total CO_2 emissions globally (World Economic Forum 2016, UNEP 2020). It is not surprising because, according to (Guo et al. 2019), there is a strong positive linear relationship between the GDP and the embodied energy use and direct energy use in the construction sector. According to (Statista 2021c), 52% of the global steel production in 2019 (978 Mt) is used in the construction sector (Statista 2021d). The extraction of raw materials such as limestone and iron ore and the production of construction materials like cement and steel are energy and carbon-intensive processes (Vitale and Arena 2017, Hammond and Jones 2008). For instance, the production of the steel sections used in the construction sector (buildings and infrastructure) accounts for around 3.5% of the annual CO_2 emissions worldwide (WSA 2012). It is noteworthy that globally, the iron and steel sector accounted for 2.6 gigatons of direct CO_2 emissions in 2019, or 7% of

[1] Considering a conversion rate of 1.277 based on the average monthly exchange rates of GBP to USD in 2020 (UK Government 2020).

total energy sector emissions (International Energy Agency 2020). Likewise, cement, another ingredient of many constructions worldwide (Vitale and Arena 2017), is accountable for around 7% of the world energy sector CO_2 emissions (International Energy Agency 2020).

Besides, construction activities produce the highest amount of waste among all other sectors (Defra 2019, Eurostat 2016, Clark, Jambeck, and Townsend 2006, UNEP 2015). In the UK, among the 223 million tons of waste generated in 2016, around 61% belonged to construction, demolition, and excavation activities (Defra 2019). The construction and demolition waste (CDW) in some parts of the world constitutes up to 40% of the total waste stream (Hoornweg and Bhada-Tata 2012). As an instance, CDW is accountable for around 36% of the total waste generated in the EU-27 (Eurostat 2020). In the OECD countries, CDW accounts for around 36% of the total waste generated annually (UNEP 2015). It is noteworthy that the OECD countries are responsible for the production of approximately 44% of the global waste (Hoornweg and Bhada-Tata 2012).

In the light of the Paris Agreement and to maintain the global temperature increase well below two degrees Centigrade, the need to decreasing the amount of CO_2 and other greenhouse gases (GHG) has become inevitable in all sectors (UN 2015). According to (International Finance Corporation 2016), 101 of the signatories of the Paris Agreement highlighted that waste is a crucial sector for fulfilling the targets set by the agreement. Moreover, 66 of the countries in the Paris Agreement confirmed that buildings are another pivotal sector for achieving the targets of sustainable development (International Finance Corporation 2016). Therefore, acknowledging the share of the construction industry in the global GDP, raw materials and energy consumption, and GHG production, it is evident that the building sector has a considerable potential to fulfil the Paris Agreement targets by improving its overall sustainability footprint. Since most of the embodied energy and CO_2 impacts of buildings are related to the load-bearing systems (Kaethner and Burridge 2012), methods for extending the life of the structure of buildings seems promising.

In recent years, new design and construction methods such as design for deconstruction (DfD) (Akinade et al. 2017, Tingley and Davison 2011), design for manufacture and assembly (DfMA) (Kalyun and Wodajo 2012), and Modular Construction (Thai, Ngo, and Uy 2020) are introduced to decrease waste and promote the reuse of the load-bearing components at the end-of-life of a building. However, most of the existing buildings are not designed based on the above techniques, which results in the generation of a considerable amount of wastes during the refurbishment or demolition phases (Chileshe et al. 2016, Rose and Stegemann 2018, Chileshe, Rameezdeen, and Hosseini 2015). Moreover, a considerable focus of the research body is on the adaptive reuse of the existing buildings (Nevzat and Atakara 2015, Sfakianaki and Moutsatsou 2015, Tan, Shen, and Langston 2014, Sanchez and Haas 2018, Bullen 2007). While adaptive reuse is the most promising option to prevent waste and promote the sustainability of the structure of a building, in many instances, it is not practical, and the removal of a building at its end-of-life becomes inevitable. In this case, if the structure of the building is not recovered and reused, it results in the loss of valuable resources (Fujita and Iwata 2008).

Reusing the recovered load-bearing building components in new constructions for aesthetic or environmental purposes has attracted different clients worldwide in the last two decades, which have resulted in various successful case-study projects. For instance, in 1997, the Udden project reused several components such as 73 concrete wall elements and 41 concrete floor beams recovered from buildings built in the 1960s. Moreover, in 2001, the Nya Udden project recovered several load-bearing building components such as 72 concrete outer-wall elements and 224 concrete beams from various 1970s buildings and reused them in new student accommodation (Addis 2006). In 2002, and in an attempt to develop an ultra-green residential and office complex, various reclaimed building components and materials were used in the construction of the Beddington Zero Energy Development, London, UK (Lazarus 2003). These include reclaimed, reused, and recycled building components such as steel (95% of the steel structure), timber for internal and external studwork, floorboards, bollards, paving slabs, and shuttering ply. While these projects show that the reuse of load-bearing building components is practical, this practice is still not mainstream due to the amplitude of prohibiting factors (Section 2.3.2) (for other examples of such case study buildings, please refer to (Addis 2006, Gorgolewski et al. 2008, Gorgolewski 2008)).

The reuse of load-bearing building components at the end-of-life of existing buildings, and the factors affecting its uptake in new constructions has been the focus of research for several years. Researchers have identified various economic, environmental, organisational, regulatory, social, and technical barriers to reuse in the building sector. From an economic perspective, barriers such as lack of an established reuse market, additional costs, and revenue were among the main factors prohibiting the reuse uptake (da Rocha and Sattler 2009, Rameezdeen et al. 2016, Dantata, Touran, and Wang 2005, Chileshe et al. 2016). From an organisational aspect, factors related to the lack of infrastructure to perform deconstruction and reuse, lack of experienced contractors, and managerial problems such as lack of ownership or systems thinking are prohibiting reuse (Arif et al. 2012, Rose and Stegemann 2018, Dunant et al. 2018, Yeung, Walbridge, and Haas 2015). From a regulatory perspective, factors such as the requirement to comply with the latest norms and standards and lack of regulatory incentives were identified as the main barriers to reuse (Rose and Stegemann 2018, Huang et al. 2018, Shaurette 2006, Chini and Acquaye 2001). On a social dimension, lack of awareness of the stakeholders, negative perception of the practitioners and clients, and various perceived risks associated with reuse were among the main barriers to reuse (Chileshe, Rameezdeen, and Hosseini 2015, Ajayi et al. 2015, Tingley et al. 2017). And from a technical perspective, lack of design for deconstruction in the existing buildings, several design challenges, and unavailability of information and details about the recovered building components were among the main factors identified as barriers to reuse (Rose and Stegemann 2018, Chini and Acquaye 2001, Sansom and Avery 2014, Gorgolewski et al. 2008) (see Chapter 2 for a complete list of factors affecting reuse).

According to (Akinade et al. 2016), the existing research on construction waste management is focused on the management strategies (waste hierarchies), waste generation (quantification, sources, etc.), performance measurement, stakeholders' attitude, regulatory environment (policies, charges, etc.), and management tools. While all the above research themes attempt to promote the success of construction waste management practices, according to Akinade et al. (2016), the waste management tools are pivotal in this endeavour. Akinade et al. (2016) identify thirty-two tools and categorise them into six groups. These include waste management plan templates and guides, waste data collection tools, waste quantification models, waste prediction tools, and Geographic Information System (GIS)-enabled waste tools. A review of these tools reveals that most of them are intended for the new buildings. An exception is a waste prediction tool developed by (Cheng and Ma 2013), where the authors explain a BIM-based system for the estimation and planning of the demolition and renovation wastes. Notwithstanding, none of the existing waste management tools provides instructions for the stakeholders on how to evaluate the reusability of building structural components at the end-of-life of the existing buildings.

1.2 Concept of the circular economy

The circular economy is a new paradigm that is focused on the management of resources to decrease (and eventually eliminate) the impact of the anthropogenic activities on the environment by decoupling the continuous economic growth from the exploitation of natural resources (Pomponi and Moncaster 2017).

The concept of the circular economy has been continuously evolving for the last sixty years. While the knowledge of the negative consequences of human inventions goes back to the ancient Greek myths (Frosch and Gallopoulos 1989), the initial effort to change the linear industrial models (take-make-waste) took place in the 60s. In 1966, Boulding (Boulding 2013) made the first attempt to increase the awareness of the need for a paradigm shift from a linear economy to an early version of the circular economy, which he called a closed sphere. In 1976, Walter Stahel and Genevieve Reday introduced the economy in loops (or the circular economy) in their report "The Potential for Substituting Manpower for Energy" to the European Commission in Brussels (The Product-Life Institute n.d.). In their work, the authors reviewed the impact of their proposed model on "job creation, economic competitiveness, resource savings, and waste prevention"(The Product-Life Institute n.d.). Next, in 1989, Frosch and Gallopoulos (Frosch and Gallopoulos 1989) called for a shift from the linear industrial model to an industrial ecosystem in which "the consumption of energy and materials is optimized, waste generation is minimized, and the effluents of one process, …, serve as the raw material for another process." They further clarified that their proposed industrial ecosystem "would function as an analogue of biological ecosystems." (Frosch and Gallopoulos 1989). Based on the concept of biological ecosystems, biomimicry (Benyus 1997) and biomimetics (Bhushan 2009) were introduced in 1997 and 2009, respectively. These new industrial concepts can be interpreted as the innovation and design of products inspired by nature (Benyus 1997, Bhushan 2009).

As a result, the Ellen MacArthur Foundation defines the circular economy as "an industrial system that is restorative or regenerative by intention and design" (Ellen MacArthur Foundation 2013, 2015). They further elaborate that the circular economy "replaces the 'end-of-life' concept with restoration, shifts towards the use of renewable energy, eliminates the use of toxic chemicals, which impair reuse, and aims for the elimination of waste through the superior design of materials, products, systems, and, within this, business models" (Ellen MacArthur Foundation 2013). Therefore, its principles lie on "designing out waste and pollution, keeping products and materials in use, and regenerating natural systems" (Ellen MacArthur Foundation n.d.).

1.3 Justification of the study

The construction industry consumes between 30% to 50% of the natural resources (Anink et al. 1996, Herczeg et al. 2014, WSA 2012), produces up to 40% of the total waste stream (excluding the excavation waste) (Eurostat 2016, Clark, Jambeck, and Townsend 2006, Defra 2019, UNEP 2015), and generates around 39% of the world's greenhouse gas emissions (Abergel, Dean, and Dulac 2017). The above facts are alarming due to the urgent need to decrease the GHGs

(UNFCCC 2015) and because we are facing landfilling restrictions (Brewer and Mooney 2008) and resource deficiency globally (Ellen MacArthur Foundation 2013, Chen et al. 2010).

The depletion of the earth's resources as a result of fast economic expansion, continuous population growth, and the drastic increase in demand for products and services has led the governments to run resource-efficient economies (Ellen MacArthur Foundation 2013). Therefore, the regulatory authorities worldwide, such as the European Commission Waste Framework Directive 2008/98/EC (European Union 2008) and the Demolition Protocol (ICE 2008), introduce waste hierarchies to improve the material efficiency across all the economic sectors, including the building industry (Figure 1.1). According to these regulations, preparing for reuse (or reuse) is the second-best solution after prevention to decrease the high level of waste generation, and to decouple the economic growth from resource consumption (European Union 2008, ICE 2008).

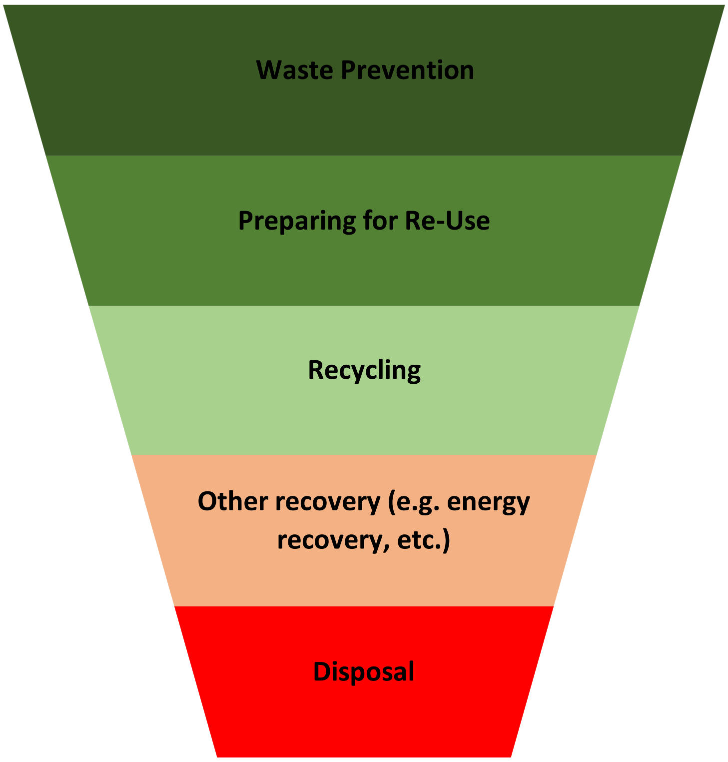

Figure 1.1 The waste hierarchy (European Union 2008, ICE 2008).

According to the waste hierarchies, reuse is preferred to recycling. However, most of the recovery of CDW happens in the form of recycling and not reuse. For example, in the UK, nearly 91% of the non-hazardous CDW is recovered through recycling (Defra 2019). While recycling can divert waste from landfills, the processes involved are energy and resource-intensive and impose a noticeable pressure on the environment in terms of GHGs and other sorts of emissions (Addis 2006, WRAP 2008). Contrarily, reused load-bearing building components (beams, columns, truss, etc.) have far lower environmental impacts when compared with recycled materials (Geyer, Jackson, and Clift 2002). For instance, when new steel sections that have around 60% recycled content are used, their environmental impacts are still twenty-five times more than reusing the equivalent reclaimed steel sections (WRAP 2008). According to (Lazarus 2003), reusing reclaimed structural steel and timber sections can decrease the environmental impacts by 96% and 83%, respectively. It is primarily due to significantly lower treatment and reprocessing required for reusing the load-bearing building components in comparison with recycling (Gorgolewski et al. 2008).

Although efforts have been made to increase the reuse rates of building structural elements in recent years, there are yet no signs of improvements. Contrarily, the reuse rates in the building sector have declined in the last two decades in countries like the UK, and only a fraction of load-bearing components at the end-of-life of a building are reused (Addis 2006, Sansom and Avery 2014). For instance, only 5% of the reclaimed steel sections in the UK are reused, and the remaining are recycled (Sansom and Avery 2014). Part of the recent efforts to promote the reuse rates includes predicting the reusability of the load-bearing building components to assist the stakeholders in making sound judgements of the reuse potentials at the end-of-life of a building and alleviate the uncertainties and perceived risks (Yeung, Walbridge, and Haas 2015, Keller et al. 2019, Fujita and Kuki 2016, Cavalli et al. 2016, Smith et al. 2013, Fujita and Masuda 2014). However, the continuous decline in reusing the structural elements of buildings shows that there is a need for the development of robust interdisciplinary reusability prediction tools to improve the reuse rates.

1.4 Research problem and gap in knowledge

While the reviewed articles (Section 1.1 and Chapter 2) show that a wide range of studies has extensively tried to identify the barriers ahead of the widespread reuse of building structural elements, they did not provide any indication of the reusability of these components based on the identified barriers. In the lack of an evaluation material to synthesise the identified barriers, find the correlations between them, and estimate the reusability of the load-bearing building components, the reuse of these elements will not grow in the building industry. It is because the fragmented body of knowledge available in the literature is unable to direct the stakeholders to take progressive steps towards the circularity of materials in this sector. Some authors recognised this gap but attempted to fill it by estimating the physical properties (dimensional or mechanical) of the recovered building structural elements as an indication of their reusability and ignored the impact of other variables. In this light, determining the reusability of the load-bearing building elements has introduced a new paradigm in the field of reuse and has been the focus of research recently (Yeung, Walbridge, and Haas 2015, Keller et al. 2019, Fujita and Kuki 2016, Cavalli et al. 2016, Smith et al. 2013, Fujita and Masuda 2014).

For instance, focusing on the dimensional aspect, (Yeung, Walbridge, and Haas 2015) studied the impact of accurate geometric characterisation of the steel structure of a building at its end-of-life on the decision process for reusing the structural components. The authors initially developed a decision-making framework to facilitate the stakeholders in identifying the reuse potentials for recovered building structural steel. They then presented an automated object recognition algorithm to identify the member cross-sections. They eventually performed a reliability analysis to evaluate the performance of the proposed geometric identification method. Based on the results of the reliability analysis, the authors proposed a semi-automatic geometric identification method to enable designers to integrate the reused structural elements in new buildings at their full capacity.

In another study focused on determining the physical properties of the structural steel, the authors developed a performance evaluation procedure to estimate the mechanical properties of reused structural elements using non-destructive testing (NDT) (Fujita and Kuki 2016). They estimated the Vickers hardness using portable ultrasonic hardness testers and rebound type portable hardness meters. They then used the estimated values as the basis to calculate the mechanical properties of the reusable elements. The results of the test specimens showed good agreement with the standard values.

Similarly, (Keller et al. 2019) used wireless sensors to monitor the stresses induced during the construction of a steel-framed building to evaluate the reusability of steel members. According to this study, the authors observed that the maximum measured stresses were almost half of the nominal yield strength, confirming that the current design practices allow the reuse of structural steel (see also (Farsi et al. 2020) for similar studies in different systems and industries).

In a relevant study focused on estimating the mechanical properties of timber, (Cavalli et al. 2016) developed linear regression models to predict the Modulus of Elasticity and Modulus of Rupture of in-use and recovered timber sections based on the NDT methods. According to this study, the developed models can be used to assess the reusability of timber structures on site. Notwithstanding, the proposed linear regression-based models are too simple to model the complex system described above, and the predicted values are not accurate. Therefore, the derived results using the linear regression models are not reliable, and considerable care should be taken to use the outcomes of this study. However, this study shows the substantial potential of the machine learning techniques in determining the reusability of the load-bearing building components.

The above studies concentrate on discovering the technical reusability of the building structure by focusing on one aspect, like determining the mechanical properties or dimensional details of potential structural components for reuse and ignored the impact of other variables. The only exception is a study performed by (Hradil et al. 2017), in which the authors developed an indicator for estimating the technical reusability of steel-framed buildings considering a combination of variables. These variables include the impact of disassembly technique, handling, availability of the earlier design documents, potential new deployment (same purpose or repurposing), and the need for quality and dimensional checks. In another study, the authors also considered the marketability of the structure and extended the index by integrating the economic prospect of the recovered components (Hradil, Fülöp, and Ungureanu 2019). Nevertheless, these two studies are limited to steel-framed industrial buildings, and the developed predictive method is not based on actual reused components. Moreover, they considered only one economic factor, ignored the impact of other variables, and did not consider the interdependencies between the affecting variables.

In brief, the deficiencies of the methods used to evaluate or predict the reusability of load-bearing building elements include:

i. Most of these methods are focused on one aspect of reusability, which is determining the mechanical properties of the elements.
ii. They are limited to a specific material.
iii. They do not consider the economic and social reusability of the elements (as essential dimensions of sustainability).
iv. Most of them are not based on real projects with reused structural components.
v. The complexity of the interactions of the affecting variables is ignored.
vi. None of these studies used advanced data analysis methods such as novel/advanced supervised machine learning techniques to reveal the sophisticated relationship between the variables and then predict the reusability of the elements using the developed probabilistic models.

The above shortcomings and the low reuse rates of the load-bearing building elements emphasise the need for the development of better tools to provide a first-hand idea about the reusability of the structural components of the buildings. Any such tool should consider the interdependencies of the multitude of factors affecting the reuse of load-bearing building components at its core and should be easy to understand and implement.

1.5 Research questions

The research questions of this study are developed considering the research problem and gap in knowledge in Section 1.4. These questions are as follows:

1) What are the factors affecting the reusability of the load-bearing building components?
2) What are the weightage and impact of the identified factors on the reusability of the building structural elements?
3) What combination of factors can contribute to the efficient development of a predictive model to assess the reusability of the load-bearing building components?
4) Which supervised machine learning technique can then be selected to accurately predict the reusability of the load-bearing building components?

1.6 Aim and objectives

This research aims to develop a model that can predict the reuse potential of structural elements at the end-of-life of a building based on professional experience. The following objectives are then considered to answer the research questions and fulfil the aim of this study.

1) To identify and assess factors affecting the reusability of a building's structural elements (reusability factors) through literature review.
2) To quantify the weightage and impact of the reusability factors based on the experience and expertise of the professionals elicited using questionnaires.
3) To determine the best combination of the identified factors (model structure) to develop the Building Structural Elements Reusability Predictive Models (BSE-RPMs).
4) To develop a best-practice of BSE-RPM using advanced supervised machine learning techniques, which provides reliable predictions.

1.7 Unit of analysis

While performing research, it is crucial to know the unit on which the data needs to be collected and analysed. The purpose of collecting data on this unit is fulfilling the aim of the research by addressing the research questions. Therefore, this unit is called the unit of analysis (Salkind 2010). According to (Addelman 1970), the unit of analysis or the experimental unit "is that entity that is allocated to a treatment 'independently' of other entities." Moreover, to collect the necessary data for research, a unit is observed, which is called the unit of observation. This entity can be similar to or different from the unit of analysis (Salkind 2010).

This research aims to develop a reliable probabilistic model using supervised machine learning techniques that can predict the reuse potential of structural elements (beams, columns, slabs, truss, etc.) at the end-of-life of a building. For this purpose, this study collects data on the factors affecting the reusability of these elements. Hence, the unit of analysis of this research is the load-bearing building components. While this research intends to consult the reuse experts to quantify the reusability factors, because it is observing the reused load-bearing building components through the senses of the experts, its unit of observation and analysis are equal. Further discussion on the unit of analysis could be found in Section 4.2.

1.8 Methodology

The first objective of this research was to identify factors affecting the reuse of load-bearing building components. Various studies have attempted to identify these factors using different methods such as interviews (da Rocha and Sattler 2009), questionnaire surveys (Chileshe et al. 2016), literature review (Tingley et al. 2017), etc. (see Tables 2.1 and 2.2 of Chapter 2 for a complete list of methods used in the literature). Moreover, most of the studies in this area are published after 2000, reflecting contemporary issues in the field of reuse. Therefore, the factors affecting the reuse of load-bearing building components can be derived from the existing body of knowledge. All these meant that there was no need to conduct interviews with the experts in this field, and a literature review could fulfil the first aim of this study. Hence, as the first step, a systematic literature review was performed to identify the reusability factors. Next, the results of the systematic review were used to develop an online questionnaire survey to fulfil the second objective of this research. In the next stage, the outcome of the survey was used to develop the BSE-RPMs.

The above discussion reveals that since the required knowledge to develop the BSE-RPMs is acquirable (first and second objectives of this research); hence, the data collection and communication approaches embrace the realism ontology (see Section 3.2.1) (Saunders, Lewis, and Thornhill 2016, Burrell and Morgan 2016). Moreover, it reveals that knowledge is objective, and the researcher is value-free because this research uses a questionnaire survey (a quantitative method) for its data collection (see Section 3.2.1) (Burrell and Morgan 2016, Chilisa and Kawulich 2012). Likewise, this study seeks generalisations by developing BSE-RPMs; hence, its approach to theory development follows a deductive pattern. Therefore, this research follows positivism as its research philosophy (Burrell and Morgan 2016). According to Crotty

(Crotty 1998), “positivism is objectivist through and through”. In the following paragraphs, the sample, data collection method, and analysis techniques used in this research are briefly discussed.

Sample: Since the unit of analysis of this research is the structural components of a building, the sample population was comprised of all recovered load-bearing building components intended for reuse (regardless of success). However, because there was no record available on the sample population, and because access to company documents was not possible, this study sought experts’ knowledge about the reused elements to develop the BSE-RPMs. Nevertheless, because there was no way to identify based on what structural element the reuse experts would complete the questionnaire, all the reused components had an equal chance for selection by the potential respondents. Hence, the sample population was the professionals with reuse experience working in construction, deconstruction, demolition, or reuse companies. The sampling methods are discussed in Chapter 4.

Literature review: A systematic literature review of the studies dealing with the factors affecting the adoption of component reuse in the building sector was performed to fulfil the first objective of this research. The identified factors from the systematic review were then used to develop an online questionnaire survey to accomplish the second objective of this research. The details of the systematic literature review are presented in Chapter 2.

Online questionnaire: The experts’ opinions were elicited by developing a comprehensive online questionnaire survey research methodology to provide a numeric description of the reusability factors and a primary evaluation of the relationship between the variables. Using the Online Surveys (Jisc 2019), an online questionnaire survey was developed based on the results of the systematic literature review (Chapter 2), and its link was shared with the potential respondents. In this research, the variables (reusability factors) identified in the questionnaire (both independent and dependent) were in the form of closed questions with the Likert-style ratings (Likert 1932). While the Likert response sets can include four or more points, this study used a five points system, which is more common (Lavrakas 2008). In total, 481 invitations were sent to the experts to complete the online questionnaire, and 90 completed surveys were received, yielding a response rate of 18.7%.

Analysis of data: The received questionnaires were initially assessed for completeness, reliability, and relevance. Next, the acceptable received questionnaires were split into three sections to develop technical, economic, and social datasets. After removing irrelevant and highly incomplete questionnaires, there were still unanswered questions in the datasets. Since these data were missing completely at random, the Multiple Imputation technique was used to estimate the missing values. Next, and after addressing the imbalance in the datasets, a three-stage feature selection using filter and wrapper methods was performed to identify the best combination of the variables to develop BSE-RPMs. This stage fulfilled the third objective of this research by identifying the list of independent variables required for developing the BSE-RPMs for each of the datasets. Then, using 70% of the data in each dataset, which was selected randomly, thirteen different supervised machine learning methods were employed to develop 13 BSE-RPMs. Next, a 10-fold Cross-Validation method was used to evaluate the performance of the models. The results show that interpretable/transparent models such as Logistic Regression and Decision Trees have poor performances (Section 6.3). Therefore, the best-practice model was selected based on their predictive performances. The result was the selection of random forest models for all three datasets as the best-practice models. Next, using sensitivity analysis and visualisation techniques, the selected black-box random forest models were opened to improve their transparency. Eventually, using rule extraction techniques, three easy-to-understand predictive models were developed that can reliably estimate the technical, economic, and social reusability of the load-bearing building components (4^{th} objective).

1.9 Novelty of research

This research, which aimed to develop BSE-RPMs to estimate the reuse potential of the structural elements of a building at its end-of-life to promote the reuse rates in the building sector, is novel in several ways. It is the first study that uses advanced supervised machine learning techniques such as random forests, K-Nearest Neighbours algorithm, Gaussian processes, support vector machines, adaptive boosting, BART machine, etc., (Section 5.5) to develop models that predict the reusability of the structural elements from technical, social, and economic perspectives. Also, it is the first study that uses advanced machine learning methods to rank the factors affecting the reuse of building structural components. A look at the literature shows that the publications in this field limit themselves to ordinary descriptive statistics and ignore the possible interdependencies of the variables. This project reveals that the relationships between variables are not linear. Moreover, it is the first study that identifies the best combination of variables to develop the BSE-RPMs. Furthermore, it is the first study that uses sensitivity analysis and visualisation techniques to interpret the selected black-box best-practice BSE-RPMs. Likewise, for the first time, this research develops a set of predictive rules that can be used by professionals in the building sector for estimating the technical, economic, and social reusability of the structural components effectively.

This research resulted in the publication of the first systematic literature review on the factors affecting the reuse of the load-bearing building components. This systematic review has contributed to identifying, categorising, and prioritising the factors affecting the reuse of components of the superstructure of a building at its end-of-life. The results of this systematic review were used to identify the reusability factors for the development of a questionnaire survey to fulfil the second objective of this research.

The easy-to-understand predictive tools developed during this research have several advantages, as follows. First, they can be used by any practitioner in the building sector, and they do not need a machine learning background. Second, they give a first-hand idea about the reusability of structural components by collecting the necessary data. Third, they have the potential to promoting reuse by increasing the reuse rates, which, in turn, can accelerate the growth of reuse markets. Considering the UK economy post-Brexit and the impact of the COVID-19 outbreak on the employment rate, the results of this project can provide new job opportunities in the building sector in the UK.

1.10 Scope and limitation

The scope of a project is dictated by its aim, objectives, unit of analysis, and unit of generalisation. This project focuses on load-bearing building component reuse, and other types of reuse, such as adaptive reuse, recycling, and non-load-bearing building material reuse, are out of the scope of this study. While adaptive reuse is the most preferred option to prevent waste, because this research focuses on the management of CDW after generation (as the result of construction, refurbishment, and demolition/deconstruction), adaptive reuse is out of the scope of this study. As explained in Section 1.1, other waste treatment options such as recycling are energy and resource-intensive (Addis 2006, WRAP 2008); therefore, not considered in the scope of this study.

The terms load-bearing building component(s) and element(s) are used interchangeably in this research. These are restricted to sections forming the superstructure of a building as defined by (BCIS 2012) that can be dismantled (through demolition, deconstruction, or selective demolition) and reused for the same function with minimum (or zero) treatments (Addis 2006, Parker and Deegan 2007). Therefore, this research does not consider substructure (foundation), plinth, finishes, fittings, furnishings, equipment, and services in its scope (BCIS 2012).

As discussed in Section 1.1, new design and construction techniques such as design for deconstruction (DfD) (Akinade et al. 2017, Tingley and Davison 2011), design for manufacture and assembly (DfMA) (Kalyun and Wodajo 2012), and Modular Construction (Thai, Ngo, and Uy 2020) could potentially promote the reuse of load-bearing building components in the long run. However, existing buildings are not designed and constructed based on these methods. Since the focus of this research is promoting the reusability of the load-bearing components of the existing buildings, buildings designed and constructed using these novel techniques are out of the scope of this research. Therefore, the results of this research could not be used to evaluate the reusability of the load-bearing structural elements of such buildings.

The most important limitation in this research is the low rate of reuse in the building sector that restricts access to more experts with such experience. Moreover, while the researcher tried to decrease error by considering a wide range of machine learning methods to develop the predictive models, there still might be some errors due to a missing key factor that has not been integrated into the questionnaire.

Likewise, the questionnaire is developed based on a systematic literature review focused on the superstructure of a building. Therefore, the results of this study cannot be generalised to the substructures. Also, while the questionnaire was not limited to any material, the responses provided were restricted to timber, steel, and concrete. Hence, the developed predictive tools in Chapter 6 can be used to determine the reusability of timber, steel, and concrete load-bearing building components.

Moreover, less than 10% of the received questionnaires used demolition to recover the structural element, out of which only one component was reusable. The remaining elements were recovered using deconstruction and components specific recovery (87.5%) or were surplus (1.4%) or reused in-situ (1.4%). Therefore, the results of this research could not be extended to evaluate the reusability of components recovered through demolition. It should be noted that while this research focuses on the building sector, the approaches used can be adapted to perform similar studies in other subsectors of the construction industry, as well.

1.11 Thesis structure

This thesis consists of seven chapters. Chapter one introduces the background, justification for the study, and the gap in the knowledge, and portrays the aim and objectives of this research. In Chapter Two, a systematic literature review focused on the factors affecting the reuse of load-bearing building components is presented. Chapter three discusses the philosophical assumptions of the research and scrutinises the potential theoretical perspectives to identify the research philosophy, and eventually identifies and justifies the choices for research methodologies and research methods. Chapter four deals with the data collection in this project. Chapter five deals with analysing the collected data using advanced supervised machine learning techniques such as random forests, K-Nearest Neighbours algorithm, Gaussian processes, support vector machines, adaptive boosting, BART machine, etc., (Section 5.5) to develop the BSE-RPMs. Chapter six is focused on selecting the best-practice technical, economic, and social BSE-RPMs and developing three easy-to-understand predictive models that can be used by the practitioners in the building sector to assess the reusability of the load-bearing building components. Findings are discussed in Chapter 6 as well. And finally, Chapter seven concludes this research (Figure 1.2).

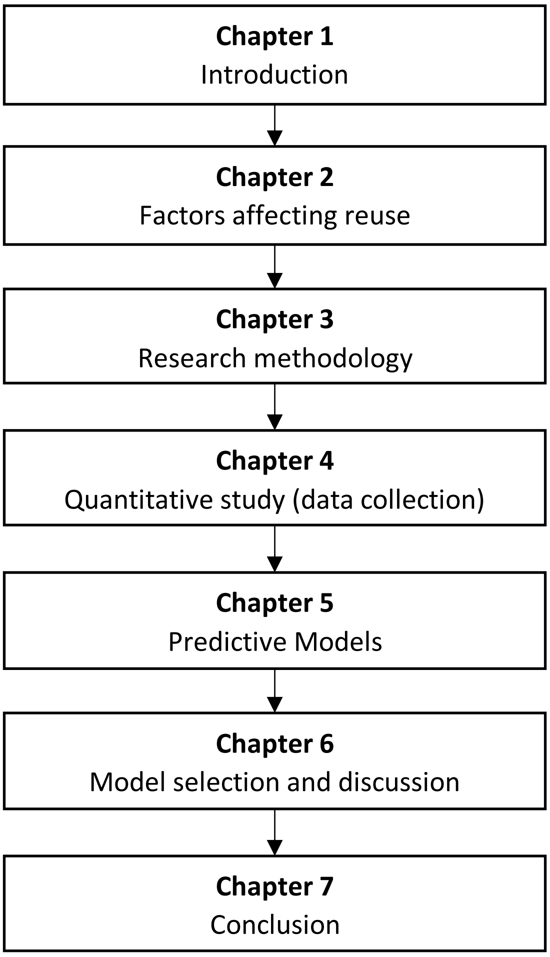

Figure 1.2 Thesis structure.

1.12 Key achievements

A significant achievement of this research is the development of three easy-to-understand predictive tools using advanced machine learning methods that can be used by practitioners in the building sector to determine the technical, economic, and social reusability of the load-bearing building components. There is only one study in this field that has tried to develop a set of rules to estimate the technical and partially economic reusability of such components (Hradil, Fülöp, and Ungureanu 2019). However, this study is not based on real reused elements and is limited to steel-framed industrial buildings. Moreover, it considers only a limited number of variables for the technical reusability assessment, has only one economic variable, and ignores the impact of other variables. Likewise, it does not integrate any social factors in its rules and does not consider the interdependencies of the affecting factors.

Another accomplishment of this research is identifying the most significant factors affecting the reuse of structural elements of a building and ranking them using advanced supervised machine learning techniques such as the Boruta method and recursive feature elimination technique. While other studies tried to identify and prioritise these factors using ordinary descriptive statistics, none used advanced machine learning techniques for this purpose.

Another achievement of this study is the successful use of advanced supervised machine learning techniques such as random forests, K-Nearest Neighbours algorithm, Gaussian processes, support vector machines, adaptive boosting, BART machine, etc., (Section 5.5) to develop BSE-RPMs. No other study has ever used such methods to predict the technical, economic, and social reusability of the load-bearing building components.

Finally, using the random forests method, this study developed best practice BSE-RPMs with a considerable overall accuracy of 96%, 89%, and 94% for the technical, economic, and social models, respectively.

1.13 Chapter summary

This chapter provided a background of the position of the construction industry in the global economy and discussed that this sector is not sustainable. Next, this chapter identified that reusing the load-bearing building components has a high potential for improving the overall sustainability footprint of the construction industry. This chapter then justified the need for this research based on the low reuse rates in the UK, and globally. The gap in the knowledge showed that the available reusability assessment tools are oversimplified. They are also limited to identifying the mechanical properties of the structural components, not considering the interdependencies between the variables, and ignoring important technical, economic, and social factors. It also showed that none of such studies used advanced data analysis methods such as supervised machine learning techniques to develop reusability assessment tools.

This chapter revealed that the unit of analysis is all load-bearing building components. This chapter further discussed the methodology adopted in this research and explained how positivism is the philosophical underpinning of the study and justified the quantitative method approach used for data collection and analysis.

The novelty of the research section highlighted that it is the first study that uses advanced supervised machine learning techniques to develop predictive models to assess the reusability of the structural elements from technical, social, and economic perspectives. Also, it shows that this research is the first study that develops three easy-to-understand predictive tools, which could assist practitioners in the building sector in evaluating the technical, economic, and social reusability of load-bearing building components.

Chapter 2 – Systematic review of factors affecting the reuse of load-bearing building components

2.1 Chapter introduction

This research aims to develop probabilistic predictive models to evaluate the reusability of the structural elements of a building at its end-of-life. In this chapter, a systematic literature review is performed to identify factors affecting the reuse of load-bearing building elements. The outcome of this chapter fulfils the first objective of this research (Section 1.6). Kindly note that the systematic literature review performed in this chapter is focused on the construction engineering journal articles and not machine learning papers. Relevant machine learning articles are referred to and discussed in Chapters 4, 5, and 6.

The scope of this chapter is limited to peer-reviewed journal articles because these types of research works are considered of high quality and validity (Schlosser 2007). This approach is in line with the advice of (Yi and Chan 2014) to investigate top-tier construction journals while performing literature reviews. As discussed in Section 1.10, the scope of this research is limited to the load-bearing building components reuse, and other types of reuse, such as adaptive reuse, recycling, and building material reuse are not considered. This trend is followed while selecting the proper search words, as well (Section 2.2). Two major examinations are performed to scrutinise the articles reviewed in this chapter. The first method (Section 2.3) is focused on identifying and analysing reuse drivers and barriers (cumulatively called factors), and the second method (Section 2.4) is focused on correlations and the possible inter-relationships between reuse barriers. This chapter concludes with the chapter summary in Section 2.5.

It is noteworthy that this chapter is published in the journal of Waste Management & Research as the first systematic literature review in this field (Rakhshan et al. 2020).

2.2 Systematic literature review approach

This chapter uses a systematic literature review method to identify various factors (drivers and barriers) affecting the reuse of load-bearing building components on a global scale. A systematic review is a comprehensive and reliable process for locating the existing body of knowledge (published scientific work) regarding a very particular research question (GET-IT Glossary n.d., Denyer and Tranfield 2009). Because this process is based on a defined search strategy with clearly specified objective(s), it can be used to analyse, synthesise and critically evaluate the existing literature identified within the context of the research question (Section 1.5) (Denyer and Tranfield 2009, Bettany-Saltikov 2016). This methodology provides a strong basis for reliable judgments about “what works” the best (Petrosino and Lavenberg 2007) and finds gaps in the literature for further research (Denyer and Tranfield 2009). The systematic literature review is a well-known methodology for the study of the existing knowledge in medical sciences because of its unique properties, as expressed above (Tranfield, Denyer, and Smart 2003). Nevertheless, the systematic literature review is acquiring its position among other research areas such as engineering and management (Hosseini et al. 2015, Alaka et al. 2018, 2016, Charef, Alaka, and Emmitt 2018).

The complete process of the systematic literature review is presented in Figure 2.1. In this research, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (PRISMA 2018) checklist is used to step-by-step perform and record the methodology. PRISMA checklist is widely used by researchers while performing systematic literature reviews (Moher et al. 2009).

A pre-requisite to conducting a systematic review is a clear research question (question 1 of Section 1.5) as well as knowing the proper keywords to perform an effective search. Because a building at the end of its lifecycle is removed through demolition (with some other variations such as selective demolition and deconstruction), to identify the proper keywords, the researcher performed an initial literature search using "deconstruction" and "demolition" search words at stage 1. Through this initial search, 11 relevant papers were identified, which helped in the selection of the search words listed in Figure 2.1 (stage 2).

At stage 2, a Boolean search criterion is followed to answer the research question of this study (question 1 of Section 1.5). At this stage, the search is limited to the “titles” of the articles. The initial search in Scopus showed that studies containing discussions on the reuse of building components focus on construction and demolition waste management. Therefore, the first set of search words intends to ensure that any article containing these words are considered. The AND combination with the second set of search words guarantees that all relevant articles dealing with reuse in the building sector are included in the search. Because the scope of this research is load-bearing building components (BCR) reuse and not building reuse or building material reuse, keywords such as “refurbish” or “refurbishment”, which primarily deal with adaptive reuse of existing buildings (particularly historic buildings) or “material”, which deals with material reuse are not included in the search words (Figure 2.1).

The cut-off date for stages 1 and 2 of the literature reviews is March 2019, whereas the cut-off date for stage 3 is January 2020. Because this research only focuses on peer-reviewed journal papers, following (Yi and Chan 2014), all other types of publications (book chapters, conference papers, trade journals, etc.) are excluded. Hence, only “Articles” and “Articles in press” published in peer-reviewed journals are considered for this research. Likewise, to limit the number of unwanted articles, irrelevant subject areas, as listed in Figure 2.1 at stage 2, are excluded from the search criterion. It is because search words such as “building”, “construction”, “structure”, “reuse”, and “recover” are found in a broad range of scientific publications. Furthermore, since most of the publications in this area are published post-2000, stage 2 considers the range of articles published between 2000 and March 2019.

Among the 2,387 article titles screened at stage two, 2,162 articles were found irrelevant and excluded. Figure 2.2 depicts the percentage of the subject areas of the excluded papers during the screening stage. The appearance of articles in areas like medical sciences (while was excluded from the subject areas) could be because of the interdisciplinary nature of some papers. The researcher then reviewed the abstracts of the remaining 226 articles during the eligibility check of stage 2 (PRISMA 2018) (Figure 2.1). At this stage, irrelevant papers such as those focusing on construction waste management other than reuse (Guo 2016, Jin et al. 2017), concentrating on other sectors like reverse logistics in the electronics industry (Sirisawat and Kiatcharoenpol 2019), or talking about reuse but dealing with recycling or down-cycling (Migliore et al. 2015) are identified and excluded. The result is the exclusion of 141 more papers from the full-text review. The researcher eventually reviewed 85 full-text articles from which 54 papers were found relevant to the objective of this chapter.

The search results from stages 1 and 2 indicate that the journal of *Resources, Conservation and Recycling (RCR)* has the highest number of publications (16 papers) among all other reviewed journals. Hence, following the framework pursued by Yi and Chan (2014), a third stage systematic literature review was performed considering all the ten first-tier construction journals plus *Resources, Conservation and Recycling (RCR)*. The complete list of all these journals are *Automation in Construction (AIC); Building and Environment (BE); Building Research and Information (BRI); Canadian Journal of Civil Engineering (CJCE); Construction Management and Economics (CME); Engineering, Construction and Architectural Management (ECAM); International Journal of Project Management (IJPM); Journal of Computing in Civil Engineering (JCCE); Journal of Construction Engineering and Management (JCEM); Journal of Management in Engineering (JME); Resources, Conservation and Recycling (RCR)*. At this stage, the identified search words were used to perform a Boolean search in the ‘title/abstract/keywords’ of each of the journals separately. Moreover, the year 2000 restriction was lifted at this stage (Figure 2.1). All the above was to overcome the restrictive nature of the stage 2 limitations (Figure 2.1), as well as to make sure that articles published in high-impact journals related to the built environment are considered.

During this process, 490 articles were excluded from abstract review for similar reasons observed in stage 2. For instance, while (Ling and Leo 2000) focuses on identifying drivers to promote timber formwork reuse, it is out of the scope of this study, which is the superstructure of a building. After reviewing 609 abstracts during the eligibility check, only 28 papers were identified for full-text review. While the reviewed full texts contained a combination of the search words, the focus of the rejected papers was not in line with the aim of this chapter. Following the same protocol pursued at stage 2, a total number of 11 more papers were identified at this stage. According to what mentioned earlier and combining the identified papers at all three stages, 76 articles were found relevant to the objective of this chapter and reviewed. Nonetheless, the identified new articles, as the result of the third stage systematic review, were all published after the year 2000, which validates the initial decision in restricting the publication date.

The presented systematic literature review framework in Figure 2.1 is highly reproducible and suitably matches objectivism, which is the epistemological stance of this research (Section 3.2.1). Nevertheless, there is always a risk of not locating some relevant articles due to the restrictions considered in performing a systematic literature review. Therefore, the researcher looked for the "grey literature" by carefully reviewing the references of the identified articles. While according to (Adams et al. 2016), "grey literature includes a range of documents not controlled by commercial publishing organisations", in this research, this term is extended to those articles that are missed out as the result of defined restrictions in the process of performing the systematic literature review (Figure 2.1).

While reviewing the references of each paper, the researcher identified potential papers for further review. Next, the researcher checked if these potential papers were already identified during the systematic literature review and excluded them with a reason (Figures 2.1 and 2.2). In most of the cases, the identified papers were already reviewed and excluded. For instance, while reviewing the references of (Chileshe, Rameezdeen, and Hosseini 2016), the authors identified seven potential papers. However, after checking the excluded papers during the systematic review process, it was observed that those papers were excluded for a reason (Figure 2.2). As an example, (Leigh and Patterson 2006) was initially identified as a potential paper. However, after reviewing the notes, it was observed that this paper was focused on identifying means that could be used by the local government to promote deconstruction.

Notwithstanding, there were still some papers that were not identified during the process of the systematic review. For instance, (Kuehlen, Thompson, and Schultmann 2014) is a relevant paper, which was identified during reviewing the references of (Dunant et al. 2018). However, this is a conference paper, and as discussed earlier, this systematic review did not include conference papers and only focused on peer-reviewed journal articles. Nevertheless, a careful review of such conference papers revealed that they were the basis for most of the articles reviewed during the systematic literature review in this chapter. For instance, the paper by (Kuehlen, Thompson, and Schultmann 2014) was referred to in (Dunant et al. 2017), which is another identified paper for review in this chapter. Another example is a CIB Report, Publication 252 (Kibert and Chini 2000), which is cited in different articles identified during the systematic literature review in this chapter, such as (Huuhka et al. 2015, Shaurette 2006, Diyamandoglu and Fortuna 2015, Chileshe et al. 2016), among others.

Therefore, the review of the grey literature revealed that no important journal articles were missing, and the located papers during the systematic literature review represent the state-of-the-art in this field.

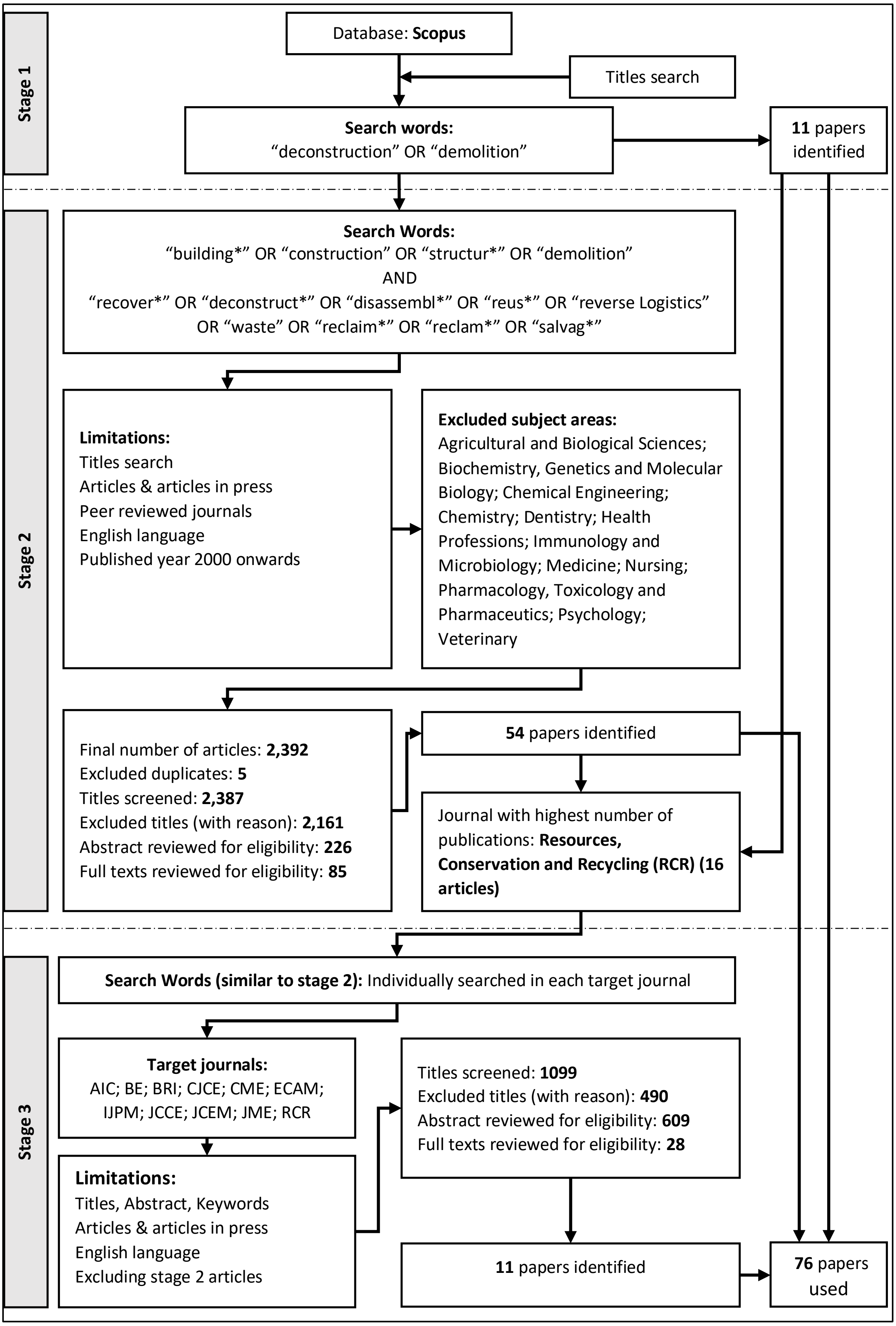

Figure 2.1 Systematic literature review framework (inspired by (Charef, Alaka, and Emmitt 2018, PRISMA 2018, Yi and Chan 2014)*)*

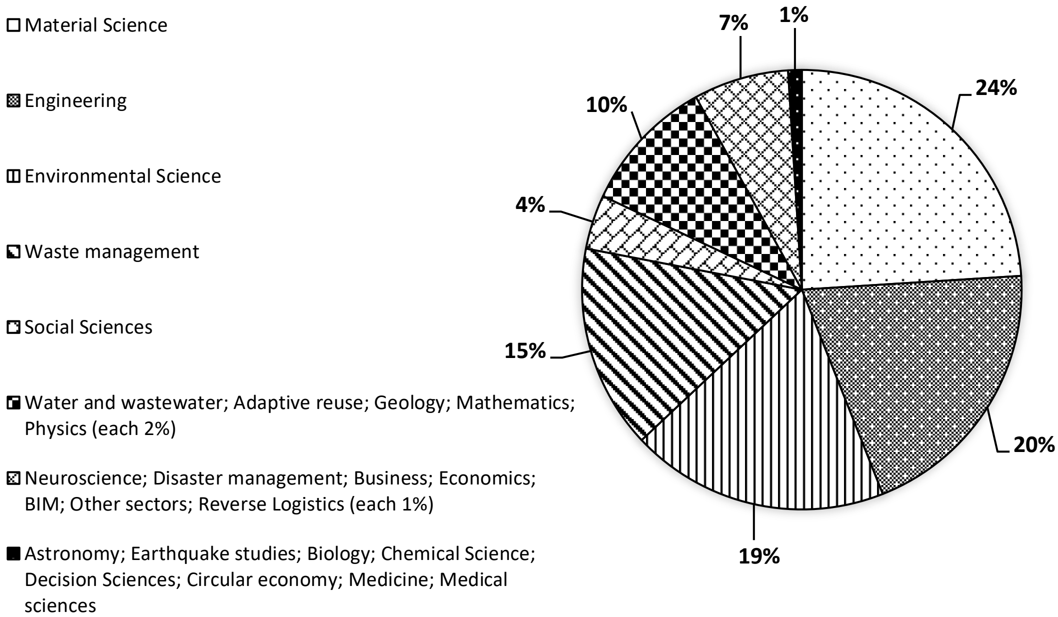

Figure 2.2 Subject area of the excluded papers during the screening process at stage 2

2.3 Results of the systematic literature review

Figure 2.3 shows the distribution of the papers reviewed in this chapter by the year of publication. According to this figure, the number of peer-reviewed journal articles has been increasing since 2014, which indicates an increasing focus on construction and demolition waste treatment through reuse. However, there is a decline in the number of publications in 2019, which needs further investigations to identify the root causes.

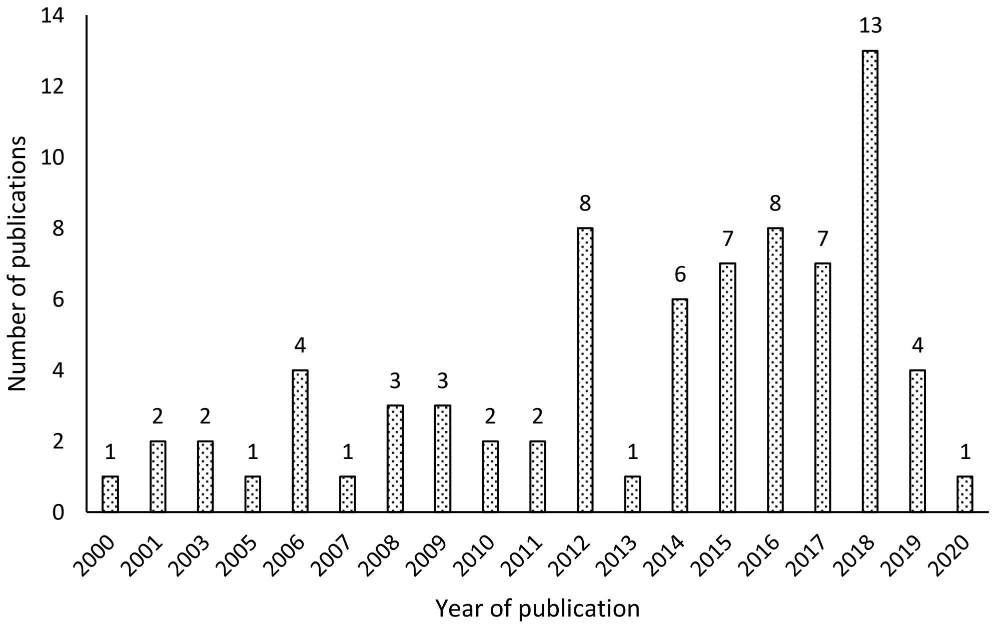

Figure 2.3 Publications by year.

Figure 2.4 shows the geographic location of the reviewed articles in this chapter. According to this figure, waste management in buildings through reuse is an international trend. It should be noted that the split of the reviewed articles based on their geographic locations is based on the focus of the research paper on the construction context of the listed countries.

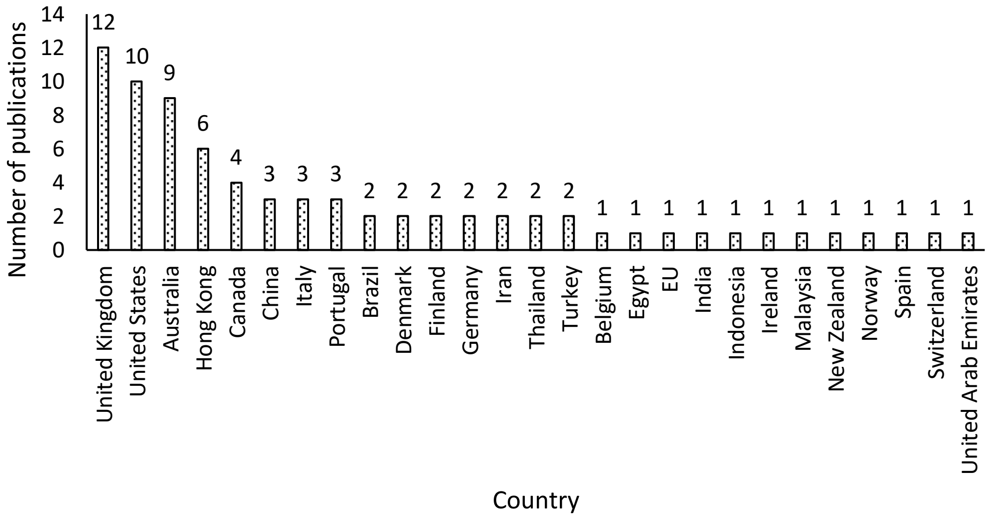

Figure 2.4 Publications by location.

Tables 2.1 and 2.2 show that the authors of the reviewed papers employed various methodologies to perform their research. These methodologies are identified for the individual papers in Table 2.1 for reuse drivers & Table 2.2 for reuse barriers. The variety of techniques used, including various qualitative and quantitative methods, show the attempts made by different authors to study different aspects of BCR, which reveals the increasing importance of this intervention among researchers. For instance, a series of studies performed in Australia employs mixed methodologies such as interviews and questionnaire surveys and targets various stakeholders to investigate drivers and barriers to reverse logistics in the South Australian construction context (Chileshe et al. 2016, Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016, Chileshe et al. 2018, Chileshe, Rameezdeen, and Hosseini 2015). These studies show the importance of a holistic approach in seeking the experts' opinions (through qualitative methods (Saunders, Lewis, and Thornhill 2016)) and the actual experiences (through quantitative methods (Saunders, Lewis, and Thornhill 2016)) to identify deficiencies in the body of knowledge and eventually promote practices like reuse in the building sector. While it is tempting to discuss different research methods and methodologies employed in the 76 papers reviewed (and compare advantages and limitations of them), the above is out of the scope of this chapter and can be investigated separately. However, the methodologies and methods applicable to this research are discussed in detail in Chapter 3.

Throughout the systematic literature review process, the researcher identified 57 drivers and 130 barriers affecting the reuse of building components. From a sustainability perspective, the reuse of building components has social, environmental, and economic advantages (Jaillon and Poon 2014); hence, certain factors can be categorised under these three groups. However, the successful implementation of any intervention (here, the reuse of building components) to promote sustainability in the building sector highly depends on the technical feasibility (such as durability), the regulatory enforcement (minimum performance requirements set by regulations), and competency and willingness of the organisations engaged (knowledge, skills, infrastructure, innovation, etc.)(Nußholz, Nygaard Rasmussen, and Milios 2019). Therefore, an interdisciplinary approach towards sustainability becomes crucial while addressing the shortcomings in the body of knowledge on reuse (Kajikawa, Tacoa, and Yamaguchi 2014). On this basis and following (Pomponi and Moncaster 2017, Tingley et al. 2017), the identified reuse drivers and barriers were grouped under economic, environmental, social, technical, regulatory, and organisational categories (Tables 2.1 and 2.2).

Besides, to better present the identified reuse drivers and barriers and to avoid congested tables, under each major category, the factors were further grouped into sub-categories, as shown in Tables 2.1 and 2.2. These sub-categories are defined based on the common characteristics of groups of factors. For instance, “Lower cost of reused components” and “Increased cost of landfilling” are economic drivers and are grouped under the sub-category “Cost” in Table 2.1. It is because in the case of the former, the lower cost of the component can decrease the total cost of the project and in the case of the latter, landfilling is expensive and reusing the element can reduce additional costs. This approach has been pursued in the case of barriers to BCR, as well.

2.3.1 Reuse drivers

The complete list of identified reuse drivers is available in Appendix A (Table A.1). Figure 2.5 shows the distribution of the observed drivers in the reviewed papers. According to this figure, the principal identified drivers are economic (25%), organisational (23%), environmental (17%), and social (15%). The sub-categories of the factors shown in this figure present a similar trend between main categories and sub-categories. Among the drivers, “cost” is the most reported sub-category, followed by “energy and GHG”, “organisational sustainability”, and “willingness” sub-category of drivers. These observations are discussed further in the following subsections.

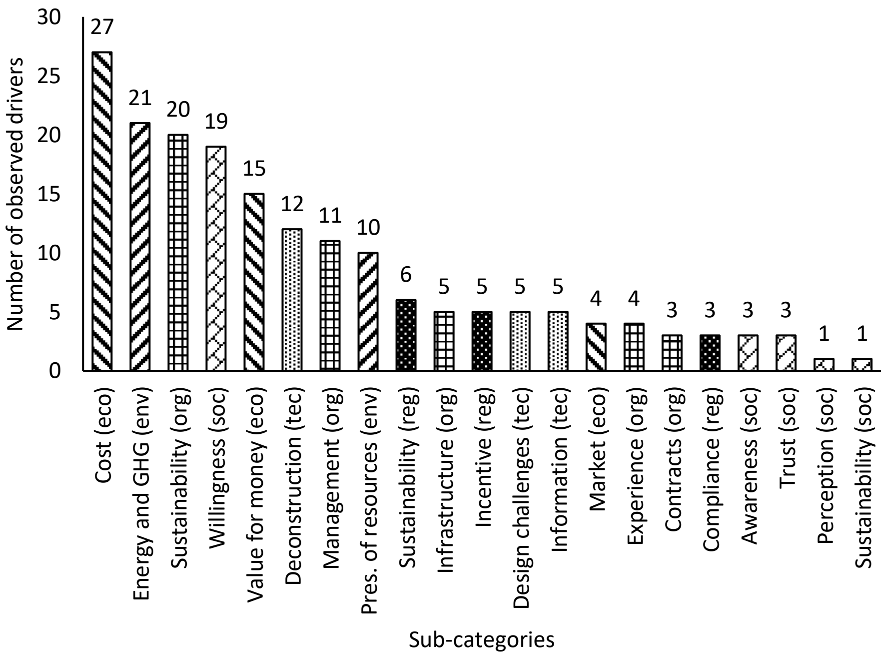

Figure 2.5 Distribution of the observed reuse drivers (eco: economic; env: environmental; org: organisational; reg: regulatory; soc: social; tec: technical)

2.3.1.1 Economic drivers

From the reviewed articles, it is observed that the potential cost savings as the result of using recovered building components can promote reuse. For example, according to (MacKinnon 2000, Klang, Vikman, and Brattebø 2003, Gorgolewski et al. 2008, da Rocha and Sattler 2009, Dunant et al. 2017, Chileshe et al. 2018), the lower price of the reused components can contribute to the cost savings in the construction projects. Likewise, according to (Cooper et al. 2016), reusing steel sections results in the purchase of fewer new steel sections. If the price for the reused components is attractive, the demand for them can increase (Klang, Vikman, and Brattebø 2003), which in the long run supports the growth of a reuse market (da Rocha and Sattler 2009, Tingley et al. 2017) and increases the revenue from the resale of these components (Klang, Vikman, and Brattebø 2003, Dantata, Touran, and Wang 2005, da Rocha and Sattler 2009, Dunant et al. 2017, Chileshe et al. 2018, Sea-Lim et al. 2018). Moreover, the increased cost of landfilling can act as a reuse driver because it increases the disposal cost of CDW (Dantata, Touran, and Wang 2005, Gorgolewski 2008, Chinda and Ammarapala 2016, Chileshe, Rameezdeen, and Hosseini 2016). By reusing the recovered building components, this extra cost can be decreased (Pun, Liu, and Langston 2006). However, these factors highly depend on the geographic location of the building, which might have an opposing effect on reuse. For instance, (Huang et al. 2018) report that the lower cost of landfilling is an impediment to reuse. The study is performed in China, where cheap landfilling discourages choosing other waste treatment options such as reuse or recycling.

2.3.1.2 Organisational drivers

According to the literature, reducing CDW generated by the firms (Pun, Liu, and Langston 2006, Guy 2006, Schultmann and Sunke 2007, Densley Tingley et al. 2012, Aye et al. 2012) (among others[2]) and promoting the green image of the companies to improve competitiveness (Rogers 2011, Durão et al. 2014, Chileshe et al. 2016, Chinda and Ammarapala 2016, Chileshe, Rameezdeen, and Hosseini 2016) (among others) rank the highest among all other organisational drivers.

One method to increase the reuse rates by the organisations is through integrating reuse in the design process of new projects (Gorgolewski et al. 2008, Gorgolewski 2008, Rogers 2011, Tingley et al. 2017) (among others). As a result and to support this idea, some articles suggest that by integrating reuse in the contractual requirements, reuse rates will increase (MacKinnon 2000, Gorgolewski et al. 2008, Gorgolewski 2008). Also, the existence of a reclaimed components management coordinator (Gorgolewski 2008, Tingley et al. 2017), and the knowledge of a known list of structural components to reuse early on in the design phase are suggested to potentially increase the adoption of reuse by the firms (Gorgolewski 2008, Rose and Stegemann 2018). The latter can be facilitated by the coordination between the owners of the demolition site and the new building. However, in many instances, this coordination never happens (Dunant et al. 2018, Nußholz, Nygaard Rasmussen, and Milios 2019). One solution, as observed by (Nußholz, Nygaard Rasmussen, and Milios 2019), is companies' entrepreneurial activities to integrate circular principles. According to this study, a Danish company involved in brick reuse could overcome certain limitations by changing its business model by integrating deconstruction into its scope to safeguard a more sustainable supply of the reused bricks.

Training operators for effective deconstruction (Dantata, Touran, and Wang 2005, Shaurette 2006, Elias Özkan 2012), availability of space for the storage of the reusable components after deconstruction (Rogers 2011), and the knowledge and experience in using reused components (Tingley et al. 2017), as well as proper separation of the reusable components after deconstruction (Rogers 2011, Elias Özkan 2012, Ding et al. 2016, Ajayi et al. 2017) are among other factors driving reuse.

[2] This term indicates that there are other references identifying the same factor (Appendix A).

2.3.1.3 Social drivers

Factors such as society's environmental concerns (Chileshe, Rameezdeen, and Hosseini 2016), or the increased awareness of the full benefits of reuse among the stakeholders (MacKinnon 2000) are identified as drivers to reuse. Nußholz et al. (2019) report recognition of reuse in the public debate can enhance public awareness and promote reuse.

Besides, from a social perspective, the positive perception and willingness of the stakeholders such as clients (Shaurette 2006, Gorgolewski et al. 2008, Gorgolewski 2008, Arif et al. 2012, Sansom and Avery 2014, Dunant et al. 2017, 2018), designers (Gorgolewski et al. 2008, Gorgolewski 2008, Rameezdeen et al. 2016, Dunant et al. 2017, Tingley et al. 2017, Dunant et al. 2018), and contractors (Gorgolewski et al. 2008, Rogers 2011, Chileshe et al. 2016, Dunant et al. 2017, Chileshe et al. 2018) to integrate reused components into their projects are determining.

Unlike new building components that can be sourced from the market with proper quality certificates, salvaged building components are usually not available off the shelf and cannot be trusted. However, according to a few articles, informality, and good relationship among the stakeholders is reported to overcome this challenge and promote reuse (Shaurette 2006, da Rocha and Sattler 2009, Chileshe et al. 2016).

Table 2.1 Summary of reuse drivers

SN	Reference	Cntr.[a]	Research method[b]	Categories of reuse drivers[c]																				
				Economic			Env		Organisational					Regul.			Social					Technical		
				A: Cost	B: Market	C: Value for money	D: Energy and GHG	E: Preservation	F: Contracts	G: Experience	H: Infrastructure	J: Management	K: Sustainability	L: Compliance	M: Incentive	N: Sustainability	O: Awareness	P: Perception	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design	X: Information
1	(MacKinnon 2000)	US	DR; GI; I(4); OBS	1					1								1							
2	(Sára, Antonini, and Tarantini 2001)	IT	CS(1); LIR				1	1																
3	(Li, Chen, and Wong 2003)	HK	CS(2); S												1									
4	(Klang, Vikman, and Brattebø 2003)	US	CS(1); I(10); Q(10/10)	2	1	1	1																	
5	(Dantata, Touran, and Wang 2005)	US	CS(5); LIR	1		1				1														
6	(Pun and Liu 2006)	AU	TF																			1		
7	(Pun, Liu, and Langston 2006)	AU	CS(1)			2							1									1		
8	(Shaurette 2006)	US	Q(296/83)							1						1				1	1			
9	(Guy 2006)	US	CS(4)										1											
10	(Schultmann and Sunke 2007)	DE	T										1									1		
11	(Gorgolewski et al. 2008)	CA	AR; CS(3)	3					1			1									3	1	2	1
12	(Gorgolewski 2008)	CA	AR; CS(2)	3					1			3				1					2		1	1
13	(Tam and Tam 2008)	HK	CS(1); I(20)												1		1							
14	(da Rocha and Sattler 2009)	BR	CD; CS(1); DO(5); GM(4); SSI(27)	2	1	1														1				
15	(Nordby et al. 2009)	NO	CS(1)																					1
16	(Dewulf et al. 2009)	BE	CS(1)					1																
17	(Denhart 2010)	US	CS(4)			1																1		
18	(Rogers 2011)	AE	CS(1)								2	1	2					1			1			
19	(Forsythe 2011)	AU	CS(9); DO; UI			1																1		
20	(Chau et al. 2012)	HK	CS(13)				1	1																
21	(Arif et al. 2012)	IN	CS(2); SSI(15)	1																	1			
22	(Lachimpadi et al. 2012)	MY	CS(8)																			1		
23	(Boyd, Stevenson, and Augenbraun 2012)	US	CS(2)				1																	
24	(Densley Tingley et al. 2012)	GB	CS(1); LIR					1					1									1		
25	(Coelho, de Brito, and Brito 2012)	PT	CS(15)				1					1												
26	(Aye et al. 2012)	AU	CS(1)				1	1					1											
27	(Elias Özkan 2012)	TR	AR; CS; DO(21); I							1	1													
28	(Hglmeier et al. 2013)	DE	CS(1)																			1		
29	(Sansom and Avery 2014)	GB	Q(160/32)																		1			
30	(Elias-Ozkan 2014)	TR	CS(2)			1	1	1					1											
31	(Pongiglione and Calderini 2014)	IT	AR; CS(1)	1																		1	1	1
32	(Durão et al. 2014)	PT	CS(2)										1											
33	(Diyamandoglu and Fortuna 2015)	US	CS(1)	1	1	1	1																	
34	(Yeung, Walbridge, and Haas 2015)	CA	DO(4)																			1		
35	(Wu et al. 2016)	CN	CA													1								

S N	Reference	Cntr.[a]	Research method[b]	Categories of reuse drivers[c]																				
				Economic			Env		Organisational					Regul.			Social					Technical		
				A: Cost	B: Market	C: Value for money	D: Energy and GHG	E: Preservation	F: Contracts	G: Experience	H: Infrastructure	J: Management	K: Sustainability	L: Compliance	M: Incentive	N: Sustainability	O: Awareness	P: Perception	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design	X: Information
36	(Cooper et al. 2016)	GB	CS(2); LIR; SSI(17)	1		1																		
37	(Rameezdeen et al. 2016)	AU	SSI(8)																		1			
38	(Ding et al. 2016)	CN	CS(1); LIR; SSI(12)								1													
39	(Chileshe et al. 2016)	AU	LIR; Q(539/49); SSI(6)										2							1	1			
40	(Ajayi et al. 2016)	GB	FGI(23)																			1		
41	(Chinda and Ammarapala 2016)	TH	CS(2); I(6); LIR	1				1					1			1								
42	(Chileshe, Rameezdeen, and Hosseini 2016)	AU	LIR; SSI(8)	1									1						1					
43	(Tatiya et al. 2017)	US	CS(1); LIR; SI(3)	1																				
44	(Ajayi et al. 2017)	GB	FS; Q(200/131)								1	1												
45	(Surahman, Higashi, and Kubota 2017)	ID	CS(2)				1						1											
46	(Chau et al. 2017)	HK	CS(1)				1																	
47	(Dunant et al. 2017)	GB	I(30); Q(24)	1		1															3			
48	(Faleschini et al. 2017)	IT	CS(1)				1																	
49	(Tingley et al. 2017)	GB	LIR; SSI(13)	1	1		1	1		1		2	1								1			
50	(Yeung et al. 2017)	CA	CS(1)				1	1																
51	(Machado, de Souza, and Veríssimo 2018)	BR	LIR				1																1	1
52	(Gottsche and Kelly 2018)	IE	ACT(1); CS(5)			1	1						1											
53	(Gálvez-Martos et al. 2018)	EU	CA										1											
54	(Brütting et al. 2019)	CH	CS(2)	2			1																	
55	(Chileshe et al. 2018)	AU	Q(260/26)	1		1							2	2	2	1					1			
56	(Sea-Lim et al. 2018)	TH	SD			1																		
57	(Mahpour and Mortaheb 2018)	IR	CS(1); Q(81/81)												1									
58	(Rose and Stegemann 2018)	GB	CD; CS(6); DO; SSI(21)									1	1											
59	(Dunant et al. 2018)	GB	I(30)	2																	2			
60	(Zaman et al. 2018)	NZ	CS(1)				1																	
61	(Dunant et al. 2019)	GB	ECOM				1																	
62	(Nußholz, Nygaard Rasmussen, and Milios 2019)	DK	CS(3); Q(3); SSI(3)	1		1	1	1				1		1		1	1				1			
63	(Brambilla et al. 2019)	GB	CS(1)				1																	
64	(Eberhardt, Birgisdóttir, and Birkved 2019)	DK	CS(1)				1																	
			Total numbers:	27	4	15	21	10	3	4	5	11	20	3	5	6	3	1	1	3	19	12	5	5

[a] ***Country:*** *According to ISO 3166*

[b] ***Research Method:*** *(ACT) Action Research (n = number of case(s), if provided); (AR) Archival research (n = number of case(s), if provided); (CA) Comparative analysis; (CD) Company documentation; (CS) Case study (n = number of case(s)); (DO) Direct observation (n = number of case(s)); (DR) Document review; (ECOM) Economic models; (EX) Experiment; (FGI) Focused-group interview (n = number of interviewee(s)); (FS) Field study; (GI) Group Interview; (GM) Group meetings (n = number of attendant(s)); (I) Unspecified type Interviews (n = number of interviewee(s)); (LIR) Literature review; (OBS) Observation; (Q) Questionnaire (n = number of sent Q / m = number of completed Q); (S) Survey (i.e. empirical survey, etc.); (SD) System dynamics; (SI) Structured interviews (n = number of interviewee(s)); (SSI) Semi-structured interviews (n = number of interviewee(s)); (T) Theoretical study; (TF) Theoretical framework; (UI) Unstructured interview*

[c] *The numbers in the table corresponds to the number of drivers grouped under each sub-category.*

2.3.1.4 Environmental drivers

One potential reuse driver is the scarcity of landfilling sites, which helps the environment by avoiding dumping the reusable waste into landfills (Chinda and Ammarapala 2016, Chau et al. 2012). According to the literature, reuse can decrease the use of virgin materials and water consumption (Tingley et al. 2017, Sára, Antonini, and Tarantini 2001, Densley Tingley et al. 2012, Aye et al. 2012, Yeung et al. 2017). As mentioned in Section 1.1, because of the considerable advantages of reuse, components reuse can improve the environmental footprint of buildings worldwide. By reusing building components embodied energy and carbon of construction can be decreased (Klang, Vikman, and Brattebø 2003, Tingley et al. 2017, Yeung et al. 2017, Brütting et al. 2019) (among others). Brütting et al. (2019) show that a structure made with the reused steel sections have considerably lower embodied energy and CO_2. In their study, the authors developed a discrete structural optimisation method to reuse the existing stock of the steel sections. They used LCA to compare the environmental impacts of conventional design with the proposed method (Brütting et al. 2019).

2.3.1.5 Other drivers

Based on the reviewed articles, deconstruction instead of demolition can enhance the reusability of the recovered components (Gorgolewski et al. 2008, Hglmeier et al. 2013, Pongiglione and Calderini 2014, Yeung, Walbridge, and Haas 2015) (among others). According to (Gorgolewski et al. 2008, Gorgolewski 2008, Pongiglione and Calderini 2014), the availability of information about the characteristics, details, certificates, and drawings of the recovered building components can positively contribute to increasing the reuse rates, as well.

In projects with recovered building components, the proper estimation of the required sizes and lengths at the beginning of the design phase is reported to promote reuse (Gorgolewski et al. 2008). Some articles advise that reusing the recovered components, such as the structural components, to serve the same purpose (for instance, similar loads) has a positive impact on the success of this intervention (Gorgolewski et al. 2008, Gorgolewski 2008, Pongiglione and Calderini 2014).

The environmental policies (Chileshe et al. 2018) and green building rating systems such as BREEAM and LEED are reported to have a positive impact on reuse rates (Shaurette 2006, Gorgolewski 2008). The availability of regulatory and financial incentives to encourage deconstruction and reuse, as well as the existence of regulations supporting these interventions can potentially promote reuse (Chileshe et al. 2018). However, according to the reviewed articles, such ordinances are currently not available (Yeung, Walbridge, and Haas 2015, Chileshe et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016, Tingley et al. 2017, Rose and Stegemann 2018).

2.3.2 Reuse barriers

The complete list of identified reuse barriers is available in Appendix A (Table A.2). Figure 2.6 shows the distribution of the observed barriers in the reviewed papers. According to this figure, the identified barriers are primarily economic barriers (39%), followed by technical (23%), and social barriers (15%). The sub-category of the factors shown in this figure reveals additional information about the observations. Among the identified factors, "cost" is the most reported sub-category of barriers followed by "design challenges", "compliance", "market", "deconstruction", and "perception". However, unlike the main categories, the third rank in sub-categories, "compliance", is a regulatory barrier. These observations are discussed further in the following sections.

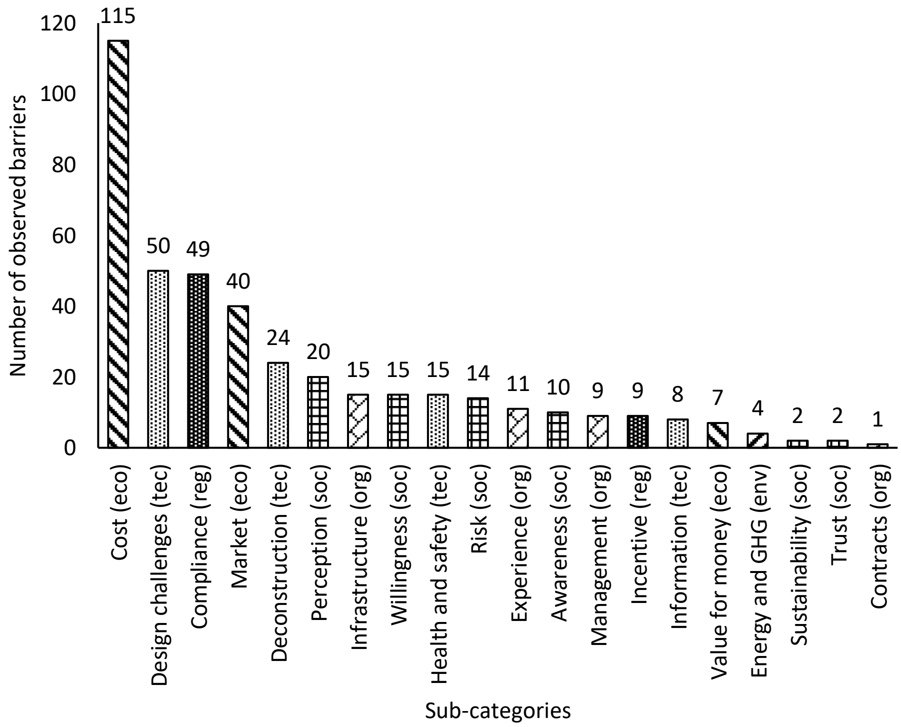

Figure 2.6 Distribution of the observed reuse barriers (eco: economic; env: environmental; org: organisational; reg: regulatory; soc: social; tec: technical).

2.3.2.1 Economic barriers

While deconstruction can increase the reusability of the recovered building components (Addis 2006, Munroe, Hatamiya, and Westwind 2006), it is associated with extra efforts (Gorgolewski et al. 2008, Chileshe, Rameezdeen, and Hosseini 2015, Rameezdeen et al. 2016). Dantata et al. (2005) highlight that the time required to deconstruct a 1000 to 2000 square foot building is three to five times higher than the time needed for the demolition of the same building. According to the reviewed articles, the time required for deconstruction and reuse, and the consequent project scheduling is one of the main barriers to reuse (MacKinnon 2000, Dantata, Touran, and Wang 2005, Shaurette 2006, Gorgolewski et al. 2008, Gorgolewski 2008) (among others). It is because there is usually a high pressure to complete construction projects as early as possible (Chinda and Ammarapala 2016). The tight project schedule negatively affects the efficient disassembly of the existing buildings and lowers the chance for the recovery of reusable building components (Sansom and Avery 2014).

During the deconstruction phase, more time is required to carefully remove and sort the recovered building components (Gorgolewski 2008), which increases the cost of sorting (Rameezdeen et al. 2016). Sometimes the deconstruction time extends beyond anticipations because of issues such as the lack of space for the equipment, complexity of the building design, and the geographic location of the building (Tatiya et al. 2017). These extra charges can yield in higher deconstruction cost (when compared to the demolition of the same building) (Dantata, Touran, and Wang 2005, Chileshe, Rameezdeen, and Hosseini 2015, Yeung, Walbridge, and Haas 2015, Tingley et al. 2017, Rose and Stegemann 2018, Dunant et al. 2018) and eventually increase the price of the recovered components (Shaurette 2006, Chileshe, Rameezdeen, and Hosseini 2015, Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016, Tingley et al. 2017, Dunant et al. 2018).

Another economic barrier to the BCR is the higher cost of design with the reused components (Gorgolewski et al. 2008, Gorgolewski 2008, Dunant et al. 2017). It is because the design team needs to put extra efforts to find the reused elements (Gorgolewski et al. 2008), and the design needs to remain as flexible as possible (Gorgolewski et al. 2008). Sometimes it is required to purchase the identified reused components early in the project (Gorgolewski et al. 2008, Gorgolewski 2008) to cope with the uncertainty about the timely availability of the desired elements (Gorgolewski et al. 2008, Chileshe, Rameezdeen, and Hosseini 2015). Consequently, this practice may raise cash flow problems and increase the overall cost of the project due to additional storage costs, which is another barrier to the BCR (Gorgolewski et al. 2008, Gorgolewski 2008, da Rocha and Sattler 2009, Yeung, Walbridge, and Haas 2015, Chinda and Ammarapala 2016) (among others).

All the above explain the increased labour cost (Klang, Vikman, and Brattebø 2003, Dantata, Touran, and Wang 2005, Shaurette 2006, Gorgolewski et al. 2008, Rameezdeen et al. 2016, Chinda and Ammarapala 2016) (among others), transportation cost (Gorgolewski et al. 2008, Gorgolewski 2008, da Rocha and Sattler 2009, Pongiglione and Calderini 2014, Yeung, Walbridge, and Haas 2015, Rameezdeen et al. 2016) (among others), and storage cost associated with deconstruction and reuse which are identified as barriers to the BCR in several articles.

In some cases, the fabrication cost of the recovered building components might be higher than the fabrication cost of the new elements (Dunant et al. 2017, Tingley et al. 2017, Dunant et al. 2018). Dunant et al. (2017) explain that because reused steel components are associated with existing connections, holes, stiffeners, welds, end-plates, etc., the preparation of these components might increase the overall cost of fabrication because of the extra time, labour and machinery required. Other additional charges which can increase the overall price of the recovered components are the cost of testing (Gorgolewski 2008, Yeung, Walbridge, and Haas 2015, Rameezdeen et al. 2016, Tingley et al. 2017, Dunant et al. 2018), cost of treatment of the salvaged parts (Chini and Acquaye 2001, Huuhka and Hakanen 2015, Dunant et al. 2018), cost of insurance (Tingley et al. 2017), and cost of marketing for the recovered building components (Dantata, Touran, and Wang 2005).

Another barrier to reuse, as reported in several articles, is the lack of an established market for the reused building components (Shaurette 2006, Gorgolewski et al. 2008, Gorgolewski 2008, Chileshe et al. 2016, Rameezdeen et al. 2016, Chinda and Ammarapala 2016, Chileshe, Rameezdeen, and Hosseini 2016) (among others). This factor, which is partially the outcome of the tight project schedules (Tatiya et al. 2017), results in the lack of sufficient supply for the reused components with the desired characteristics (dimension, quality, etc.) (Gorgolewski 2008, da Rocha and Sattler 2009, Dunant et al. 2017, Tingley et al. 2017, Brütting et al. 2019, Rose and Stegemann 2018). According to (Dunant et al. 2018), the above restriction encourages the contractors to sell their reusable waste to the recycling companies regardless of their high quality (Sansom and Avery 2014, Huuhka and Hakanen 2015, Yeung, Walbridge, and Haas 2015, Tingley et al. 2017, Yeung et al. 2017). If the demand for the reused building components increases (Chileshe et al. 2016), the market for these products can grow sustainably. In contrast, lack of demand (Shaurette 2006, Rogers 2011, Huuhka and Hakanen 2015, Chileshe et al. 2016, Tingley et al. 2017) or uncertainty about the need for the reused components (Rose and Stegemann 2018) causes the scepticism about the revenue from the reused components resale (Yeung, Walbridge, and Haas 2015, Chileshe, Rameezdeen, and Hosseini 2016, Rose and Stegemann 2018, Dunant et al. 2018). All the above negatively affects the chance for the growth of a reuse market. With an underdeveloped reuse market, the supply chain remains fragmented, and the information about the supply and demand cannot be shared, which further decreases the reuse rates (Gorgolewski et al. 2008, Rameezdeen et al. 2016, Rose and Stegemann 2018).

According to the literature, higher deconstruction costs can hinder its application (Dantata, Touran, and Wang 2005, Chileshe, Rameezdeen, and Hosseini 2015, Yeung, Walbridge, and Haas 2015, Tingley et al. 2017, Rose and Stegemann 2018, Dunant et al. 2018, Tatiya et al. 2017) and might elevate the financial risks associated with deconstruction and reuse (Rameezdeen et al. 2016). However, this finding is in contrast with the observations in (da Rocha and Sattler 2009). According to this study, in Brazil, the cost of deconstruction is lower than demolition due to the low cost of manual labour and the high demand for demolition products (da Rocha and Sattler 2009). In a separate study, Dantata et al. (2005) suggest that if the productivity of the deconstruction team increases or the wages decreases or the disposal cost rises, the overall cost of deconstruction decreases, and it becomes a desirable option in Massachusetts. Therefore, it can be concluded that the socio-economic context of the location of a building can convert some barriers to drivers and vice-versa.

2.3.2.2 Technical barriers

Ajayi et al. (2015) suggest that by integrating design for deconstruction (DfD) during the design stage of a building, recovery of building components for reuse would be facilitated. According to the literature, the lack of such intervention is a barrier to reuse (Chileshe, Rameezdeen, and Hosseini 2015, Huuhka and Hakanen 2015, Ajayi et al. 2015, Chileshe et al. 2016, Tatiya et al. 2017, Dunant et al. 2017) (among others). Some outcomes of this design gap are permanent joints (welding, etc.) (Gorgolewski 2008, Pongiglione and Calderini 2014, Tingley et al. 2017), composite joints (Tingley et al. 2017), and hard to access connections (Tingley et al. 2017), which can negatively affect deconstruction and make the recovery of the building components challenging (Huuhka et al. 2015).

Because deconstruction is not considered at the design stage, building components are prone to more damage during the deconstruction phase (Chini and Acquaye 2001, Gorgolewski 2008, Pongiglione and Calderini 2014). Damages to the reused building components can decrease the quality of the elements and affect their reusability (da Rocha and Sattler 2009, Durão et al. 2014, Huuhka and Hakanen 2015, Tatiya et al. 2017). Damages can also happen as the result of corrosion (Chini and Acquaye 2001, Huuhka et al. 2015, Yeung, Walbridge, and Haas 2015), post-production modifications (holes for ductwork, etc.) (Chini and Acquaye 2001, Yeung, Walbridge, and Haas 2015), presence of water (Yeung, Walbridge, and Haas 2015, Tatiya et al. 2017), exposure to weather conditions (Huuhka and Hakanen 2015), fire (Yeung, Walbridge, and Haas 2015, Tatiya et al. 2017), during refurbishment (nail removal, etc.) (Chini and Acquaye 2001), by the living organisms (termite, bacterial attack, etc.) (Chini and Acquaye 2001), fatigue (Yeung, Walbridge, and Haas 2015), frost (Huuhka et al. 2015), degradation (Durão et al. 2014), type of joints (Gorgolewski 2008), during storage and transportation of the recovered components (Gorgolewski 2008), and due to impact (Yeung, Walbridge, and Haas 2015), etc.

Difficulty in designing with the reused components is another barrier to the widespread reuse of the building components (Gorgolewski et al. 2008, Pongiglione and Calderini 2014, Tingley et al. 2017, Brütting et al. 2019). As discussed earlier, the design of the new buildings with reused building components needs to remain flexible. It is because the design should be able to accommodate alternative dimensions of the reused components due to the uncertainty in the availability of the desired sections (Gorgolewski et al. 2008, Gorgolewski 2008). Brütting et al. (2019) argue that unlike structures made out of new steel sections where components with different cross-sections and lengths can be fabricated to the required shape, in the case of the reused steel sections, this luxury doesn't exist and the properties of the available components dictate the structure geometry.

Pongiglione and Calderini (2014) discuss that in the process of designing a new structure using the recovered components, due to architectural and structural reasons, new structural elements should be used as well. However, to secure the safety of such structures, the new components should be over-dimensioned, which eventually results in overdesigned structures (Gorgolewski et al. 2008, Gorgolewski 2008, Pongiglione and Calderini 2014, Brütting et al. 2019). It is either because of the lower strength of the reused components or when the remaining capacity of the reused components is unknown (Huuhka and Hakanen 2015, Yeung, Walbridge, and Haas 2015). The latter happens when the information about the characteristics, details, certificates, and drawings of the reused components are not available (Gorgolewski et al. 2008, Gorgolewski 2008, Huuhka and Hakanen 2015, Yeung, Walbridge, and Haas 2015, Tingley et al. 2017, Rose and Stegemann 2018). Other design challenges while reusing recovered building components are designing with long spans (because such elements might not be readily available) (Gorgolewski et al. 2008, Huuhka and Hakanen 2015, Brütting et al. 2019), the difference in the loading requirements of the old and the new buildings (Gorgolewski et al. 2008), and the mismatch between the old spans and the new features (Huuhka and Hakanen 2015).

Additional health and safety precautions necessary for deconstruction, component recovery, and reuse are some other technical barriers to reuse (Sansom and Avery 2014, Chileshe, Rameezdeen, and Hosseini 2015, Huuhka and Hakanen 2015, Yeung, Walbridge, and Haas 2015, Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016, Tingley et al. 2017). It is because, during the deconstruction of a building, or while treating a component for reuse, there is a risk of encountering hazardous, banned or contaminating coatings on the reused components (Rameezdeen et al. 2016, Tatiya et al. 2017, Tingley et al. 2017). In case of facing hazardous materials such as lead or asbestos, specific procedures and licensed contractors are required (Rameezdeen et al. 2016).

2.3.2.3 Social barriers

The negative perception of the stakeholders about the reused building components can act as a barrier to reuse (MacKinnon 2000, Klang, Vikman, and Brattebø 2003, Chileshe, Rameezdeen, and Hosseini 2015, Huuhka and Hakanen 2015, Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016) (among others). One reason behind this is the visual appearance of the reused components that might be interpreted as lower quality when compared with a new element (Durão et al. 2014, Tingley et al. 2017, Dunant et al. 2017). For instance, Durão et al. (2014) report that the architects refuse to use recovered wood in visible places due to its poor appearance. However, the visual appearance can be a point of further discussion since it is highly subjective and can be attractive to some people (Nußholz, Nygaard Rasmussen, and Milios 2019). Another reason behind this negative perception, and at a larger scale the construction sector resistance against reuse (Gorgolewski 2008, Durão et al. 2014, Rameezdeen et al. 2016, Tingley et al. 2017), stems from the potential risks perceived by the stakeholders during deconstruction or while using the recovered building components (Shaurette 2006, Gorgolewski 2008, Chileshe, Rameezdeen, and Hosseini 2015, Rameezdeen et al. 2016, Dunant et al. 2017, Tingley et al. 2017).

The occupational health concerns (Klang, Vikman, and Brattebø 2003, Rameezdeen et al. 2016), liability and fear (da Rocha and Sattler 2009), lack of trust in the supplier of the reused components (Dunant et al. 2017, 2018), and unsatisfactory working environment during the treatment of the reused components (Klang, Vikman, and Brattebø 2003) can all worsen the lack of interest to integrate the reused components in the projects (Chileshe et al. 2016, Rameezdeen et al. 2016). Among the stakeholders, perception of clients (da Rocha and Sattler 2009, Chileshe, Rameezdeen, and Hosseini 2015, Dunant et al. 2017, Rose and Stegemann 2018), contractors (Shaurette 2006, Gorgolewski 2008), and designers (Gorgolewski 2008) have a higher impact on the successful integration of recovered components into a new building. However, if the client does not support reuse (Huuhka and Hakanen 2015, Rameezdeen et al.

2016, Tingley et al. 2017, Rose and Stegemann 2018), there is very little chance that designers or contractors risk the project by introducing such components. On the other hand, according to (Gorgolewski 2008), if the client is motivated to use the reused building components, the barriers such as the unwillingness of the design team (Chileshe, Rameezdeen, and Hosseini 2015, Rameezdeen et al. 2016) or the contractors (Gorgolewski 2008) can be handled effectively. Nevertheless, the inequality in the distribution of risk among the stakeholders (Dunant et al. 2018) can yet challenge motivated clients and architects.

Gorgolewski (2008) argues that while choosing deconstruction to remove the existing buildings improves the supply of the reused components, due to the perceived economic and programming reasons, it is not yet a preferred option among the contractors (Gorgolewski 2008). One reason for such reluctance is because the stakeholders are unaware of the full benefits of deconstruction and reuse (Gorgolewski 2008, Chileshe, Rameezdeen, and Hosseini 2015, Huuhka and Hakanen 2015, Chileshe et al. 2016, Rameezdeen et al. 2016). As mentioned earlier, some of the benefits of deconstruction and reuse are cost savings and less pollution to the environment. Therefore, educating the stakeholders on the advantages of deconstruction and reuse, as identified by (Gorgolewski 2008, Chileshe, Rameezdeen, and Hosseini 2015), could be an effective measure to cope with some social resistance against reuse.

2.3.2.4 Regulatory barriers

One of the challenges ahead of reuse is that the existing regulations do not support deconstruction and reuse (Gorgolewski 2008, Hglmeier et al. 2013, Chileshe, Rameezdeen, and Hosseini 2015, Huuhka and Hakanen 2015, Huuhka et al. 2015, Chileshe et al. 2016, Rameezdeen et al. 2016) (among others). Rameezdeen et al. (2016) argue that bureaucracy is a barrier ahead of necessary approvals for deconstruction projects in South Australia. According to this study, even after getting approvals for deconstruction, since existing regulations do not allow the storage of the salvaged components and consider them as waste (Rameezdeen et al. 2016), the reuse of the recovered components is hindered. This study suggests that governments should support the reuse of recovered components in the new constructions (Rameezdeen et al. 2016); however, in reality, it is not the case (Chileshe et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016). Rameezdeen et al. (2016) further discuss that, while regulations support recycled-content products, due to the inconsistency and the lack of coordination among the regulatory bodies (Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016), regulatory agencies have a prohibitive approach towards deconstruction and reuse. It should be noted that these studies focus on the Australian construction sector, and the results should be considered cautiously (Chileshe et al. 2016, Rameezdeen et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016).

Lack of quality certificates for the reused components can negatively affect reuse (Chini and Acquaye 2001). Dunant et al. (2017) explore this barrier by highlighting the need for the traceability of the steel sections, which is essential to certify, fabricate, and erect the segments. Usually, the traceability of the reused steel sections cannot be guaranteed (Dunant et al. 2017, Tingley et al. 2017), and in many instances, all the segments need to be tested to certify their properties and assure the quality. However, according to this study, in case of stricter requirements on CE marking (Dunant et al. 2017, Tingley et al. 2017), even the individual testing fails to certify the reused components.

Lack of confidence in the quality of the reused components negatively affects reuse in new constructions (Shaurette 2006, Chileshe, Rameezdeen, and Hosseini 2015, Ajayi et al. 2015, Chileshe et al. 2016, Chileshe, Rameezdeen, and Hosseini 2016) (among others). Huang et al. (2018) observed that there is a negative attitude towards using recovered construction and demolition waste among the building construction companies because of the lack of guarantees for these components. According to the reviewed articles, currently, there are no standards to certify the quality of the reused components (Chini and Acquaye 2001, Dunant et al. 2017, Huang et al. 2018). Therefore, the lack of procedures to evaluate and guarantee the performance of reused components (Shaurette 2006, Tingley et al. 2017), and the fact that the existing codes, standards, and procedures do not consider BCR (Gorgolewski 2008, Huuhka and Hakanen 2015, Rameezdeen et al. 2016, Tingley et al. 2017) further decrease the reuse rate in buildings.

Table 2.2 Summary of reuse barriers

S N	Reference	Cntr.[a]	Research method[b]	Categories of reuse barriers[c]																			
				Economic			Env	Organisational				Regul.		Social						Technical			
				A: Cost	B: Market	C: Value for money	D: Energy and GHG	F: Contracts	G: Experience	H: Infrastructure	J: Management	L: Compliance	M: Incentive	O: Awareness	P: Perception	Q: Risk	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design	W: Health and safety	X: Information
1	(MacKinnon 2000)	US	DR; GI; I(4); OBS	1											1								
2	(Chini and Acquaye 2001)	US	EX	1								2									5		
3	(Klang, Vikman, and Brattebø 2003)	US	CS(1); I(10); Q(10/10)	1											1	1	1						
4	(Dantata, Touran, and Wang 2005)	US	CS(5); LIR	4																			
5	(Pun and Liu 2006)	AU	TF		3																		
6	(Pun, Liu, and Langston 2006)	AU	CS(1)	4	3							1		1								1	
7	(Shaurette 2006)	US	Q(296/83)	3	2				1	2		1			1	1							
8	(Guy 2006)	US	CS(4)	4								2									5	1	
9	(Gorgolewski et al. 2008)	CA	AR; CS(3)	8	2						1										5		1
10	(Gorgolewski 2008)	CA	AR; CS(2)	6	2				1	1		2		1	3	1			2	1	5		1
11	(da Rocha and Sattler 2009)	BR	CD; CS(1); DO(5); GM(4); SSI(27)	2	1							2			1	1					1		
12	(Nordby et al. 2009)	NO	CS(1)	2								1					1			2	1		
13	(Jaillon and Poon 2010)	HK	AR; CS(7); DO(7); I(35); Q(84)																	1			
14	(Rogers 2011)	AE	CS(1)		1																		
15	(Forsythe 2011)	AU	CS(9); DO; UI	3		1															1	2	
16	(Arif et al. 2012)	IN	CS(2); SSI(15)								2		1										
17	(Coelho, de Brito, and Brito 2012)	PT	CS(15)						1														
18	(Elias Özkan 2012)	TR	AR; CS; DO(21); I							2		1								1			
19	(Hglmeier et al. 2013)	DE	CS(1)									1											
20	(Gangolells et al. 2014)	ES	Q(658/74)						1														
21	(Sansom and Avery 2014)	GB	Q(160/32)	2																		1	
22	(Jaillon and Poon 2014)	HK	CS(2); LIR																	2			
23	(Pongiglione and Calderini 2014)	IT	AR; CS(1)	1																1	3		
24	(Durão et al. 2014)	PT	CS(2)												1				1		2		
25	(Chileshe, Rameezdeen, and Hosseini 2015)	AU	LIR; Q(539/49); S	4					1		1	2		1	3	1			2	1		1	
26	(Ferreira, Duarte Pinheiro, and De Brito 2015)	PT	CS(1); LIR																		2		
27	(Huuhka and Hakanen 2015)	FI	Q(11/11)	3	2		1					5		1	1	1			1	1	3	1	2
28	(Huuhka et al. 2015)	FI	AR(276); LIR									1								1	2		
29	(Yeung, Walbridge, and Haas 2015)	CA	DO(4)	6		1			1		1		1								5	1	2
30	(Ajayi et al. 2015)	GB	FGI(25); LIR													1				1			
31	(Cooper et al. 2016)	GB	CS(2); LIR; SSI(17)	5																			
32	(Rameezdeen et al. 2016)	AU	SSI(8)	9	2							5		2	1	2			4			2	
33	(Chileshe et al. 2016)	AU	LIR; Q(539/49); SSI(6)		2				2	1		3	2	3					1	1			
34	(Chinda and Ammarapala 2016)	TH	CS(2); I(6); LIR	4	1					2													
35	(Chileshe, Rameezdeen, and Hosseini 2016)	AU	LIR; SSI(8)	4	1	1						2	1		1	1						1	
36	(Tatiya et al. 2017)	US	CS(1); LIR; SI(3)	5	1															1	2	1	

SN	Reference	Cntr.[a]	Research method[b]	Categories of reuse barriers[c]																			
				Economic			Env	Organisational				Regul.		Social						Technical			
				A: Cost	B: Market	C: Value for money	D: Energy and GHG	F: Contracts	G: Experience	H: Infrastructure	J: Management	L: Compliance	M: Incentive	O: Awareness	P: Perception	Q: Risk	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design	W: Health and safety	X: Information
37	(Dunant et al. 2017)	GB	I(30); Q(24)	5	2				1	1		6			2	1		1		1			
38	(Tingley et al. 2017)	GB	LIR; SSI(13)	9	3	1		1				6	2	1	2	1			2	3	1	3	1
39	(Yeung et al. 2017)	CA	CS(1)	2																1			
40	(Machado, de Souza, and Veríssimo 2018)	BR	LIR							1										1	3		
41	(Gálvez-Martos et al. 2018)	EU	CA		2																		
42	(Huang et al. 2018)	CN	CD; LIR; SSI(40)	1	1							2											
43	(Brütting et al. 2019)	CH	CS(2)		1																3		
44	(Sea-Lim et al. 2018)	TH	SD	2						1													
45	(Rose and Stegemann 2018)	GB	CD; CS(6); DO; SSI(21)	3	4	1				1	2	1	1		1				1	1			1
46	(Dunant et al. 2018)	GB	I(30)	9	1	1			1	1	1					2		1					
47	(Mahpour 2018)	IR	LIR; Q(6/6)												1								
48	(Zaman et al. 2018)	NZ	CS(1)	1		1			1	1		1											
49	(Nußholz, Nygaard Rasmussen, and Milios 2019)	DK	CS(3); Q(3); SSI(3)	1	3		1			1	1	2	1						1				
50	(Brambilla et al. 2019)	GB	CS(1)				2													1			
51	(Basta, Serror, and Marzouk 2020)	EG	CS(1); TF																	2	1		
			Total number:	115	40	7	4	1	11	15	9	49	9	10	20	14	2	2	15	24	50	15	8

[a] ***Country:*** *According to ISO 3166*

[b] ***Research Method:*** *(ACT) Action Research (n = number of case(s), if provided); (AR) Archival research (n = number of case(s), if provided); (CA) Comparative analysis; (CD) Company documentation; (CS) Case study (n = number of case(s)); (DO) Direct observation (n = number of case(s)); (DR) Document review; (ECOM) Economic models; (EX) Experiment; (FGI) Focused-group interview (n = number of interviewee(s)); (FS) Field study; (GI) Group Interview; (GM) Group meetings (n = number of attendant(s)); (I) Unspecified type Interviews (n = number of interviewee(s)); (LIR) Literature review; (OBS) Observation; (Q) Questionnaire (n = number of sent Q / m = number of completed Q); (S) Survey (i.e. empirical survey, etc.); (SD) System dynamics; (SI) Structured interviews (n = number of interviewee(s)); (SSI) Semi-structured interviews (n = number of interviewee(s)); (T) Theoretical study; (TF) Theoretical framework; (UI) Unstructured interview*

[c] *The numbers in the table corresponds to the number of drivers grouped under each sub-category.*

2.3.2.5 Organisational barriers

Because deconstruction and reuse are still uncommon practices (Dunant et al. 2017, 2018), the number of companies with experience in deconstruction and reuse is low (Chileshe et al. 2016). According to the literature, the lack of skills, experience, and knowledge in deconstruction, salvage, and using reused components negatively affect the reuse of the building components (Shaurette 2006, Gorgolewski 2008, Chileshe, Rameezdeen, and Hosseini 2015, Yeung, Walbridge, and Haas 2015, Chileshe et al. 2016). Unlike demolition, deconstruction requires enough space for the storage, sorting, and treatment of the recovered building components. However, an inexperienced contractor cannot correctly estimate the space required for the storage of the recovered components after deconstruction. This lack of space for storage (Shaurette 2006, Gorgolewski 2008, Chinda and Ammarapala 2016, Dunant et al. 2017, Rose and Stegemann 2018, Dunant et al. 2018) results in the transportation and storage of the recovered components at a different location and would increase the overall cost of the reused elements.

Lack of systems thinking (Rose and Stegemann 2018), ownership (Arif et al. 2012), and integration of reuse in the design process of the new projects (Rose and Stegemann 2018) are identified to decrease the reuse rates in the building sector. Yeung et al. (2015) highlight the importance of a decision-making framework in informing the contractors and the client regarding when alternative reuse options should be investigated. According to this study, this decision-making framework helps in making informed decisions about deconstruction and reuse and maximises the advantages of potential reuse by identifying the necessary steps to be taken by the stakeholders (Yeung, Walbridge, and Haas 2015). Other observed organisational barriers are proprietary lock-in (Tingley et al. 2017), the need for infrastructure and equipment to perform deconstruction (Shaurette 2006, Chileshe et al. 2016, Sea-Lim et al. 2018), and inconsistency in waste management practices (Arif et al. 2012).

2.3.2.6 Environmental barriers

While component reuse is identified as a sustainable end-of-life treatment of the superstructure of a building (Klang, Vikman, and Brattebø 2003, Tingley et al. 2017, Yeung et al. 2017, Brütting et al. 2019), there are concerns regarding the adverse effects of this practice due to increased GHG emissions related to deconstruction activities and transportation of the recovered elements (Brambilla et al. 2019, Nußholz, Nygaard Rasmussen, and Milios 2019, Huuhka and Hakanen 2015).

Brambilla et al. (2019) performed a study to evaluate the environmental impacts of various steel-concrete composite floor systems. In this study, the authors performed a comparative LCA and compared the four composite connections, including a novel demountable steel-concrete composite floor system and three conventional systems. The authors concluded that a transport distance between 20 km and 200 km has no significant impact on environmental advantages achieved by the demountable system. However, they concluded that a distance of 1000 km could diminish the environmental benefits achieved by this system. The authors also discussed that deconstruction of the demountable composite structure takes more time compared to demolition, which results in the emission of higher amounts of GHGs since the heavy machinery and equipment need to operate longer (Brambilla et al. 2019).

2.4 Prioritising reuse barriers

Previous observations in Section 2.3.2 provide an insight into the challenges ahead of component reuse in the building sector; however, prioritising them needs a further investigation about the inter-dependency of these factors. Reviewing the co-occurrences of data is a way to identify the impact of various variables of a research topic on one another and to reveal their potential correlations. And identifying the correlation between the key variables helps in better devising solutions to achieve the objectives of the research (Rameezdeen et al. 2016, Eck and Waltman 2009). The next section develops the co-occurrence of all the 19 sub-categories available in Table 2.2 to analyse the relationship between identified barriers.

2.4.1 Co-occurrence of reuse barriers

In this section, a binary approach for the presence (1) or the absence (0) of the sub-category of barriers in Table 2.2 is considered to identify their co-occurrences and eventually develop their correlations. It means that if in Table 2.2, under a particular sub-category for a specific paper, no barrier is observed, value 0, which means absence, is considered. On the other hand, the available observations (regardless of their number) are converted to 1.

Table 2.3 shows the co-occurrence of the sub-categories of reuse barriers in the reviewed articles. For example, sub-category A & sub-category B (AB) appear 15 times together in all the articles reviewed in this chapter. To analyse the correlation between the sub-categories, the researcher also developed the co-occurrence index (C-Index) of the pairs of the sub-categories. In this section, the C-Index is calculated using the software "R" (R Core Team 2020) through the "jaccard" package (Chung et al. 2018), which is based on Eq. (2.1) (Atlas.ti 2014). In Eq. (2.1), n_{12} is the co-occurrence frequency of the two sub-categories (the number of times the two sub-categories show up together; hence is not equal to $n_1 + n_2$), and n_1 & n_2 are the total numbers of occurrences of each of the sub-categories in all the studies. C-Index varies from 0 to 1, with 1 showing the highest correlation and 0 indicating no relationship. The null hypothesis is that there is no correlation between the pairs of the sub-categories. To test the null hypothesis, the p-value through the embedded test in the "jaccard" package (jaccard.test.exact) is used (Chung et al. 2018). If the p-value is less than 0.05, then the null hypothesis is false, and statistically, there is a correlation between the pairs of the sub-categories (James et al. 2017).

$$C - Index = \frac{n_{12}}{(n_1 + n_2) - n_{12}} \qquad (2.1)$$

In Table 2.3, the highlighted cells represent the high levels of co-occurrence between the sub-categories. The corresponding C-Index of these pairs of sub-categories of the barriers are sorted and listed in Table 2.4. Also, the p-value, which indicates if the correlation is significant or not (James et al. 2017), is listed against each of the pairs.

According to Table 2.4, there is a significant correlation between perception and risk, with the C-Index of 0.63, ranking the highest among other sub-categories. It indicates that the perception of the stakeholders about reuse is affected by the potential risks associated with this intervention. Perception co-occurs with compliance, cost, and market, as well (all are significant with p-values 0.004, 0.02, and 0.02, respectively). It reveals the importance of addressing the economic and regulatory obstacles to promote reuse among the stakeholders. The second and third highest ranks belong to the cost and compliance as well as market and compliance, with the C-Indices of 0.49 and 0.45, respectively. It shows that an established reuse market requires to offer products at reasonable prices complying with state-of-the-art codes and regulations. On the other hand, the existence of ordinances, as well as the best practices on the reused components, would help the growth of a reuse market.

Table 2.3 Co-occurrence of sub-categories of reuse barriers

Code	Economic			Env	Organisational				Regul.		Social						Technical			
	A: Cost	B: Market	C: Value for money	D: Energy and GHG	F: Contracts	G: Experience	H: Infrastructure	J: Management	L: Compliance	M: Incentive	O: Awareness	P: Perception	Q: Risk	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design challenges	W: Health and safety	X: Information
A	-	16	7	2	1	7	9	6	17	5	6	12	11	2	2	7	10	12	11	6
B		-	4	2	1	5	8	4	13	5	6	9	9	0	2	7	7	7	6	5
C			-	0	1	3	3	3	4	4	1	3	3	0	1	2	2	3	4	3
D				-	0	0	1	1	2	1	1	1	1	0	0	2	2	1	1	1
F					-	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1

Code	Economic			Env	Organisational				Regul.		Social						Technical			
	A: Cost	B: Market	C: Value for money	D: Energy and GHG	F: Contracts	G: Experience	H: Infrastructure	J: Management	L: Compliance	M: Incentive	O: Awareness	P: Perception	Q: Risk	R: Sustainability	S: Trust	T: Willingness	U: Deconstruction	V: Design challenges	W: Health and safety	X: Information
G						-	6	3	6	2	3	4	5	0	2	3	4	2	2	2
H							-	3	8	3	2	4	4	0	2	4	6	2	0	2
J								-	3	4	1	2	2	0	1	3	2	2	2	3
L									-	5	7	10	9	1	1	8	10	8	7	4
M										-	2	3	2	0	0	4	3	2	3	3
O											-	5	5	0	0	6	5	3	5	3
P												-	10	1	1	7	6	5	5	4
Q													-	1	2	5	6	4	5	3
R														-	0	0	1	1	0	0
S															-	0	1	0	0	0
T																-	6	4	4	4
U																	-	9	4	4
V																		-	6	5
W																			-	3

Table 2.4 C-Indices of the correlation between major sub-categories.

Seq. No	Code	Sub-category pair	C-Index	P-value
1	PQ	Perception & Risk	0.63	<0.00001*
2	AL	Cost & Compliance	0.49	0.007*
3	BL	Market & Compliance	0.45	0.006*
4	AB	Cost & Market	0.44	0.04*
5	LP	Compliance & Perception	0.40	0.004*
6	BQ	Market & Risk	0.38	0.004*
7	LQ	Compliance & Risk	0.38	0.004*
8	AP	Cost & Perception	0.36	0.02*
9	AW	Cost & Health and safety	0.35	0.001*
10	BP	Market & Perception	0.35	0.02*
11	AQ	Cost & Risk	0.34	0.007*
12	LU	Compliance & Deconstruction	0.33	0.2
13	AV	Cost & Design challenges	0.32	0.5
14	UV	Deconstruction & Design challenges	0.32	0.1
15	AH	Cost & Infrastructure	0.26	0.2
16	AU	Cost & Deconstruction	0.25	0.4
*Denotes a significant correlation (less than 0.05)				

The fourth highest rank belongs to cost and market with a C-Index of 0.44. It indicates that without a competitive price, a well-established market for reused elements is unlikely to grow.

Moreover, it depicts that the growth of the reused components market can help to make the cost of reused components more competitive. However, the correlation between these two sub-categories is not very significant (p-value close to 0.05). It is interesting because, in most of the reviewed papers, both sub-categories are repeated. It can be further interpreted that these two sub-categories are similar, and no special consideration for prioritising this pair is required as the improvement in one promotes the other one.

From Table 2.4, it can be observed that the social, economic, and regulatory barriers co-occur frequently. Therefore, it seems that any further action to promote reuse should prioritise actions to be taken under these themes. Notwithstanding, this result is different from the initial observation in Figure 2.6, where the economic factors were ranked the highest, followed by the technical, social, regulatory, and organisational barriers.

2.4.2 Discussion

The observed environmental advantages of reuse indicate that this intervention is an effective strategy that should receive more attention to reduce the environmental footprint of the building sector.

From an economic perspective, the advantages of reuse in terms of cost savings and profit are key drivers. According to the reviewed articles, economic barriers can be categorised into supply chain level, component level, and project level. At the supply chain level, in the absence of a mature reuse market, the sustainable supply of recovered components for use in the superstructure of a building is challenging. While some innovative companies such as Gamle Mursten in Denmark integrate deconstruction into their core business (Nußholz, Nygaard Rasmussen, and Milios 2019), most companies are reluctant to change their business model. Hence, as advised by (Dunant et al. 2018, Nußholz, Nygaard Rasmussen, and Milios 2019), close cooperation between construction and demolition companies can address this barrier. At the component and project levels, a strict financial risk assessment at the beginning of the project should be performed. Because this intervention is rather new, the availability of resources to decrease the financial risks would be helpful (Gorgolewski 2008, Tingley et al. 2017). Such financial incentives have the potential to promote deconstruction and reuse activities and could help the growth of reuse markets, and potentially make the price of the recovered elements more competitive (Table 2.4).

Notwithstanding, other attempts could be made to make the cost of the recovered components competitive. One possible solution is following the successful example of increasing the landfilling tax in the UK (Defra 2007, 2019). Considering the waste hierarchy, if the cost of other waste treatment options increases in favour of reuse, the additional costs due to deconstruction, treatment, and testing could be compensated. However, there are reports of illegal landfilling in reaction to the increased landfilling taxes (da Rocha and Sattler 2009, Rameezdeen et al. 2016). Therefore, further research in different geographical locations should be conducted to recognise the mechanisms leading to such behaviour and provide guidelines to prevent it.

From a social perspective, the factors affecting reuse can be categorised into perception, awareness, and risks. Most of the discussions in the literature from a social perspective are focused on the perception and willingness of the stakeholders regarding reuse and are less focused on the advantages of reuse for the general public. Therefore, further research should be conducted to establish the benefits of reuse for society. Nevertheless, the negative perception of the stakeholders towards reuse is recognised in the literature as an impediment to its adoption in the building sector. Based on Table 2.4, this negative perception is associated with the perceived risks at different stages of projects with recovered building components as well as the need for compliance to the regulatory requirements and is fuelled by the concerns about the health and safety of the stakeholders. Therefore, steps should be taken to improve the perception of the stakeholders about the recovered building components. For instance, the development of standard test procedures to test, evaluate, and certify the recovered building components can positively contribute to this attempt. Such standards and guidelines can address the reported concerns and resistances in the construction sector against the recovered building components and help the growth of a reuse market by offering quality products.

The regulatory barriers can be categorised into incentive level and compliance level, for which, the advantages of the availability of regulatory incentives were discussed earlier. At the policy level, the reported regulatory barriers highlight that the existing codes and regulations do not consider deconstruction and reuse, which, in the long run, inhibits the integration of the recovered building components in the superstructure of the buildings. Moreover, as discussed earlier, the existing standards only certify new components and not the recovered elements. According to Section 2.4.1, the capability of suppliers in offering second-hand components with proper quality certificates and guarantees could potentially help the growth of a reuse market (Table 2.4). In this regard, one possible solution is the development of new standards to certify recovered building components. An example of the successful development of certifying standards is provided by (Nußholz, Nygaard Rasmussen, and Milios 2019). In this study, the case study companies developed certifying standards to assure the quality of their products. Moreover, proper standards and procedures should be developed for the effective deconstruction of the existing buildings and guide designers to integrate the recovered building components into the design of new buildings. Because of the variety of building designs in different periods and locations, proper databases for the existing buildings should be developed to assist such guidelines. These databases should contain the historical reports for each building, including the refurbishments, fire, extensions, and potential end-of-life treatment plans.

According to the literature, the advantages of reuse in reducing the CDW and increasing the competitiveness of the firms are key organisational drivers. However, most of the companies in the building sector do not have enough experience in deconstruction and reuse, which results in following other end-of-life treatment options such as demolition and recycling. Therefore, companies should take necessary actions to train the workforce to improve the productivity of their deconstruction activities and increase the reusability of the recovered building components. As discussed earlier, one possible driver to encourage companies to change their business model is the availability of regulatory incentives. However, further research should be performed to analyse the driving forces, which would help companies to integrate circularity in their business models.

The technical barriers can be categorised into deconstruction level, performance level, and health and safety level. As observed in the reviewed literature, at the deconstruction level, the biggest challenge to recover building components is that buildings are not designed for deconstruction. While innovative design techniques can address this barrier in new buildings, it remains a significant challenge ahead of deconstruction of the existing built stock. At the performance level, one of the barriers to the reuse of building components after recovery is the reusability of the element (due to damages, availability of information, design challenges, etc.). According to the definition of reuse, reusability can be defined as the extent to which the recovered building component in its new life could perform similarly to its earlier life. It is because most of the existing buildings are not designed for deconstruction, details about the existing buildings are unavailable, and proper guidelines and skills for effective deconstruction do not exist. As mentioned earlier, deconstruction can increase the reuse rate; however, there is no available guideline to help the practitioners to estimate the reuse potential of the building components before deconstruction. Therefore, further research to develop cheap and reliable techniques to investigate the reusability of building components is necessary. Moreover, while the DfD is identified as a solution to the end-of-life treatment of buildings, this design method is based on new building components. Hence, further research should be conducted to integrate the recovered building components into this design technique. At the health and safety level, as observed in Table 2.4, there is a strong correlation between cost and health and safety requirements of a project with deconstruction and reuse. It indicates that the increased health and safety precautions necessary for deconstruction and reuse activities (as the result of the presence of hazardous materials, etc.) could potentially increase the overall cost of the project.

2.5 Chapter summary

Chapter 2 fulfilled the first objective of this research by identifying factors affecting the reuse of load-bearing building components through a systematic literature review. Initially, a Boolean search focused on peer-reviewed articles in top-tier journals was performed in Scopus to identify the papers for review. This stage resulted in identifying 76 journal papers. Since these papers are derived from top-tier construction journals, they represent the state-of-art in the body of knowledge. Next, these papers were scrutinised to identify the factors affecting reuse. In total, 57 drivers and 130 barriers were recognised in these articles. Consequently, these factors were classified into six major categories and twenty-three sub-categories. Then, the inter-dependencies between the barriers were studied by developing the correlation indices between the sub-categories. Results indicate that addressing the economic and social barriers should be prioritised. According to this chapter, the impact of barriers under perception, risk, compliance, and market sub-categories are very pronounced. However, perception and risk show the highest inter-dependency among the sub-categories of variables. This observation suggests that the stakeholders' perceptions are affected by the potential risks of reusing load-bearing building components.

Chapter 3 – Research methodology

3.1 Chapter introduction

This research aims to develop a model that can predict the reuse potential of the structural elements at the end-of-life of a building. Therefore, it is essential to identify what information is required, plan to collect them, and eventually analyse the collected data to fulfil the aim of the study through the development of the first Building Structural Elements Reusability Predictive Model (BSE-RPM). Crotty (1998) emphasises the need to develop a research process that fulfils the aim of the research and answers the associated research questions. However, because research is done by human beings, the approach adopted by the researcher is inevitably affected by the assumptions he/she makes throughout the process of knowledge development (Burrell and Morgan 2016). These assumptions, which are affected by the researcher's knowledge, values and beliefs, form his/her conception about the nature of being (ontological assumptions (*Oxford English Dictionary* n.d.)), reflect his/her understandings about the nature, limitation and validity of knowledge (epistemological assumptions (*Merriam-Webster Dictionary* n.d.)), and explain the extent the researcher believes his/her values should and might affect the research process (axiological assumptions (Saunders, Lewis, and Thornhill 2016)). It is these assumptions and beliefs that shape the researcher's theory for the study at hand and consequently form his/her philosophy of how to performing the research (Crotty 1998). The research philosophy then shapes the research strategies (Crotty 1998), which lead the researcher to his/her choices of data collection techniques and analysis approaches (Crotty 1998), and eventually, develops his/her research design that is coherent at all the stages (Saunders, Lewis, and Thornhill 2016).

This chapter, therefore, discusses the philosophical assumptions underpinning the research (Section 3.2), introduces the strategies (or methodologies) and methods adopted (Sections 3.3 & 3.4, respectively), and elaborates the reasons behind these decisions (Crotty 1998). It is noteworthy that the path followed in this chapter is in the reverse order of the approach suggested by Crotty (1998). However, it serves the same purpose and makes this study sound and credible.

3.2 Research Philosophy

The research philosophy is the backbone of the methodological choice(s) a researcher makes to conduct research (Crotty 1998). Therefore, justifying the choice of the research philosophy among the available alternatives becomes a crucial aspect of any study (Crotty 1998, Johnson and Clark 2006). Moreover, as explained earlier, the research philosophy is the reflection of the ontological, epistemological, and axiological assumptions made by the researcher at the onset of the research (Crotty 1998). The combination of these philosophical assumptions underlies the theoretical perspective (or paradigm) of a study which represents "the frame of reference, mode of theorising and ways of working in which a group operates" (Burrell and Morgan 2016). Therefore, this section initially discusses the theoretical assumptions of the study, then it introduces different research philosophies and evaluates their suitability to this study, and eventually justifies the selected research philosophy based on the assumptions made.

3.2.1 Theoretical assumptions

In the approach to performing research, a researcher has certain ontological assumptions about the nature of the events under investigation (Burrell and Morgan 2016). Ontology is the science or study of being (Crotty 1998, Burrell and Morgan 2016). According to the Oxford English Dictionary (*Oxford English Dictionary* n.d.), it is a "branch of metaphysics concerned with the nature or essence of being or existence". Ontology means whether there exists a unique and generalisable reality or different realities that are socially constructed coexist (Patton 2002). Burrel and Morgan (2016) explain the challenge ahead of a social scientist to distinguish if the "reality" under investigation is external to the individual, i.e., it is real, objective, and exists regardless of personal awareness, or it is the result of the social consciousness or perceptions and therefore is relative.

In this study, to understand the underlying ontological assumptions of the research, it is necessary to have a thorough understanding of the status of the reuse of building elements in the body of knowledge. According to the literature, different studies use different approaches to analyse why the reuse of building elements is not a widespread practice. These studies list numerous factors in terms of drivers and barriers (Section 2.3) and discuss that by addressing specific barriers and providing proper incentives reuse rate of the building elements could be increased. On the other hand, some studies suggest that without radical changes in the design of the buildings (e.g. integration of interventions such as Design for Deconstruction (DfD), Design for Manufacture and Assembly (DfMA), etc.), the reuse of building structural elements (RBSE) cannot widespread (Iacovidou and Purnell 2016, Akinade et al. 2017, Tingley and Davison 2011, Kalyun and Wodajo 2012). As discussed in Section 1.1, these design features are only suitable in the case of new buildings and fail the existing stock of buildings that do not have such interventions at their core. Moreover, several successful case-study buildings with reused building structural elements show that the RBSE is a reality and can happen even in the case of buildings not designed for deconstruction (Section 1.1). The above discussion yield an assumption that there exists another approach (a reality) which have never been embarked on and can promote the RBSE within the scope of the existing buildings. Based on the above discussion, it can be concluded that the “reality” under investigation, i.e., an alternative approach to determine the reusability of the building structural elements at any time, is external to the individuals and is “real”. However, there might be alternative approaches to achieve this reality.

The above ontological discussion is accompanied by a set of epistemological assumptions (Burrell and Morgan 2016). Epistemology, which is “the study or a theory of the nature and grounds of knowledge especially with reference to its limits and validity” (*Merriam-Webster Dictionary* n.d.), reveals the assumptions about the nature of knowledge and truth, and how this knowledge can and should be conveyed to the peer, and at a larger scale, to the world (Burrell and Morgan 2016). The researcher’s epistemologies, which reflect his/her ontological assumptions, discuss the reliable sources of knowledge and excavates if the knowledge is objective and can be collected by correct tools, or is subjective, and needs to be experienced (Burrell and Morgan 2016, Chilisa and Kawulich 2012). It means that if the researcher concludes that knowledge is acquirable, the data collection and communication approaches embrace realism, and the nature of the reality being investigated has a physical property that can be experienced by all the social actors equally (Saunders, Lewis, and Thornhill 2016, Burrell and Morgan 2016). This epistemology is known as objectivism. On the other hand, if the researcher reasons that knowledge is produced individually by the social actors (including the researcher), the data collection and communication techniques lean toward nominalism, and there is no unique truth but there exist multiple realities produced by the individuals (or groups of individuals) which cannot be experienced equally (Saunders, Lewis, and Thornhill 2016, Burrell and Morgan 2016). This epistemology is called subjectivism.

This research considers objectivism as its epistemological ground of knowledge based on the ontological assumptions of the study. The epistemological stance of this study requires the results to be generalisable and reproducible, which suitably matches the aim of this study. While interviewing with the experts (which is a subjective approach) at the beginning of this study could provide an understanding of the underlying factors affecting reuse in the UK, it will be limited and cannot fulfil the global perspective of this study. It is because the reuse rates in the UK have been declining continuously (Addis 2006, Sansom and Avery 2014); hence, a sampling frame was developed to target a broader population to improve the response rate (see Sections 4.3 and 4.7). Therefore, in this study, a systematic literature review (SLR) at a global scale is performed at the beginning of the research to provide a profound understanding of the factors affecting the reuse of building components (Chapter 2). This approach suitably matches objectivism epistemology because the SLR is a reproducible approach and can provide an in-depth understanding of the research subject (Denyer and Tranfield 2009, Bettany-Saltikov 2016).

The above discussion sheds light on the axiological assumptions of this study. Axiology, which is the role of ethics, the researcher's values, and the values of the research participants in the process of research (Saunders, Lewis, and Thornhill 2016), plays a crucial role in guiding the researcher's actions (Heron 1996). According to Heron (1996), the researcher's axiological assumptions correlate his/her values to the research subject and the methodological choices (s)he makes. As discussed earlier, in this research, the researcher remains detached from the research participants through following an objective approach. Remaining value-free is of utmost importance in this research because it guarantees the generalisability of the research results (Saunders, Lewis, and Thornhill 2016).

The above discussion clarifies the assumptions about the nature of the science (Figure 3.1) (Burrell and Morgan 2016). However, Burrell and Morgan (2016) suggest another bi-polar dimension, the sociology of regulation-sociology of radical change, which helps to better analyse different research philosophies by showing the political or ideological assumptions of the researchers about the nature of society (Saunders, Lewis, and Thornhill 2016). Briefly, the sociology of regulation tries to explain the reasons behind the success of a social entity (an organisation) and focuses on the improvement of the existing regulations (Saunders, Lewis, and Thornhill 2016). On the other hand, the sociology of radical change questions the existing regulations and focuses on finding alternatives (often Utopian) for the social unit under investigation (Saunders, Lewis, and Thornhill 2016).

The current study deals with predicting the reusability of the building structural elements from social, economic, and technical perspectives. The successful development of various case-study buildings using reused structural components (Section 1.1) shows that the members of the construction sector can integrate these components in new buildings. Therefore, there is no need to radically change the way organisations work. Hence, this study embraces the sociology of regulation (or in short, regulation) as its ideological orientation.

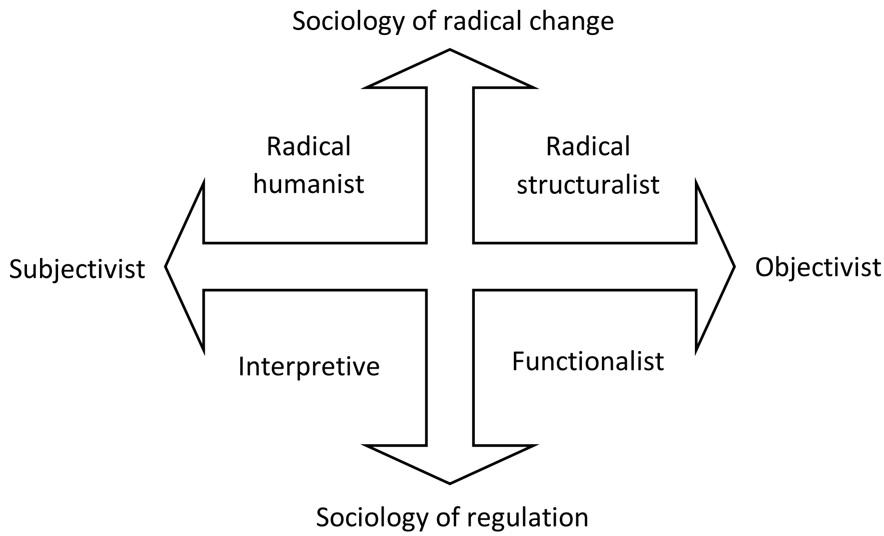

Figure 3.1 Research paradigms (Burrell and Morgan 2016).

3.2.2 Research paradigm

Section 3.2.1 clarifies the theoretical assumptions of the research. These assumptions are then able to inform the theoretical perspectives (or paradigms) of the research (Crotty 1998). To develop a suitable research process, as is the ultimate goal of this chapter, understanding the existing philosophies and paradigms is of great help (Crotty 1998). It is because knowing these theoretical perspectives enlightens the research methodology and eventually justifies the research method. Crotty (1998) explains that the established paradigms should be used to describe and demonstrate the research philosophy. He argues that this approach makes the research process transparent and accountable (Crotty 1998).

Burrell and Morgan (2016), combined the objectivist-subjectivist and regulation–radical change dimensions and developed their four research paradigms (Figure 3.1). These research paradigms are functionalist, interpretive, radical structuralist, and radical humanist (Burrell and Morgan 2016). Burrell and Morgan (2016) define a research paradigm as a set of assumptions which "underwrite the frame of reference, mode of theorising and modus operandi" (ways of working) in which a group of researchers work. This definition is more or less similar to the definition of Kuhn for paradigm which is "universally recognised scientific achievements that for a time provide model problems and solutions to a community of practitioners" (Kuhn 1970). Therefore, a paradigm organises similar types of assumptions that a group of researchers makes about the nature of science and society under investigation, which leads them to consider a specific mode of data collection, analysis, and validation to conduct their research.

The functionalist paradigm, which represents the objectivist-regulation dimensions, deals with the research subject from an objectivist perspective (Burrell and Morgan 2016). It tries to provide rational explanations for the existing social phenomena and seeks practical solutions for the real-world problems within the existing structures (Burrell and Morgan 2016). Positivism is the dominant research philosophy of the studies within this research paradigm (Saunders, Lewis, and Thornhill 2016).

The interpretive paradigm, which is a product of subjectivist-regulation dimensions, concerns with understanding the reality of a social entity (for example an organisation) or in general the social world from the perspective of its members (Burrell and Morgan 2016). Interpretivism is the dominant research philosophy of the studies within this research paradigm (Saunders, Lewis, and Thornhill 2016).

The radical structuralist paradigm, which represents the objectivist-radical change dimensions, focuses on radically changing the existing social structures through analysing the human relationships in social entities (organisations) such as structural power relationships and hierarchies from an objectivist perspective (Burrell and Morgan 2016). Critical realism is the dominant research philosophy of the studies within this research paradigm (Saunders, Lewis, and Thornhill 2016).

The radical humanist paradigm, which results from the combination of subjectivist-radical change dimensions, concerns radically changing the existing social structures in organisations such as power relationships and hierarchies, however, from a subjectivist perspective emphasising human consciousness (Burrell and Morgan 2016).

Based on these discussions, the next sections introduce the above mentioned three research philosophies, interpretivism, critical realism, and positivism, and discuss their suitability for the current research.

3.2.3 Positivism

Positivism, which deals with what is posited (given) (Crotty 1998), seeks to explain and predict the social phenomena through identifying regulations and cause-and-effect interactions between its constituent elements (Burrell and Morgan 2016). Positivists follow the scientific method approach and perform their research through direct experiences (Crotty 1998) to achieve data and facts about the subject of study, which is uninfluenced by human consciousness or bias (Saunders, Lewis, and Thornhill 2016). A positivist researcher may then develop generalisations based on the observed causal relationships between the observed facts (Saunders, Lewis, and Thornhill 2016).

The positivist's ontology is real and independent and, he/she performs value-free research, and is detached from the subject of the study and keeps an objective stance throughout the study (Saunders, Lewis, and Thornhill 2016). The epistemology of the positivist researcher is, therefore, objectivism. According to Crotty (1998), "positivism is objectivist through and through". Moreover, in addition to drawing generalised conclusions, positivists endeavour to verify or falsify the existing theories rather than seeking new hypotheses (Burrell and Morgan 2016, Saunders, Lewis, and Thornhill 2016). On this basis, positivism tends towards deduction in its approach to theory development (Saunders, Lewis, and Thornhill 2016). Studies with positivism research philosophy might embrace survey research as their methodology and use quantitative research and statistical analysis tools to analyse the collected data (Crotty 1998).

3.2.4 Critical realism

Critical realism seeks the reality of an observable event by identifying the underlying structures and mechanisms resulting in the known regularity (Denzin editor and Lincoln editor 2018). Ontologically, critical realism recognises that there is a single, independent reality. However, this reality is different from empirical experiences and, there is a chance that it may never be fully understood because of the hidden aspects of the generative mechanisms of reality (Denzin editor and Lincoln editor 2018, Collier 1994). As such, since it is not possible to know the reality through senses and, one should return to his/her experiences and further investigate the root cause(s), critical realism tends towards induction (retroduction) in its approach to theory development (Reed 2005). As explained earlier, it is this generative mechanism that is the focus of critical realism (Collier 1994). Therefore, the ontology of the critical realism research philosophy is stratified (Collier 1994) and comprises of three layers: the empirical, the actual, and the real (Saunders, Lewis, and Thornhill 2016). Collier (1994) further explains that the ontology of critical realism emphasises that the world is comprised of transitive and intransitive objects. Transitive objects are the "theories about the nature of the world" at any given time that the researcher makes to deepen his/her knowledge about the intransitive objects (the reality), which exist independently of one's consciousness (Collier 1994).

From the above, it can be concluded that while ontologically critical realism acknowledges that there exists a single reality, epistemologically, it leans toward subjectivism because the nature of human beings' knowledge is transitive, temporary, and challengeable and can change in the future. This conclusion is in agreement with Saunders et al. (2016) where they consider relativism (a mildly subjectivist approach) as the epistemology of critical realism. The temporary nature of human knowledge means that the social realities change with time, and social facts are generated and agreed upon by the social members at any given time (Saunders, Lewis, and Thornhill 2016). Therefore, the subjective nature of the critical realism philosophy and its ontological assumption about truth allows using both qualitative and quantitative methods to arrive at reality (Reed 2005, Healy and Perry 2000, Given 2008). From an axiological perspective, while the objectivist epistemology urges the researcher to be value-free, because a critical realist engages with the members of a social entity (like an organisation) to uncover the underlying mechanisms leading to the reality, it cannot stay completely objective and there is a chance of bias based on the socio-cultural background of the researcher (Saunders, Lewis, and Thornhill 2016).

3.2.5 Interpretivism

Interpretivism seeks the realities about a social phenomenon through interpreting the cultural and historical perceptions of the society about that phenomenon (Crotty 1998, Denzin editor and Lincoln editor 2018). Ontologically, interpretivism declares that the reality is not absolute, but it is relative and is constructed by one's actual experiences as well as through interactions with others in the society (Crotty 1998, Denzin editor and Lincoln editor 2018). Therefore, instead of a single, generalisable truth, there exist multiple realities about an event, a phenomenon, or a social entity. It means that the interpretivist researcher considers the perspectives of different members of a society or an organisation to create a new and richer understanding of the subject under investigation (Saunders, Lewis, and Thornhill 2016); hence, developing new theories.

Epistemologically, interpretivism embraces subjectivism, and the researcher and the subjects interact closely to create knowledge (Burrell and Morgan 2016, Denzin editor and Lincoln editor 2018). Consequently, the interpretivist's values play a crucial role during the research process and, from an axiological perspective, the researcher becomes a part of the research and stays reflexive throughout the study (Saunders, Lewis, and Thornhill 2016). On this basis, interpretive research is an inductive process in its approach to theory development, and researchers majorly use qualitative research methods to develop their theories (Crotty 1998, Saunders, Lewis, and Thornhill 2016).

Crotty (1998) introduces various themes of interpretivism and emphasises that the participants' lived experience (phenomenology), cultural artefacts such as texts and symbols (hermeneutics), and inter-subjectivity (symbolic interactionism) form the researcher's knowledge about the research subject. It is noteworthy that phenomenologists and symbolic interactionists act oppositely in dealing with culture as the former considers culture as a potential barrier for making new meanings in the social life and the latter considers culture as a guide to a comprehensive set of concepts (Crotty 1998).

3.2.6 Research philosophy of the study

Considering critical realism as the potential research philosophy, while it is epistemologically objective, it seeks the causal explanation of the underlying reasons for a phenomenon and does not intend to predict or generalise the results. Moreover, a critical realist, while tries to be as objective as possible, because of the transitive nature of the world he/she investigates cannot remain fully value-free in his/her research. Hence, the critical realism research philosophy does not match the requirements of this research because it contradicts the philosophical assumptions underpinning the study of BSE-RPM.

Considering interpretivism as the potential research philosophy, because the study of BSE-RPM seeks the experts' opinion to develop its predictive model(s), it depends on human beings during the data collection stage. However, this study investigates the reusability of the structural elements (physical objects), and the experts provide facts and figures about a component they reused in the past. Therefore, since these facts and figures are measurable and reproducible independently, the research philosophy of this study cannot follow interpretivism.

In contrast, as discussed in Section 3.2.1, the study of developing BSE-RPM embraces realism (ontological assumption), objectivism (epistemological assumption), and remains value-free (axiological assumption). Moreover, this study seeks generalisations by developing predictive models; hence, its approach to theory development follows a deductive pattern. Likewise, based on Section 3.2.2, this study follows the sociology of regulation; and consequently, it falls under the functionalist paradigm. Therefore, positivism is the appropriate research philosophy for this study.

3.3 Research Methodology

Research methodology (or research strategy) refers to an overall framework determining the research strategy, plan of action, and direction which leads to the choice and rationale of research method(s), and analysis technique(s) to answer the research questions and meet the project objectives (Saunders, Lewis, and Thornhill 2016, Crotty 1998). This procedure that guides the researcher on how to collect data and how to analyse them depends on the theoretical perspectives (research philosophy) of the study, philosophical assumptions of the research, and the type of research questions under investigation (*Oxford English Dictionary* n.d., Saunders, Lewis, and Thornhill 2016, Crotty 1998).

While, as discussed in Section 3.2, there are some philosophical assumptions and several research philosophies, a considerable number of methodologies and countless methods exist in various textbooks (Crotty 1998, Burrell and Morgan 2016, Saunders, Lewis, and Thornhill 2016). However, Section 3.3 limits itself to the discussion of those methodologies related to the philosophical stance of this research, which is positivism. Therefore, this section initially introduces the research strategies related to positivism, then evaluates their suitability to this study and eventually justifies the selected research strategy based on the philosophical assumptions, research philosophy, and research questions.

3.3.1 Experimental research

(Nesselroade and Cattell 1988) define experiment as “a recording of observations, quantitative or qualitative, made by defined and recorded operations and in defined conditions, followed by an examination of the data, by appropriate statistical and mathematical rules, for the existence of significant relations.” Experimental research, which is commonly used in various sciences such as natural sciences, sociology, and psychology, is a set of procedures in which a set of independent variables are manipulated to assess their effect on a dependent variable (Saunders, Lewis, and Thornhill 2016). The experimental research, which is a sub-division of empirical research (Cash, Stanković, and Štorga 2016), seeks verification (or falsification) of a prediction (the hypothesis) and generalisation; hence, looking for the causal relationship between the dependent and independent variables (Srinagesh 2006). Therefore, experimental research can address the exploratory and explanatory research question.

The study of BSE-RPM looks for causal relationships between factors affecting reuse (independent variables), and the reusability of the structural elements (dependent variables) in terms of technical, social, and economic aspects. Moreover, the research questions are explanatory (see Section 1.5) for which makes experimental research a potential methodology for this study.

However, experimental research, in terms of laboratory testing of the recovered structural elements, is limited to identifying their characteristics and cannot fulfil the aim of this project for the following reasons. Firstly, these types of tests can only determine the physical properties of an element and at the most can provide a narrow idea about the technical reusability of the component when compared to an equivalent new one (the control). Even in this case, because the researcher tests the element in a laboratory condition, and it is not going to be installed in a new building, its technical reusability in combination with other components in a new installation cannot be verified. Secondly, experimental research requires a considerable number of reused structural elements for testing to provide enough data to develop a predictive model. However, considering the required time and the availability of funding, laboratory testing is not feasible. Thirdly, laboratory tests cannot provide any insights into the economic and social aspects of reusability as are within the scope of this research. Therefore, experimental research methodology is not pursued.

3.3.2 Archival and documentary research

Archives are a set of documents collected by an organisation, individual, or government (Frisch et al. 2012) which are a good source of secondary data for research (Saunders, Lewis, and Thornhill 2016). In archival and documentary research, the researcher looks for answers to the research questions using documents archived by organisations and individuals. Because organisations record all their activities in both digital and paper formats, it is possible to get a rich insight into the reusability of building structural elements by searching project documents such as test certificates, tender bulletins, standard specifications, etc. However, accessing such documents, if not impossible, would be very hard because most of them contain sensitive data (Saunders, Lewis, and Thornhill 2016). Therefore, in this study, archival research methodology is not considered.

3.3.3 Case study research

Case study research methodology is an in-depth investigation of a case (a phenomenon, event, an organisation, a group, etc. (Saunders, Lewis, and Thornhill 2016)) in its real-world settings (Yin 2018). A researcher may perform a case study research when he/she wants to investigate a real-life situation as well as when the distinction between the case and the context are not distinguishable (Yin 2018). Case study research may employ qualitative, quantitative, or mixed-mode research to achieve an in-depth understanding of the phenomenon (Yin 2018, Saunders, Lewis, and Thornhill 2016). Moreover, case studies can be designed to answer exploratory, explanatory, and descriptive research questions (Yin 2018).

While the study of BSE-RPM seeks the experts' opinion on the reusability of building structural elements, case study research is unable to fulfil the aims of the project for the following reasons. The study of BSE-RPM would eventually develop a best practice predictive model based on statistical generalisations; however, in doing a case study research, the researcher looks for expanding and generalising the theories (Yin 2018). Moreover, a case study does not represent a population; however, the study of BSE-RPM develops predictive models from a representative of the structural elements of a building (Section 4.2) based on the real experience of the experts from a global perspective (Yin 2018). Therefore, case study research methodology is not pursued.

3.3.4 Survey research as the methodological choice of the study

Survey research is “a systematic set of methods used to gather information to generate knowledge and to help make decisions” (Lavrakas 2008). It involves the collection of quantifiable data from a population by selecting a sample of individuals voluntarily answering a set of questions (Check and Schutt 2011, De Vaus 2014, Sapsford 2007). Therefore, the survey research methodology usually takes a deductive reasoning technique (Saunders, Lewis, and Thornhill 2016). Moreover, it can employ closed questions (like Likert scale questions) or open-ended questions (that can be coded by the researcher later) or a combination of both in a questionnaire (Ponto 2015).

Because surveys gather data with the intention to elaborating the characteristics, attitudes, experience or opinion of a group or population, or identifying standards to compare the existing conditions, as well as determining the potential correlations between specific events (Cohen, Manion, and Morrison 2018), they can address explanatory, exploratory, and descriptive research questions. While different types of surveys exist, survey research using self-administered questionnaires (both online and paper-based) are widespread because they can collect standardised data from a wide range of population at a low cost (Saunders, Lewis, and Thornhill 2016) and allow generalisations of the results through statistical, or machine learning techniques (Yin 2018).

The study of BSE-RPM seeks the experts’ opinion on the reusability of building structural elements and intends to develop a best-practice predictive model using machine learning techniques. Therefore, it needs input from a wide range of professionals to be able to perform generalisation. Moreover, the research questions are explanatory, and the research takes a deductive reasoning technique. Furthermore, this study aims to determine the reusability of the structural elements in technical, social, and economic aspects. Therefore, based on the above discussion, this study follows the survey research methodology as its research strategy.

3.4 Research Method

The philosophical assumptions of a study play a pivotal role in determining its philosophical stance, which leads to the methodological choice of the research (Crotty 1998). As discussed in Section 3.2, positivism is the research philosophy of this study, which leads to the selection of the survey research as the methodological choice of the research (Section 3.3). In this section, the research design or the research method, which is the choice of the techniques and procedures to provide the required inputs (data collection) and analyse them, are introduced (Crotty 1998). However, the justification for the selection of the data collection and data analysis techniques are provided in Chapters 4 and 5, respectively.

Depending on the research questions, research philosophy, and research methodology, the type of data necessary for a study may vary between numerical (numbers), non-numerical (texts, images, etc.), or a mixture of the two. Therefore, a quantitative method, a qualitative method, or a mixed-method research design might serve the aims and objectives of any research (Saunders, Lewis, and Thornhill 2016). This study seeks numerical data to develop the first BSE-RPM using advanced supervised machine learning techniques. Therefore, this study chooses a quantitative research method for its data collection and data analysis, because, this study aims to develop models that can efficiently and accurately predict the reuse potential of structural elements at the end-of-life of a building based on the experts' opinions using several advanced supervised machine learning methods.

Since this study seeks to quantify the qualitative variables affecting the reusability of the structural elements (independent variables) based on the experts' opinions, there is a doubt whether a qualitative approach to identify these factors should be followed at the inception of the research. While the identification of the independent variables, which is an exploratory attempt, can be performed using various techniques such as literature review or interview with the experts in the form of unstructured (in-depth) individual or group meetings (Saunders, Lewis, and Thornhill 2016), interviewing does not suit this study for the following reasons. First, the study of BSE-RPM is objective from an epistemological perspective (Section 3.2.1) and has value-free axiology. However, in-depth interviewing is a purely subjective approach, and the researcher cannot stay value-free, which in turn increases the risk of bias in the research. Moreover, in-depth interviews are not reproducible; however, a systematic literature review approach, which is used in this study to explore the independent variables (Chapter 2), is a highly reproducible research method, which produces unbiased reports to enlighten the existing knowledge about the particular research question, provides a robust basis for reliable judgments about "what works" the best (Petrosino and Lavenberg 2007), and finds gaps in the literature for further research (Denyer and Tranfield 2009). Therefore, this study does not integrate a qualitative research method to identify the independent variables in its design.

3.4.1 Quantitative research method

Quantitative research is “an approach for testing objective theories by examining the relationship among variables” (Creswell and Creswell 2018). However, quantitative research can be used to develop new theories, as well (Crotty 1998, Saunders, Lewis, and Thornhill 2016). Therefore, in the case of the former, where data are used to test the theories, quantitative research follows a deductive approach, and in the case of the latter, it is associated with inductive reasoning (Creswell and Creswell 2018, Saunders, Lewis, and Thornhill 2016). Moreover, the variables in quantitative research can be either measured experimentally (using instruments) or can be collected objectively using survey research (Section 3.3.1 and Section 3.3.4) (Creswell and Creswell 2018, Saunders, Lewis, and Thornhill 2016). In either way, quantitative research embraces Objectivism as its epistemological assumption and positivism as its research philosophy (Saunders, Lewis, and Thornhill 2016). Furthermore, in a quantitative study, the measured variables are mostly analysed using statistical and graphical procedures (Creswell and Creswell 2018, Saunders, Lewis, and Thornhill 2016). Since the survey research methodology for collecting data has an objective nature, the researcher remains value-free and has no impact on the respondents.

According to Section 3.3.4, this study follows a survey research methodology as its research strategy. Survey research can provide a numeric description of the relationship between variables by studying a representative sample of the population (Creswell and Creswell 2018). While the survey research can be conducted in the form of questionnaires or structured interviews to collect necessary data with the intent of generalising the results (Fowler 2014), this study uses a questionnaire survey for its data collection. Eventually, this study uses statistical and supervised machine learning methods (using SPSS and R, respectively) to analyse the collected data and develop predictive models (R Core Team 2020). The justification for using supervised machine learning methods is discussed in Chapter 5.

3.5 Chapter Summary

According to Crotty (1998), at the inception of developing a research proposal, it is necessary that the researcher identifies and justifies the choices for research methodologies and research methods, which lead to choosing the data collection and analysis techniques. However, before discussing these crucial features, this chapter initially clarified the philosophical assumptions and scrutinised the potential theoretical perspectives to identify the research philosophy.

According to Section 3.2, this study follows realism as its ontology, embraces Objectivism as its epistemology, and the researcher remains value-free, which is essential to minimise the bias and guarantee the generalisability of the research results. In the next step, the four paradigms, as identified by Burrell and Morgan (2016), were introduced, and the researcher justified that, according to the philosophical assumptions of the research, this study follows the functionalist paradigm. It is noteworthy that the term "paradigm" is used in the concept defined by Burrell and Morgan (2016) because it adds a new philosophical assumption to the earlier assumptions discussed, which is the political or ideological assumptions of the researcher about nature of society. It is because ontology, epistemology, and axiology only deal with the nature of science and do not reveal the researcher's assumptions about society. This approach, while on the surface deviates from most research books, does not differ in context. The reason is what is called a paradigm generally is called a research philosophy in this study. It is because there is no unanimity in the naming approaches in the social sciences. For instance, Crotty (1998) follows a different naming philosophy and uses the term "theoretical perspectives" for what is called research philosophies in this research.

Next, the researcher introduced positivism, critical realism, and interpretivism and compared these research philosophies. In Section 3.2, it was concluded that positivism is the research philosophy of this study. Consequently, the researcher introduced the research methodologies embracing positivism and reasoned that this study follows survey research as its methodology. Moreover, by analysing the potential research designs (quantitative, qualitative, or mixed-method), it was concluded that this study follows a quantitative research design and uses questionnaires to collect data. Likewise, the researcher introduced statistical (using SPSS) and supervised machine learning techniques (through R) as the methods used to analyse the data and develop the first building structural elements reusability predictive model. However, it was emphasised that the justification for the choice of data collection and analysis techniques would be provided in Chapters 4 and 5, respectively. The below table is the summary of all the above discussions that will be referred to throughout this work.

Table 3.1 Research design essentials

Title	Potential options	Selected options
Theoretical assumptions: Ontology	Realism Relativism	Realism
Theoretical assumptions: Epistemology	Objectivism Subjectivism	Objectivism
Theoretical assumptions: Axiology	Value-free Value-bounded	Value-free
Research approach	Deduction	Deduction

Title	Potential options	Selected options
	Induction Retroduction (abduction)	
Research paradigm	Functionalist Interpretive Radical structuralist Radical humanist	Functionalist
Research philosophy	Positivism Critical realism Interpretivism	Positivism
Research methodology	Experimental Archival and documentary Case study Survey	Survey
Research method	Quantitative Qualitative Mixed	Quantitative
Data collection	Structured interview Questionnaire	Questionnaire
Data analysis tools	Many	Statistical Supervised machine learning

Chapter 4 – Quantitative study

4.1 Chapter introduction

In this chapter, the technique used to collect the required data to develop the first Building Structural Elements Reusability Predictive Model (BSE-RPM) is discussed in detail. As discussed in Section 3.4.1, this study uses a self-completed questionnaire as its data collection method. Online questionnaires are distributed among a sample of experts with previous experience in structural elements reuse at a global scale. The introduction section of this chapter includes the justification for using the questionnaire as the data collection technique, and the advantages and practical limitations of using questionnaires. Section 4.2 introduces the unit of analysis of this research. Section 4.3 discusses the sampling process, and the following sections discuss the process of designing and testing the questionnaire (Sections 4.4 and 4.5, respectively) before sharing it with the experts (Section 4.6). The study then discusses the response rate of the survey (Section 4.7), analyses the missing data (Section 4.8), reviews the validity and reliability of the questionnaire based on the received responses (Section 4.9), performs statistical analysis on the collected data (Sections 4.10 and 4.11) and concludes with the chapter summary (Section 4.12).

4.1.1 Justification for using a questionnaire

Chapter four aims to quantify the weightage and impact of the reusability factors based on the experts' opinions using questionnaires, which is the second objective of this research (Section 1.6). Because each expert (respondent) replies to the same set of questions, it is possible to approach a large sample to collect the necessary data (Saunders, Lewis, and Thornhill 2016). Moreover, using a questionnaire provides an efficient way of accessing and quantifying the professionals' knowledge regarding the factors affecting the reusability of the structural element(s) that they reused in the past. Accordingly, this chapter enables achieving the second objective of this research (Section 1.6). Therefore, the variables (factors) identified in the questionnaire (both independent and dependent) are in the form of closed questions using a five-point Likert-scale rating, which enables quantifying the qualitative variables. Moreover, this feature facilitates the respondents with easy-to-understand questions and eventually increases the response rate.

4.1.2 Advantages and disadvantages of questionnaire

A self-administrated questionnaire, as used in this research, is very efficient because the respondent can spend enough time answering each question and can complete the survey at a later time (Brace 2013). Even the respondent may refer to different documents and consult with others to provide more accurate answers. Moreover, the bias is reduced because firstly, the respondent remains anonymous, and he/she can feel very safe while answering the questions, and secondly, the researcher has no direct influence on the answers the respondent provides. Likewise, as discussed in Section 4.1.1, the web-based questionnaire can be shared with a large sample to increase the response rate.

On the other hand, there is a limitation to the number of questions that a questionnaire can contain (Yin 2018). Therefore, the researcher needs to make sure that the questions can provide reliable answers to the research questions to fulfil the aim of the study (Saunders, Lewis, and Thornhill 2016). Moreover, the questionnaire provides only one chance to collect data from a potential respondent (Saunders, Lewis, and Thornhill 2016); hence, if any question is missing, or any parts of the survey is unclear or biased, there is no chance of going back to the respondent to rectify the error. Therefore, to overcome these limitations, the questions were re-written several times and discussed with the supervisory team to improve the quality of the survey. Moreover, the pilot study helped to achieve the final shape of the questionnaire, which will be discussed later in Section 4.5.

4.2 Unit of analysis

While performing research, it is critical to know the aim(s) and objective(s) of the study. It is because, throughout the research, the collected data need to be analysed continuously to ensure that the research question(s) can be addressed and eventually the project's aim(s) can be achieved. Therefore, it is vital to distinguish the unit on which the researcher needs to collect the necessary data.

The unit that the researcher collects data about and then performs his/her analyses is called the unit of statistical analysis, or simply the unit of analysis (Salkind 2010). Therefore, the unit of analysis of a study depends on the aim and objectives of the research. According to Addelman (1970), the unit of analysis or the experimental unit "is that entity that is allocated to a treatment 'independently' of other entities." The unit of analysis may be the same or different from the unit of observation, the unit of sampling, the unit of generalisation, and the unit of measurement (Salkind 2010, Decarlo 2018). While the unit of analysis is the entity that the researcher collects data about, the unit of observation is the item that is observed to collect the required data for the study (Decarlo 2018). For instance, if the researcher intends to collect data about a neighbourhood based on observing people living there, the unit of analysis would be that neighbourhood while the unit of observation would be the residents (Lavrakas 2008).

Moreover, the unit of sampling is the entity the researcher chooses the samples to perform the observations. In addition, the unit of sampling is also directed by the unit of generalisations that the researcher generalises about (Salkind 2010). And last but not least, the unit of measurement is the unit on which the impacts of a measure is studied, which can be similar or different from the unit of generalisation or sampling (Salkind 2010).

This study aims to develop a reliable model to predict the reuse potential of the structural elements of a building based on the experience of the professionals who worked with such components. Therefore, the focus of the study is on the reusability of the structural components used in a building such as beams, columns, slabs, truss, etc. Referring to Section 1.5, the research questions concentrate on collecting data about the reused structural elements to determine the best combination of factors that can help in the development of a predictive model to assess the reusability of these elements at the end-of-life of a building. Therefore, in this study, the unit of analysis, which is the subject of advanced statistical and machine learning analyses, is the structural elements of a building (Salkind 2010). Moreover, the unit of observation is the same as the unit of analysis because the research intends to develop a robust predictive model to determine the reusability of the structural elements of a building. Consequently, the unit of sampling, the unit of generalisations, and the unit of measurement are the same as the unit of analysis.

4.3 Sampling

According to the unit of analysis of this study (Section 4.2), the population from which the sample needs to be selected is the structural elements of a building. Since this research intends to determine the reusability of the structural elements of a building and eventually generalise the results about them, the target population is all of the recovered building structural elements intended for reuse (regardless of success) (Figure 4.1). However, since this study seeks the experts' knowledge about the reused elements to develop its predictive model(s), it is necessary to sample from the experts with experience in reusing building structural elements. Nevertheless, because there is no way to identify based on what structural element the potential respondent would complete the questionnaire, all the reused components have an equal chance for selection by an expert.

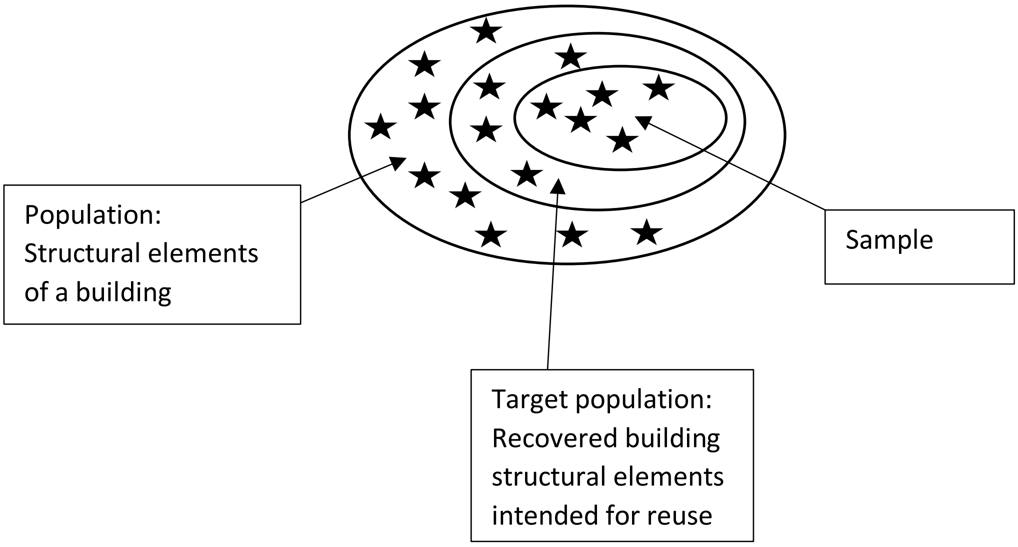

Figure 4.1 Population, sampling frame and sample at the structural elements level

For sampling from the experts with reuse experience, the sampling frame is depicted in Figure 4.2. As shown in this figure, the target population is professionals with reuse experience working in construction, deconstruction, demolition, or reuse companies in the construction sector. Moreover, the sample consists of all the identified experts in the target population.

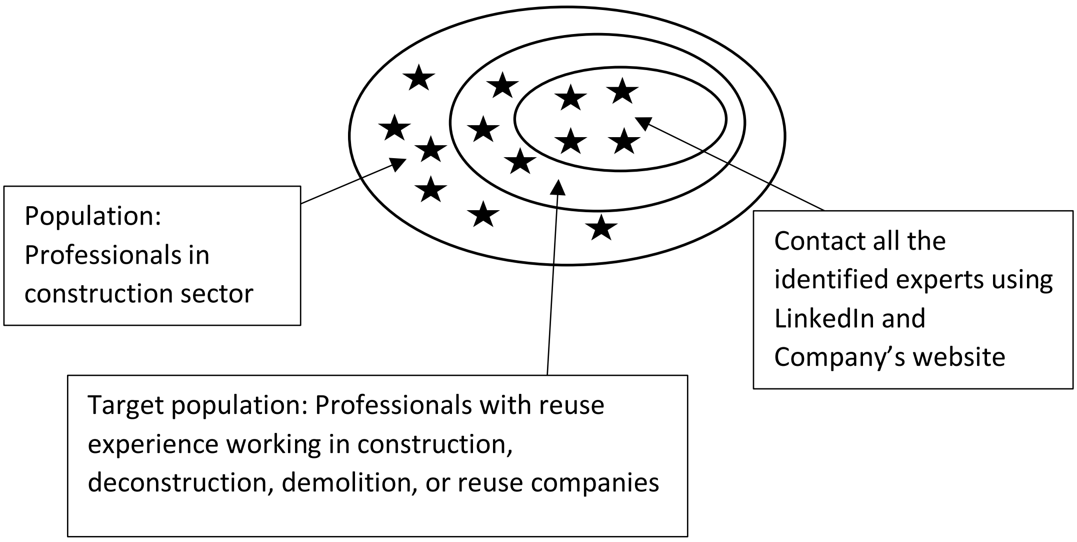

Figure 4.2 Population, sampling frame and sample at the construction professionals' level

In this study, the purposive sampling technique is used to select professionals from the target population. It is because reuse is not a commonplace practice, and no database or list of experts with reuse experience is available to perform a probability sampling. Therefore, the researcher developed a sampling frame to reach out to the target population. However, because the reuse rates have been continuously declining in the UK (Addis 2006, Sansom and Avery 2014), the number of available professionals with the required profile is not sufficient in the UK. Hence, the sampling frame is developed with a global perspective. Therefore, this study uses the following resources for developing the sampling frame, which includes experts with experience in reusing building structural elements (Figure 4.2). The first reference referred to for developing the sampling frame is the list of top 100 demolition companies worldwide in 2018, which is used to locate the experts (KHL Group 2018). Moreover, using LinkedIn, a list of companies in the construction sector with experience in building component reuse or deconstruction is prepared. These two lists are then merged, and any duplications are removed.

In this study, all the experts are located using the companies' websites and LinkedIn. While a company's website gives a general overview of the top management team (this depends on their privacy policy) and the types of services the company offers, most of the time, it does not provide any details about the employees recruited by the company. On the other hand, LinkedIn provides a platform for accessing the professionals and their profiles and level of experiences at no cost. According to (Duffy 2015), LinkedIn is "the most important cross-industry professional network around" with more than 645 million members from 200 countries worldwide (LinkedIn 2019) and a high growth rate in the number of experts joining this social media (Dusek, Yurova, and Ruppel 2015). After the sampling frame was developed, all the located experts were contacted (Figure 4.2). As a result, a total number of 481 invitations are sent to the experts to complete the online questionnaire.

The limitation of developing the sampling frame using LinkedIn is that LinkedIn is not an exhaustive list of all the population of the professionals and if an expert is not registered online, he/she cannot be located and therefore, is not included in the sampling frame (Dusek, Yurova, and Ruppel 2015). Moreover, as discussed earlier, the limitation of using the Companies' website as the other source for locating the experts is that they are not giving any details about their employees or their level of experiences. Therefore, to decrease bias, the snowballing technique to locate more respondents is used as well. It means that if an expert provided an email address after completing the questionnaire, he/she was requested to share the survey with the other experts that he/she knows. Nonetheless, since these experts did not provide any feedback on the number of times, they shared the questionnaire link, there is no way to judge the number of invitations sent based on the snowballing technique.

4.4 Questionnaire design

The theory behind this study is that by quantifying the impact of factors affecting the reusability of the building structural elements (the independent variables), it is possible to predict the technical, economic, and social reusability of these components (the dependent variables) through developing predictive models using advanced supervised machine learning techniques. Therefore, after performing a thorough literature search (including a systematic literature review discussed in Chapter 2), and identifying, analysing, and categorising the independent variables, these factors are used to develop a self-completed online questionnaire. This questionnaire is then shared with a representative sample of the experts discussed in Section 4.3. Therefore, this questionnaire aims to address the second objective of this study, and as discussed in Section 4.1.1, to provide the required data to achieve the third objective of this research.

Initially, and based on the identified independent variables, a paper-based questionnaire was developed, which included 125 questions. However, after several rounds of reviewing the questionnaire, consulting with the supervisory team, conducting a self-check (Section 4.4.3 and Appendix B), and finally performing a pilot study (Section 4.5), the total number of questions decreased to 72.

4.4.1 Sections of the questionnaire

This questionnaire consists of six sections and 72 questions (see Appendix C). Section A contains demographic questions and asks five questions about the details of the respondents and the years of experience in the construction sector. While the initial purpose of this section is to acquire a general overview of the respondents, the details will be further used as an additional checkpoint to evaluate the validity of the responses (Section 4.9).

Section B deals with the structural element that the respondent used in the past and would complete the rest of the questionnaire by referring to it. This section contains 11 questions and is in two parts. Questions 1 to 6 seek the details of the reused element, and questions 7 to 11 compare the current use of the component (or use after deconstruction) with its previous deployment before it was removed/deconstructed from a building. The purpose of questions 7 to 11 is twofold. First, to understand the current application of the element and second, to determine the changes in its performance.

Section C is concerned with the barriers to reuse, as identified during the literature review. This section intends to quantify the impact of the identified barriers on the reusability of the structural elements from social, economic, and technical perspectives. Further details about the identified barriers are available in Section 2.3.2 and Appendix A.

Section D contains those factors that can act as either a barrier or a driver in different circumstances. For instance, according to Gorgolewski et al. (2008), the purchasing price of the reused building components is a driver to reuse; however, according to Tingley et al. (2017) and Dunant et al. (2018), the cost of these elements is a barrier to reuse. Therefore, Section D lists the variables for which their impact on the reusability of the structural elements are unknown. Like Section C, this section also includes technical, social, and economic variables that affect reuse.

Section E inquires the reusability of the structural element that the respondent used before and based on that replied to the questions in Sections B, C & D. In total, there are three questions in this section, which together form the dependent variables of this study. These questions aim to understand the respondent's evaluation of the reusability of the structural element. These questions are very important to achieve the third objective of this study. Through using advanced supervised machine learning techniques, the impact of the independent variables (Sections B, C, & D) on the dependent variables (Section E) would be analysed, and the best combination of the independent variables that can predict the reusability of the structural elements of a building would be developed.

In this questionnaire, to avoid any misinterpretation by the respondent, the dependent variables are defined before the questions as follows:

<u>*Technical reusability:*</u>

The extent to which the reused structural element in its new life could perform similarly to its earlier life.

<u>*Economic reusability:*</u>

The cost savings in the project as the result of using the reused structural element when compared to a similar project using a new structural element with the same performance.

<u>*Social reusability:*</u>

The acceptance level of the stakeholders (clients, CEO, designers, construction team, occupants, etc.) about using the reused structural element in the new building.

After this section, the respondents are free to add any additional comments if they wish. Moreover, to incentivise the respondents to answer the questionnaire (as a bonus for taking part), the survey encourages the respondents to provide their contact details if they wish to receive the results of the study upon publication. Notwithstanding, all the above is optional.

4.4.2 Types of questions and scales

Excluding Section A, which includes demographic questions, and except Section B, where the respondent has the option to respond other than the pre-determined answers for questions 1 to 6, the rest of the questionnaire contains closed questions. Moreover, questions 1 to 6 in Section B seek the details of the reused element and are factual, while others inquire a fact based on the respondent’s experience about the impact of the variables on the reusability of the structural component. It is noteworthy that questions B1 to B6 are nominal, and the possible answers have no correct order. On the other hand, the rest of the questions are ordinal (or categorical), and the answers are ordered. For the details about nominal and ordinal variables, please refer to Section 12.2 of (Saunders, Lewis, and Thornhill 2016). In this research, ordinal and categorical variables are used interchangeably.

For the closed questions in this questionnaire, the Likert-style ratings are used (Likert 1932). By developing this rating measurement system, Rensis Likert intended to measure different aspects of an attitude, opinion, or a belief by requiring the respondents to express their level of agreement or disagreement to a question or a statement (Brace 2013). While the Likert response sets can include four or more points (Lavrakas 2008), this study uses a five points system, which is more common (Lavrakas 2008). As discussed in Section 4.1.1, the above rating system is very important because it helps to quantify the independent variables identified during the literature review phase of this research.

While, in this study, all the closed questions follow a five categories Likert scale, the wording labels and the purpose for different sets of questions vary. Table 4.1 summarises various Likert scale ratings used in this study.

Table 4.1 Likert-style ratings used in this study

Section	Question(s)	Reason	Scale				
			1	2	3	4	5
B	7 to 9	To assess the physical changes	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
B	10 to 11	To assess the functional changes	Much lower	Lower	Equal	Higher	Much Higher
C	All	To determine the level of the	Very High	High	Moderate	Low	Very Low

Section	Question(s)	Reason	Scale				
			1	2	3	4	5
		negative impact of the barrier					
D	All	To determine the effect of the variable	Very negatively	Negatively	No real effect	Positively	Very Positively
E	All	To determine the reusability level	Very low	Low	Moderate	High	Very High

4.4.3 Self-checking the questionnaire

Before launching the online questionnaire for pilot testing, the survey was thoroughly checked for layout, question order, and question-wording. For this purpose, three checklists (inspired by Saunders et al. (2016)) were prepared and used to develop the online survey for pilot testing (Appendix B). The checklists contain 5, 7, and 18 questions for checking the questionnaire layout, questions order, and questions wording, respectively. After self-checking, the questionnaire was reviewed for the wording and grammatical errors by an advisor at the Centre for Academic Writing (CAW) at Coventry University. Upon the incorporation of the comments by the CAW advisor, the questionnaire was launched online for pilot testing.

4.5 Pilot study

One of the problems with self-completed questionnaires is that, unlike in-depth or semi-structured interviews, it is not possible to modify or alter it after it is launched (Saunders, Lewis, and Thornhill 2016). Therefore, if the information necessary to address the research objectives are missing, or if the questionnaire or the questions are biasing or biased, the collected data cannot be trusted, which can cause serious risk to the project (Brace 2013). Moreover, there is a risk that the researcher and the respondent might interpret the questions and answers in different ways, which again makes the collected data unreliable (Saunders, Lewis, and Thornhill 2016). Therefore, care should be taken in designing the questionnaire and the questions to ensure the validity and reliability of the responses (Saunders, Lewis, and Thornhill 2016, Brace 2013). While, according to Section 4.4.3, self-checking using the recommended checklist by Saunders et al. (2016), and grammar and wording review by an expert can mitigate some of the above risks, it is always advised to pilot the questionnaire before performing the data collection (Saunders, Lewis, and Thornhill 2016, Brace 2013).

Following the above discussion, the link for the online questionnaire was shared with a group of 12 experts and non-experts for pilot testing. Unlike the sampling process explained in Section 4.3, the pilot study followed convenience sampling and snowballing techniques to locate the respondents, which is an acceptable approach (Saunders, Lewis, and Thornhill 2016). According to Saunders et al. (2016), non-probability sampling can be performed during the pilot testing of a questionnaire as well. Moreover, to collect the feedback of the participants, an online feedback form was also shared with the respondents. This feedback form contained the following questions as advised by (Bell and Waters 2014):

- How long the questionnaire took to complete?
- Were the instructions clear? If not, what are your suggestions to improve them?
- Were any of the questions unclear or ambiguous? If yes, kindly provide more details (question number, etc.).
- Were there any questions you felt uneasy about answering? If yes, kindly provide more details (question number, etc.).
- In your opinion, were there any major topic omissions? If yes, please provide further details.
- Was the layout clear and attractive? If not, what is your suggestion to improve it?
- Are there any other comments you wish to share?

Based on the completed feedback forms, the maximum time to complete the questionnaire was reported to be 15 minutes except for one respondent who reported one hour to complete the survey. The pilot study and the associated feedback forms helped in improving the quality of the questionnaire through rephrasing some questions, modifying the scales, and removal of some unnecessary questions.

4.6 Data collection

After improving the quality of the survey based on the feedback forms (Section 4.5), the final revision of the questionnaire was developed and shared online. The online questionnaire was developed using the Online surveys (formerly BOS) platform (Jisc 2019), and the link was shared with the potential respondents identified in Section 4.3.

In this study, an online questionnaire is used to collect the required data, and other possible varieties such as post, delivery and collection, telephone, and structured interviews are not used. The reason for this choice was because of the low cost of using the online questionnaire (a free account for using Online surveys was provided by the Coventry University) and its capability to be shared with a large sample (Saunders, Lewis, and Thornhill 2016).

While face-to-face data collection (also known as structured interview), guarantees that the respondents represent the targeting population (Szolnoki and Hoffmann 2013), in this research, since the professionals are geographically dispersed and travel within the UK or abroad was not feasible, the researcher did not consider this method. Moreover, postal or telephone questionnaires were not considered because of the additional costs involved and the limited number of potential respondents within the UK. Likewise, for the respondents within the UK, it was possible to send the survey to their email addresses (see Appendix C for the sample email). Notwithstanding, whenever the email address of a potential respondent was not available, the messaging facility of LinkedIn was used. In total, 481 invitations are sent to the experts to complete the online questionnaire.

4.7 Response rate

To increase the response rate, the author sent out several reminders in fixed intervals to the potential respondents. As advised by Saunders et al. (2016), the first reminder was sent one week after sending the questionnaire link to the recipient. A second reminder was sent after three weeks, and, a third follow-up email was sent after another two weeks. After all the above steps, the total number of received questionnaires reached 90, yielding a response rate of 18.7%. After careful review of the responses, 18 questionnaires disqualified due to either being irrelevant to the focus of the study or being incomplete. Based on the above, 72 questionnaires are used for statistical analysis and predictive model development.

While the above response rate may look rather low, considering the nature of the data collection instrument (online questionnaire), the level of qualification expected from the respondents (Section 4.3) and that there is no way to force the experts to complete the questionnaire, this value is still within the threshold for online surveys performed outside an organisation (Saunders, Lewis, and Thornhill 2016). According to (Baruch and Holtom 2008), several reasons are contributing to the nonresponses including being too busy (28%), irrelevant (14%), unavailability of the return address (12%) (in the case of mail surveys), and company policy restrictions on participation (22%) (Johnson and Owens 2003). The first justification (being too busy) is a primary cause of nonresponse in this research based on several refusal emails received by the researcher declining due to the high workload. However, other reasons like the company's privacy policy requirements and security reasons were among other justifications provided by some respondents who declined to participate. One respondent, who is the Head of Agency at an overseas company, refused to take part with the following justification: “Sorry, but our codes of conduct for incoming items prevent us from clicking on unknown links”. Therefore, in the case of this research, being too busy, and company policy restrictions are identified to be the reasons for nonresponses.

This chapter seeks the professional opinion of building experts with experience in reusing load-bearing building components. In this research, an expert is defined as someone with six years and above of professional experience in the building sector. Hence, if a respondent does not match the required profile, he/she is automatically filtered out and could not complete the survey. As discussed in Section 4.4.1, these questions are highly technical and, the respondent needs in-depth knowledge about the element to complete the questionnaire. It means that if the respondent has not done this practice recently, then he/she may need to refer back to the archives or consult with colleagues, which can result in nonresponse. Considering the above facts about the potential respondents and the fact that the data collection happens at the international level (Section 4.3), the target population becomes a hard-to-reach population (Harzing 1997). According to Harzing (1997), the international sampling frame decreases the response rates of a survey.

As shown in Table 4.2, 67.7% of the respondents are managers and top managers, 10.8% are architects, 7.7% are engineers, 4.6% are consultants, 4.6% are deconstruction experts, and others are reuse experts and construction waste prevention experts (7 respondents did not answer to this question). According to Table 4.3, 39.1% of the respondents work in deconstruction/demolition companies, 29% in consultancy, 8.7% in contracting organisations, and the rest in universities or supplier/stockiest firms (3 respondents did not answer this question) (Table 4.3). The respondents also worked in the construction sector from 6-10 years (33.3%) to over 40 years (11.1%), with 66.7% having above 10 years of experience (Table 4.4).

Table 4.2 Position of the research respondents

Position of the respondent	Frequency	Percentage (%)
Architect	7	10.8
Consultant	3	4.6
Deconstruction expert	3	4.6
Designer	1	1.5
Engineer (Civil/Structural)	5	7.7
Manager (e.g. project managers, design managers, marketing manager, etc.)	15	23.1
Reuse expert	1	1.5

Position of the respondent	Frequency	Percentage (%)
Top manager (e.g. head managers, owner of companies, executive managers, managing director, CEO, etc.)	29	44.6
Waste prevention specialist	1	1.5
The above percentages are based on 65 respondents.		

Table 4.3 Type of organisation the research respondents work in

Type of the organisation	Frequency	Percentage (%)
Client	3	4.3
Consultancy (architectural, structural, etc.)	20	29
Contractor	6	8.7
Deconstruction/Demolition	27	39.1
Supplier/Stockiest	5	7.2
University/Academic institution	2	2.9
Other	6	8.7
The above percentages are based on 69 respondents.		

Table 4.4 Years of experience of the research respondents in the construction sector

Years of experience	Frequency	Percentage (%)
6-10	24	33.3
11-15	10	13.9
16-20	10	13.9
21-25	8	11.1
26-30	9	12.5
31-35	3	4.2
36-40	0	0
over 40	8	11.1

4.8 Missing data analysis

Missing values or item nonresponse in survey research happens when a respondent does not provide an answer to one or more questions of a questionnaire (Allison 2001, Graham 2012). While there are several reasons for item nonresponse in survey research (Graham 2012), the missing values can have a significant impact on the conclusions of the research (Graham 2009). It should be noted that almost all of the statistical and machine learning methods do not consider missing values in a dataset while analysing research data. Therefore, item nonresponse can decrease the statistical power of the research in testing the null hypothesis correctly. Moreover, it can cause bias in both the dependent (DV) and independent variables (IV), negatively affect the representativeness of the sample, and result in inaccurate conclusions (Kang 2013, Allison 2001). It is because as defined by (Little and Rubin 2019), ‘missing data are unobserved values that would be meaningful for analysis if observed; in other words, a missing value hides a meaningful value‘. Therefore, missing data analysis is a crucial aspect of any research (Graham 2009).

There are different assumptions on the pattern of item nonresponse in survey research, and the researcher should identify the dominant assumption before adopting any technique to rectifying the missing values. According to Graham (2012), depending on reasons for the item nonresponse, the missing data can be categorised into missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). Allison (2001) considers the first two missing data mechanisms as ‘ignorable’ and the third one as nonignorable. In the case of MCAR, the probability of a missing value on any of the variables (both dependent and independent) is completely at random and is unrelated to the value of the variable itself, nor any other variable in the dataset. This pattern is ideal for any research because there is no bias in the analysis of the estimated values (Kang 2013, Allison 2001). However, MCAR does not happen all the time (Allison 2001). A more realistic missingness mechanism is when data are missing at random (MAR) (Allison 2001). In this case, the probability of missing data on a given variable can be determined based on one or more predictors (variables), and within the categories of each predictor, the probability of missing data is at random and is unrelated to the value of the variable itself (Allison 2001). If the missingness is neither MCAR nor MAR, it is said that the data is not missing at random (NMAR) (Murphy 2012). In this case, the missing data mechanism must be either modelled (following the definition of ‘missing data creation model’ in (Graham 2012) and considering that item nonresponse mechanism in NMAR is nonignorable) or deleted to avoid bias (Kang 2013, Allison 2001).

There are various techniques to handle missingness including listwise deletion, pairwise deletion, dummy variable adjustment, marginal mean imputation, regression imputation, maximum likelihood estimation (MLE), and multiple imputation (MI) (Kang 2013, Allison 2001). According to Allison (2001), if the missing data mechanism depends on the values of the independent variables only, then the listwise deletion technique results in unbiased estimations. (Allison 2001, Kang 2013, Graham 2012) advise using MLE and MI techniques to tackle the missing values in survey research. According to (Graham and Schafer 1999), the MI technique provides very good results with sample sizes as small as 50 cases. In this research, no dependent variable is missing. Moreover, the respondents provided their answers to the questionnaire before reaching the dependent variables, meaning that if any independent variable was missing, it was regardless of their response to the dependent variables (Appendix C.2). Therefore, listwise deletion was initially used to remove responses with high number of missing values. Next, MI technique was used to estimate the value of the remaining missing values in the dataset.

Before implementing MI, and after performing the listwise deletion, the SPSS expectation-maximisation (EM) algorithm (a type of MLE method (Dempster, Laird, and Rubin 1977)) was used to evaluate the mechanism of missingness. In this study, 90 completed questionnaires were initially received (Section 4.7). After a careful review, 7 responses were found irrelevant (talking about material reuse such as recycled concrete aggregate, etc.), and 11 were removed due to the high number of missing values (above 20% of the independent variables were missing), leaving 72 questionnaires for further analysis. This research aims to develop models to predict the reusability of the structural elements of a building in terms of technical, economic, and social aspects (Section 1.6). So, the survey collected the necessary data to fulfil the aims of the study. In the next stage, and to analyse the responses, the questionnaires were split into three separate datasets (each containing 72 cases) focusing on technical, economic, and social reusability of the structural elements (the dependent variables) based on the relevant independent variables. These datasets are called technical (TEC), economic (ECO), and social (SOC) focusing on technical, economic, and social aspects of this study, respectively. Then, Little's MCAR test using IBM SPSS version 25 with EM algorithm was performed for the complete list of dependent and independent variables for each dataset. The null hypothesis was that the data were missing completely at random. To reject the null hypothesis, the chi-square should be significant at 0.05. However, the significance of the chi-square test for all datasets was higher than the 0.05 threshold (insignificant), meaning that the null hypothesis cannot be rejected, and the missing values were MCAR (see Appendix D for the test results).

The TEC dataset contains 42 IV and 1 DV and has 2% of values missing, involving 29% of respondents. The ECO dataset contains 12 IV and 1 DV and has 1% of values missing, involving 11% of respondents. The SOC dataset contains 10 IV and 1 DV and has 2% of values missing, involving 18% of respondents. In this research, to perform MI for the missing variables, R system packages 'MissMDA' and 'mice' were employed (see Appendix E, Script E.1 for a copy of the R code used) (Josse and Husson 2016, Audigier, Husson, and Josse 2017, van Buuren and Groothuis-Oudshoorn 2011, R Core Team 2020). In the data augmentation process, as advised by Allison (2001), both independent and dependent variables were used to impute the missing independent variables. Each dataset was handled separately and the updated datasets with no missing values were extracted for further analysis.

4.9 Validity and reliability

The quality of the collected data is one of the major concerns of a researcher. Therefore, the data collection instrument, as well as the collected data, should be verified in terms of validity and reliability.

In this study, the validity of the questionnaire is evaluated through self-checking the questionnaire (Section 4.4.3) and the pilot study (Section 4.5). According to the feedback forms (Section 4.5), the average time for answering the questions was less than 15 minutes, which is a reasonable time for completing a technical questionnaire. Moreover, there were a few comments about the clarity of the instructions that were rectified. In Section E of the questionnaire, the definitions were initially after the questions; however, most of the respondents suggested moving the definitions before the questions. Furthermore, one respondent recommended clarifying the focus of the questionnaire on the front page of the survey. And finally, few respondents recommended making shorter questions. However, all of the respondents agreed that there were no missing questions/sections in the questionnaire. After incorporating the comments, the questionnaire was shared with the available respondents (those known by the researcher and were accessible), and they all agreed with the final design and content of the online survey.

The reliability of the responses refers to the capacity of the results to be reproduced by other researchers. Therefore, reliability is linked with the respondents being representative of the targeting population. While one of the indicators of reliability is the response rate, it is still possible to have a low response rate with a sample that is representative of the population (Dusek, Yurova, and Ruppel 2015). Because this study targets experts with previous experience in reusing the structural elements of a building, the chance that the questionnaire is completed by an inexperienced respondent is low. Moreover, at the beginning of the online questionnaire, and after elaborating the focus of the research, the respondent should answer a question about his/her previous experience with reused building structural elements. All the respondents confirmed that they have this experience. Hence, they are representative of the target population.

Another option to check the reliability of a questionnaire is through checking the internal consistency of the responses by calculating Cronbach's alpha value (Saunders, Lewis, and Thornhill 2016). In this study, this is done by evaluating the consistency of the independent variables by calculating Cronbach's alpha value using SPSS. If the Cronbach's alpha value is equal to or greater than 0.7, then the combination of the questions measures the same thing

(Saunders, Lewis, and Thornhill 2016). Nevertheless, while 0.7 depicts acceptable reliability, higher values up to 0.9 are more desirable (Tavakol and Dennick 2011).

This study assesses the reusability of the structural elements; hence, to calculate Cronbach's alpha, all the items identified to have a potential impact on the dependent variables were considered. This has been performed and shown in Tables 4.5, 4.6, and 4.7 for TEC, ECO, and SOC datasets, respectively (Also discussed later in Section 4.11). These tables present the overall Cronbach's alpha for each set of questions (depending on the scale used, see Section 4.4.2). The value of Cronbach's alpha when the item is deleted is listed against each variable in these tables, as well. When this value is higher than the overall Cronbach's alpha value, it means that the variable does not contribute to the overall reliability of the questionnaire, and by its removal, the internal consistency of the data will increase (Field 2009). According to Field (2009), if a questionnaire is reliable, the removal of a variable should not considerably affect the reliability of the survey.

For the TEC dataset, questions in Section B consist of details about the element and the state of the component in its new application (Section 4.4.1). Therefore, only questions B7 to B11 are checked for their internal consistency. Because questions B7 to B9 have different scales than questions B10 and B11, their internal consistencies were checked separately. It is because, as advised by Field (2009), for checking the reliability of a questionnaire, if different subscales exist in a survey (Table 4.1), the Cronbach's alpha value should be calculated separately for these subscales. According to Table 4.5, while the Cronbach's alpha for questions B7 to B9 is above the minimum threshold value of 0.7, by deleting question B9, the reliability of this section increases slightly by 0.03, which according to Field (2009) is not substantial and can be ignored. However, questions B10 & B11 have a Cronbach's alpha of 0.263, which is below the minimum acceptable value of 0.7 (Field 2009). Hence, questions B10 and B11 will not be involved in model building. In Section C, removing question C12 improves the internal consistency of the TEC dataset slightly by a 0.001 increase in the overall Cronbach's alpha value. Nevertheless, since this increase is negligible, these questions will be included in model building. For the ECO dataset, the overall Cronbach's alpha value for Section D is 0.916 and removing question D1 increases the reliability of the dataset by 0.01, which is negligible (Table 4.6). Therefore, D1 will also be included in the model development. Regarding the SOC dataset (Table 4.7), Cronbach's alpha value is above 0.8 for both Sections C & D, and the removal of none of the variables in this dataset will not increase the reliability.

4.10 Preliminary statistical analysis of the survey

In this section, descriptive statistics are used to rank the technical (TEC dataset), economic (ECO dataset), and social (SOC dataset) factors based on the mean value of the variables. Descriptive statistics are a set of statistical approaches, including measures of central tendency (mean, median, mode, etc.) and measures of variability (standard deviation, variance, minimum/maximum, skewness, etc.), to quantitatively summarise a given data set (Field 2009, Bryman and Cramer 2005).

The results of the descriptive statistics for the TEC, ECO, and SOC datasets are presented in Tables 4.5 to 4.7 and discussed in subsections 4.10.1, 4.10.2, and 4.10.3, respectively. While according to (Stevens 1946) the permissible statistics for ordinal scales (questions B7 to B11, Section C, D, and E of the questionnaire, See Appendix B.2) are the median and percentiles, other statisticians such as (Lord 1953, Labovitz 1970, Sauro and Lewis 2016) allow the use of statistics applicable to interval and ratio scales for ordinal values such as Likert scales used in this research. It is noteworthy that the latter has been adopted in this research.

4.10.1 Descriptive statistics for TEC dataset

Figures 4.3 to 4.8 show the distribution of the answers provided by the respondents to questions B1 to B6.

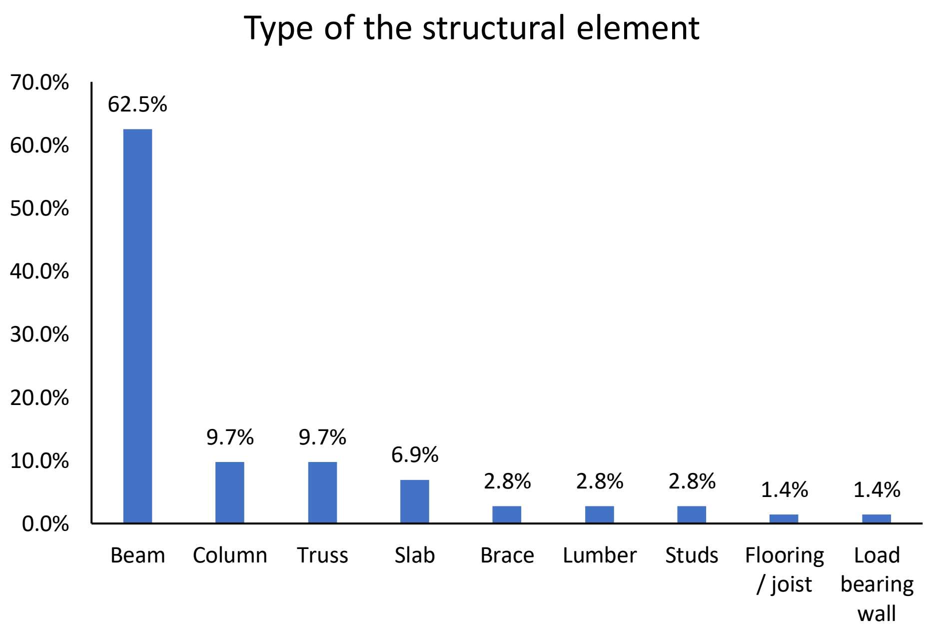

Figure 4.3 Type of the structural element used to complete the questionnaire (question B1)

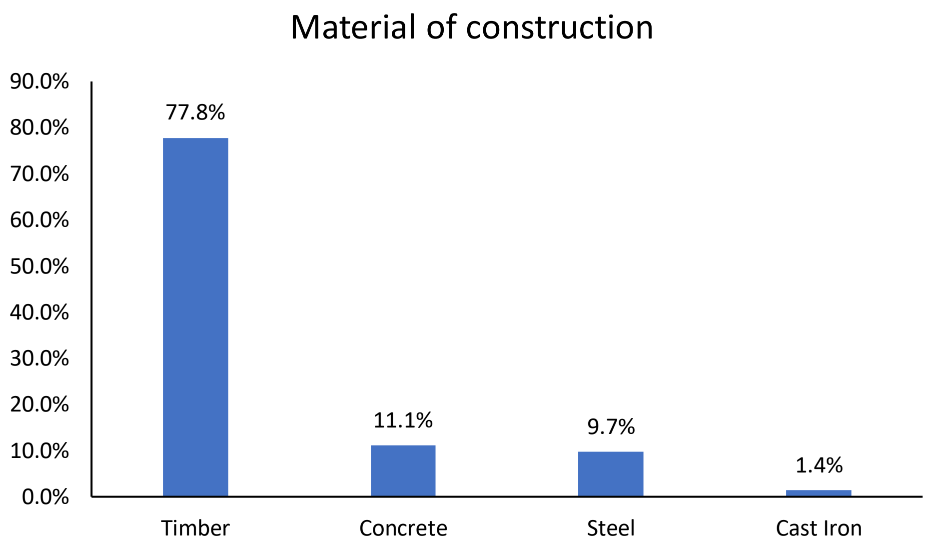

Figure 4.4 Material of the structural element used to complete the questionnaire (question B2)

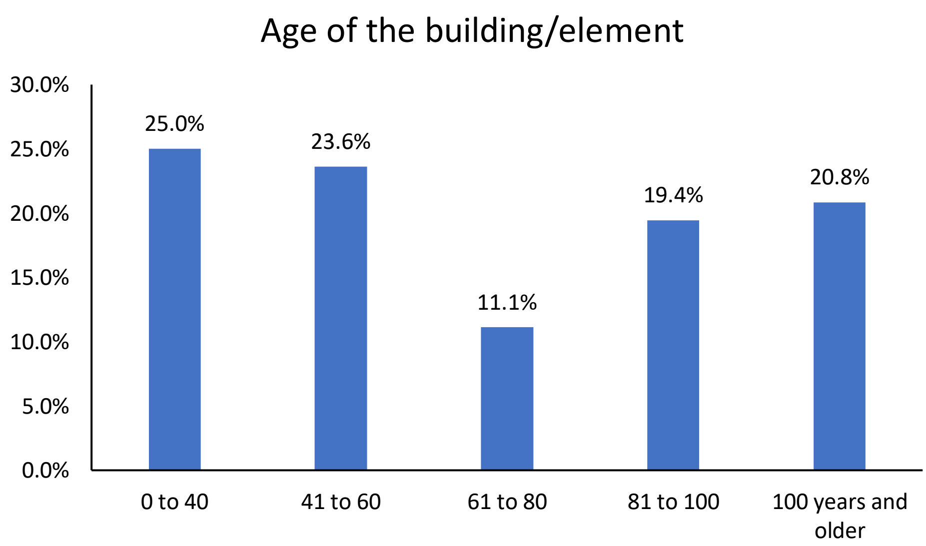

Figure 4.5 Age of building/element (question B3)

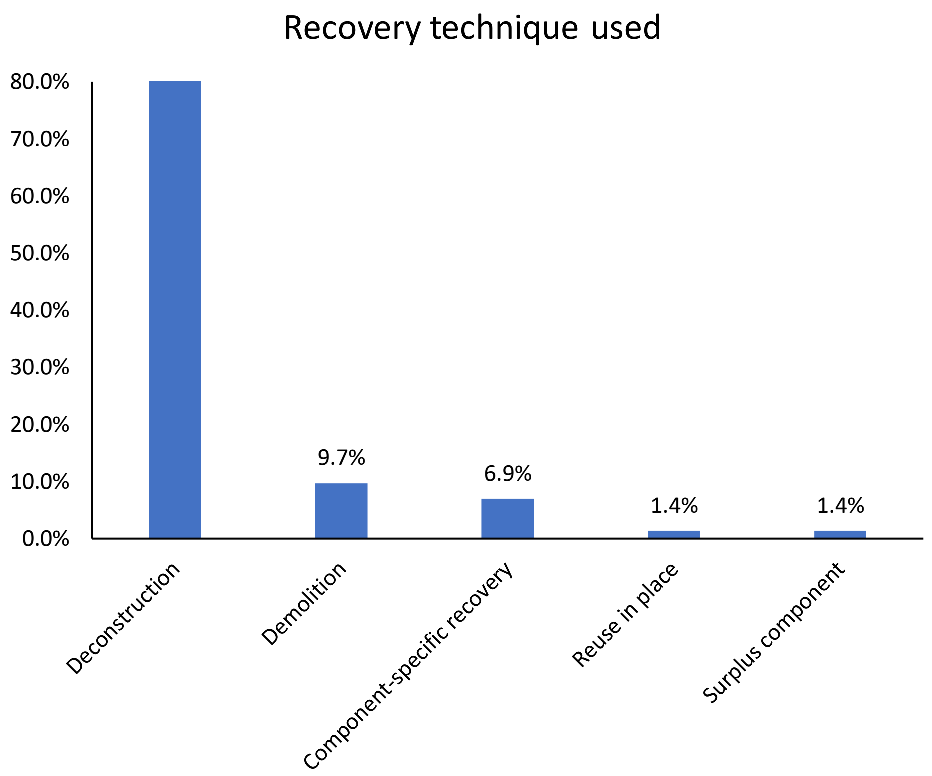

Figure 4.6 The recovery technique used to recover the element (question B4)

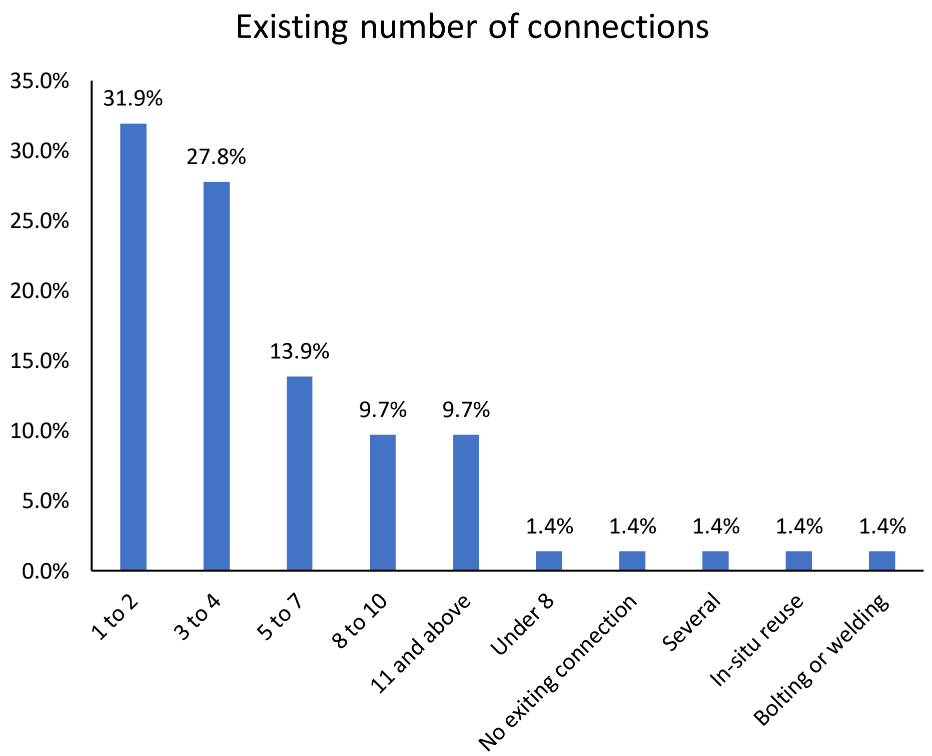

Figure 4.7 The number of existing connections of the recovered element (question B5)

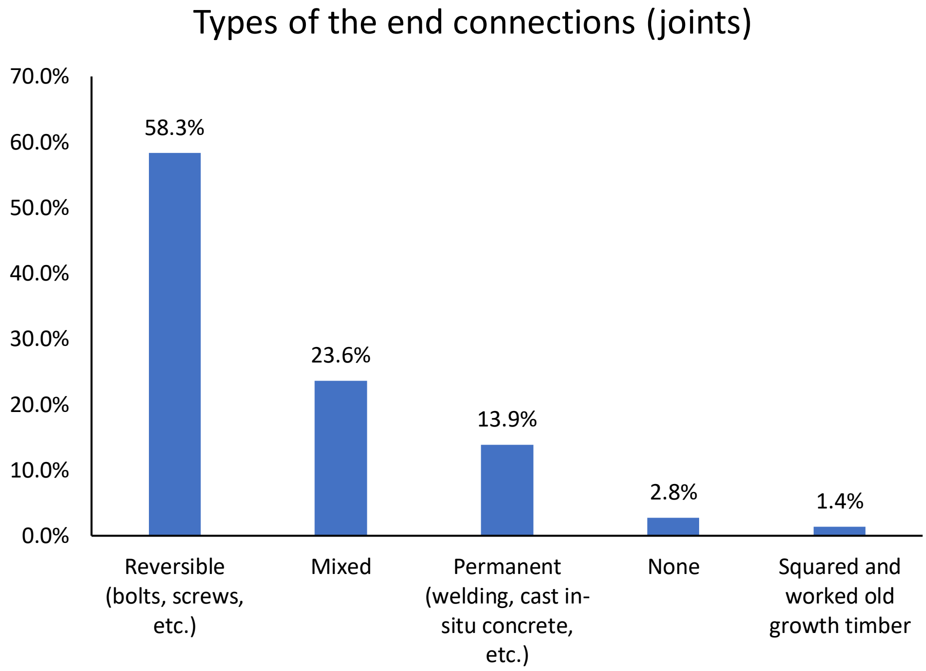

Figure 4.8 Type of the end-connections of the recovered element (question B6)

The results of the descriptive statistics for the TEC dataset for questions B7 to B11 and Sections C & D of the questionnaire is available in Table 4.5. The mean, median, and standard deviations are developed using SPSS 25 version. According to this table, for Section C, the following barriers are identified as the top technical factors negatively affecting the reusability of the building structural components.

1- Lack of certificates of quality for the element when acquired
2- Damage during deconstruction/demolition
3- Lack of standards to certify the element
4- The potential risk associated with the structural integrity
5- Damage due to water penetration/presence

Moreover, for Section D of this dataset, the following are identified as the top-ranked barriers:

1- Matching the original design with the dimensions of the reused element
2- Changes in the design codes (BS codes to Eurocodes, etc.)
3- CE marking
4- Matching the original design with the strength of the reused element
5- Presence of fire protection on the element

4.10.2 Descriptive statistics for ECO dataset

On the economic dimension, Table 4.6 shows the mean, median, and standard deviations of the factors affecting reuse. Section C comprises only one variable; hence, the rankings were only performed for Section D.

According to Section D, the purchasing price of the reused element (variable D1) is the only driver to reuse, and the other variables act as reuse barriers. The rankings of these barriers are as follows:

1- Cost of testing
2- Cost of insurance
3- Storage cost
4- Cost of refurbishment (sandblasting, treatment, etc.)
5- Cost of design with the reused element

4.10.3 Descriptive statistics for SOC dataset

On the social dimension, and based on the SOC dataset, Sections C and D contain two and eight variables, respectively (Table 4.7). The rankings of the barriers in Section C are as follows:

1- Potential liability risks
2- Potential health and safety risks

In Section D, among the eight variables, five are drivers, and three are barriers. The rankings of these factors are as follows:

Top drivers:

1- Perception of the client/top management team about the element
2- Perception of the end-users (when it is not the client) about the element
3- Perception of the designers about the element
4- Visual appearance
5- Perception of the builders/contractors about the element

Top barriers:

1- Changes in the health and safety regulations (fire, etc.)
2- Perception of the stockiest about the element
3- Perception of the regulatory authorities about the element

4.11 Test for significant difference

According to Section 4.2, the unit of analysis of this study is the structural elements of a building. Therefore, in this section, a non-parametric test is used to evaluate if there are statistically significant differences between the types of structural elements (question B1) regarding the ordinal independent and dependent variables asked in the questionnaires. Using SPSS version 25, the Kruskal-Wallis H test is performed on each of the TEC, ECO, and SOC databases of this study (Section 4.8) at a 5% significance level (Field 2009, Bryman and Cramer 2005). The results of these tests are discussed in subsections 4.11.1, 4.11.2, and 4.11.3, respectively. The null hypothesis is that there is no difference between the groups of structural elements. The purpose of this test is to make sure that combining the responses for all the elements for further analysis will not affect the overall reliability of the TEC, ECO, and SOC datasets.

4.11.1 Kruskal-Wallis H test on TEC dataset

The Kruskal-Wallis H test was used on the technical factors (TEC dataset) to determine if the type of the element affects the scores provided for the factors affecting the reusability of the structural components. As presented in Table 4.5, the Kruskal-Wallis H test results indicate that none of the p-values of the technical factors is less than 0.05 and that there is not enough evidence to reject the null hypothesis. It means that the TEC dataset can be safely used to develop BSE-RPMs, which is the fourth objective of this study (Section 1.6).

4.11.2 Kruskal-Wallis H test on ECO dataset

The Kruskal-Wallis H test was performed to understand whether the variables affecting the economic reusability of the building structural elements, measured on an ordinal scale (Section 3.3.4), differed based on the type of the component. The results indicate that there is no statistical difference between the groups of the structural elements at a 95% confidence level, which means that the null hypothesis discussed in Section 4.11 is valid (Table 4.6). Therefore, using the combination of the responses for further developing the predictive models does not affect the overall reliability of the ECO dataset.

4.11.3 Kruskal-Wallis H test on SOC dataset

Like the TEC and ECO datasets, the same non-parametric test was performed on the SOC dataset. The results indicate that the p-value of all applicable independent variables is more than 0.05, which means that there is no significant difference in the independent variables between the various group of the structural elements (Table 4.7). In other words, the data can be combined and used for further development of the predictive models.

Table 4.5 Preliminary statistical analysis of the survey: TEC dataset

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test p-value
	Section B						
B	**Details about the reused structural element**						
	Overall Cronbach's alpha for Section B, questions B7 to B9 = 0.780						
B7	The structural element is serving the same purpose (i.e. as a beam, slab, column, etc.) in its new installation as in its previous installation.	0.648	4	3.71	1.192	2	0.480
B8	The cross-section/thickness dimensions of the structural element in its new installation are equal or nearly equal to the cross-section/thickness dimensions of the element in its previous installation.	0.641	4	3.93	1.012	1	0.388
B9	The length dimensions of the structural element in its new installation are equal or nearly equal to the length dimensions of the element in its previous installation.	0.814[a]	3	3.25	1.207	3	0.085
	Overall Cronbach's alpha for Section B, questions B10 to B11 = 0.263						
B10	The amount of load supported by the structural element in its new installation compared to the amount of load supported by the element in its previous installation.	N/A[b]	2	2.31	0.762	2	0.720

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test p-value
B11	The life expectancy of the structural element in its new installation compared to the life expectancy of the element in its previous installation.	N/A[b]	3	2.88	0.821	1	0.386
	Section C						
C	**Factors affecting the reusability of the structural element**						
	Overall Cronbach's alpha for Section C = 0.891						
C1	Damage during deconstruction/demolition	0.887	3	2.99	1.295	2	0.364
C2	Damage due to fatigue	0.888	4	3.85	1.070	14	0.276
C3	Damage due to fire	0.889	5	4.26	1.245	20	0.406
C4	Damage during transportation	0.888	5	4.31	0.944	21	0.635
C5	Damage during storage	0.889	5	4.21	1.061	19	0.116
C6	Damage due to the type of joints	0.885	4	3.78	1.178	10	0.185
C7	Damage due to corrosion	0.884	5	4.19	1.121	18	0.307
C8	Damage due to frost	0.888	5	4.58	0.707	24	0.213
C9	Damage due to water penetration/presence	0.885	4	3.53	1.267	5	0.405
C10	Damage during refurbishment (nail removal, etc.)	0.887	4	3.85	1.016	13	0.342
C11	Damage due to exposure to wind, acidic rain, etc.	0.890	5	4.42	0.946	23	0.499
C12	Damage caused by living organisms (termite, bacterial attack, etc.)	0.892[a]	4	3.88	1.310	16	0.919
C13	Damage due to earthquake	0.891	5	4.85	0.494	25	0.559
C14	Damage due to impact	0.888	5	4.35	1.064	22	0.160
C15	Damage due to post-production modifications (e.g. holes, etc.)	0.888	4	3.76	1.081	9	0.322
C16	Lack of certificates of quality for the element when acquired	0.884	3	2.97	1.472	1	0.505
C17	Lack of standards to certify the element	0.886	3	3.04	1.505	3	0.652

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test p-value
C18	Lack of the original drawings	0.881	5	3.76	1.477	8	0.130
C19	Lack of the original design calculations	0.885	5	3.81	1.469	12	0.351
C20	Lack of earlier certificates (inspection, material, etc.)	0.883	4.5	3.72	1.465	7	0.273
C21	Lack of traceability of the element	0.882	5	3.88	1.433	15	0.324
C25	The potential risk associated with the structural integrity	0.886	4	3.43	1.276	4	0.090
C26	The potential risk of damage to the machinery (nails in timber, etc.)	0.885	4	3.79	1.113	11	0.572
C27	A potential problem with collateral warranties	0.888	4	3.97	1.162	17	0.167
C28	Presence of hazardous, banned or contaminating coatings	0.885	4	3.6	1.241	6	0.875
	Section D						
D	**Other factors affecting the reusability of the structural element**						
	Overall Cronbach's alpha for Section D = 0.847						
D18	Presence of fire protection on the element	0.820	3	2.72	1.129	5	0.325
D19	Changes in the design codes (BS codes to Eurocodes, etc.)	0.843	3	2.63	1.106	2	0.552
D21	CE marking	0.814	3	2.64	1.104	3	0.884
D22	Matching the original design with the dimensions of the reused element	0.822	3	2.53	1.138	1	0.282
D23	Matching the original design with the strength of the reused element	0.795	3	2.71	1.261	4	0.761
D24	Other design challenges with the reused element	0.836	3	2.79	1.100	6	0.341

[a] Since the Cronbach's alpha value increases negligibly, this variable will be used for model development.
[b] At least three variables are required to calculate this value.

Table 4.6 Preliminary statistical analysis of the survey: ECO dataset

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test
	Section C						
C	**Factors affecting the reusability of the structural element**						
	Cronbach's alpha cannot be calculated because only one variable exists						
C24	Potential financial risks	N/A[a]	4	4.01	1.132	1	0.419
	Section D						
D	**Other factors affecting the reusability of the structural element**						
	Overall Cronbach's alpha for Section D = 0.916						
D1	The purchasing price	0.927[b]	4	3.68	1.265	1	0.684
D2	Cost of insurance	0.907	3	2.69	0.959	2	0.664
D3	Cost of testing	0.910	3	2.58	1.058	1	0.615
D4	Cost of refurbishment (sandblasting, treatment, etc.)	0.905	3	2.81	1.043	4	0.746
D5	Cost of design with the reused element	0.903	3	2.82	1.079	5	0.523
D6	Storage cost	0.903	3	2.78	1.213	3	0.634
D7	Transportation cost	0.906	3	2.82	1.179	6	0.828
D8	Cost of labour	0.910	3	2.94	1.112	10	0.448
D9	Cost of fabrication	0.904	3	2.89	1.082	9	0.459
D10	Cash flow (need to purchase the element early, etc.)	0.908	3	2.86	1.154	8	0.727
D25	Sourcing/procurement process	0.903	3	2.83	1.187	7	0.931

[a] At least three variables are required to calculate this value.
[b] Since the Cronbach's alpha value increases negligibly, this variable will be used for model development.

Table 4.7 Preliminary statistical analysis of the survey: SOC dataset

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test
	Section C						
C	**Factors affecting the reusability of the structural element**						
	Overall Cronbach's alpha for Section C = 0.857						
C22	Potential liability risks	N/A[a]	3	3.24	1.399	1	0.572
C23	Potential health and safety risks	N/A[a]	4	3.81	1.274	2	0.816
	Section D						
D	**Other factors affecting the reusability of the structural element**						
	Overall Cronbach's alpha for Section D = 0.899						
D11	Perception of the client/top management team about the element	0.884	4	3.46	1.162	1	0.955
D12	Perception of the designers about the element	0.879	4	3.36	1.248	3	0.322
D13	Perception of the builders/contractors about the element	0.877	3	3.10	1.302	5	0.595
D14	Perception of the end users (when it is not the client) about the element	0.880	4	3.44	1.433	2	0.414
D15	Perception of the stockiest about the element	0.891	3	2.71	0.985	2	0.797

Section / Question	Variables	Cronbach's alpha if item deleted	Median	Mean	Standard Deviation	Rank	Kruskal-Wallis H test
D16	Perception of the regulatory authorities about the element	0.894	3	2.97	1.150	3	0.638
D17	Visual appearance	0.887	3	3.32	1.309	4	0.157
D20	Changes in the health and safety regulations (fire, etc.)	0.895	3	2.65	0.981	1	0.434
[a] At least three variables are required to calculate this value.							

4.12 Chapter summary

Using the results of the review of the literature (including a systematic review, see Chapter 2), a self-completed online questionnaire was developed, pilot tested and shared with the professionals with reuse experience to collect the data necessary to fulfil the aim of this research. Since this study seeks the experts' opinions to identify the reusability factors, it followed the purposive sampling technique in choosing the professionals because there is no way to have access to all the population to perform probability sampling. Therefore, to develop a sampling frame, the list of top 100 demolition companies worldwide along with the list of construction companies experienced in deconstruction and reuse was used to locate the experts with reuse experience. To do so, companies' website and LinkedIn were employed, and all the located experts were contacted through email or LinkedIn direct messaging service. Moreover, to facilitate the sampling technique and to access those who are not reachable using the internet, companies' websites, or LinkedIn, the snowballing technique was used as well. In total, 481 questionnaires were sent, and 90 responses were received, yielding a response rate of 18.7%. After the evaluation of the responses, 72 valid questionnaires were identified and used for a preliminary statistical analysis.

Since both above techniques fall under non-probability sampling, there might be a concern that the results of this study cannot be generalised. However, because the unit of analysis and the unit of generalisation are the structural elements of a building, this concern is not valid. It is because the generalisations happen at the elements' level, and not the experts' level. Moreover, while there might be a small level of subjectivity in answering the questions by the respondents, the variation in responses would be negligible because the respondents are asked to provide facts about a structural element they reused in the past. Nevertheless, checking the reliability of the responses using Cronbach's alpha value revealed high consistency and reliability of the received questionnaires.

Chapter 5 – Predictive models

5.1 Chapter introduction

This chapter discusses all steps taken to develop the BSE-RPMs based on the results of the quantitative study elaborated in the previous chapter. This chapter focuses on the third and fourth objectives of this study (Section 1.6). Sections 5.2 discusses the process of oversampling using the Synthetic Minority Oversampling Technique (SMOTE) to address the class-imbalances problem in the dependent variables (responses) of the TEC, ECO, and SOC datasets. Section 5.3 describes the process of dividing the datasets into training and testing sets and the justifications behind it. Section 5.4 deals with the third objective of this project and focuses on determining the best combination of the identified factors to develop the BSE-RPM. And Section 5.5 applies the results of Sections 5.2 to 5.4 and focuses on the development of the predictive models using various powerful machine learning methods employing the software 'R' (version 4.0.2) (R Core Team 2020). Therefore, Section 5.5 partially fulfils the fourth objective of this study (Section 1.6), whereas, the selection of the best-practice BSE-RPM is performed in Chapter 6. Finally, Section 5.6 summarises this chapter.

5.1.1 Justification for using supervised machine learning techniques

Statistical analysis of collected data in Chapter 4 provides an overview of the most significant factors affecting the reuse of load-bearing building components. However, to determine the reusability of these elements based on the affecting variables, which is the aim of this research (Section 1.6), it is essential to learn from the collected data. The intention of learning from data is to uncover the relationships among the affecting variables, "and understand what data says" (Hastie, Tibshirani, and Friedman 2009). The collection of data analysis methods that automatically learn from data is called machine learning (Murphy 2012). According to Murphy (2012), machine learning is defined as "a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!)." Since this research aims to predict the reusability of the load-bearing building components given factors affecting reuse, it adopts the prediction aspect of machine learning, which is known as supervised machine learning techniques (Hastie, Tibshirani, and Friedman 2009, Murphy 2012).

5.2 Oversampling

This study comprises three datasets, and each dataset has a unique dependent variable (response). The responses for the TEC, ECO, and SOC datasets are the technical reusability (E1), economic reusability (E2), and social reusability (E3), respectively (please refer to Appendix C, Section C.2, for a copy of the questionnaire used in this research). E1, E2, and E3 are based on a five-point Likert scale (Table 4.1). This study aims to develop a model to predict the reusability of the structural elements of a building. Therefore, following the approach adopted by (Jang et al. 2015), the responses were converted to a binary scale with 0, non-reusable, and 1, reusable. The responses with Likert scale values of 1 to 3 are considered non-reusable (represented by 0), and the remaining responses (Likert scale values 4 and 5) are identified as reusable (converted to 1). Consequently, the dependent variables are transformed from multi-scale responses to binary responses. While this conversion simplifies the interpretation of the results of the predictive models, the proposed methodology in this research can be conveniently generalised to multi-class response variables. Instead of relying on five points to decide if a component is reusable or not, the stakeholders have a straightforward basis for deciding on the fate of a structural element. Likewise, for a supervised machine learning method to perform effectively with a multi-class response, a large sample size is required. However, since the reuse of the load-bearing components of a building is not a widespread practice, collecting more data was not possible. Moreover, the uncertainties in the assessment of the reusability factors (features or independent variables), which is based on expert opinion, limits the effectiveness of a multi-scale response.

After converting the multi-scale responses to binary values, it was observed that the new binary classes were considerably imbalanced. In the case of the TEC dataset, 24 elements are non-reusable, and 48 are reusable. In the case of the ECO dataset, this imbalance changes to 22 non-reusable and 50 reusable components. And in the case of the SOC dataset, these figures are 16 and 56 for non-reusable and reusable elements, respectively. From the above figures, it is evident that the datasets are unbalanced and contain more reusable components than non-reusable elements. Consequently, after the dataset is divided into training and testing sets, due to the different number of reusable and non-reusable elements in the original observations, the training and testing sets will also have imbalanced responses. It can be argued that the initial data collection could be continued to have more balanced responses; however, due to the time constraints, as well as the limitations explained in Sections 4.3 and 4.7, this option was not practical. Moreover, even if the data collection continues, since the respondents are free to choose any structural component with any level of reusability to complete the questionnaire, it is impossible to guess the outcome of the new survey, which might end up with a similar imbalanced dataset.

According to (Torgo 2016, He and Ma 2013, Fernández et al. 2018), imbalanced datasets negatively affect predictive methods during model development and performance assessment stages. One of the metrics used to assess the performance of a machine learning method is its accuracy in predicting correct answers (Chawla et al. 2002). When one class is dominant (due to the imbalance in the dataset), the predictions are inherently biased towards that, yielding an unrealistic accuracy (Torgo 2016). It is because the predictive methods look for the rules and regulations in a dataset, and imbalanced datasets make this task difficult (Torgo 2016).

In supervised machine learning, different methods can be used to address the issues caused by an imbalanced dataset (He and Ma 2013, Fernández et al. 2018). These include cost-sensitive learning (manipulating the threshold values, etc.), pre-processing the imbalanced dataset (oversampling, under-sampling, SMOTE, etc.), algorithm level approaches (active learning, kernel modifications, etc.), and ensemble learning (cost-sensitive boosting, etc.) (Fernández et al. 2018). It should be noted that according to Fernández et al. (2018), there is no best strategy to deal with the issues caused by imbalanced datasets. For a comprehensive discussion over various methods to handle imbalanced datasets, please refer to (Fernández et al. 2018).

In this study, the oversampling technique developed by Chawla et al. (2002) is employed to pre-process the datasets and minimise the class imbalance impact. This technique is known as Synthetic Minority Over-sampling Technique (SMOTE). Unlike other oversampling techniques that rely on replacement in data space (Japkowicz 2000), SMOTE creates synthetic examples of the minority class in feature space using the k-nearest neighbours (KNN) algorithm (with the default value for k=5) without duplicating any data (Chawla et al. 2002, Bischl et al. 2016).

In this study, following the approach adopted by (Agrawal et al. 2018, Naseriparsa and Kashani 2013, Taft et al. 2009, Al-Bahrani, Agrawal, and Choudhary 2013), the SMOTE was performed on the TEC, ECO, and SOC datasets. The results of oversampling on these datasets are presented in Table 5.1. A comparison between the oversampled and the original datasets reveals the following. For the TEC dataset, the imbalance has improved from 34% (non-reusable) and 66% (reusable) to 50% (non-reusable) and 50% (reusable). For the ECO dataset, the imbalance has improved from 31% (non-reusable) and 69% (reusable) to 51% (non-reusable) and 49% (reusable). For the SOC dataset, the imbalance has improved from 23% (non-reusable) and 77% (reusable) to 53% (non-reusable) and 47% (reusable). Before developing the predictive models, the oversampled datasets are split into training and testing data to assess the initial performance of the developed fits (Section 5.3).

Table 5.1 Oversampled datasets

	TEC dataset	ECO dataset	SOC dataset
Non-reusable (0)	96	93	59
Reusable (1)	96	91	53
Total number of elements	192	184	112

In this study, R package mlr (Bischl et al. 2016) is used to perform SMOTE-NC (Synthetic Minority Oversampling Technique for Nominal and Continuous) (Chawla et al. 2002) for each dataset separately. The script used to perform oversampling is available in Appendix E (Script E.2).

5.3 Training and testing datasets

The accuracy and interpretability of any machine learning model play an important role in choosing the best predictive model for the study at hand (James et al. 2017). The above two metrics are also used in Chapter 6 to further examine the thirteen BSE-RPMs that are developed in Chapter 5. In general, the machine learning methods are assessed in terms of their capability in predicting the responses to previously unseen data (test or out-of-sample data) (James et al. 2017). In the current research, and as a preliminary metric, the validation set approach is employed for determining the performance of the developed predictive models by developing training and testing data for the TEC, ECO, and SOC datasets separately. The available data in each oversampled dataset from Section 5.2 is divided on a 70/30 basis considering 70% of the dataset for the training purpose and 30% for the testing purpose. Script E.3 (Appendix E) is used to perform the data split using the caTools package in R (Tuszynski 2020). For further details about the validation set approach, please refer to Section 6.2 and Chapter 5 of James et al. (2017).

Table 5.2 shows the result of splitting the datasets into training and testing sets. However, before developing the predictive models, the available features need to be assessed to choose the best combination of the independent variables to generate the BSE-RPMs (Section 5.4).

Table 5.2 Split of the oversampled data into training and testing sets

	TEC dataset			ECO dataset			SOC dataset		
	Train	Test	Total	Train	Test	Total	Train	Test	Total
Non-reusable (0)	67	29	96	65	28	93	41	18	59
Reusable (1)	67	29	96	64	27	91	37	16	53
Total number of elements	134	58	192	129	55	184	78	34	112

5.4 Feature selection

Feature selection is a vital stage in supervised machine learning (Torgo 2016). It includes selecting a subset of features (independent variables) in a dataset for efficient and optimum analysis of the problem at hand (Torgo 2016, Ding and Peng 2003). In supervised machine learning, there is always a chance that some variables are irrelevant to the response or redundant. In such cases, their presence negatively affects the performance of a predictive model. Proper feature selection results in the development of predictive models that perform optimally on both seen and unseen data. Therefore, feature selection focuses on identifying relevant features and discards irrelevant or redundant independent variables (Urbanowicz et al. 2018). This process fulfils the third objective of this research (Section 1.6), which is selecting the best combination of the identified factors to develop the BSE-RPMs. It is noteworthy that in the process of selecting variables, only the training datasets (Table 5.2) are considered to avoid inaccurate estimates of the test errors (James et al. 2017, Urbanowicz et al. 2018).

There are three methods for selecting a subset of features (Guyon et al. 2006, Saeys, Inza, and Larrañaga 2007). Filter methods (or simply filters) use statistical properties of the features (like correlation coefficients, F-test, T-test, etc.) or information-theory based measures (such as mutual information, interaction information, etc.) to rank features based on their relevance to the response and other features (Torgo 2016, Guyon et al. 2006, McGill 1954, Iguyon and Elisseeff 2003). These methods can be grouped into univariate and multivariate filter methods. Univariate filter methods rank features only based on their relevance to the response, whereas multivariate filter methods consider the interaction between features as well (Guyon et al. 2006).

Wrappers are the second method for feature selection (Torgo 2016). In this group of techniques, a machine learning model is used to score subsets of features based on the predictive power of the method. The process of feature selection can be categorised into forward selection, backward elimination, and mixed selection (James et al. 2017). Forward selection methods start modelling with zero predictors (a base model), select features step-by-step and evaluate the performance. Whereas, backward feature elimination methods start with the complete set of independent variables and look for an optimum subset of variables with the best performance through stepwise elimination of non-informative features (James et al. 2017). Wrappers use cross-validation to optimise the performance of the learning method to select the optimum subset of variables (Guyon et al. 2006).

A third approach that is sometimes grouped with wrappers (Torgo 2016) is the embedded or intrinsic method (Guyon et al. 2006, Kuhn and Johnson 2020). Similar to wrappers that select a subset of variables based on a learning model, an embedded method embeds this process into its predictive model development (Guyon et al. 2006). For instance, if the variable importance measure of the random forest method is used to improve the performance of a random forest model, then it is an embedded method. Whereas, if this capacity is used to select features and develop predictive models with methods other than the random forest, then it is a wrapper method. In this research, the process of feature selection is not integrated with the model developments, so the embedded methods are not used.

For the TEC dataset, feature selection is performed at three stages. At stage 1, features are ranked using various methods embedded in the mlr package in R (filter methods) (Bischl et al. 2016). At stage 2, the "Boruta" method is used to select a subset of features (a backward variable elimination wrapper technique) (Kursa and Rudnicki 2010). This method will be explained in detail in Section 5.4.2. At stage 3, using recursive feature elimination (RFE) methods embedded in the caret package, subsets of variables are selected (wrapper methods with repeated cross-validation) (Kuhn 2008). Eventually, the results of the above three stages are compared to determine the final subset of variables for model development in the TEC dataset.

In the case of the ECO and SOC datasets, the process of variable selection is limited to the first two stages used for the TEC dataset (see the previous paragraph). This decision was made based on the total number of variables in each dataset. The TEC dataset has a considerable number of thirty-nine predictors, whereas the ECO and SOC datasets comprise only twelve and ten features, respectively.

Section 5.4.1 explains the first stage of feature selection (used for the TEC, ECO, and SOC datasets). Section 5.4.2 discusses the Boruta method (used for the TEC, ECO, and SOC datasets). And Section 5.4.3 introduces the RFE method that is only used for the TEC dataset.

5.4.1 Features ranking methods

In this research, ten methods are used to rank features in each dataset using the integrated filter methods in the mlr package (Bischl et al. 2016). Each of these methods is briefly introduced in the following sub-sections.

5.4.1.1 cForest importance (cF)

This method uses the permutation accuracy importance measure to determine the importance of variables using the cforest function in the party package (Strobl et al. 2007, Hothorn, Hornik, and Zeileis 2006). In random forests, an ensemble of classification trees is generated from the original sample (training dataset). This is done either by bootstrapping (drawing several samples with replacement) or subsampling without replacement. Random forests assign a small random of features to each decision tree. In this method, the value of an independent variable in a decision tree is randomly permuted to separate its correlation from the dependent variable. It is the impact of the permuted variable on the prediction accuracy of the model that determines its importance (Strobl et al. 2007, 2008).

5.4.1.2 Chi-squared (CHIS)

The chi-squared (CHIS) method is a univariate filter that ranks features based on the strength of the association between the independent variables and the response. This correlation-based filter performs the feature ranking by evaluating Pearson's χ^2 (Guyon et al. 2006). This method has a low accuracy because it makes simplistic assumptions about feature independence (Guyon et al. 2006, Saeys, Inza, and Larrañaga 2007).

5.4.1.3 Information gain (IG)

The information gain (IG) method is an entropy-based metric that quantifies the expected amount of information held in a random feature on the response. Information gain has two entropy measures. One is the class entropy, which is the information available on the response classes, and the other one is the conditional class entropy, which is the information available on the response classes given the values of a random feature. Information gain is calculated by subtracting the latter from the former (Torgo 2016, Cover and Thomas 2005). In this study, the information gain measure of the independent variables are used to rank features in each of the datasets through packages FSelector and mlr (Bischl et al. 2016, Romanski and Kotthoff 2018).

5.4.1.4 Gain ratio (GR)

This method is a variation of the information gain of a feature. This method has an additional entropy measure, which is the feature entropy. The gain ratio (GR) of an independent variable is determined by dividing the information gain of the feature by its entropy (Torgo 2016). In this study, the gain ratio of the independent variables are used to rank features in each of the datasets through packages FSelector and mlr (Bischl et al. 2016, Romanski and Kotthoff 2018).

5.4.1.5 Kruskal test (KT)

Similar to the chi-squared method, the Kruskal test (KT) is a univariate filter, as well (Saeys, Inza, and Larrañaga 2007). It is a non-parametric test and evaluates if the values of a feature affect the reusability of an element (Kruskal 1952, Kruskal and Wallis 1952, Ruxton and Beauchamp

2008). The null hypothesis is that the distribution of the values of a feature is the same for the response classes, based on a median rank. If the p-value is significant, it rejects the null hypothesis. In this method, features are ranked based on the significance of their p-values. The closer the p-value to zero, the higher the rank of a feature.

5.4.1.6 Minimum-redundancy-maximum-relevance (MRMR)

MRMR is a multivariate filter method developed by Ding and Peng (2003). This method provides a feature set with the highest relevance to the response and the lowest collinearity among the independent variables. Therefore, the identified feature set is a true representative of the original feature space covered by the dataset. This property improves the generalisability of the selected feature set, and it results in the selection of a smaller number of independent variables with the same performance.

5.4.1.7 oneR

One rule (oneR) method is a univariate filter that ranks features according to their classification error rate. It works by developing a base model by assigning the most frequent class of the response as the one rule to each of the values of a feature. This model is then used to predict the class of the response for each feature. The feature with the lowest error rate ranks the highest, followed by features with higher error rates (Jamjoom 2020).

5.4.1.8 Random forest (RF)

This method is like the cForest importance measure (subsection 5.4.1.1). However, it uses a different measure to assess the importance of a feature. This method uses the Gini importance measure, which is the outcome of the Gini impurity index used in the RandomForest package (Nembrini, König, and Wright 2018, Liaw and Wiener 2002, Breiman et al. 2017). Moreover, the method develops decorrelated trees, which result in a considerable decrease in the variance of the model compared to a single decision tree. A further explanation of the RF method is available in Section 5.5.1.7.

5.4.1.9 Relief

Relief is a non-parametric multivariate filter method that ranks individual features using an approach based on the K-Nearest-neighbour (KNN) method (Guyon et al. 2006, Urbanowicz et al. 2018, Kira and Rendell 1992). In this study, the RReliefF filter through packages FSelector and mlr is used to rank features (Bischl et al. 2016, Romanski and Kotthoff 2018). According to Urbanowicz et al. (2018), this method ranks features in the context of other features. However, it does not remove redundant independent variables while ranking the features (Urbanowicz et al. 2018). A further explanation of the KNN method is available in Section 5.5.1.1.

5.4.1.10 Symmetrical uncertainty (SU)

Symmetrical uncertainty (SU) is an entropy-based measure and is a variation of the gain ratio method. This method has one additional entropy measure in its denominator, which is the class entropy. The symmetrical uncertainty of an independent variable is determined by dividing twice the information gain of the feature by the sum of its entropy and the class entropy (Sarhrouni, Hammouch, and Aboutajdine 2012). In this study, the symmetrical uncertainty of the independent variables are used to rank features in each of the datasets through packages FSelector and mlr (Bischl et al. 2016, Romanski and Kotthoff 2018).

5.4.1.11 Implementation of the features ranking methods

A filter method produces a score for each of the features in the datasets. The higher the score of a predictor, the more important is the variable according to the selected filter method. However, the raw values produced by different filters are not having the same scale and cannot be compared. Therefore, after identifying the raw scores of the features using a filter method, these values are converted into percentage values by dividing them by the sum of the quantities of all variables. These percentages represent the level of importance of each feature in a ranking method (filter) and provide a baseline for comparing the results of different techniques. For the final ranking, the percentage values of all ten filter methods for each independent variable are summed up together to create a new metric. This new metric is then used to rank the features in each dataset. Tables 5.3 to 5.6 are the results of the feature ranking methods for the TEC, ECO, and SOC datasets, respectively.

The listed packages in Script E.4 (Appendix E) were initially installed to perform feature ranking methods discussed in this section.

In the next stage, Script E.5 (Appendix E) was used to determine the rank of the features in each dataset based on the discussion in Section 5.4.1.

Table 5.3 Raw scores for the features' ranks in the TEC dataset

Var.[1]	cF	CHIS	IG	GR	KT	MRMR	oneR	RF	Relief	SU
B2	0.000	0.157	0.013	0.157	0.117	0.795	0.343	4.455	0.060	0.021
B3	0.009	0.323	0.055	0.323	5.532	0.923	0.440	13.872	0.180	0.048
B4	0.001	0.306	0.060	0.306	0.095	0.949	0.366	2.985	0.060	0.106
B5	0.013	0.337	0.060	0.337	0.217	0.821	0.448	12.742	0.040	0.056
B6	0.001	0.113	0.006	0.113	0.831	0.641	0.351	3.955	0.100	0.009
B7	0.003	0.000	0.000	0.000	4.248	0.462	0.306	6.993	0.080	0.000
B8	0.007	0.000	0.000	0.000	6.394	0.744	0.306	8.434	0.025	0.000
B9	0.002	0.000	0.000	0.000	0.455	0.333	0.306	7.989	0.050	0.000
C1	0.004	0.000	0.000	0.000	0.000	0.256	0.306	10.286	0.045	0.000
C2	0.002	0.000	0.000	0.000	2.737	0.385	0.306	6.197	0.015	0.000
C3	0.002	0.000	0.000	0.000	2.444	0.410	0.306	6.641	-0.004	0.000
C4	0.002	0.000	0.000	0.000	0.309	0.359	0.306	4.904	-0.014	0.000
C5	0.002	0.000	0.000	0.000	1.280	0.718	0.306	6.735	0.090	0.000
C6	0.005	0.000	0.000	0.000	4.125	0.538	0.306	8.740	0.035	0.000
C7	0.004	0.000	0.000	0.000	5.590	0.231	0.306	5.531	0.110	0.000
C8	0.000	0.000	0.000	0.000	0.343	0.667	0.306	3.383	0.027	0.000
C9	0.004	0.000	0.000	0.000	0.955	0.103	0.306	8.619	0.125	0.000
C10	0.005	0.000	0.000	0.000	3.227	0.487	0.306	8.170	0.045	0.000
C11	0.001	0.000	0.000	0.000	1.209	0.308	0.306	3.406	0.015	0.000
C12	0.003	0.000	0.000	0.000	0.109	0.846	0.306	9.709	0.030	0.000
C13	0.001	0.000	0.000	0.000	2.163	0.897	0.306	4.065	0.050	0.000
C14	0.002	0.000	0.000	0.000	4.235	0.564	0.306	6.185	0.107	0.000
C15	0.003	0.000	0.000	0.000	2.300	0.590	0.306	6.997	0.033	0.000
C16	0.014	0.377	0.076	0.377	12.895	0.692	0.470	12.758	0.115	0.120
C17	0.011	0.000	0.000	0.000	9.501	0.615	0.306	9.197	0.095	0.000
C18	0.003	0.000	0.000	0.000	2.974	0.282	0.306	7.763	0.025	0.000
C19	0.006	0.000	0.000	0.000	5.455	0.077	0.306	7.639	0.160	0.000
C20	0.005	0.000	0.000	0.000	5.927	0.179	0.306	9.136	0.130	0.000
C21	0.001	0.000	0.000	0.000	1.995	0.026	0.306	5.994	0.075	0.000
C25	0.013	0.357	0.067	0.357	16.205	0.872	0.463	10.233	-0.009	0.106
C26	0.002	0.000	0.000	0.000	4.646	0.128	0.306	7.265	-0.039	0.000
C27	0.014	0.000	0.000	0.000	9.680	0.513	0.306	8.198	0.045	0.000
C28	0.040	0.413	0.090	0.413	29.194	1.000	0.500	19.762	0.130	0.135
D18	0.001	0.000	0.000	0.000	0.127	0.436	0.306	5.802	0.020	0.000
D19	0.002	0.000	0.000	0.000	0.991	0.051	0.306	6.945	0.050	0.000
D21	0.000	0.000	0.000	0.000	0.235	0.154	0.306	4.388	0.065	0.000
D22	0.003	0.000	0.000	0.000	0.325	0.205	0.306	6.334	0.090	0.000
D23	0.003	0.335	0.063	0.335	2.318	0.769	0.425	10.420	0.110	0.114
D24	0.011	0.381	0.097	0.381	10.517	0.974	0.433	13.438	0.055	0.181

[1] The details of the features are available in a copy of the survey in Section C.2 (Appendix C)

Table 5.4 Percentages, and the final ranking of features in the TEC dataset

Var.[1]	cF	CHIS	IG	GR	KT	MRMR	oneR	RF	Relief	SU	% sum (new metric)	Final rank
B2	0.0%	5.1%	2.2%	5.1%	0.1%	4.0%	2.6%	1.5%	2.5%	2.3%	25.2%	9
B3	4.6%	10.4%	9.3%	10.4%	3.4%	4.6%	3.4%	4.5%	7.4%	5.4%	63.5%	5
B4	0.3%	9.9%	10.3%	9.9%	0.1%	4.7%	2.8%	1.0%	2.5%	11.8%	53.2%	8
B5	6.5%	10.9%	10.2%	10.9%	0.1%	4.1%	3.4%	4.2%	1.7%	6.3%	58.1%	7
B6	0.4%	3.6%	1.1%	3.6%	0.5%	3.2%	2.7%	1.3%	4.1%	1.0%	21.6%	12
B7	1.4%	0.0%	0.0%	0.0%	2.6%	2.3%	2.3%	2.3%	3.3%	0.0%	14.3%	19
B8	3.5%	0.0%	0.0%	0.0%	3.9%	3.7%	2.3%	2.8%	1.0%	0.0%	17.3%	15
B9	0.9%	0.0%	0.0%	0.0%	0.3%	1.7%	2.3%	2.6%	2.1%	0.0%	9.9%	29
C1	1.9%	0.0%	0.0%	0.0%	0.0%	1.3%	2.3%	3.4%	1.9%	0.0%	10.7%	27
C2	0.7%	0.0%	0.0%	0.0%	1.7%	1.9%	2.3%	2.0%	0.6%	0.0%	9.3%	30
C3	0.9%	0.0%	0.0%	0.0%	1.5%	2.1%	2.3%	2.2%	-0.2%	0.0%	8.7%	33
C4	0.9%	0.0%	0.0%	0.0%	0.2%	1.8%	2.3%	1.6%	-0.6%	0.0%	6.2%	39
C5	1.1%	0.0%	0.0%	0.0%	0.8%	3.6%	2.3%	2.2%	3.7%	0.0%	13.8%	20
C6	2.5%	0.0%	0.0%	0.0%	2.5%	2.7%	2.3%	2.9%	1.4%	0.0%	14.4%	18
C7	2.0%	0.0%	0.0%	0.0%	3.5%	1.2%	2.3%	1.8%	4.6%	0.0%	15.3%	16
C8	0.0%	0.0%	0.0%	0.0%	0.2%	3.3%	2.3%	1.1%	1.1%	0.0%	8.0%	34
C9	1.8%	0.0%	0.0%	0.0%	0.6%	0.5%	2.3%	2.8%	5.2%	0.0%	13.2%	22
C10	2.3%	0.0%	0.0%	0.0%	2.0%	2.4%	2.3%	2.7%	1.9%	0.0%	13.6%	21
C11	0.4%	0.0%	0.0%	0.0%	0.7%	1.5%	2.3%	1.1%	0.6%	0.0%	6.8%	38
C12	1.4%	0.0%	0.0%	0.0%	0.1%	4.2%	2.3%	3.2%	1.2%	0.0%	12.4%	23
C13	0.3%	0.0%	0.0%	0.0%	1.3%	4.5%	2.3%	1.3%	2.1%	0.0%	11.8%	24
C14	0.9%	0.0%	0.0%	0.0%	2.6%	2.8%	2.3%	2.0%	4.4%	0.0%	15.1%	17
C15	1.5%	0.0%	0.0%	0.0%	1.4%	2.9%	2.3%	2.3%	1.4%	0.0%	11.8%	25
C16	6.7%	12.2%	12.9%	12.2%	8.0%	3.5%	3.6%	4.2%	4.8%	13.4%	81.4%	3
C17	5.3%	0.0%	0.0%	0.0%	5.9%	3.1%	2.3%	3.0%	3.9%	0.0%	23.6%	10
C18	1.7%	0.0%	0.0%	0.0%	1.8%	1.4%	2.3%	2.5%	1.0%	0.0%	10.8%	26
C19	2.9%	0.0%	0.0%	0.0%	3.4%	0.4%	2.3%	2.5%	6.6%	0.0%	18.1%	13
C20	2.5%	0.0%	0.0%	0.0%	3.7%	0.9%	2.3%	3.0%	5.4%	0.0%	17.7%	14
C21	0.5%	0.0%	0.0%	0.0%	1.2%	0.1%	2.3%	2.0%	3.1%	0.0%	9.3%	31
C25	6.6%	11.5%	11.4%	11.5%	10.0%	4.4%	3.5%	3.3%	-0.4%	11.8%	73.7%	4
C26	0.8%	0.0%	0.0%	0.0%	2.9%	0.6%	2.3%	2.4%	-1.7%	0.0%	7.4%	36
C27	7.0%	0.0%	0.0%	0.0%	6.0%	2.6%	2.3%	2.7%	1.9%	0.0%	22.4%	11
C28	19.9%	13.3%	15.3%	13.3%	18.0%	5.0%	3.8%	6.5%	5.4%	15.1%	115.6%	1
D18	0.4%	0.0%	0.0%	0.0%	0.1%	2.2%	2.3%	1.9%	0.8%	0.0%	7.7%	35
D19	1.2%	0.0%	0.0%	0.0%	0.6%	0.3%	2.3%	2.3%	2.1%	0.0%	8.7%	32
D21	-0.1%	0.0%	0.0%	0.0%	0.1%	0.8%	2.3%	1.4%	2.7%	0.0%	7.3%	37
D22	1.3%	0.0%	0.0%	0.0%	0.2%	1.0%	2.3%	2.1%	3.7%	0.0%	10.7%	28
D23	1.5%	10.8%	10.8%	10.8%	1.4%	3.8%	3.2%	3.4%	4.6%	12.7%	63.1%	6
D24	5.5%	12.3%	16.6%	12.3%	6.5%	4.9%	3.3%	4.4%	2.3%	20.2%	88.2%	2

[1] The details of the features are available in a copy of the survey in Section C.2 (Appendix C)

Table 5.5 Raw scores, percentages, and the final ranking of features in the ECO dataset

Var.[1]	Value	cF	CHIS	IG	GR	KT	MRMR	oneR	RF	Relief	SU	% sum (new metric)	Final rank
C24	raw	0.064	0.422	0.093	0.422	21.777	0.917	0.496	20.973	0.120	0.136	216.2	1
	%	30.6	27.1	28.8	27.1	16.3	14.1	11.8	12.4	20.9	27.2		
D1	raw	0.033	0.000	0.000	0.000	7.650	0.833	0.295	21.733	0.035	0.000	60.1	5
	%	15.6	0.0	0.0	0.0	5.7	12.8	7.0	12.9	6.1	0.0		
D2	raw	0.014	0.000	0.000	0.000	13.057	0.417	0.295	11.966	-0.014	0.000	34.4	10
	%	6.7	0.0	0.0	0.0	9.8	6.4	7.0	7.1	-2.6	0.0		
D3	raw	0.004	0.000	0.000	0.000	5.980	0.750	0.295	12.030	0.040	0.000	39.2	8
	%	2.1	0.0	0.0	0.0	4.5	11.5	7.0	7.1	7.0	0.0		
D4	raw	0.002	0.000	0.000	0.000	2.034	0.333	0.295	9.174	0.095	0.000	36.4	9
	%	0.8	0.0	0.0	0.0	1.5	5.1	7.0	5.4	16.5	0.0		
D5	raw	0.004	0.000	0.000	0.000	4.394	1.000	0.295	18.452	0.055	0.000	47.9	6
	%	1.7	0.0	0.0	0.0	3.3	15.4	7.0	10.9	9.6	0.0		
D6	raw	0.003	0.355	0.070	0.355	7.820	0.250	0.426	11.197	0.025	0.119	123.4	4
	%	1.6	22.8	21.6	22.8	5.9	3.8	10.1	6.6	4.3	23.7		
D7	raw	0.002	0.000	0.000	0.000	6.871	0.167	0.295	6.075	0.005	0.000	20.0	12
	%	0.8	0.0	0.0	0.0	5.1	2.6	7.0	3.6	0.9	0.0		
D8	raw	0.010	0.000	0.000	0.000	11.422	0.583	0.295	11.762	0.055	0.000	45.7	7
	%	4.7	0.0	0.0	0.0	8.6	9.0	7.0	7.0	9.6	0.0		
D9	raw	0.002	0.000	0.000	0.000	7.843	0.083	0.295	8.083	0.035	0.000	26.2	11
	%	1.2	0.0	0.0	0.0	5.9	1.3	7.0	4.8	6.1	0.0		
D10	raw	0.036	0.387	0.078	0.387	26.697	0.500	0.473	18.424	0.025	0.117	168.6	3
	%	17.1	24.9	24.2	24.9	20.0	7.7	11.2	10.9	4.3	23.4		
D25	raw	0.036	0.392	0.082	0.392	18.024	0.667	0.465	18.750	0.100	0.129	181.9	2
	%	17.1	25.2	25.4	25.2	13.5	10.3	11.0	11.1	17.4	25.7		

[1] The details of the features are available in a copy of the survey in Section C.2 (Appendix C)

Table 5.6 Raw scores, percentages, and the final ranking of features in the SOC dataset

Var.[1]	Value	cF	CHIS	IG	GR	KT	MRMR	oneR	RF	Relief	SU	% sum (new metric)	Final rank
C22	raw	0.025	0.000	0.000	0.000	0.148	0.600	0.295	26.128	0.140	0.000	75.0	5
	%	14.5	0.0	0.0	0.0	0.2	10.9	8.4	18.9	22.0	0.0		
C23	raw	0.023	0.380	0.094	0.380	7.569	0.900	0.410	17.871	0.030	0.180	160.6	3
	%	12.9	22.9	21.8	22.9	11.2	16.4	11.7	13.0	4.7	23.2		
D11	raw	0.002	0.000	0.000	0.000	7.224	0.300	0.295	6.596	0.055	0.000	39.0	8
	%	1.0	0.0	0.0	0.0	10.7	5.5	8.4	4.8	8.7	0.0		
D12	raw	0.006	0.000	0.000	0.000	7.399	0.200	0.295	8.810	0.070	0.000	43.8	7
	%	3.4	0.0	0.0	0.0	11.0	3.6	8.4	6.4	11.0	0.0		
D13	raw	0.003	0.000	0.000	0.000	4.909	0.100	0.295	8.620	0.040	0.000	31.8	10
	%	1.7	0.0	0.0	0.0	7.3	1.8	8.4	6.2	6.3	0.0		
D14	raw	0.090	0.455	0.113	0.455	19.956	1.000	0.474	21.681	0.115	0.184	250.9	1
	%	51.2	27.4	26.1	27.4	29.6	18.2	13.5	15.7	18.1	23.7		
D15	raw	0.003	0.000	0.000	0.000	8.036	0.700	0.295	9.438	0.025	0.000	45.5	6
	%	1.6	0.0	0.0	0.0	11.9	12.7	8.4	6.8	3.9	0.0		
D16	raw	0.014	0.471	0.143	0.471	5.554	0.800	0.462	17.673	0.060	0.250	187.9	2
	%	7.8	28.3	33.0	28.3	8.2	14.5	13.1	12.8	9.4	32.2		
D17	raw	0.010	0.356	0.083	0.356	6.187	0.500	0.397	9.726	0.045	0.162	132.5	4
	%	5.9	21.4	19.1	21.4	9.2	9.1	11.3	7.0	7.1	20.9		
D20	raw	0.000	0.000	0.000	0.000	0.350	0.400	0.295	11.411	0.055	0.000	33.0	9
	%	-0.1	0.0	0.0	0.0	0.5	7.3	8.4	8.3	8.7	0.0		

[1] The details of the features are available in a copy of the survey in Section C.2 (Appendix C)

5.4.2 Feature selection using the Boruta method

At this stage, the Boruta method through the Boruta package in R is used to identify relevant features in each of the datasets (Kursa and Rudnicki 2010). Filter methods mainly consider a direct correlation with the response as an essential step in feature selection. However, wrappers can find valuable features, even in the absence of such a correlation (Kursa and Rudnicki 2010). Unlike computationally inexpensive filter methods, wrappers use classifiers to rank features and take more time because they are more demanding (Kursa and Rudnicki 2010). Most of the wrapper methods use a random subset of features during variable selection. However, the Boruta method, which is a backward variable elimination wrapper, uses an all-relevant feature selection method. It means that this method minimises the random selection of features (Kursa and Rudnicki 2010, Sarkar et al. 2020).

The Boruta package uses a random forest classifier based on the RandomForest package in R (Kursa and Rudnicki 2010, Liaw and Wiener 2002). However, instead of using the Gini impurity index (Nembrini, König, and Wright 2018), it uses the permutation accuracy importance measure to determine the importance of variables (Kursa and Rudnicki 2010, Strobl et al. 2007). Initially, the original dataset is expanded by introducing a copy of all features in the dataset. Then, the values of the copied features are randomly manipulated to eliminate their correlation with the response. These shuffled independent variables are called shadows (Kursa and Rudnicki 2010). Next, the random forest classifier is applied to the extended dataset to gather the Z-statistics. The method then looks for the maximum Z-statistic among the shadow features (MZSF) and assigns a hit to the independent variables with higher z-score than the MZSF. If the importance of a feature is undecided, a two-sided test of equality with the MZSF will be performed. Features with z-scores considerably higher than MZSF are important (Confirmed), and those with significantly lower z-scores are unimportant (Rejected). For the undecided variables, the whole process is repeated. However, before reiterating, the shadow features created in the previous stage are removed. Nevertheless, if the model cannot decide to accept or reject some variables, their status will be reported as Tentative. In such circumstances, the researcher might choose to increase the maximum iterations or decide to reject or retain the variable based on his/her intuition or expert opinion. In the Boruta package (Kursa and Rudnicki 2010), the former could be done by increasing the maximal number of importance source runs (the “maxRuns” argument in the Boruta function, see Script E.6, Appendix E).

The results of feature selection using the Boruta method are presented in Figures 5.1 to 5.3 for the TEC, ECO, and SOC datasets, respectively.

Script E.6 (Appendix E) is used to derive the results of feature selection using the Boruta method for each of the datasets.

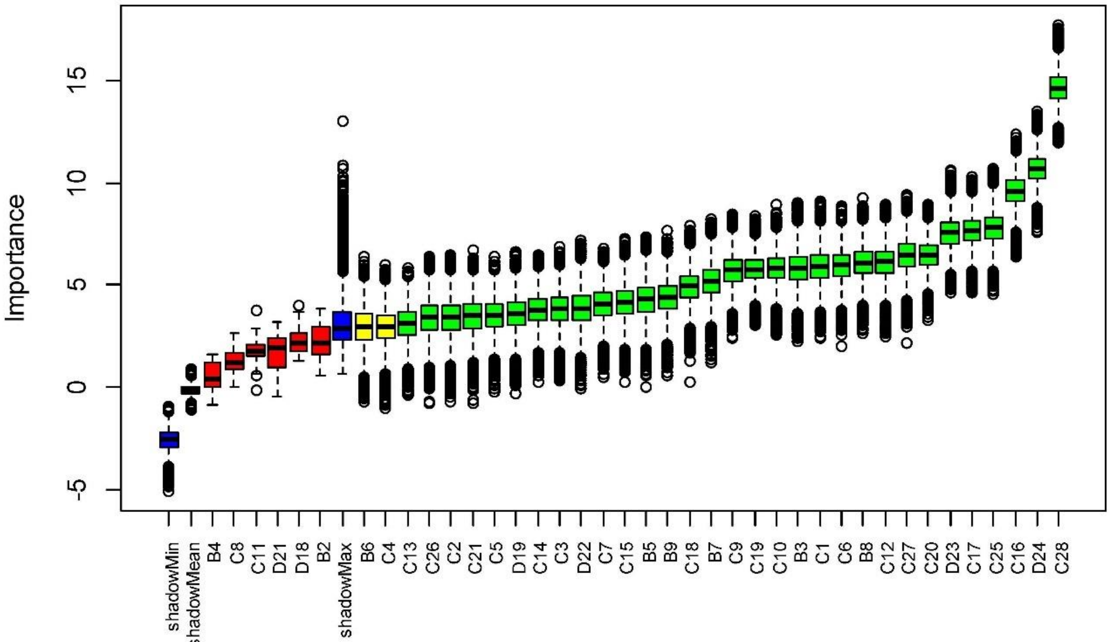

Figure 5.1 Importance of the features in the TEC dataset using the Boruta method

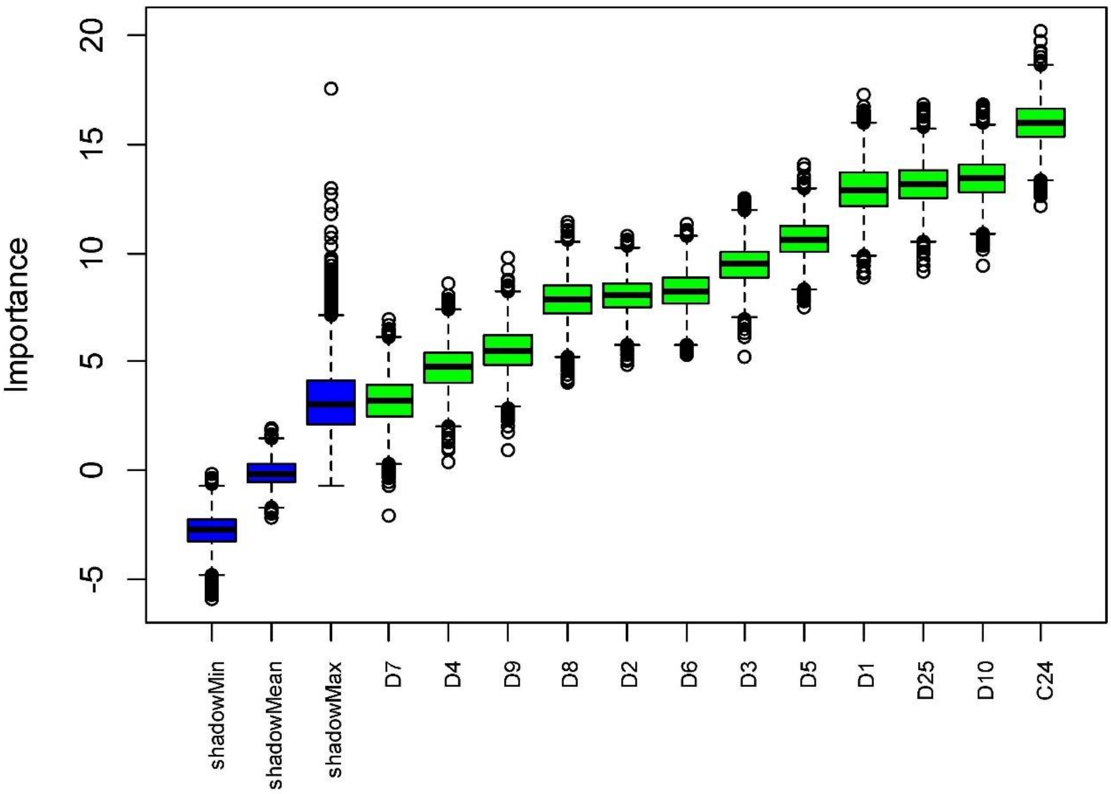

Figure 5.2 Importance of the features in the ECO dataset using the Boruta method

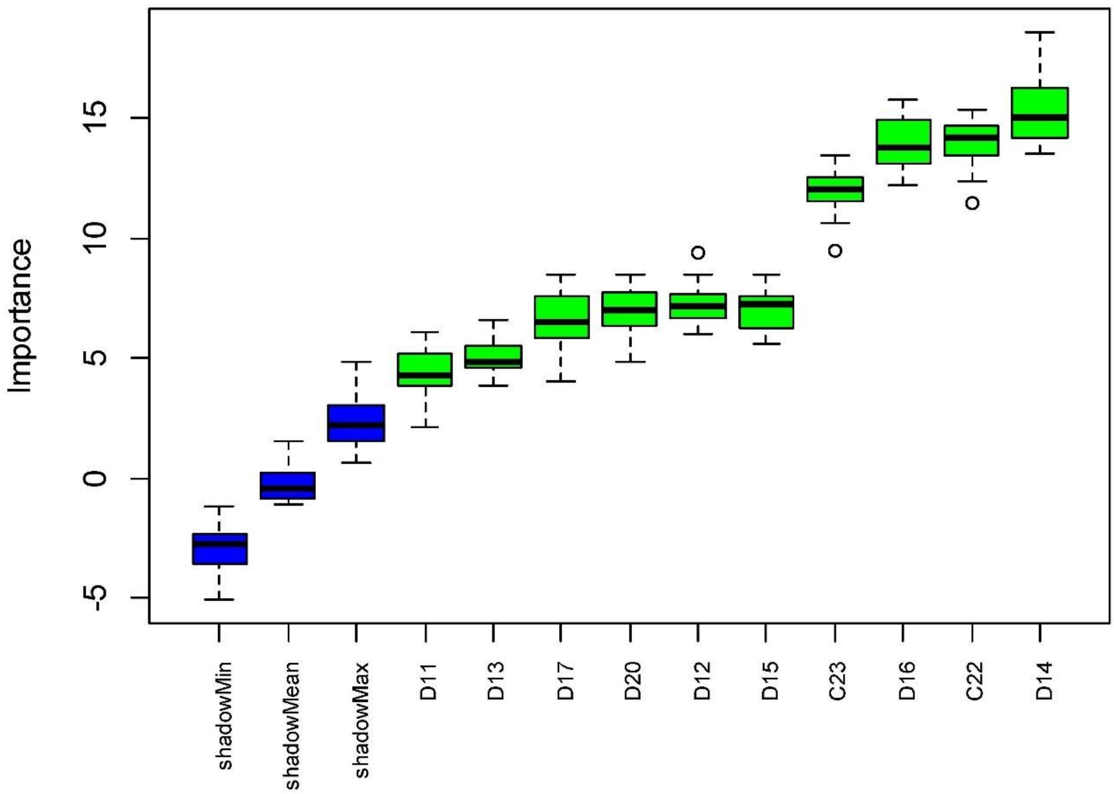

Figure 5.3 Importance of the features in the SOC dataset using the Boruta method

In these figures, from left to right, the blue boxes represent the minimum, mean, and maximum Z-statistics among the shadow features, respectively. The variable boxes in red represent rejected variables, those in green are the confirmed variables, and the yellow boxes are tentative variables. In the case of ECO and SOC datasets, the choice of maxRuns equal to 10,000 is to make sure no feature remains undecided (tentative). However, in the case of the TEC dataset, after increasing this value to 30,000, the method was still unable to categorise B6 and C4.

5.4.3 Variable selection using recursive feature elimination (RFE)

At this stage, recursive feature elimination (RFE) is performed using various methods through the caret package in R (Kuhn 2008). The methods used include Random Forests (randomForest package (Liaw and Wiener 2002)), Naïve Bayes (klaR package (Weihs et al. 2005)), Decision Trees (Bagging; ipred package (Peters and Hothorn 2019)), and Random Forests through the Caret Function (caret package (Kuhn 2008)).

RFE is a backward variable selection wrapper technique (Kuhn and Johnson 2020). Initially, a dedicated method is used to develop a model with all available independent variables and rank the features based on a measure of importance. Next, the least important feature is eliminated, and a new model is developed based on a smaller number of variables. Then, the remaining independent variables are re-ranked (Kuhn and Johnson 2020). In this method, the model identifies two parameters. The first parameter is the number of subsets to evaluate. The second parameter is the number of features in each subset. For each subset, the method continues to eliminate the least-important features until it reaches the determined subset size. Next, it compares the performance of each subset and determines the best subset size with the best accuracy (Kuhn and Johnson 2020). The latter is presented by plotting the number of features (based on their importance) against accuracy (Figures 5.4 to 5.7). In this section, the performance of the wrappers is assessed using k-fold cross-validation (k=10), which repeats five times. Script E.7 (Appendix E) is used to perform RFE.

The results of variable selection in each dataset using RFE (as applicable), along with the results of the earlier feature selection methods presented in Sections 5.4.1 and 5.4.2, are displayed in Tables 5.7 to 5.9 for the TEC, ECO, and SOC training data, respectively. Moreover, Figures 5.4 to 5.7 show the plots representing the performance of these RFE models based on the ranks of the variables for the TEC dataset. For instance, while in Figure 5.4, the model's accuracy based on C28 only is around 65%, the accuracy improves by adding variables based on their rank (Table 5.7). In the case of Figure 5.4, after adding D24, the accuracy increases to 71%, and so on.

Table 5.7 Status and rank of the variables in the TEC training dataset using the RFE method, the Boruta method, and filters

Variable	Random Forests	Naïve Bayes	Decision Trees (Bagging)	Caret Functions (Random Forests)	The Boruta method (Section 5.4.2)	Filters (Section 5.4.1)	Final decision
B2	38	Rejected	26	Rejected	Rejected	9	
B3	4	8	2	2	14	5	Selected
B4	37	38	Rejected	Rejected	Rejected	8	
B5	5	33	3	3	20	7	Selected
B6	28	27	20	29	Tentative	12	Selected
B7	18	13	14	14	17	19	Selected
B8	11	7	8	16	10	15	Selected
B9	20	29	18	19	19	29	Selected
C1	12	37	12	10	12	27	Selected
C2	27	17	23	23	29	30	Selected
C3	24	22	24	30	24	33	Selected
C4	33	34	25	Rejected	Tentative	39	
C5	29	26	28	28	27	20	Selected
C6	16	14	9	13	11	18	Selected
C7	30	11	Rejected	27	22	16	
C8	39	32	Rejected	Rejected	Rejected	34	
C9	15	25	19	15	16	22	Selected
C10	14	15	15	17	13	21	Selected

Variable	Random Forests	Naïve Bayes	Decision Trees (Bagging)	Caret Functions (Random Forests)	The Boruta method (Section 5.4.2)	Filters (Section 5.4.1)	Final decision
C11	35	24	Rejected	33	Rejected	38	
C12	9	36	16	11	9	23	Selected
C13	32	28	Rejected	Rejected	31	24	
C14	31	16	Rejected	34	25	17	
C15	21	20	17	22	21	25	Selected
C16	3	3	5	4	3	3	Selected
C17	7	5	11	7	5	10	Selected
C18	19	18	27	21	18	26	Selected
C19	17	10	22	20	15	13	Selected
C20	10	9	13	12	7	14	Selected
C21	26	21	Rejected	26	28	31	
C25	6	2	6	5	4	4	Selected
C26	25	12	Rejected	24	30	36	
C27	13	6	10	9	8	11	Selected
C28	1	1	1	1	1	1	Selected
D18	34	35	Rejected	31	Rejected	35	
D19	23	23	Rejected	25	26	32	
D21	36	31	Rejected	32	Rejected	37	
D22	22	30	21	18	23	28	Selected
D23	8	19	7	8	6	6	Selected
D24	2	4	4	6	2	2	Selected

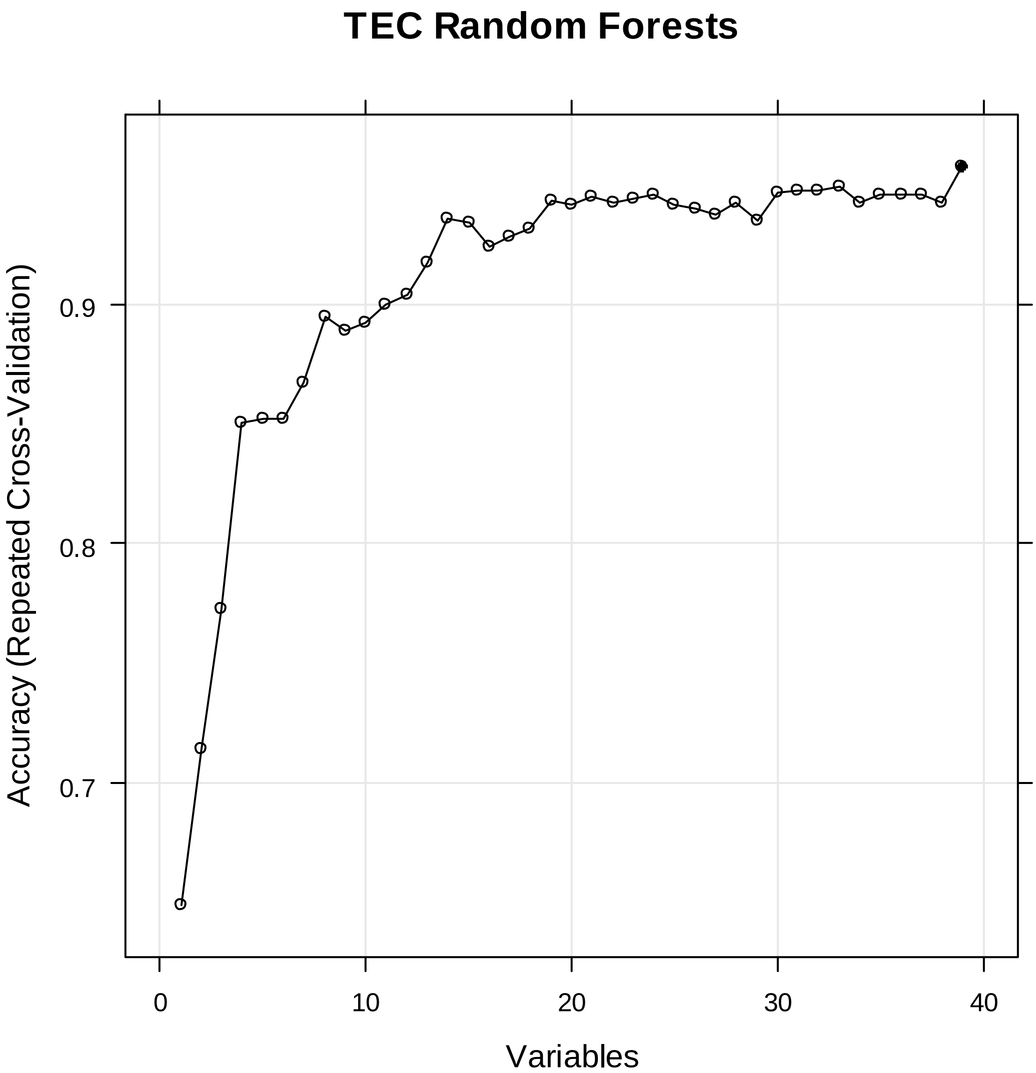

Figure 5.4 Performance of the RFE and Random Forests based on the ranks of the features (TEC dataset)

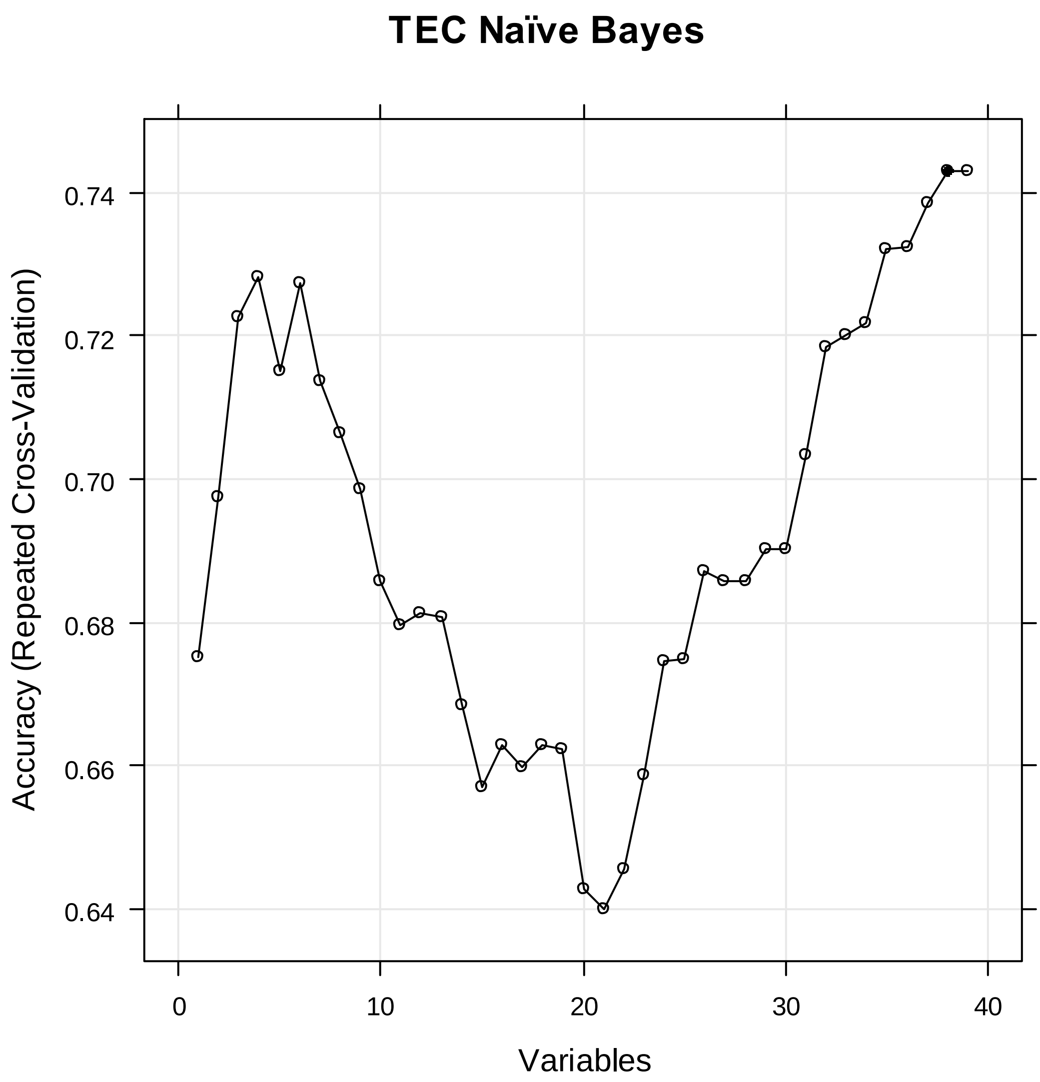

Figure 5.5 Performance of the RFE and Naïve Bayes based on the ranks of the features (TEC dataset)

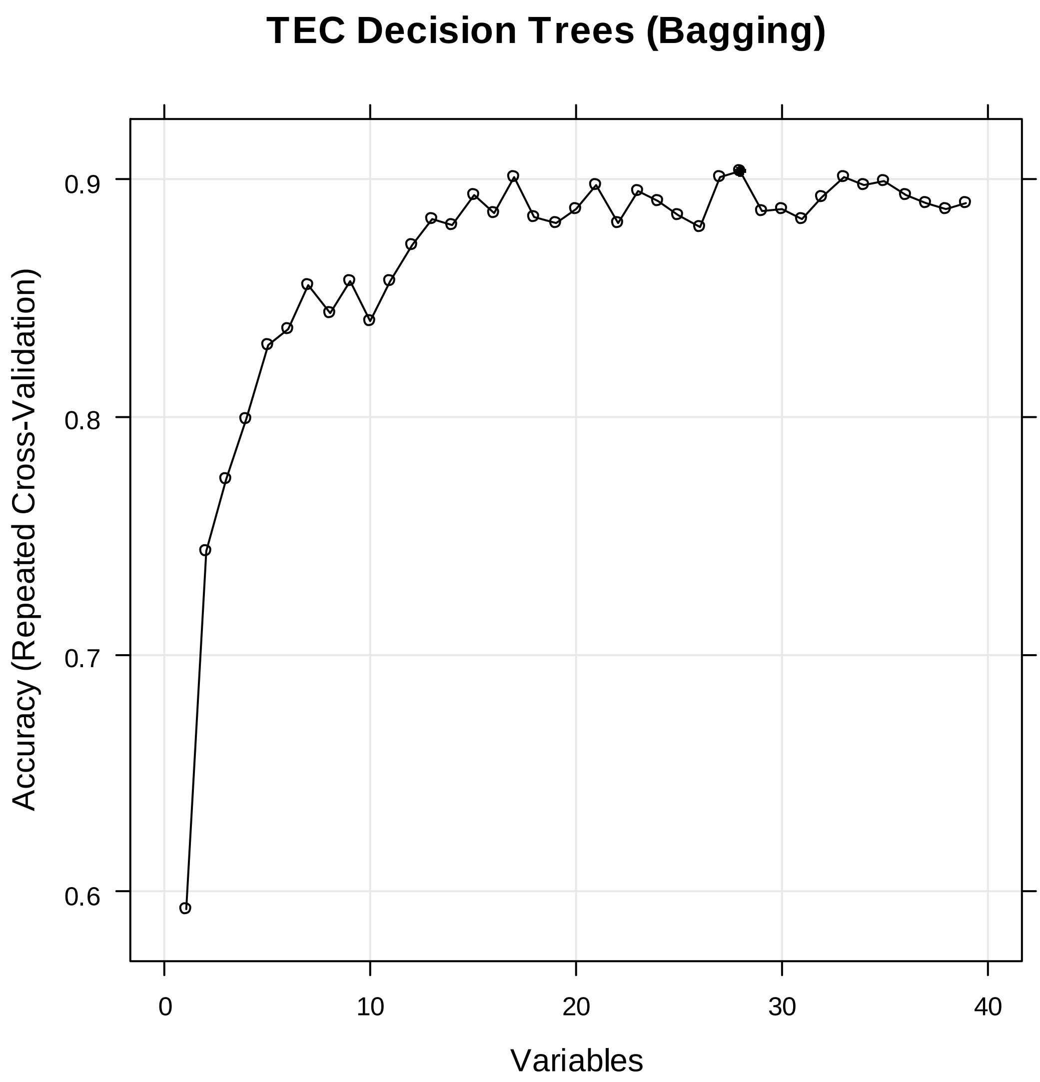

Figure 5.6 Performance of the RFE and Decision Trees (Bagging) based on the ranks of the features (TEC dataset)

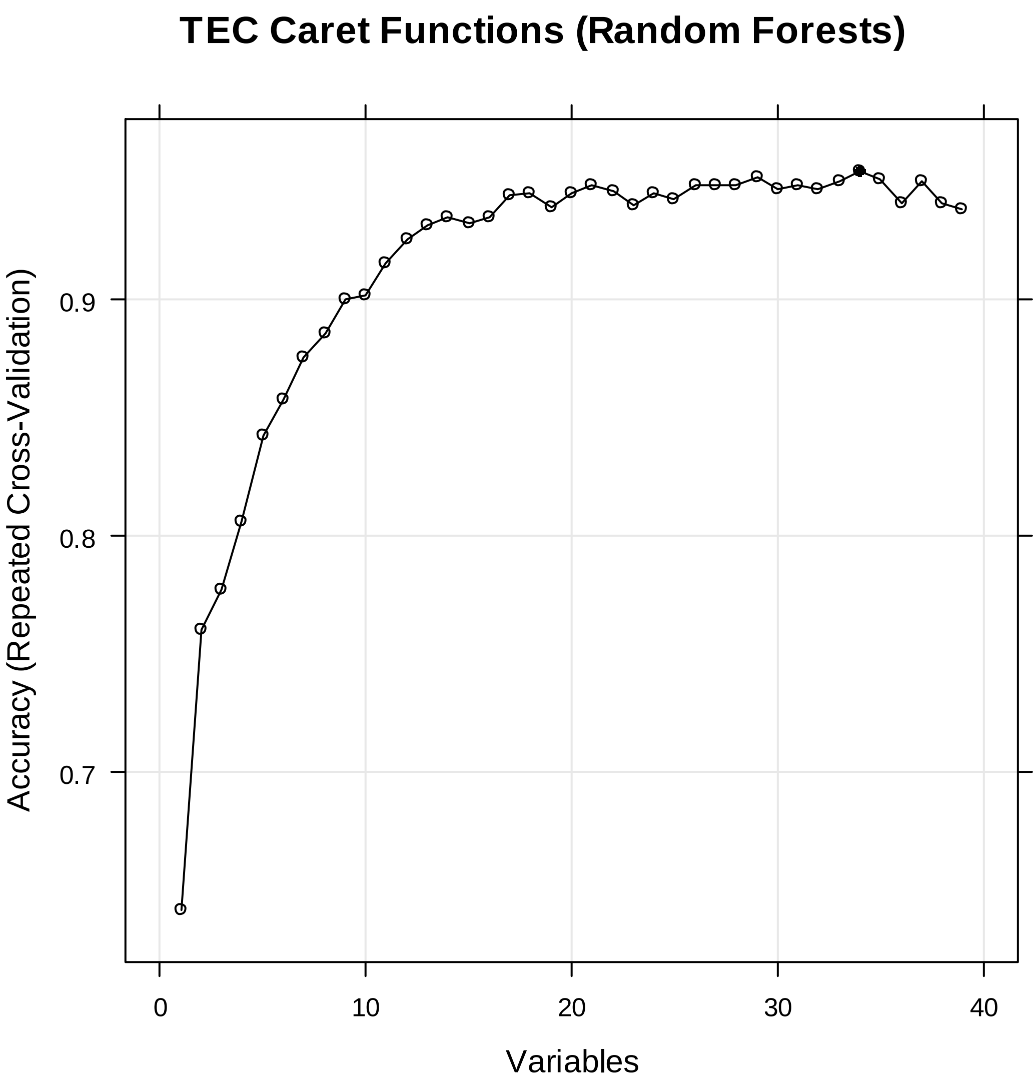

Figure 5.7 Performance of the RFE and Caret Function (Random Forests) based on the ranks of the features (TEC dataset)

Table 5.8 Status and rank of the variables in the ECO training dataset using the Boruta method and filters

Variable	The Boruta method (Section 5.4.2)	Filters (Section 5.4.1)	Final decision
C24	1	1	Selected
D1	4	5	Selected
D2	8	10	Selected
D3	6	8	Selected
D4	11	9	Selected
D5	5	6	Selected
D6	7	4	Selected
D7	12	12	Selected
D8	9	7	Selected
D9	10	11	Selected
D10	2	3	Selected

Variable	The Boruta method (Section 5.4.2)	Filters (Section 5.4.1)	Final decision
D25	3	2	Selected

Table 5.9 Status and rank of the variables in the SOC training dataset using the RFE method, the Boruta method, and filters

Variable	The Boruta method (Section 5.4.2)	Filters (Section 5.4.1)	Final decision
C22	3	5	Selected
C23	4	3	Selected
D11	10	8	Selected
D12	5	7	Selected
D13	9	10	Selected
D14	1	1	Selected
D15	8	6	Selected
D16	2	2	Selected
D17	6	4	Selected
D20	7	9	Selected

5.4.4 The final list of features

5.4.4.1 List of selected features for the TEC dataset

A comparison between different variable selection techniques performed in Sections 5.4.1, 5.4.2, and 5.4.3 for the TEC dataset reveals that the result of the rankings achieved from filter methods is slightly different from the other two techniques (Table 5.7). However, there is a good agreement between the Boruta method and the RFE variable selection methods.

According to Figure 5.4, while the method selects all the available 39 variables, it can be observed that with as low as 19 variables, a high level of accuracy is attainable. More precisely, with the top 19 variables identified using RFE plus Random Forests (Table 5.7), an accuracy of 94% is achievable, while with all 39 variables, this value improves to 96% (Figure 5.4). Similar trends can be observed in Figure 5.6, where the top 17 features result in an accuracy of 90%, which is equal to the performance of all 28 variables selected by the method. Figure 5.7 also shows the same trend with the top 18 variables selected by the method. Hence, referring to these figures, and based on Table 5.7, a list of all top variables was developed. It resulted in the selection of 16 variables that were common between all three RFE methods used to develop Figures 5.4, 5.6, and 5.7. Then, those variables that were not rejected by any of the methods were selected (Table 5.7), which was the result of comparing RFE methods and the Boruta method. It yields in the selection of 10 more variables, resulting in a total number of 26 variables to be used for the development of BSE-RPMs for predicting the technical reusability of building structural elements. The complete list of all selected variables are as follows: B3; B5; B6; B7; B8; B9; C1; C2; C3; C5; C6; C9; C10; C12; C15; C16; C17; C18; C19; C20; C25; C27; C28; D22; D23; D24. The selected independent variables are marked in Table 5.7.

While the rejection of B4 (the technique used to recover the element) seems counterintuitive, looking at the answers provided by the respondents (Figure 4.6) reveals that only less than 10% of the elements were recovered through demolition and the remaining were recovered using deconstruction (80.6%) and component-specific recovery (6.9%). The rest were reused in-situ (1.4%) or were surplus components (1.4%). Moreover, among the components recovered through demolition, only one (1) was reusable, and the remaining were non-reusable. Therefore, the results of this research would be limited to load-bearing building components recovered using deconstruction technique or its variations such as component-specific recovery.

5.4.4.2 List of selected features for the ECO dataset

For the ECO dataset, Tables 5.5 and 5.8 and Figure 5.2 show a good agreement between all the observations. According to Table 5.5, the least important feature is D7 (transportation cost), and the most affecting variable is C24 (potential financial risks). This trend is observed in Figure 5.2 and Table 5.8. Moreover, according to Figure 5.2, all the variables are relevant. Therefore, all the predictors listed in Table 5.8 are considered for the development of predictive models.

5.4.4.3 List of selected features for the SOC dataset

Analysing the results of variable selection for the SOC dataset reveals that, according to Figure 5.3, all the features are necessary. Therefore, all predictors listed in Table 5.9 will be considered for developing the predictive models for the SOC dataset.

5.5 Models development

The process of selecting an appropriate method for developing a predictive model using machine learning techniques is of ample importance because there is not a unique best model available for all problems (James et al. 2017). This study intends to develop BSE-RPMs to estimate the technical, economic, and social reusability of the structural elements at the end-of-life of a building with the highest possible accuracy. While accuracy is a driving metric in choosing a model, the interpretability of the selected model plays an important role, as well (Guidotti et al. 2018). It is because this study intends to provide an easy-to-understand model that can be used by various stakeholders in the building sector who necessarily might not be able to use complex predictive models (Guidotti et al. 2018). The above property is essential for the selected predictive model because it encourages the stakeholders to use the model effectively.

Based on the above discussion, it seems reasonable to choose interpretable methods such as logistic regression to develop the BSE-RPMs (Molnar 2020). Nevertheless, interpretable models are not always accurate and might have a high bias in their predictions (James et al. 2017). It is because these models are mostly less flexible, and some of them consider a functional form for the relationship between the predictors and the response (parametric models) (James et al. 2017). On the other hand, there are very flexible models such as the support vector machines or K-nearest neighbours (KNN) classifier (mostly, nonparametric methods) that produce very accurate models on the training dataset (James et al. 2017, Cortez and Embrechts 2013, Murphy 2012). However, this flexibility comes at the cost of losing interpretability, high variance, and sometimes overfitting, which results in inaccurate predictions on unseen data (James et al. 2017). Therefore, in selecting the proper method for developing a predictive model, this trade-off between bias and variance should be considered (Murphy 2012, Hastie, Tibshirani, and Friedman 2009, Geman, Bienenstock, and Doursat 1992).

Besides, constraints such as the limited number of observations in each dataset (Section 5.3), and unawareness of the nature of the relationship between the predictors and the responses brought new dimensions to the challenge of selecting a proper machine learning method. Therefore, it was decided to study a wide range of machine learning methods to develop an optimum predictive model that fulfils the fourth objective of this research (Section 1.6). The above decision is in line with the 'no free-lunch' theorems suggested by (Wolpert and Macready 1997). These models are listed in Table 5.10. The listed packages in Script E.8 (Appendix E) were initially installed to develop the models listed in Table 5.10.

Table 5.10 List of machine learning methods used to develop BSE-RPMs (Murphy 2012)

Model	Parametric / Non-parametric	Section	Script used (Appendix E)
K-Nearest Neighbours (KNN)	Non-parametric	5.5.1.1	E.9
Logistic Regression (LR)	Parametric	5.5.1.2	E.10
Linear Discriminant Analysis (LDA)	Parametric	5.5.1.3	E.11
Quadratic Discriminant Analysis (QDA)	Parametric	5.5.1.4	E.12
Naïve Bayes (NB)	Parametric	5.5.1.5	E.13

Model	Parametric / Non-parametric	Section	Script used (Appendix E)
Decision Trees (DT)	Non-parametric	5.5.1.6	E.14
Random Forests (RF)	Non-parametric	5.5.1.7	E.15
Adaptive Boosting (AB)	Non-parametric	5.5.1.8	E.16
BART Machine (BM)	Non-parametric	5.5.1.9	E.17
Artificial Neural Networks (ANN)	Parametric	5.5.1.10	E.18
Gaussian Processes (GP)	Non-parametric	5.5.1.11	E.19
Propositional Rule Learner (PRL)	Non-parametric	5.5.1.12	E.20
Support Vector Machine (SVM)	Non-parametric	5.5.1.13	E.21

5.5.1 Predictive models

Due to the binary nature of the responses (either reusable or non-reusable), the process of predicting the reusability of the structural elements of a building is a classification problem. In a classification setting, the classifier would predict if an element is reusable (1) or not (0). An optimum classifier is one that can classify unseen observations with the minimum incorrect classifications (James et al. 2017). In this study, thirteen different methods are used to develop the BSE-RPMs (Table 5.10). These models are fitted to the training sets of the TEC, ECO, and SOC datasets (Section 5.5) and then used to predict the technical, economic, and social reusability of the elements in the testing sets to evaluate the performance of the fits. In the next subsections, each of these methods is discussed briefly.

It should be noted that this research adopts a probabilistic approach, meaning that a predictive model selects the label with maximal probability given the features. This rule, which is known as conditional probability, is defined as follows.

$$pr(A|B) = \frac{pr(A \cap B)}{pr(B)} \; if \; pr(B) > 0 \tag{5.1}$$

In (5.1), $pr(A|B)$ is "the conditional probability of event A, given that event B is true", $pr(A \cap B)$ is the joint probability of both events, and $pr(B)$ is the probability of event B (Murphy 2012).

It should be noted that this research considers the Bayes classifier threshold value of 0.5 for the probability of an element being reusable or not. It means that if the conditional probability of an element being reusable given the features is being calculated (i.e., $pr(reusability = 1|\boldsymbol{X} = \boldsymbol{x})$), the probabilities above 0.5 conclude that the item is reusable. Otherwise, it would be classified as non-reusable.

The Bayes classifier is a very simple classifier that assigns an observation to the most probable response class based on the values of its feature (James et al. 2017). This classifier works based on the conditional distribution of the response given the features and results in the highest theoretical accuracy (James et al. 2017). In this study, the conditional probability of the reusability (response) equal to one (reusable) can be presented as below:

$$pr(reusability = 1 \,\middle|\, \boldsymbol{X} = \boldsymbol{x}) \tag{5.2}$$

In the conditional probability (5.2), $\boldsymbol{x} = (x_1, x_2, \dots, x_p)$ represents all applicable features in each dataset for every datapoint (Section 5.4.4). If the value of conditional probability given in (5.2) is higher than 0.5, then the Bayes classifier classifies the observation as reusable, otherwise, non-reusable ('pr' means probability) (James et al. 2017). The left-hand panel of Figure 5.8 shows a simplified classification problem with two features (x_1 & x_2) using the Bayes theorem (Scutari and Denis 2015, Witten et al. 2017). The black dashed line is the Bayes decision boundary. The black circles correspond to reusable training structural elements, and the plus signs represent non-reusable training structural components. For each of the values of x_1 and x_2, the probability of an element to be reusable or non-reusable is different. It is imagined that the exact location of the Bayes decision boundary is known because it is assumed that the conditional distribution of the reusability of the elements is known. For an unseen observation, based on the values of x_1 and x_2, if the element falls on the left-hand side of the Bayes decision boundary, the component is reusable; otherwise, it is non-reusable. For those elements falling on the decision boundary, the component is considered non-reusable.

In theory we would always like to predict qualitative responses using the Bayes classifier. But for real data, we do not know the conditional distribution of Y given X, and so computing the Bayes classifier is impossible. Therefore, the Bayes classifier serves as an unattainable gold standard against which to compare other methods. Many approaches attempt to estimate the conditional distribution of Y given X, and then classify a given observation to the class with highest estimated probability. One such method is the K-nearest neighbours (KNN) classifier.

5.5.1.1 K-Nearest Neighbours (KNN)

The K-nearest neighbours (KNN) classifier is a method that attempts to estimate the Bayes classifier (James et al. 2017).

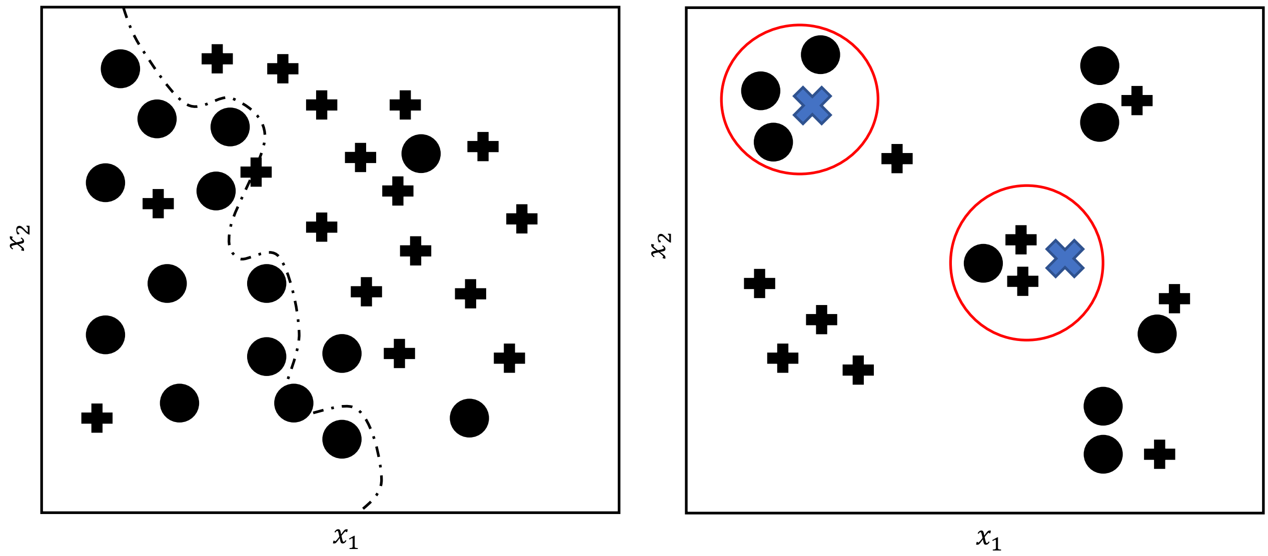

Figure 5.8 The Bayes classifier (left) and K-Nearest Neighbours (KNN) classifier (right)

However, the conditional distributions of the technical, social, and economic reusability of the structural elements of a building are unknown. Therefore, for an unseen data point, the KNN classifier looks for the K closest data points to the new observation in the training set (K is an arbitrary positive integer) and classifies the test observation to the class with the highest probability (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The KNN method assumes that the reusability of a new recovered structural element is like its nearest neighbours in the training dataset. This process is shown on the right-hand panel of Figure 5.8. If K=3, then the KNN classifier classifies the new observation (shown with a cross sign) on the top-left corner as reusable because the three nearest neighbours in the training dataset are reusable, yielding a class probability of 100%. However, the new observation in the centre is adjacent to two non-reusable and one reusable element in the training dataset. In this case, this new element would be classified as non-reusable since two-third of its nearest neighbours in the training dataset are non-reusable, and only one-third is reusable.

The choice of the number of neighbours has a considerable impact on the prediction results (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). While the number of K depends on the sample size, theoretically, it is possible to assign any positive integer to K (James et al. 2017). However, if K is too small (for instance, equal to one), the classifier would strictly follow the training observations and becomes highly flexible, it might overfit, and potentially results in a model with high variance and low bias (James et al. 2017). On the other hand, large values of K can potentially make the classifier less flexible, which results in a low variance model with a high bias (James et al. 2017). In this study, using the standard holdout method (equal to two-third of the training observations), the value of the number of neighbours was estimated. Accordingly, the value of K used for modelling is equal to six, five, and eight for the TEC, ECO, and SOC datasets, respectively (See Appendix E, Script E.9).

In this study, the mlr and kknn packages are used to develop the predictive model based on the KNN classifier (Bischl et al. 2016, Schliep, Hechenbichler, and Lizee 2016, Hechenbichler and Schliep 2004). It is noteworthy that the KNN classifier is a nonparametric method and does not assume any functional form for the relationship between the response and features.

5.5.1.2 Logistic Regression (LR)

Logistic regression (LR) directly models the probability that an element is reusable or not (James et al. 2017). Unlike the KNN method, LR assumes a functional form for the relationship between the response and factors affecting reuse (features) in its attempt to predict the reusability; hence, it is a parametric machine learning approach (James et al. 2017, Murphy 2012). So, the conditional probability (5.2) can be written in the following form.

$$p(\boldsymbol{X}) = \mathrm{pr}(reusability = 1|\boldsymbol{X} = \boldsymbol{x}) \tag{5.3}$$

LR uses (5.4), the logistic function, to calculate $p(\boldsymbol{X})$ and employs the Maximum Likelihood estimation method to fit the model based on the training observations (James et al. 2017, Murphy 2012).

$$p(\boldsymbol{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}} \tag{5.4}$$

It is noteworthy that the logistic function (5.4) results in values between zero and one. In (5.4), the β_p (betas) are unknown constants that should be identified (James et al. 2017). Hence, in LR, the problem of identifying the relationship between $p(\boldsymbol{X})$ and $\boldsymbol{X}$ in the training set is reduced to estimating these coefficients (James et al. 2017). In this case, the Maximum Likelihood (MLE) seeks estimates of these betas, so (5.4) yields a probability close to one for reusable elements, and to zero for non-reusable components (James et al. 2017).

After estimating the unknown constants in (5.4) using the training data, this classifier assigns a new observation given its feature values to one of the two classes based on the quantity of $p(\boldsymbol{X})$ and a threshold value (James et al. 2017, Murphy 2012). If the Bayes classifier threshold value of 0.5 is assumed, then for $p(\boldsymbol{X}) > 0.5$, the classifier predicts the element reusable (James et al. 2017). However, a conservative designer might choose a higher threshold value to decrease the probability of making a false positive error (Section 5.5.2.3) (James et al. 2017).

In this study, the mlr package is used to develop the predictive model based on the LR classifier (Bischl et al. 2016).

5.5.1.3 Linear Discriminant Analysis (LDA)

Like the KNN method, linear discriminant analysis (LDA) attempts to estimate the Bayes classifier (James et al. 2017). The LDA method considers a functional form (the discriminant function) for the relationship between the response and factors affecting reuse; hence, like the logistic regression, it is a parametric machine learning approach (Murphy 2012). However, unlike the LR, LDA does not directly estimate the conditional probability (5.2) (James et al. 2017).

Using the Bayes' theorem (Scutari and Denis 2015, Witten et al. 2017), (5.2) can be written as follows, where k corresponds to non-reusable (0) or reusable (1) classes.

$$\mathrm{pr}(reusability = k|\boldsymbol{X} = \boldsymbol{x}) = \frac{pr(\boldsymbol{X} = \boldsymbol{x}|reusability = k)pr(reusability = k)}{pr(\boldsymbol{X})} \tag{5.5}$$

In (5.5), $pr(\boldsymbol{X}|reusability = k)$ is known as the density function of $\boldsymbol{X}$ for a structural element that belongs to class k, $pr(reusability = k)$ is the prior probability which is the probability that a given observation belongs to class k, and $pr(\boldsymbol{X})$ is the overall probability of $\boldsymbol{X}$ in the dataset (James et al. 2017). In (5.5), the prior probability is simply the result of the number of elements in each of the training classes divided by the total number of components in the training dataset (James et al. 2017). The conditional probability in (5.5) can be re-written as follows (James et al. 2017):

$$p_{\mathrm{k}}(\boldsymbol{X}) = \frac{f_{\mathrm{k}}(\boldsymbol{x})\pi_{\mathrm{k}}}{\sum_{s=1}^{k} \pi_s f_{\mathrm{s}}(\boldsymbol{x})} \tag{5.6}$$

In (5.6), $p_k(\boldsymbol{X})$ is the posterior probability that an observation is reusable or not, given the values of its features (James et al. 2017). Therefore, the LDA classifier needs to estimate the value of $f_k(\boldsymbol{x})$ (the density function) and π_k (the prior probability) and plug them into (5.6) to evaluate the posterior probability (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The LDA method assumes a one-dimensional normal distribution for each independent variable in (5.6) (a multivariate Gaussian distribution) and equal variance for the class responses (James et al. 2017). The density function in (5.6) can be then converted to the following (for further details, refer to (James et al. 2017, Hastie, Tibshirani, and Friedman 2009)):

$$\delta_{\mathrm{k}}(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\Sigma^{-1}\mu_{\mathrm{k}} - \frac{1}{2}\mu_{\mathrm{k}}{}^{\mathrm{T}}\Sigma^{-1}\mu_{\mathrm{k}} + log\,\pi_{\mathrm{k}} \tag{5.7}$$

The above is known as the discriminant function (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The LDA method estimates Σ (the covariance matrix that is common to reusable and non-reusable components), and μ_k (mean vector of the features in each class) to evaluate $\delta_k(\boldsymbol{x})$ in the training dataset (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The Bayes classifier then classifies a new observation as reusable or non-reusable for which the value of the corresponding $\delta_k(\boldsymbol{x})$ is higher (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The word 'linear' in this method stems from the fact that the discriminant function is a linear function of $\boldsymbol{x}$ (James et al. 2017, Hastie, Tibshirani, and Friedman 2009).

5.5.1.4 Quadratic Discriminant Analysis (QDA)

Quadratic discriminant analysis (QDA) is a similar approach to the LDA with the exception that, in the QDA method, each class has its covariance matrix (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). Moreover, in QDA, the discriminant function is a quadratic function of predictors $\boldsymbol{x}$ (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). The QDA method is more flexible and can handle the possible non-linear relationship between the features and the response in each dataset (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). For further details, please refer to (James et al. 2017, Hastie, Tibshirani, and Friedman 2009).

5.5.1.5 Naïve Bayes (NB)

The Naïve Bayes (NB) classifier is a non-parametric method that attempts to estimate the conditional probability of the reusability of a structural element given its features by making the naïve assumption that these features are independent (Murphy 2012, Hastie, Tibshirani, and Friedman 2009). Considering a conditional probability where there is only one independent variable X, (5.2) can be written as:

$$pr(Y = k \,|\, X = x) \qquad (5.8)$$

(5.8) can be calculated by identifying the portion of the response (a common area) for which the independent variable X is equal to x using the MLE method.

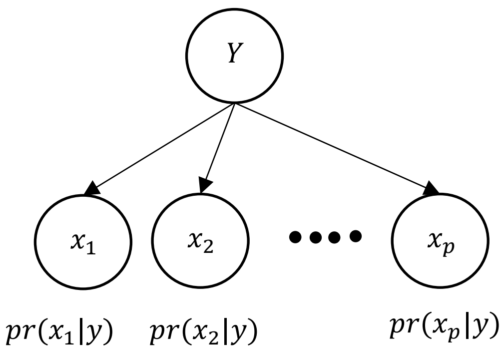

Figure 5.9 The independence of features assumed in the Naïve Bayes (NB) classifier

However, considering all the applicable reusability factors in (5.2), this common area would be very close to zero; hence, the classifier cannot make predictions (Weinberger 2018). The NB method addresses this problem by using (5.5), the Bayes' theorem (Scutari and Denis 2015, Witten et al. 2017), and making the naïve assumption that all the features are independent, given the response (Hastie, Tibshirani, and Friedman 2009, Witten et al. 2017). The independence of features assumed in the NB classifier is illustrated in Figure 5.9. Therefore, considering the above assumption, the density function $pr(\boldsymbol{X} = \boldsymbol{x}|reusability = k)$ in (5.5) can be written as follows.

$$pr(\boldsymbol{X} = \boldsymbol{x}|reusability = k) = \prod_{a=1}^{p} pr(\boldsymbol{X} = x_{\mathrm{a}}|reusability = k) \tag{5.9}$$

As discussed in Section 5.5.1.1, the Bayes classifier then assigns an observation to the most likely response label (here, reusable or non-reusable) using the Bayes' theorem (5.5) (Scutari and Denis 2015, Witten et al. 2017).

In this study, the mlr and e1071 packages are used to develop the predictive model based on the Naïve Bayes classifier (Bischl et al. 2016, Dimitriadou et al. 2019).

5.5.1.6 Decision Trees (DT)

Decision trees are machine learning methods that include stratifying the feature space of the training set into a smaller number of regions (known as terminal nodes or leaves) with similar class labels. In this method, to classify a new observation belonging to a terminal node, the mean or mode of the training observations in that leaf is considered (James et al. 2017).

The set of possible values of the 'p' predictors $(x_1, x_2, \ldots, x_p)$ of the structural elements in the training data is divided into K number of leaves $(R_1\ to\ R_k)$, which are not overlapping (James et al. 2017). Then, for an unseen observation that satisfies R_k, the DT classifier classifies a new structural element to the most commonly occurring class response of the training set in R_k (James et al. 2017). This process is shown in Figure 5.10. The left-hand panel of Figure 5.10 shows the entire dataset with the class labels and splits. In this figure, the training observations are marked with black circles (reusable) and black plus signs (non-reusable). The complete dataset is the combination of regions $R_{1.1}$, $R_{1.2}$, $R_{2.1}$, and $R_{2.2}$. Initially, the dataset was split into two regions or leaves, R_1 and R_2 (James et al. 2017). Next, to increase the purity of the regions, R_1 was divided into $R_{1.1}$ and $R_{1.2}$, and R_2 was split into $R_{2.1}$ and $R_{2.2}$ (James et al. 2017). The DT method then classifies a new observation (shown as a cross) as reusable because it is the most frequent class label in region $R_{1.1}$. The right-hand panel of Figure 5.10 shows the process of classifying a new observation using the DT method.

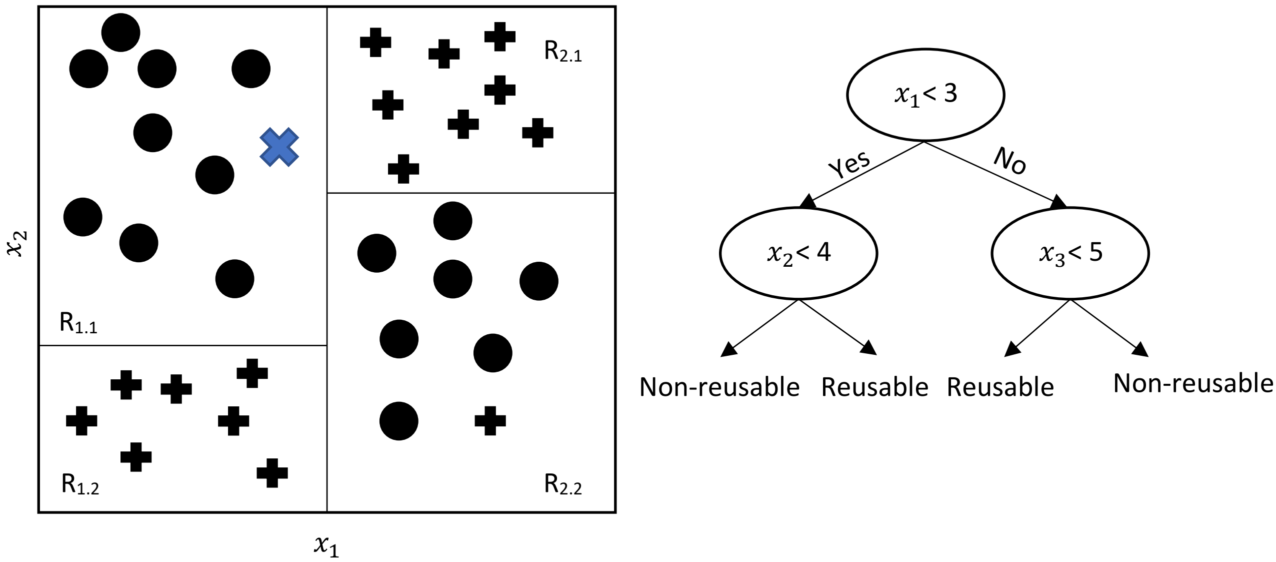

Figure 5.10 The Decision Trees (DT) method

The DT method attempt to create a set of leaves for which the resulting splits have the lowest class impurity (James et al. 2017). For this purpose, the DT method employs recursive binary splitting, which is a top-down greedy approach (James et al. 2017). At each stage, the recursive binary splitting method selects an independent variable x_j with a cut-point value of s (s is any value belong to x_j) and splits the feature space of an existing node into the new terminal nodes $\{x \mid x_j < s\}$ and $\{x \mid x_j \geq s\}$ with the highest possible purity in response classification (James et al. 2017). It is noteworthy that the split happens on the training observations available in a region and not the entire training dataset. The DT method uses the Gini index or the entropy impurity function measures to assess the purity of the splits at each stage (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). After each split, if the resulting purity of the new leaves is not satisfactory, the splitting continues to decrease the impurity of the new terminal nodes (James et al. 2017). This process continues until no further improvement is possible, resulting in a deep tree (James et al. 2017). Alternatively, the process can be stopped by setting a termination condition, such as reaching a minimum number of observations in a region (James et al. 2017). For further details on Gini and entropy impurity functions, please refer to (James et al. 2017, Murphy 2012, Hastie, Tibshirani, and Friedman 2009).

5.5.1.7 Random Forests (RF)

Decision trees (DT) explained in section 5.5.1.6 suffer from high variance, which means any change in the training dataset can potentially affect the resulting predictions (James et al. 2017). One reason is that during the first split, the dataset is roughly divided into two sections (James et al. 2017, Cortez and Embrechts 2013). Hence, if a predictive model is fit to each of the splits, the resulting predictions are not necessarily the same (James et al. 2017). One way to address this problem is by decreasing the depth of a DT model (James et al. 2017). However, this method increases the bias in the model and consequently decreases its accuracy (James et al. 2017). Another solution is to create an ensemble of decision trees using different datasets drawn from a population and averaging the results to decrease the variance (James et al. 2017). This notion is the result of the weak law of large numbers (de Alencar and de Alencar 2016). According to this law, averaging various independent observations decreases variance (James et al. 2017). Ideally speaking, by increasing the number of observations to infinity, the variance should diminish (de Alencar and de Alencar 2016). Nonetheless, this method is also not practical because of the limited access to many training datasets (for this study, the reasons are explained in Section 4.7) (James et al. 2017).

Random forests (Figure 5.11) are machine learning methods that try to address the above issue by creating many trees with maximum depth (yielding in low bias but high variance) and averaging the resulting variance through bagging (bootstrap aggregation) (James et al. 2017, Hastie, Tibshirani, and Friedman 2009). Bagging is an ensemble method that draws many samples with replacement from a dataset $D = (D_1, D_2, \dots, D_m)$ (Murphy 2012, Hastie, Tibshirani, and Friedman 2009). The replacement in this process means that one structural element in the training set can appear more than once in the bootstrap dataset (Torgo 2016). Then, the RF method fits a decision tree with the maximum possible depth to each of the new datasets, creating an ensemble of bagged trees (James et al. 2017). Before dividing the feature space at every stage, a random number of $m \approx p^{1/2}$ (p is the number of predictors in the dataset) independent variables are selected as eligible predictors from which one is picked by the method to split (without replacement) (James et al. 2017). The lack of replacement in this process makes sure that the method does not pick a specific predictor repeatedly. This approach is highly advantageous because it makes sure that the bagged trees remain uncorrelated (James et al. 2017, Murphy 2012). Whereas without this limitation, there is a high chance that all the developed trees become highly correlated, which results in a small improvement in the variance of the final model, compared to a single tree (James et al. 2017). It is because, in the presence of an influential independent variable, there is a high chance that each tree chooses that strong predictor as its root node, resulting in a similar and highly correlated ensemble of trees (James et al. 2017).

The RF method uses the ensemble of bagged trees to make predictions (James et al. 2017). While the way every single tree predicts the class of a new observation is like the DT method (Section 5.5.1.6) (James et al. 2017), the RF method predicts if a new structural element is reusable or non-reusable based on the class label with the highest number of records.

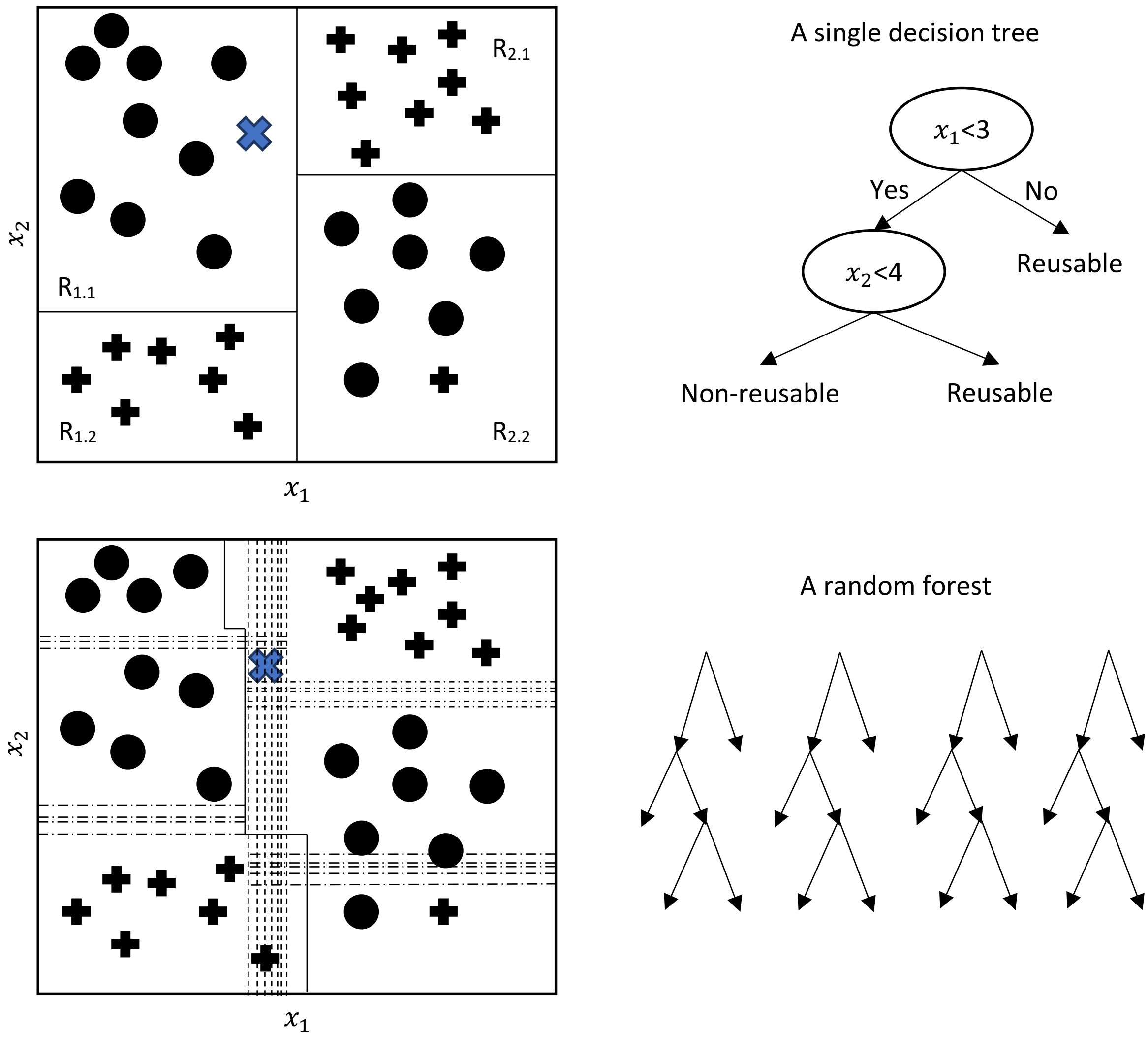

Figure 5.11 A simplified Random Forest. Top: A Decision Tree (top right) divides the feature space (top left). Bottom: A Random Forest which is a group of Decision Trees (bottom right) divide the feature space (bottom left). The cross is a new observation.

5.5.1.8 Adaptive Boosting (AB)

Boosting methods can be employed to improve the predictions from any machine learning method with high bias and high training error rate (weak learners) (James et al. 2017, Murphy 2012, Hastie, Tibshirani, and Friedman 2009). In this study, the 'AdaBoost' methods introduced by (Freund and Schapire 1997) is employed to decrease the bias in decision trees with a limited number of nodes (resulting in low variance and high bias) and increase the accuracy of predictions on unseen observations. Like random forests, adaptive boosting is an ensemble technique; however, it works quite differently (James et al. 2017). Instead of creating an ensemble of decision trees through bootstrapping, adaptive boosting creates $M-1$ new decision trees sequentially, resulting in M number of ensembled decision trees (Hastie, Tibshirani, and Friedman 2009). The first classifier is a conventional decision tree, like the one explained in Section 5.5.1.6 (Hastie, Tibshirani, and Friedman 2009). However, in creating the

$M - 1$ decision trees, the AdaBoost method alters the original dataset by weighting observations in the main dataset so that the misclassified observations are weighted higher and the correctly predicted data points are weighted lower (Hastie, Tibshirani, and Friedman 2009). Hence, the next stage decision tree focuses on those observations with the wrong classification in the previous stage (Hastie, Tibshirani, and Friedman 2009). Finally, the predictions from the ensemble of the decision trees are weighted by the AdaBoost method, so those highly accurate decision trees on the training data are weighted higher than those with poor performance (Hastie, Tibshirani, and Friedman 2009). For further details on the AdaBoost method, refer to Section 16.4 of (Murphy 2012).

5.5.1.9 BART Machine (BM)

BART (Bayesian additive regression trees) is an ensemble of decision trees with an arbitrary number of trees to be decided by the researcher (Chipman, George, and McCulloch 2010). Unlike random forests (Section 5.5.1.7) or adaptive boosting (Section 5.5.1.8) where a structural element is classified based on the most commonly occurring class response, it relies on the Bayesian probability model (Murphy 2012, Chipman, George, and McCulloch 2010). Therefore, it consists of priors for the structure and the terminal node parameters and a likelihood for data in the leaves (Chipman, George, and McCulloch 2010). The priors considered guarantee no single decision tree dominates the total model; hence, regularising the ensemble of trees (Chipman, George, and McCulloch 2010). It is noteworthy that according to the developers, the optimum number of trees is around 200 (Chipman, George, and McCulloch 2010). To predict an observation, BART uses the posterior average probability to classify a structural element as reusable or not (Chipman, George, and McCulloch 2010, Kapelner and Bleich 2016). For further details on the BART method, refer to (Chipman, George, and McCulloch 2010).

5.5.1.10 Artificial Neural Networks (ANN)

Neural networks are machine learning methods working based on the way the human brain works (Ciaburro and Venkateswaran 2017). Neural networks attempt to develop new features based on linear combinations of the input variables (reusability factors) and then predict the probabilities of the responses (reusable or non-reusable) using a nonlinear function of the newly extracted predictors (Hastie, Tibshirani, and Friedman 2009). Therefore, neural networks can be categorised as nonlinear parametric models (Hastie, Tibshirani, and Friedman 2009, Murphy 2012).

In machine learning, the architecture of any neural network (Figure 5.12) consists of a set of inputs (reusability factors), a processing unit (which includes a single or multiple hidden layers), and output(s) (reusable or not-reusable) (Hastie, Tibshirani, and Friedman 2009). There are two main groups of neural networks, feed-forward, and feed-backward neural networks (Ciaburro and Venkateswaran 2017). In feed-forward neural networks, the signal can only move in one direction from the input layer to the hidden layer(s), and finally to the output layer. However, in feed-backward neural networks, before a signal reaches the next level, it can go back to the previous level (Ciaburro and Venkateswaran 2017). Artificial neural networks (ANNs) fall under the former category, while recurrent neural networks (RNNs) fall under the latter (Ciaburro and Venkateswaran 2017). In this study, the reusability of building structural elements are assessed using a special case of ANNs.

An ANN can be a single layer perceptron (with only one hidden layer) or a multiple layer perceptron (Hastie, Tibshirani, and Friedman 2009). The architecture of a double layer perceptron is shown in Figure 5.12. According to this figure, the units in the middle layer (hidden units) develop new features. These new features are then used to determine the reusability probability of a structural element at the end-of-life of a building (5.10) (Hastie, Tibshirani, and Friedman 2009).

$$\begin{aligned} D_k &= \sigma\left(\alpha_{0k} + \alpha_k^T X\right) \\ T_l &= \beta_{0l} + \beta_l^T D \\ f_l(X) &= g_l(T) \end{aligned} \qquad (5.10)$$

In (5.10), $X = (X_1, X_2, \dots, X_p)$ denotes the input variables, $k = 1,2,\dots,K$, $l = 1,2,\dots,L$, $D = (D_1, D_2, \dots, D_K)$ represents the derived features, $T = (T_1, T_2, \dots, T_L)$ is the vector of outputs, and α_{0k} and β_{0l} are the intercepts. In (5.10), the output function $g_l(T)$ is the softmax function, which transforms the vector of outputs T and produces positive estimates that sum to one. Other than the three layers explained earlier (inputs layer, hidden layer(s), and output layer), an ANN consists of weights, biases, and an activation function, as well. In (5.10), $f_l(X)$ calculates the probability that a structural element is reusable or not, and σ is the activation function, which in the case of this study (classification problem), is a Sigmoid (Hastie, Tibshirani, and Friedman 2009). The weights are the unknowns in (5.10) and are summarised in (5.11) (Hastie, Tibshirani, and Friedman 2009). In (5.10) and (5.11), p is the number of independent variables. The goal is to estimate these weights so that the ANN model fits the training dataset well (Hastie, Tibshirani, and Friedman 2009). Therefore, to guarantee an accurate model, a measure of fit is required to evaluate the quality of the model. The measure-of-fit is calculated using the squared error or cross-entropy (Hastie, Tibshirani, and Friedman 2009). For further details about the measure-of-fit please refer to (Hastie, Tibshirani, and Friedman 2009).

$$\{\alpha_{0k}\,, \alpha_k; k = 1,2, \dots, K\}\; K(p+1)\; weights,$$
$$\{\beta_{0l}, \beta_l; l = 1,2, \dots, L\}\; L(K+1)\; weights \tag{5.11}$$

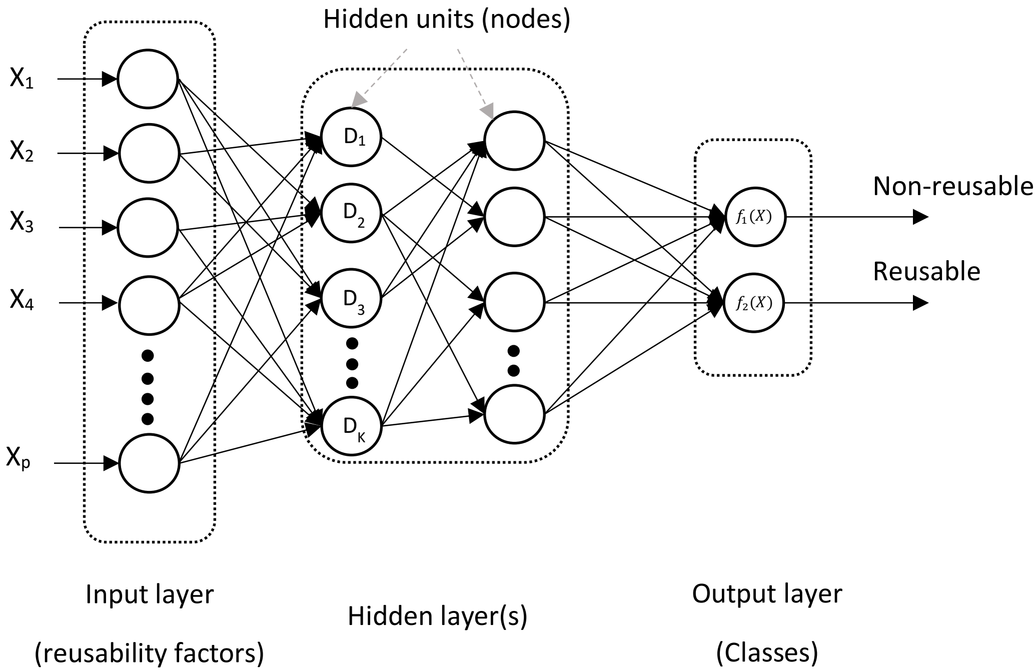

Figure 5.12 The Artificial Neural Networks (ANN) architecture (two hidden layers)

The role of an ANN model is then reiterating two major stages until it reaches a minimum training set error rate. Firstly, estimating the reusability of the building structural elements based on weighted inputs, biases, and a specific activation function in the forward propagation stage. Next, determining the error rates and estimating the weights and biases using the backward propagation algorithm (Ciaburro and Venkateswaran 2017). One of the most common problems that one could encounter while training an ANN is overfitting (Murphy 2012). Because the predicted responses/trends of an overfitted model do not follow the reality present in the data, such a model is inaccurate. There are various techniques to prevent overfitting while training neural networks. One of the widely used solutions is early stopping. Early stopping is a form of regularisation while training a model with an iterative method, such as gradient descent. This method updates the model to make it better fit the training data with each iteration. Up to a point, this improves the model's performance on data on the test set. Past that point, however, improving the model's fit to the training data leads to increased generalisation error. Regularisation is an alternative method that is commonly used to overcome the overfitting problem. This method introduces a weight decay (a penalty term) to the loss function to reduce the model's complexity.

According to Hastie et al. (2009), training neural networks requires pre-processing and extra precautions. This can be done by determining an optimum weight decay, scaling of the inputs, and assigning the number of hidden layers and nodes. The neural network method employed in this study is a single layer perceptron that uses the Sigmoid function to activate the neurons in the network. Moreover, the input variables are scaled, and two hyperparameters (size of the hidden nodes, and weight decay) are evaluated using ten-fold cross-validation on the training set considering AUC as the determining metric. The estimated hyperparameters (size and decay) for each of the datasets are as follows: TEC ($size = 9$, $decay = 0.09$), ECO ($size = 9$, $decay = 0.08$), SOC ($size = 8$, $decay = 0.04$) (Script E.18).

In this study, the reusability of building structural elements is assessed using an ANN through the nnet and rminer packages (Venables and Ripley 2002, Cortez 2020).

5.5.1.11 Gaussian Processes (GP)

Gaussian processes are nonparametric supervised machine learning methods that can be used for both regression and classification problems. In this study, Gaussian processes for classification (GPC) are used to predict the reusability probabilities of the recovered building structural elements. A GPC is a function approximation task where instead of directly estimating the class probabilities considering a predetermined functional form (such as LDA), the functional relationship is determined through a multivariate Gaussian distribution.

We consider a data set D=$\{(x_i, y_i | i = 1,2, \dots, n\}$, consisting of n samples, wherein x_i denotes the vector of input data taken from the input space, and $y_i = f(x_i)$ denotes the corresponding output (dependent variable) observation. Following (Rasmussen and Williams 2006), the GP prior model is given by (5.12):

$$f(\boldsymbol{x}) \sim GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \tag{5.12}$$

where $m(\boldsymbol{x})$ is the mean function, which is commonly and without loss of generality considered to be zero and the kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$; where $\boldsymbol{x}$ represents the training datapoint in each dataset of the structural elements. We use the radial basis function (or squared exponential) as the kernel function, see (Rasmussen and Williams 2006, Daneshkhah, Hosseinian-Far, and Chatrabgoun 2017) for the details of this kernel, including the functional form, and how the hyperparameters (smoothness parameters) of this kernel can be estimated in the light of the observed data.

The joint prior distribution of the training outputs, $\boldsymbol{f}$, and the predicted output $\boldsymbol{f}_*$ (corresponding to the test input $\boldsymbol{x}_*$), according to GP definition given in Eq. (5.12) and the properties of multivariate normal distribution, is given by (5.13):

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} \sim N(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}) \tag{5.13}$$

Where $K = k(\boldsymbol{X}, \boldsymbol{X})$, $K_* = k(\boldsymbol{X}_*, \boldsymbol{X})$, $K_*^T = k(\boldsymbol{X}, \boldsymbol{X}_*)$, $K_{**} = k(\boldsymbol{X}_*, \boldsymbol{X}_*)$, and $\boldsymbol{X}_{n\times p}$ denotes an $n \times p$ matrix of the training inputs $\{\boldsymbol{x}_i\}_i^n$ (also known as the design matrix), p stands for the dimension of input space $\boldsymbol{X}$, and $\boldsymbol{X}_*$ is the matrix of test inputs. We use the subscript $*$ to differentiate the test/predicted data from the training ones.

The posterior distribution of $\boldsymbol{f}_*$ can be obtained/derived by conditioning the joint prior distribution, given in Eq. (5.13) on the training datapoint (5.14):

$$\boldsymbol{f}_* | \boldsymbol{f}, X, X_* \sim N(K_*^T K^{-1} \boldsymbol{f}, K_{**} - K_* K^{-1} K_*^T) \tag{5.14}$$

The mean and covariance of this posterior distribution can be used as an estimate of the predicted value of $\boldsymbol{f}_*$, and uncertainty/sensitivity (Daneshkhah and Bedford 2008).

The GP that is briefly explained above, can be used as an efficient classifier by computing predictions in form of class probabilities of $y_* = f(\boldsymbol{x}_*)$ for the new test input $\boldsymbol{x}_*$. This can be done by squashing the output of a regression model through a logistic function (e.g. sigmoid function, $\sigma(.)$) to transform it from a domain of $(-\infty, +\infty)$ to $[0, 1]$ (Rasmussen and Williams 2006). For a new observation $\boldsymbol{x}_*$, the distribution of the latent variable f_* is calculated using (5.15):

$$pr(f_* | X, \boldsymbol{y}, \boldsymbol{x}_*) = \int pr(f_* | X, \boldsymbol{y}, \boldsymbol{x}_*) pr(\boldsymbol{f} | X, \boldsymbol{y}) d\boldsymbol{f} \tag{5.15}$$

Then, using the above distribution, the probabilistic prediction is performed using (5.16):

$$pr(y_* = reusable | X, \boldsymbol{y}, \boldsymbol{x}_*) = \int \sigma(f_*) pr(f_* | X, \boldsymbol{y}, \boldsymbol{x}_*) df_* \tag{5.16}$$

However, since (5.15) is non-Gaussian (response is discrete), the above integrals are approximated using the Laplace approximation method (Rasmussen and Williams 2006).

5.5.1.12 Propositional Rule Learner (PRL)

Propositional rule learner (PRL) is a classification machine learning method that finds patterns in each dataset and expresses them in terms of a set of if-then rules (Fürnkranz, Gamberger, and Lavrač 2012). These rules are then used to classify new structural elements that satisfy a rule condition. The method develops a predictive model in three stages. A PRL method first converts the features in the training dataset into sets of binary features (Fürnkranz, Gamberger, and Lavrač 2012). Then it constructs the individual rules, each covering a part of the training dataset using a covering method (Fürnkranz, Gamberger, and Lavrač 2012). At this stage, the method learns a rule that covers a part of the training observations. Then it removes those covered datapoints and learns a new rule based on the remaining observations (Fürnkranz, Gamberger, and Lavrač 2012). The method recursively performs these tasks until all training observations are covered by a rule (Fürnkranz, Gamberger, and Lavrač 2012). Finally, it combines all the learned rules and forms the predictive model (Fürnkranz, Gamberger, and Lavrač 2012). For further details about this method please refer to (Fürnkranz, Gamberger, and Lavrač 2012).

In this study, the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) method (Cohen 1995) through the RWeka package is used to develop the predictive rule learning model (Hornik, Buchta, and Zeileis 2009).

5.5.1.13 Support Vector Machines (SVM)

Support vector machines (SVM) are machine learning methods that convert a linear classifier (known as support vector classifier) in a way to produce a non-linear decision boundary between classes (two-class responses) (James et al. 2017).

A support vector classifier is a computationally efficient method for developing linear decision boundaries between two-class responses (James et al. 2017). The support vector classifier develops a hyperplane to split the observations in the training dataset into two classes (Figure 5.13) (James et al. 2017). This classifier depends only on the training observations close to the hyperplane known as the support vectors (James et al. 2017). In the left-hand panel of Figure 5.13, the left-hand side of the hyperplane represents the circle responses (reusable), and the right-hand side of the decision boundary corresponds to the plus class (non-reusable). The dashed lines in this figure are margins for the hyperplane. In Figure 5.13, only the observations on the margin or crossing the margin but on the proper side of the decision boundary are the support vectors (James et al. 2017). Therefore, training data far from the margins (and the hyperplane) do not play any role in predicting the class response for a new observation (James et al. 2017).

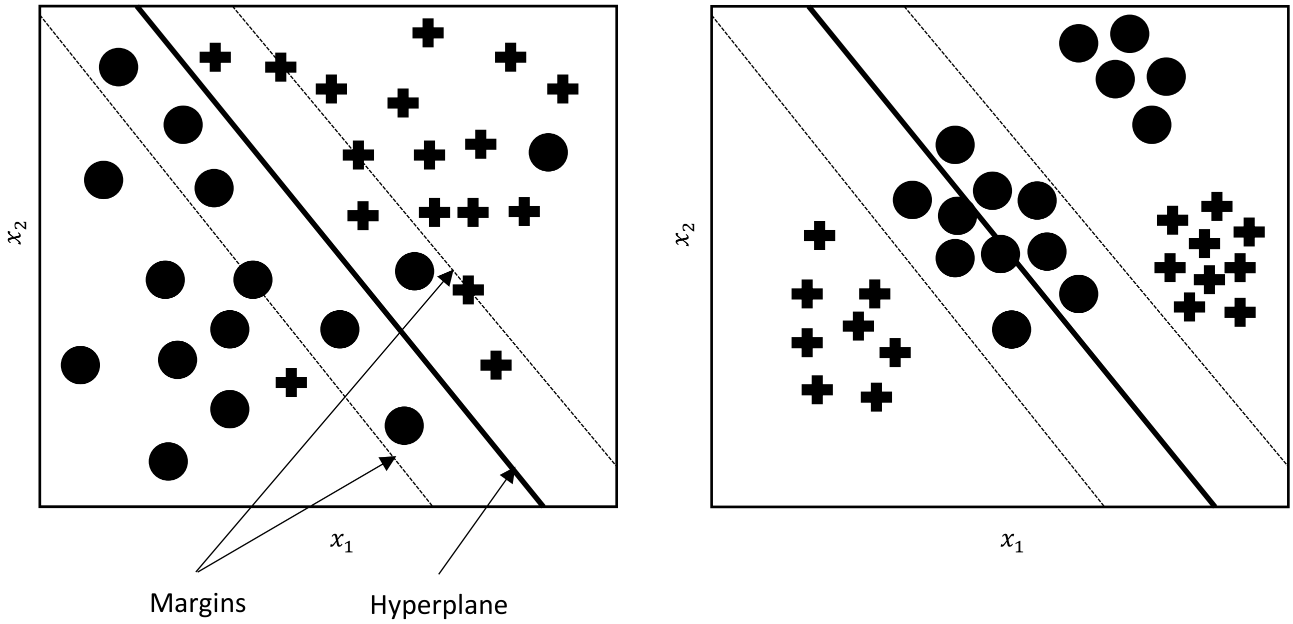

Figure 5.13 The Support Vector Classifier

The support vector classifier can be represented as follows (James et al. 2017):

$$f(x) = b_0 + \sum_{i \in S} a_i < x, x_i >, \tag{5.17}$$

$$(S = \ indices\ for\ the\ support\ vectors)$$

In (5.17), $< x, x_i >$ is the inner product of the new observation x with all support vectors, b_0 is an intercept, and a_i is a parameter required for each of the support vectors (James et al. 2017). Function (5.17) is the solution function for an optimisation problem for the support vectors. The details of the optimisation problem are available in Section 9.2.2 of (James et al. 2017). Moreover, the solution to the optimisation problem can be found in Section 12.2.1 of (Hastie, Tibshirani, and Friedman 2009).

The left-hand panel of Figure 5.13 represents a classification problem with separable (almost) class responses where the hyperplane does a reasonable job in classifying the non-reusable and reusable classes. However, in many instances, the relationship between the predictors and the responses are not linear (James et al. 2017). The right-hand panel of this figure shows an example of such a problem. As can be observed, the separating hyperplane is useless in this situation. In this case, no linear classifier can effectively separate the two classes, as the relationship between the predictors and the responses are non-linear.

The support vector machine method attempts to overcome the above limitation by enlarging the feature space using kernel functions; hence, creating non-linear decision boundaries (James et al. 2017). Kernel functions quantify the similarity of two observations and can have various forms, including radial, polynomial, hyperbolic, Laplacian, etc. (James et al. 2017). By replacing the inner product in (5.17) with the kernel, the solution function (5.17) can be re-written as (5.18), where $K(x, x_i)$ is the kernel function (James et al. 2017):

$$f(x) = b_0 + \sum_{i \in S} a_i K(x, x_i), \tag{5.18}$$

$$(S = \ indices\ for\ the\ support\ vectors)$$

In this study, a radial kernel is used to expand the feature space, and eventually develop non-linear decision boundaries between the classes. Therefore, (5.19) formulates the radial kernel.

$$K(x_i, x_{i'}) = \exp\left(-\sigma \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2\right), \tag{5.19}$$

$$where\ \sigma\ is\ a\ positive\ constant$$

In (5.19), x_i and $x_{i'}$ indicate two different observations in the training set, p is the number of predictors, and σ (sigma) controls the non-linearity of the kernel function (James et al. 2017). By increasing the value of σ, the fit becomes more non-linear (James et al. 2017). While this increased non-linearity can decrease the variance on the training dataset, it might increase the chance of overfitting (James et al. 2017). Hence, care must be taken while choosing the correct value for σ (James et al. 2017). Another hyperparameter that is required to be selected is known as cost (represented by C) (James et al. 2017). This quantity determines the width of the margin in Figure 5.13, and correspondingly the number of support vectors (James et al. 2017). This tuning parameter is used to determine a_i in (5.17) and (5.18) (see Section 12.2.1 of (Hastie, Tibshirani, and Friedman 2009)). In this study, the hyperparameters (C and sigma) are calculated using ten-fold cross-validation on the training set (Murphy 2012). According to this method, the estimated hyperparameters for each of the datasets are as follows: TEC (C= 1.601470833, sigma= 0.047078172), ECO (C= 322303.3297, sigma= 0.000226155), SOC (C=1.45e9, sigma= 0.366348636) (Script E.21).

Support vector machines inherit the properties of the support vector classifier, so in predicting the response class of a new observation, only those training observations close to the decision boundary play a role (James et al. 2017).

5.5.2 Potential metrics to interpret the predictive models

The fourth objective of this study is to develop best-practice BSE-RPMs using advanced supervised machine learning techniques, which provide reliable predictions. Therefore, to compare the performance of the predictive methods explained in Section 5.5.1 and select a best-practice BSE-RPM for each dataset (performed in Chapter 6), specific metrics should be used. The following sub-sections introduce the potential metrics that could be used for this purpose. After introducing these metrics, this section justifies the selected metrics used to compare the predictive models' performances in this research. It worth noting that the selected metrics would be used to compare the performance of the models in predicting the reusability of the unseen observations (testing sets) of the TEC, ECO, and SOC datasets.

5.5.2.1 Confusion matrix

In a binary classification problem such as the ones of interest in this study, where the methods classify the test observations to one of the two classes as reusable (1) or non-reusable (0), the outcomes (predictions) fall under one of the following categories. To evaluate whether the selected classifier correctly predicts and classifies the reusable and non-reusable items into correct classes, the true negative (TN) and true positive (TP) criteria, as represented in the confusion matrix (Table 5.11) will be used. The confusion matrix provides additional information about the rates of the predicted responses that were misclassified, which is a reusable item is classified as non-reusable (false negative or FN) or vice-versa (false positive or FP) (James et al. 2017). It should be noted that the rows and columns of Table 5.11 represent the actual and predicted values of the responses, respectively.

Table 5.11 Confusion matrix

		Predicted response values	
		Non-reusable (0)	Reusable (1)
True response values	Non-reusable (0)	True negative (TN)	False positive (FP)
	Reusable (1)	False negative (FN)	True positive (TP)

5.5.2.2 False positive error (Type I error)

Based on Table 5.11, there are two types of misclassification. The first one, which is called Type I error, is when a non-reusable item is by mistake classified as reusable. As explained earlier, many classification methods (directly or indirectly) estimate the probability of a class given the features. Bayes classifier considers a threshold of 0.5 for allocating a class to an observation. Hence, considering (5.2), if $pr(reusability = 1) > 0.5$, then the observation is classified as reusable, otherwise non-reusable. However, a conservative designer might prefer a higher threshold value to decrease the Type I error and prevent the risk of using a non-reusable component that is by mistake classified as reusable. While lower values of this metric are preferred, to compare different models, the false positive error rate is used. This ratio is calculated as follows (James et al. 2017).

$$False\ positive\ error\ rate = \frac{FP}{TN + FP} \tag{5.20}$$

5.5.2.3 False negative error (Type II error)

The second type of mistake in a binary classification problem is when the classifier predicts a component as non-reusable when it is reusable. This type of error is called a false negative or type II error. As explained above, this type of error might be preferred by a conservative client or designer, so by allocating different threshold values other than 0.5, type II error is increased. Like the type I error, it is the false negative rate that is used as another metric to compare the performance of a classifier (James et al. 2017). The false negative error rate is calculated using (5.21).

$$False\ negative\ error\ rate = \frac{FN}{TP + FN} \tag{5.21}$$

5.5.2.4 Specificity

Another important metric is the rate of non-reusable components that are correctly classified. This is called specificity and is calculated as follows (James et al. 2017). Specificity (5.22) is equal to one minus false positive rate.

$$Specificity = \frac{TN}{TN + FP} \tag{5.22}$$

5.5.2.5 Sensitivity

The rate of reusable components that are correctly categorised as reusable by the classifier is another metric that is called sensitivity (5.23) (James et al. 2017). It is equal to one minus false negative rate.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.23}$$

5.5.2.6 Overall accuracy

To calculate the overall accuracy of a classifier, the total number of correct classifications are divided by the total number of observations in the test dataset (5.24) (James et al. 2017).

$$Overall\ accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{5.24}$$

5.5.2.7 Overall error rate

To calculate the overall error rate of a classifier, the total number of false classifications are divided by the total number of observations in the test dataset (5.25) (James et al. 2017).

$$Overall\ error\ rate = \frac{FP + FN}{TN + FP + FN + TP} \quad (5.25)$$

5.5.2.8 The receiver operating characteristics (ROC) curve

The above metrics are very helpful in comparing the performance of different classifiers. However, they are restricted to a pre-determined threshold value. To be able to observe the performance of a classifier with different threshold values and to decide which threshold value works the best for a classifier, a graph, known as the receiver operating characteristics (ROC) curve is used. The Y-axis of this graph is the sensitivity or true positive rate and the X-axis is the false positive rate or one minus specificity. Then, for different values of threshold, these two metrics are calculated, and a graph is drawn by connecting the identified points on the X-Y plane (James et al. 2017).

As discussed in Section 5.1, this research considers the threshold value of 0.5 for the probability of an element to be reusable or not (see conditional probability 5.1). However, higher threshold values could be considered at the cost of decreased sensitivity to reduce Type-I error (see Figure 5.14).

5.5.2.9 The area under the ROC curve (AUC)

The area under the ROC curve (also known as the AUC), is a very important and useful metric because it shows the overall performance of a classifier considering all possible threshold values (James et al. 2017). Ideally speaking, if an AUC value is close to 1, it is preferred. The baseline value for the AUC is 0.5 and a classifier should always perform higher than this minimum value. In this study, the AUC values of the classifiers are used as one of the most important metrics to compare the performance of the predictive models on the test datasets.

5.5.2.10 Selected metrics to compare the predictive models' performances

This chapter uses the Type-I error rate, overall accuracy, and AUC to compare the predictive models' performances.

While it is desired that a model makes the least number of misclassifications, in the case of this research, Type-I error is more significant than Type-II error because of considerable economic and logistic implications of the former. If the model misclassifies a reusable element as non-reusable, while environmentally significant, it would have minor economic implications compared to a false positive error because the designer would have enough time to source other suitable recovered or new elements. However, if a model misclassifies a non-reusable component as reusable during the design phase, it would have considerable economic and logistic consequences because sourcing alternative suitable new or recovered elements would be challenging.

Regarding using other metrics, by using the Type-I error rate, it would be unnecessary to use specificity (identical to one minus false-positive error rate). Moreover, since this research focuses on the Type-I error rate, the use of sensitivity (equal to one minus false-negative error rate) as a model performance metric will be pointless.

Moreover, since it is preferred that a model makes the highest number of correct classifications, the overall accuracy provides a reliable basis to understand the overall performance of a predictive model. It also provides a basis to evaluate the overall error rate (equal to one minus overall accuracy), which includes both types of errors explained earlier.

Regarding using the AUC, it is a significant and useful metric because it shows the overall performance of a classifier considering all possible threshold values (James et al. 2017). Ideally speaking, if an AUC value is close to 1, it is preferred. The baseline value for the AUC is 0.5, and a classifier should always perform higher than this minimum value.

5.5.3 Summary of the results

The summary of the metrics used to compare the models' performances is provided in Tables 5.12 to 5.14 for the TEC, ECO, and SOC datasets, respectively. Likewise, these tables show the best performing models based on the following threshold values. In this research, following (Holdnack et al. 2013), a maximum threshold of 10% is considered acceptable for the Type-I error rate. Moreover, the minimum threshold values of 85% and 90% are considered acceptable for the models' overall accuracy and AUC, respectively.

The complete set of outputs of the models used in this study (Table 5.10) are available in Appendix F. Moreover, the scripts used to develop each of the models are available in Appendix E (Script E.8 to E.21).

According to Table 5.12, for the TEC dataset, KNN, QDA, RF, and SVM have the highest performance among all other TEC BSE-RPMs because they satisfy the considered threshold values in this research. Among these models, the RF model has the best performance because it makes no Type-I error and has the highest values for its overall accuracy and AUC.

Table 5.12 Summary of the results of the TEC BSE-RPMs developed (the validation set approach method)

Predictive model	Type-I error	Overall accuracy	AUC	High-performance models
KNN[1]	0.03	0.85	0.95	Yes
LR[*]	0.28	0.78	0.81	
LDA	0.14	0.83	0.86	
QDA	0.07	0.88	0.96	Yes
NB	0.24	0.71	0.82	
DT	0.10	0.74	0.76	
RF[2]	0.00	0.91	0.98	Yes
AB	0.07	0.81	0.93	
BM	0.07	0.78	0.91	
ANN[3]	0.14	0.86	0.90	
GP	0.14	0.78	0.91	
PRL	0.21	0.81	0.84	
SVM[4]	0.07	0.90	0.97	Yes

[*] The TEC-LR BSE-RPM did not converge. Hence, this model is excluded from further analysis.
Hyperparameters (calculated using 70% of the dataset that was selected randomly):
[1] k = 6
[2] ntree = 500, mtry = 5, nodesize = 1
[3] Size = 9, Decay = 0.09
[4] Cost = 1.601470833, Sigma = 0.047078172

For the ECO dataset, KNN, RF, ANN, and SVM have the highest performance among all other ECO BSE-RPMs (Table 5.13). Among these models, the KNN, RF, and ANN models make no false-positive errors. Likewise, the RF model has the highest AUC. However, it is the ANN and SVM models that have the highest accuracy. According to Table 5.13, none of the high-performance models could be ranked the highest based on the considered metrics.

Table 5.13 Summary of the results of the ECO BSE-RPMs developed (the validation set approach method)

Predictive model	Type-I error	Overall accuracy	AUC	High-performance models
KNN[1]	0.00	0.86	0.96	Yes
LR	0.21	0.75	0.81	
LDA	0.25	0.69	0.79	
QDA	0.21	0.76	0.83	
NB	0.32	0.69	0.77	
DT	0.25	0.78	0.80	
RF[2]	0.00	0.86	0.98	Yes
AB	0.04	0.82	0.94	
BM	0.00	0.84	0.90	
ANN[3]	0.00	0.89	0.96	Yes

Predictive model	Type-I error	Overall accuracy	AUC	High-performance models
GP	0.07	0.78	0.86	
PRL	0.25	0.71	0.72	
SVM⁴	0.07	0.89	0.95	Yes
Hyperparameters (calculated using 70% of the dataset that was selected randomly): ¹ k = 5 ² ntree = 500, mtry = 3, nodesize = 1 ³ Size = 9, Decay = 0.08 ⁴ Cost = 322303.3297, Sigma = 0.000226155				

In the case of the SOC dataset, the models with the highest performance are RF, BM, and GP (Table 5.14). According to Table 5.14, among the best performing models, it is the RF model that has the highest performance considering all three metrics.

Table 5.14 Summary of the results of the SOC BSE-RPMs developed (the validation set approach method)

Predictive model	Type-I error	Overall accuracy	AUC	High-performance models
KNN[1]	0.06	0.79	0.95	
LR	0.11	0.77	0.76	
LDA	0.11	0.74	0.77	
QDA	0.11	0.91	0.97	
NB	0.22	0.85	0.97	
DT	0.33	0.77	0.88	
RF[2]	0.00	0.91	0.99	Yes
AB	0.11	0.91	0.94	
BM	0.06	0.88	0.98	Yes
ANN[3]	0.11	0.88	0.92	
GP	0.06	0.85	0.96	Yes
PRL	0.17	0.85	0.85	
SVM[4]	0.11	0.94	0.97	
Hyperparameters (calculated using 70% of the dataset that was selected randomly): [1] k = 8 [2] ntree = 500, mtry = 3, nodesize = 1 [3] Size = 8, Decay = 0.04 [4] Cost = 1.45e9, Sigma = 0.366348636				

In this section, and to elaborate on the results presented in Tables 5.12 to 5.14, the best performing model in the TEC dataset (the RF model) is further discussed. Other results could be found in Appendix F.

Table 5.15 shows the results of the classification of the unseen observations (testing set) made by the RF model developed using the training set of the TEC dataset (also known as the TEC-RF BSE-RPM). According to Table 5.15, the TEC-RF BSE-RPM makes zero false positive errors (Section 5.5.2.2) and five false negative errors (Section 5.5.2.3). It means that the Type-I error rate is equal to zero, and the overall accuracy equal to 91% (Table 5.12) calculated using (5.24) as follows:

$$\frac{24 + 29}{29 + 0 + 5 + 24} = 0.91$$

Table 5.15 The confusion matrix of the RF BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	29	0
Actual reusable (1)	5	24

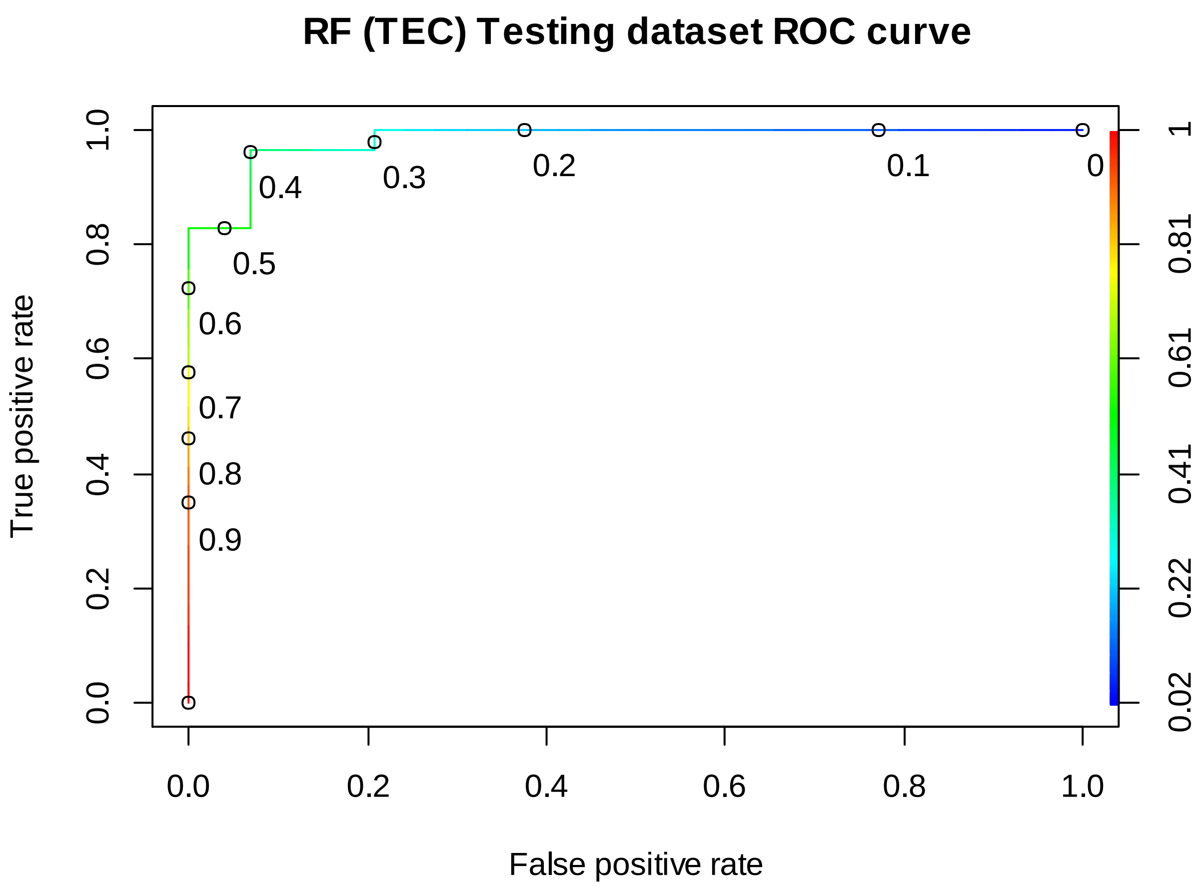

Figure 5.14 The ROC curve of the RF BSE-RPM (TEC dataset) (AUC = 0.98)

The ROC curve of the random forest model (TEC-RF BSE-RPM) for the testing set developed based on the validation set approach (Section 5.3, Table 5.2) is shown in Figure 5.14. This curve is used to observe the performance of a classifier with different threshold values and to decide which threshold value works the best for a classifier. The Y-axis of this graph shows the sensitivity or true positive rate (the number of correctly classified reusable items by a model divided by the total number of reusable components in the test dataset), and the X-axis shows the false positive or Type-I error rate. Then, for different values of threshold, these two metrics are calculated, and a graph is drawn by connecting the identified points on the X-Y plane (James et al. 2017). According to this figure, the threshold value of 0.5 (Section 5.5.1) works optimally for the classifier; hence, no need to alter it.

5.6 Chapter summary

Chapter 5 was focused on fulfilling the third and fourth objectives of this study. Following the results of the previous chapter, the final list of the reused structural elements, including their independent and dependent variables, were used to develop the BSE-RPMs. Initially, and to avoid biased predictions, the class imbalances in all datasets were addressed using the SMOTE. This measure yielded synthetically increasing the sample size in all three datasets without duplicating the observations. In the next stage, and to achieve the third objective of this study, advanced machine learning methods were used to select the applicable list of variables for developing the predictive models. This feature selection resulted in the omission of some of the independent variables. It is noteworthy that even after restricting to the listed variables in Section 5.4.4, the observations remained unique, and this practice did not result in any duplications in the TEC, ECO, and SOC datasets.

In this study, thirteen different models were used to predict the technical, economic, and social reusability of building structural elements in the TEC, ECO, and SOC datasets, respectively. These models include KNN, LR, LDA, QDA, NB, DT, RF, BM, AB, ANN, GP, PRL, and SVM. One of the reasons for using various parametric and non-parametric models is because there is no single machine learning method suitable for all types of datasets. Moreover, constraints such as the limited number of observations in each dataset, and unawareness of the nature of the relationship between the predictors and the responses brought new dimensions to the challenge of selecting a proper machine learning method. Therefore, it was decided to study a wide range of machine learning methods to develop an optimum predictive model that fulfils the fourth objective of this research.

In this chapter, to develop the predictive models, the validation set approach was used. Therefore, each of the newly developed datasets was split into a training and testing set with a 70/30 split ratio. Next, the training datasets were used to develop the predictive models, whereas the testing datasets were used to evaluate the performance of the fitted models in handling unseen data. In this research, the Type-I error rate, overall accuracy, and AUC with acceptable threshold values of 10% (maximum), 85% (minimum), and 90% (minimum) are considered to compare the models' performances and identify the best-performing ones. According to Table 5.12 and 5.14, the TEC-RF BSE-RPM and the SOC-RF BSE-RPM are the best models for the TEC and SOC dataset, respectively. For the ECO dataset, both ECO-RF BSE-RPM and ECO-ANN BSE-RPM perform optimally.

During the process of model development, the Logistic Regression model (LR) did not converge in the TEC dataset, which will be excluded from the model selection process in Chapter 6. The entire process of model development using the validation set approach is presented in Figure 5.15.

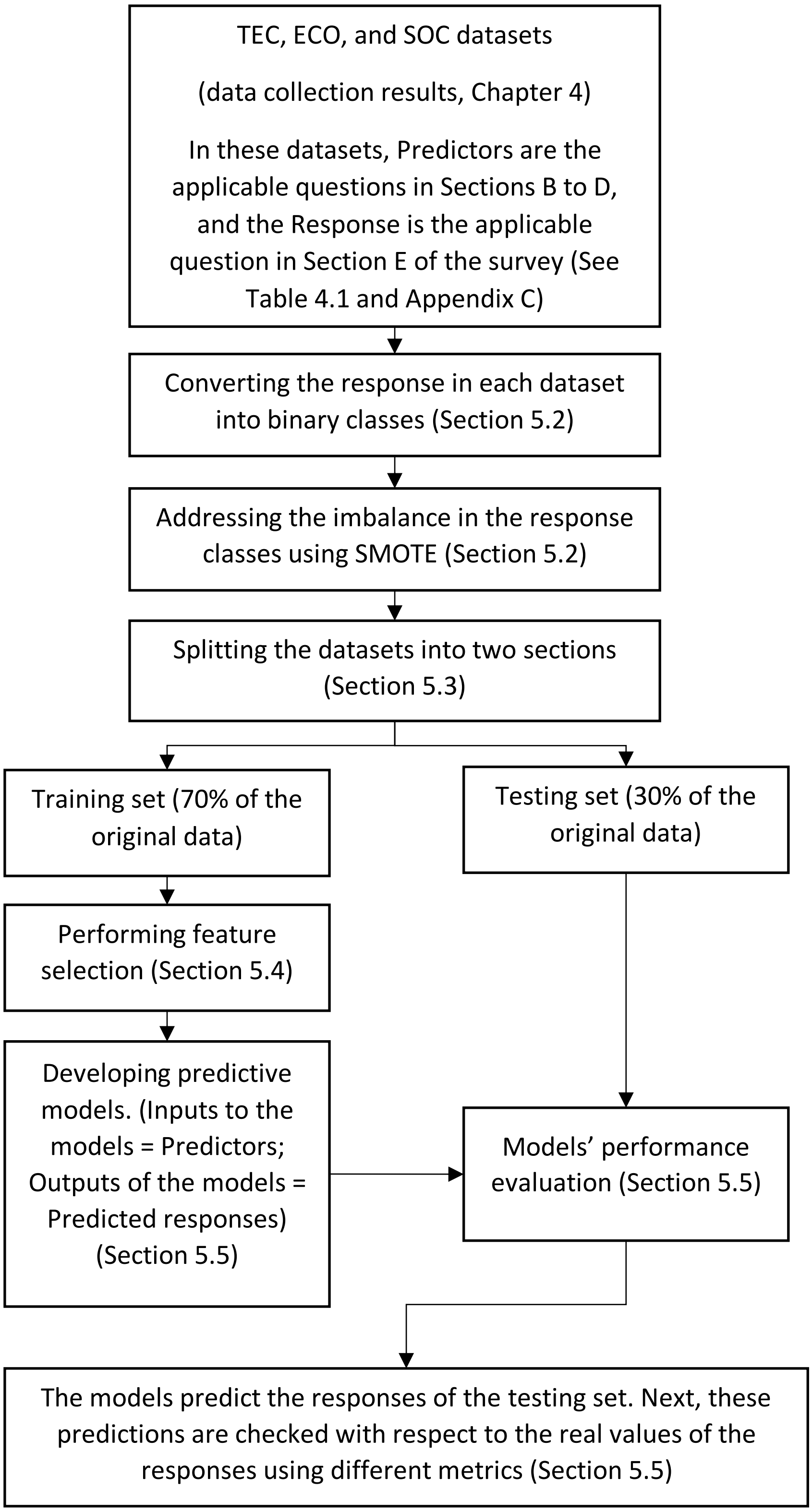

Figure 5.15 Summary of the process of developing the predictive models in Chapter 5

Chapter 6 – Model selection: Results and discussion

6.1 Chapter introduction

Chapter 6 focuses on the fourth objective of this study and discusses the process of selecting the best-practice BSE-RPMs based on the results of the developed models in Section 5.5. Therefore, initially, Section 6.2 assesses the performance of the developed models using the k-fold Cross-Validation method. Next, the outcome of this section is used to select the best-practice BSE-RPMs based on the Type-I error rate, model accuracy, and the AUC of the developed models (Section 6.3). In Section 6.4, attempts are made to clarify the selected models and develop a set of easy-to-understand rules, so the practitioners in the building sector would be able to use the results of this research effectively. Section 6.5 provides instructions for using the developed learners in Section 6.4. And eventually, Sections 6.6, 6.7, and 6.8 discuss the technical, economic, and social reusability factors based on the outcome of the methods used to clarify the selected models. This chapter concludes by summarising the results in Section 6.9.

6.2 Performance assessment for the developed BSE-RPMs

The results of the fitted models reported in Tables 5.12, 5.13, and 5.14 for the TEC, ECO, and SOC datasets are based on the validation set approach method elaborated in Section 5.3. In Chapter 6, the performances of the developed models in Section 5.5 are evaluated through the k-fold Cross-Validation (kfCV) method with $k = 10$.

The validation set approach method used to estimate the test error rates in Chapter 5 (and other metrics explained in Section 5.5.2) is an acknowledged method to assess the performance of a given machine learning technique (James et al. 2017). However, because it is based on randomly splitting the dataset into a training set (70%), and a testing set (30%) (See Section 5.3), there is a high chance of getting different performance measures if the process is repeated (James et al. 2017). This variability in the performance metrics is because the results highly depend on which observations are randomly held out for testing the fit and which are used for training the model (James et al. 2017). Moreover, because only 70% of the observations are used to fit a model, and since the performance of the predictive models improves by increasing the number of data points used to train them, the validation set approach tends to underestimate the performance of the fit by producing worse test error rates (James et al. 2017).

In this chapter, and to overcome the drawbacks of the validation set approach mentioned earlier, a kfCV method with $k = 10$ is performed to assess the performance of the developed BSE-RPMs in Section 5.5. In the kfCV method, the original dataset is randomly divided into k folds (k groups of observations) with approximately equal size (James et al. 2017). Then, the first fold is used as the testing set, and the $k - 1$ remaining folds are used to train a predictive model. Next, the performance of the fit is determined using the held-out set (James et al. 2017). The process repeats k times with all folds, and each time a different group of observations is considered as the validation set (James et al. 2017). Simultaneously, the performance results are recorded for all k folds, and eventually, the performance of the predictive model is determined using the mean performances of the k folds (James et al. 2017). According to James et al. (2017), while k can take any number less than n (n is the number of observations in a dataset), values of k equal to 5 or 10 have empirically shown resistance against high bias or variance. The choice of $k = 10$ in this study enables a higher number of training observations at each fold, which improves the performance of the classifiers (James et al. 2017).

Another method that could be used to address the drawbacks of the validation set approach in determining the test error rate is the leave-one-out cross-validation method (LOOCV) (James et al. 2017). This approach repeats n times, and at each stage, the LOOCV method performs by considering one of the observations as the testing set and the remaining as the training set (James et al. 2017). The resulting performance of the predictive model is the average performance of the n models (James et al. 2017). This method has the advantage of considering a higher number of observations for training a model and can potentially improve the performance of the modeller by eliminating the bias of the test error estimates (James et al. 2017). However, the test error estimates using the LOOCV method tend to have a higher variance than the kfCV method (James et al. 2017). Therefore, in this research, it was decided to use the kfCV method for estimating the performance of the developed models, and eventually selecting the best practice BSE-RPMs. For further details about the LOOCV and kfCV methods, please refer to Section 5.1 of (James et al. 2017).

The results of the ten-fold CV used to assess the performance of the BSE-RPMs of the TEC, ECO, and SOC datasets are presented in Sections 6.3.2, 6.3.3, and 6.3.4, respectively. The assessment results are then used to select the best-practice model in each dataset. Script E.22 (Appendix E) is used to assess the performance of the BSE-RPMs using the kfCV method.

6.3 Selection of the best-practice BSE-RPMs

This study aims to develop models that can predict the technical, economic, and social reusability of structural elements at the end-of-life of a building. Therefore, one of the key features in selecting such models is how accurately they can classify a recovered structural element in one of the two classes as reusable or not. An accurate model that makes minimum classification errors is then desirable and helps the stakeholders to decide whether to integrate recovered building structural elements in their new developments or not. On the other hand, the transparency of the results plays a vital role in encouraging the stakeholders to employ the outcomes of such machine learning techniques in their day-to-day activities (Scutari and Denis 2015).

6.3.1 Metrics used to select the best practice models

In this research, Type I error rate (Sections 5.5.2.2), overall accuracy (Section 5.5.2.6), and AUC (Section 5.5.2.9) are used to compare the performance of different models. The values of these metrics are reported in Tables 6.1, 6.3, and 6.5 for the TEC, ECO, and SOC datasets, respectively. Moreover, model transparency is considered as another metric to choose between the potential best-practice BSE-RPMs.

6.3.1.1 Model Type I error rate

Model classification error rates or Type I (5.20) and Type II (5.21) errors are significant indicators of the performance of a predictive model. According to James et al. (2017), low error rates on a given dataset guarantees the safe use of a particular supervised learning model. While both error rates should be minimum, Type I error has a pronounced impact on the success of a project with recovered building structural elements. As discussed in Section 5.5.2.2, a Type I error happens when a BSE-RPM classifies a non-reusable component as reusable. This mistake causes several logistic, financial, and technical costs by providing a false indication about the reusability of an element, which could risk the entire project. However, the consequences of a Type II error are manageable. While reuse aims to improve the circularity of materials in the building sector, a Type II error only troubles the design team to focus on other available recovered structural components. It is because by making a Type II error, a reusable section is discarded, and a designer needs to look for other recovered elements or purchase a new component. While this is not favourable in terms of the circularity of materials, unlike a Type I error, it doesn't jeopardise the entire project. Either way, by integrating proper waste management plans considering sustainability at their core, elements misclassified as non-reusable will still go through recycling or down-cycling processes, which are still far better solutions than landfilling.

In this study, following Holdnack et al. (2013), a maximum threshold of 10% is considered acceptable for the Type I error rate. Accordingly, for the TEC dataset, KNN, QDA, RF, AB, and SVM BSE-RPMs are eligible candidates on this metric (Table 6.1). Regarding the economic reusability, only KNN and RF BSE-RPMs are within the acceptable range (Table 6.3). And regarding the social reusability, KNN, RF, AB, BM, GP, PRL, and SVM fulfil the maximum allowable Type I error rate (Table 6.5).

6.3.1.2 Model accuracy

The accuracy of a predictive model to correctly identifying the reusability of the recovered building structural elements is of pronounced importance for the designers. In this study, the available datasets are approximately having an equal number of reusable and non-reusable observations. According to Table 5.1, a baseline model based on the portion of reusable and non-reusable building component elements can be developed for each of the datasets. A baseline model assigns the most frequent response (either reusable or non-reusable) for all observations. The baseline model for the TEC dataset has a 50% accuracy. It is because, if it is used to predict the technical reusability of the elements, only half of its predictions would be correct (based on Table 5.1, in the TEC dataset, the number of reusable and non-reusable components are equal). In the case of the ECO dataset, the baseline model always predicts non-reusable because 51% of the elements are non-reusable, yielding an accuracy of 51%. And for the SOC dataset, the baseline model has 53% accuracy because it always predicts non-reusable for every observation (53% of the elements are non-reusable). Therefore, the accuracy of the predictive models should be far better than the baseline models for making the best practice BSE-RPMs reliable.

In this research, a minimum threshold of 85% is considered acceptable for the predictive models' overall accuracy (see Tables 6.1, 6.3, and 6.5). Therefore, KNN (92%), QDA (91%), RF (96%), AB (87%), ANN (88%), and SVM (93%) BSE-RPMs for the TEC dataset fulfil the minimum threshold requirements on model accuracy (Table 6.1). Regarding the BSE-RPMs developed based on the ECO dataset, KNN (86%), RF (89%), AB (86%), BM (86%), ANN (86%), PRL (86%) and SVM (87%) are the acceptable models (Table 6.3). Moreover, RF (94%) and SVM (87%) are the only acceptable models for the SOC dataset (Table 6.5).

6.3.1.3 AUC

While the overall accuracy of a model is an essential metric to choose a classifier, it is limited to a fixed threshold value (in this study equal to 0.5, see Section 5.5.1), hence not comprehensive. To overcome this barrier, the area under the ROC curve (AUC) (Section 5.5.2.9), which portrays the overall performance of a classifier based on all possible threshold values (James et al. 2017), is considered as another metric for model selection.

In this study, the minimum acceptable value for the AUC is set to 90% (Section 5.5.2.9). As the result, KNN (98%), QDA (96%), RF (100%), BM (94%), AB (95%), ANN (93%), GP (92%), and SVM (98%) have high performance among all BSE-RPMs developed for the TEC dataset (Table 6.1). Moreover, KNN (93%), RF (98%), AB (92%), BM (90%), ANN (93%), GP (91%) and SVM (91%) fulfil the minimum performance requirement for the ECO dataset (Table 6.3). And finally, on the SOC dataset, KNN (92%), QDA (91%), RF (96%), AB (91%), BM (91%), GP (92%), SVM (94%) fulfil the threshold requirement on the AUC (Table 6.5).

6.3.1.4 Model transparency

The developed models in this study cover both parametric and non-parametric methods (Table 5.10). These models have different levels of transparency, ranging from transparent models (LR, LDA, DT, and PRL) to hard-to-interpret (QDA) and black-box models (KNN, NB, RF, BM, AB, ANN, GP, and SVM). While it is preferable to choose a transparent model, in some cases, such models do not yield acceptable levels of accuracy in correctly classifying reusable and non-reusable elements, and the selection of a black-box model becomes inevitable. In the case of the latter, other tools, such as the sensitivity analysis and visualisation techniques introduced by (Cortez and Embrechts 2013), can be used to open a black-box model and make the results transparent.

6.3.2 Best practice BSE-RPM for the TEC dataset (TEC BSE-RPM)

Table 6.1 reports the summary of the results of a ten-fold CV used to assess the performance of the BSE-RPMs of the TEC dataset. As mentioned in Section 5.5.3, the TEC Logistic Regression (LR) BSE-RPM did not converge. Hence, this model is not considered.

Table 6.1 Mean values of the metrics used to assess the performance of TEC BSE-RPMs (10-fold CV method)

Predictive model	Type-I error	Overall accuracy	AUC
KNN	0.03	0.92	0.98
LDA	0.18	0.81	0.90
QDA	0.09	0.91	0.96
NB	0.28	0.72	0.82
DT	0.29	0.71	0.73

Predictive model	Type-I error	Overall accuracy	AUC
RF	0.01	0.96	1.00
AB	0.08	0.87	0.95
BM	0.11	0.85	0.94
ANN	0.13	0.88	0.93
GP	0.12	0.84	0.92
PRL	0.19	0.80	0.83
SVM	0.07	0.93	0.98

For the TEC dataset, none of the developed models could satisfy all the four metrics for choosing the best-practice TEC BSE-RPM (Table 6.2). Hence, in the process of selecting the best model for the TEC dataset, only the Type-I error rate, overall accuracy, and AUC are considered. Based on Table 6.1, the random forests model (TEC-RF BSE-RPM) has the lowest Type I error rate (0.01), the highest overall accuracy (0.96), and the highest AUC (1.00) among all other models. So, the random forest model (TEC-RF BSE-RPM) is selected as the best-practice model to predict the technical reusability of the building structural elements (Table 6.2).

Table 6.2 Selecting the best-practice TEC BSE-RPM

Predictive model	Type-I error (≤10%)	Overall accuracy (≥ 85%)	AUC (≥90%)	Transparency	Selected model
KNN	Yes	Yes	Yes		
LDA				Yes	
QDA	Yes	Yes	Yes		
NB					
DT				Yes	
RF	Yes	Yes	Yes		Yes
AB	Yes	Yes	Yes		
BM			Yes		
ANN		Yes	Yes		
GP			Yes		
PRL				Yes	
SVM	Yes	Yes	Yes		

6.3.3 Best practice BSE-RPM for the ECO dataset (ECO BSE-RPM)

Table 6.3 reports the summary of the results of a ten-fold CV used to assess the performance of the BSE-RPMs of the ECO dataset.

Table 6.3 Mean values of the metrics used to assess the performance of ECO BSE-RPMs (10-fold CV method)

Predictive model	Type-I error	Overall accuracy	AUC
KNN	0.08	0.86	0.93

Predictive model	Type-I error	Overall accuracy	AUC
LR	0.26	0.73	0.82
LDA	0.26	0.74	0.83
QDA	0.22	0.77	0.88
NB	0.32	0.73	0.84
DT	0.16	0.82	0.84
RF	0.06	0.89	0.98
AB	0.13	0.86	0.92
BM	0.11	0.86	0.90
ANN	0.11	0.86	0.93
GP	0.10	0.83	0.91
PRL	0.12	0.86	0.86
SVM	0.10	0.87	0.91

For the ECO dataset, none of the developed models could satisfy all the four metrics for choosing the best-practice ECO BSE-RPM (Table 6.4). Hence, in the process of selecting the best model for the ECO dataset, only the Type-I error rate, overall accuracy, and AUC are considered. Based on Table 6.3, the random forests model (ECO-RF BSE-RPM) has the lowest Type I error rate (0.06), the highest overall accuracy (0.89), and the highest AUC (0.98) among all other models. So, the random forest model (ECO-RF BSE-RPM) is selected as the best-practice model to predict the economic reusability of the building structural elements (Table 6.4).

Table 6.4 Selecting the best-practice ECO BSE-RPM

Predictive model	Type-I error (≤10%)	Overall accuracy (≥ 85%)	AUC (≥90%)	Transparency	Selected model
KNN	Yes	Yes	Yes		
LR				Yes	
LDA				Yes	
QDA					
NB					
DT				Yes	
RF	Yes	Yes	Yes		Yes
AB		Yes	Yes		
BM		Yes	Yes		
ANN		Yes	Yes		
GP			Yes		
PRL		Yes		Yes	
SVM		Yes	Yes		

6.3.4 Best practice BSE-RPM for the SOC dataset (SOC BSE-RPM)

Table 6.5 reports the summary of the results of a ten-fold CV used to assess the performance of the BSE-RPMs of the SOC dataset.

Table 6.5 Mean values of the metrics used to assess the performance of SOC BSE-RPMs (10-fold CV method)

Predictive model	Type-I error	Overall accuracy	AUC
KNN	0.03	0.81	0.92
LR	0.26	0.70	0.85
LDA	0.25	0.72	0.85
QDA	0.28	0.76	0.91
NB	0.14	0.80	0.86
DT	0.16	0.77	0.87
RF	0.00	0.94	0.96
AB	0.07	0.81	0.91
BM	0.04	0.82	0.91
ANN	0.10	0.80	0.84
GP	0.03	0.81	0.92
PRL	0.04	0.85	0.88
SVM	0.10	0.87	0.94

For the SOC dataset, none of the developed models could satisfy all the four metrics for choosing the best-practice SOC BSE-RPM (Table 6.6). Hence, in the process of selecting the best model for the SOC dataset, only the Type-I error rate, overall accuracy, and AUC are considered. Based on Table 6.5, the random forests model (SOC-RF BSE-RPM) has the lowest Type I error rate (0.00), the highest overall accuracy (0.94), and the highest AUC (0.96) among all other models. So, the random forest model (SOC-RF BSE-RPM) is selected as the best-practice model to predict the social reusability of a building's structural elements (Table 6.6).

Table 6.6 Selecting the best-practice SOC BSE-RPM

Predictive model	Type-I error (≤10%)	Overall accuracy (≥ 85%)	AUC (≥90%)	Transparency	Selected model
KNN	Yes		Yes		
LR				Yes	
LDA				Yes	
QDA			Yes		
NB					
DT				Yes	
RF	Yes	Yes	Yes		Yes
AB	Yes		Yes		
BM	Yes		Yes		
ANN					
GP	Yes		Yes		
PRL	Yes			Yes	
SVM	Yes	Yes	Yes		

6.4 Improving the transparency of the selected best-practice models

While the selected TEC-RF BSE-RPM, ECO-RF BSE-RPM, and SOC-RF BSE-RPM models in Sections 6.3.2, 6.3.3, and 6.3.4 have high overall accuracy, high AUC, and low Type-I error rate, they lack transparency. It is because random forest models are categorised under black-box methods, and they cannot be interpreted easily (Breiman 2001). As discussed in Section 6.3, the transparency of the results of the selected predictive models is essential to encourage the stakeholders to employ the outcome of such models for assessing the reusability of building structural elements at the end-of-life of a building. Therefore, when such easy-to-understand models are not available, it is necessary to make the results of the selected models transparent.

In this research, two techniques are used to improve the transparency of the selected models. First, the sensitivity analysis and visualisation techniques suggested by Cortez and Embrechts (2013) are employed to identify the importance of the variables and open the black box models. Next, using the rule extraction method suggested by (Deng 2014) and based on the results of the previous technique, a set of decision rules was produced to explain the ensemble of trees developed in the selected RF model. While both techniques fulfil the aim of this study, the latter provides a simple and understandable set of rules for the stakeholders to estimate the reusability of building structural elements at the end-of-life of a building.

According to Cortez and Embrechts (2013), to perform the sensitivity analysis (SA), a sensitivity method needs to be identified first. A sensitivity method performs by varying a given reusability factor from its minimum to maximum possible values while conditioning the remaining independent variables and observations (Cortez and Embrechts 2013). For the nominal features (B3 and B5), the sensitivity method alters the values of the variables based on the variable levels (B3 has three levels, and B5 has five levels, see Section C.2, Appendix C). For the categorical features, following Cortez and Embrechts (2013), the sensitivity method varies the value of the predictors from one to five in seven intervals (see Table 4.1 for the Likert scale used). As recommended by Cortez and Embrechts (2013), in this research, data-based SA (DSA) was used as the sensitivity method. The DSA method randomly selects several samples from the dataset and alters the values of an independent variable for all data points and records the responses while not changing other features (Cortez and Embrechts 2013). This process is performed for all independent variables (reusability factors) in the TEC, ECO, and SOC datasets. The sensitivity responses identified using the DSA method can be used to determine the feature importance using a sensitivity measure (Cortez and Embrechts 2013). This research uses the Average

Absolute Deviation (AAD) from the Median as the sensitivity measure, as advised by Cortez and Embrechts (2013). According to Cortez and Embrechts (2013),

$$AAD = \frac{\sum_{j=1}^{L} \left| \hat{y}_{a_j} - \tilde{y}_a \right|}{L} \tag{6.1}$$

where $L = 7$ (seven intervals between one to five), $\hat{y}_{a_j}$ is the sensitivity response for $x_{a_j} \in \{1, 1.67, 2.33, 3, 3.67, 4.33, 5\}$ (j^{th} level of input $\boldsymbol{x}_a$: $a \in \{1, \dots, p\}$ for p features), and $\tilde{y}_a$ is the median of the responses. The higher the value of the AAD for an independent variable, the more important is the feature (Cortez and Embrechts 2013). This measure is then used to develop the relative importance of the input variables (Cortez and Embrechts 2013). It is noteworthy that following Cortez and Embrechts (2013), this research uses the complete TEC, ECO, and SOC datasets to perform the SA. For further details about the SA and visualisation methods used in this study, please refer to (Cortez and Embrechts 2013). Script E.23 (Appendix E) is used to perform the SA in this study.

While the sensitivity analysis and visualisation techniques presented above help in opening the selected models (TEC-RF BSE-RPM, ECO-RF BSE-RPM, and SOC-RF BSE-RPM), it still lacks the clarity level required by the stakeholders to make sound judgments about the reusability of the structural elements of a building at its end-of-life phase. Hence, as mentioned earlier, the results of the SA are used to develop a set of easy-to-understand rules that can be effectively used by the practitioners.

Figure 6.1 displays different stages of the method used to extract rules from the selected BSE-RPMs (Deng 2014).

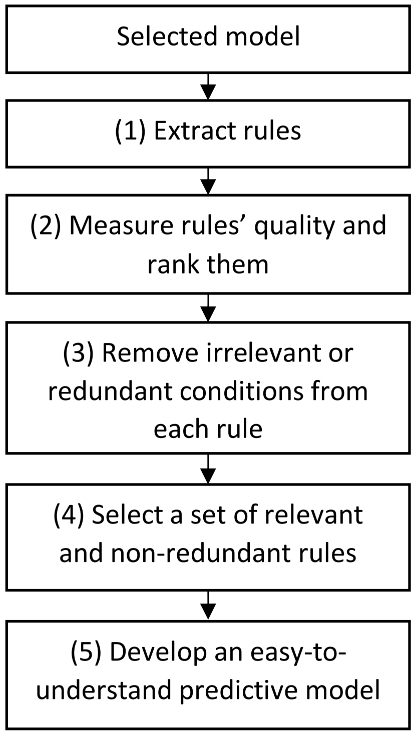

Figure 6.1 The process of developing the rules set from the selected BSE-RPMs (Deng 2014)

The results of the sensitivity analysis and rule extraction methods that were performed for improving the transparency of the selected models are provided in Sections 6.4.1, 6.4.2, and 6.4.3.

6.4.1 Improving the transparency of the TEC-RF BSE-RPM

Figure 6.2 shows the results of the feature importance for the TEC-RF BSE-RPM. In this figure, the X-axis shows the relative importance of the variables, and the Y-axis shows the features. Based on Figure 6.2, only some of the variables are relevant, and others have negligible importance. In this study, features with relative importance greater than 2% are considered for further review and development of the rules, and the remaining are ignored. It results in a total number of fourteen independent variables including, B3, B5, B7, B8, C6, C12, C15, C16, C20, C25, C27, C28, D23, and D24.

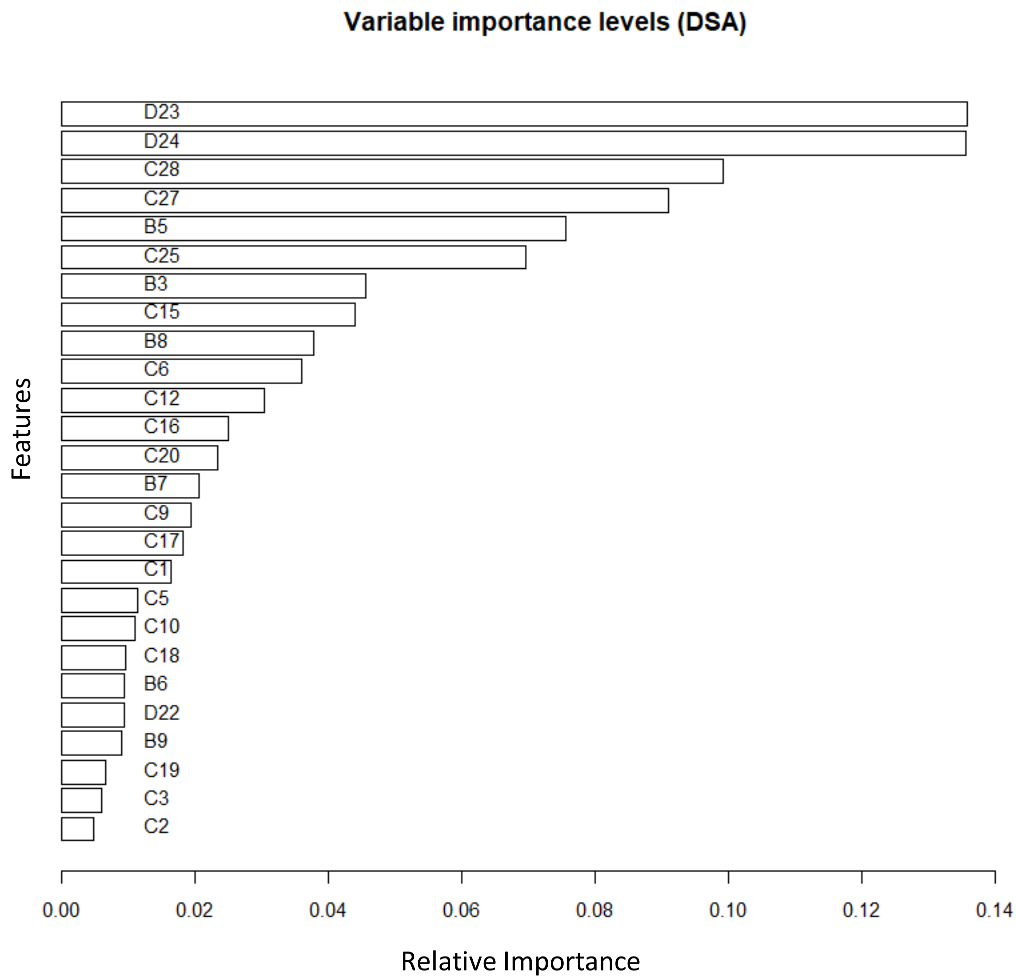

Figure 6.2 Bar plot with DSA and AAD relative feature importance for the TEC dataset based on the TEC-RF BSE-RPM

In the next stage, and to present how different values of a feature affect the technical reusability of building structural elements on average, a set of variable effect characteristic (VEC) curves are plotted for the identified fourteen variables. A VEC curve plots the average impact of different values of a reusability factor (X-axis) on the probability that a structural element is reusable (Y-axis).

Figure 6.3 shows the sensitivity analysis of the top-four factors based on Figure 6.2. According to Figure 6.3, the reusability probabilities of a building's structural elements improves when the values of these variables increase.

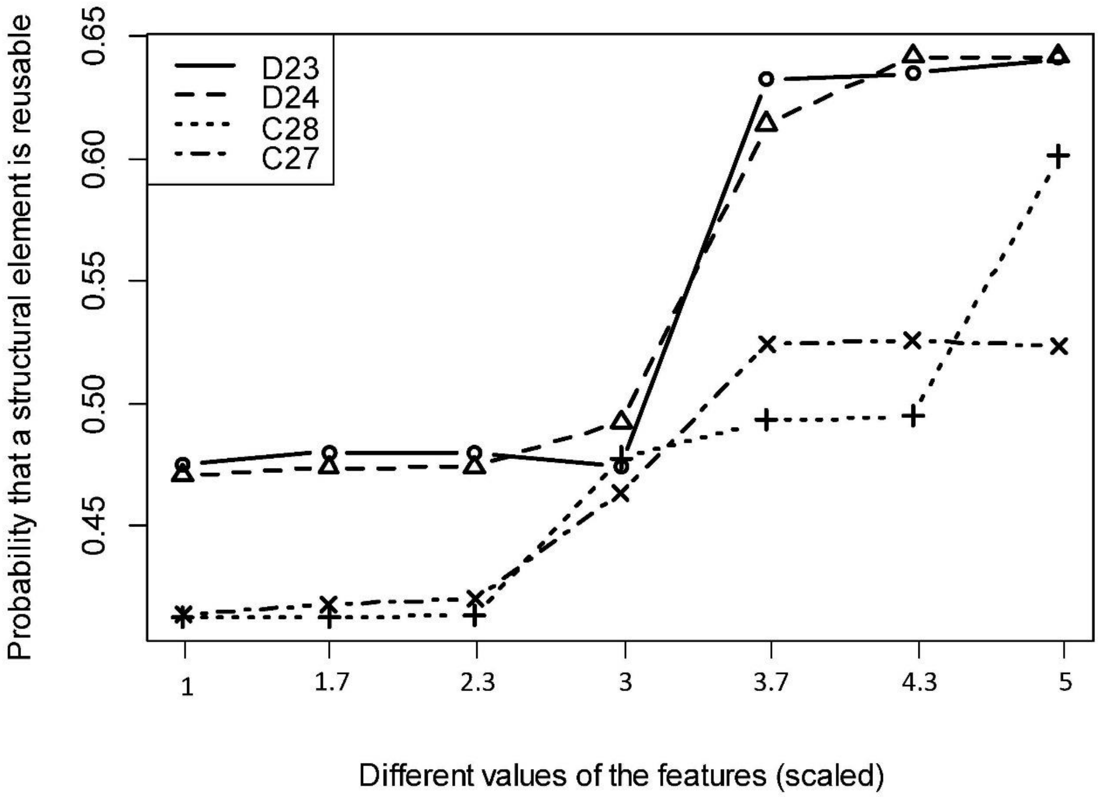

Figure 6.3 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D23, D24, C28, and C27 (the top-four variables in TEC-RF BSE-RPM)

Figure 6.4 shows the impact of different values of the listed features on the technical reusability of building structural elements. Because B3 (age of the building) and B5 (number of existing connections) are nominal variables, four separate graphs are drawn for clarity. While the decrease in the reusability probability due to reduced effects of the damage due to post-production modifications (variable C15) looks counterintuitive, the observed behaviour should not be evaluated in solitude, and the impact of the interactions with other independent variables needs to be considered, as well. It is noteworthy that in performing the sensitivity analysis for a feature, the value of other variables is not altered. Whereas, in real cases, the values of other variables might change due to the interdependencies of the features. The same applies to B5 (number of existing connections) and B3 (age of the building), as well. In the case of the former, it seems that by increasing the number of existing connections, the reusability decreases. The above observation is only correct for options three and four on the questionnaire survey, where the number of existing connections increases from five to ten. However, reusability improves for a higher number of connections, which is again counterintuitive. Notwithstanding, it can be concluded that while the limited number of existing connections is favourable, this factor cannot be considered on its own, and the interaction with other variables should be considered. The above fact applies to all other variables, as well.

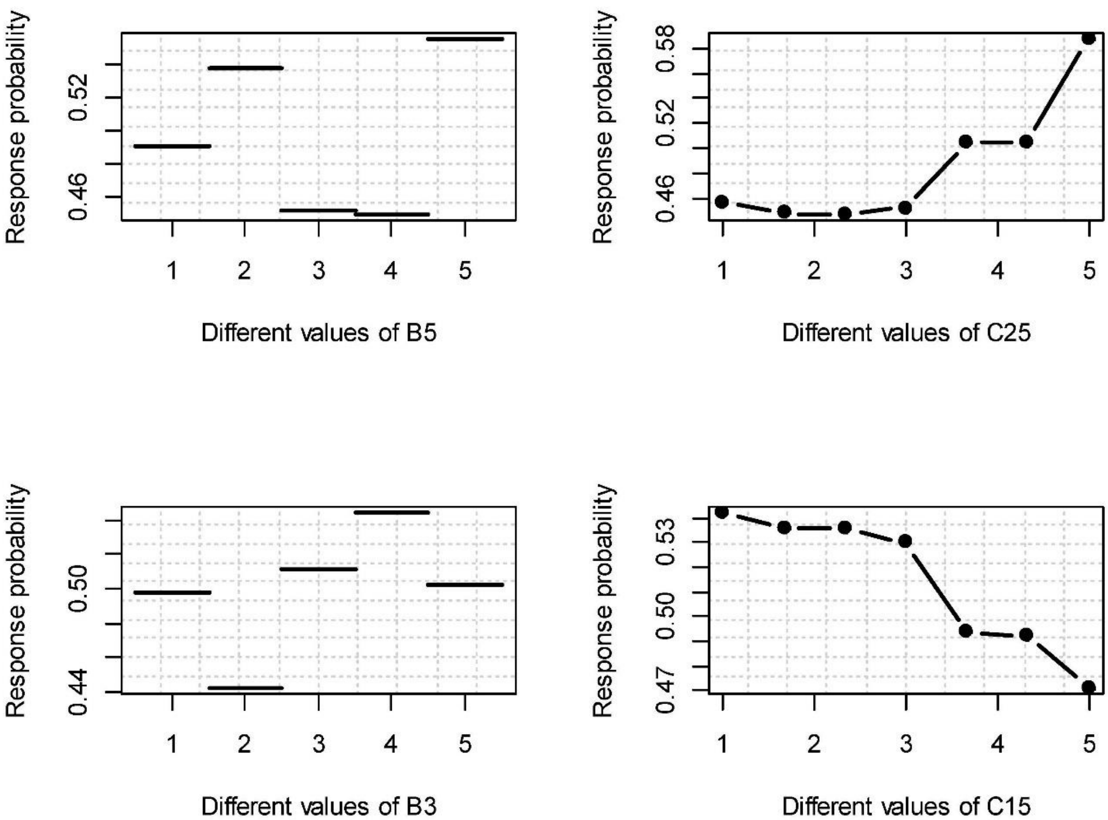

Figure 6.4 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for B5, C25, B3, and C15 (TEC-RF BSE-RPM)

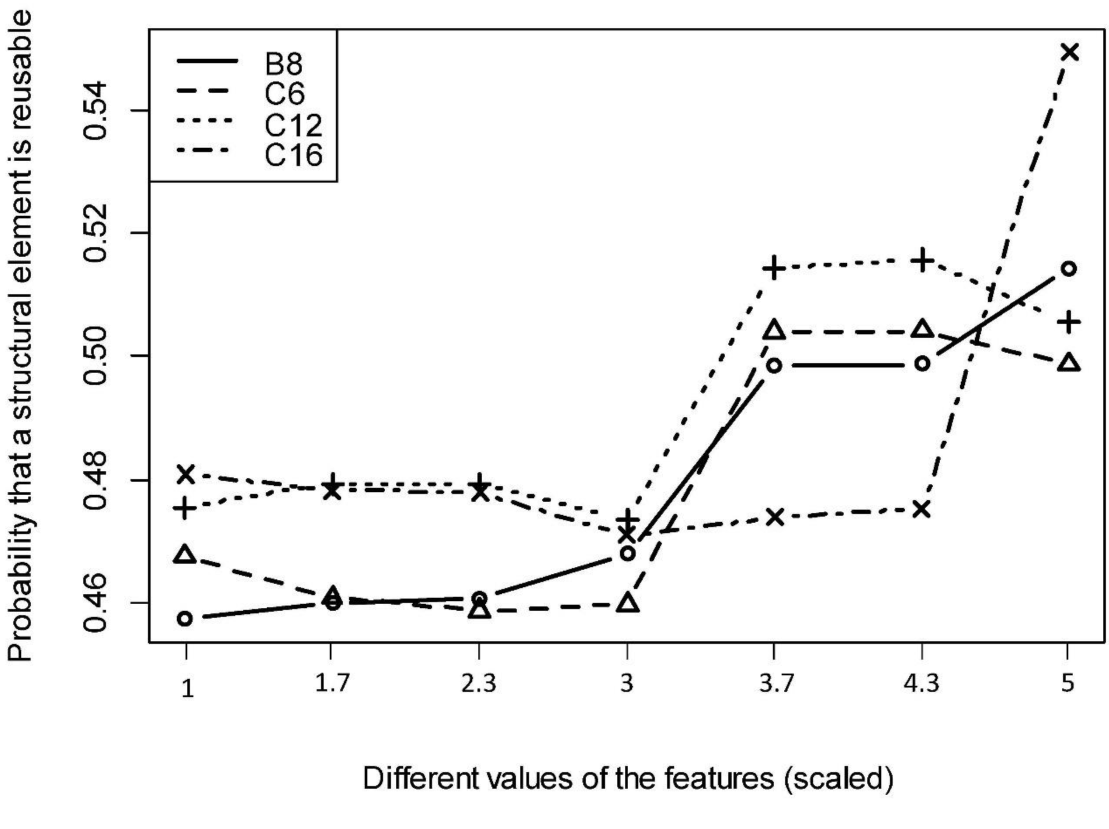

Figure 6.5 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for B8, C6, C12, and C16 (TEC-RF BSE-RPM)

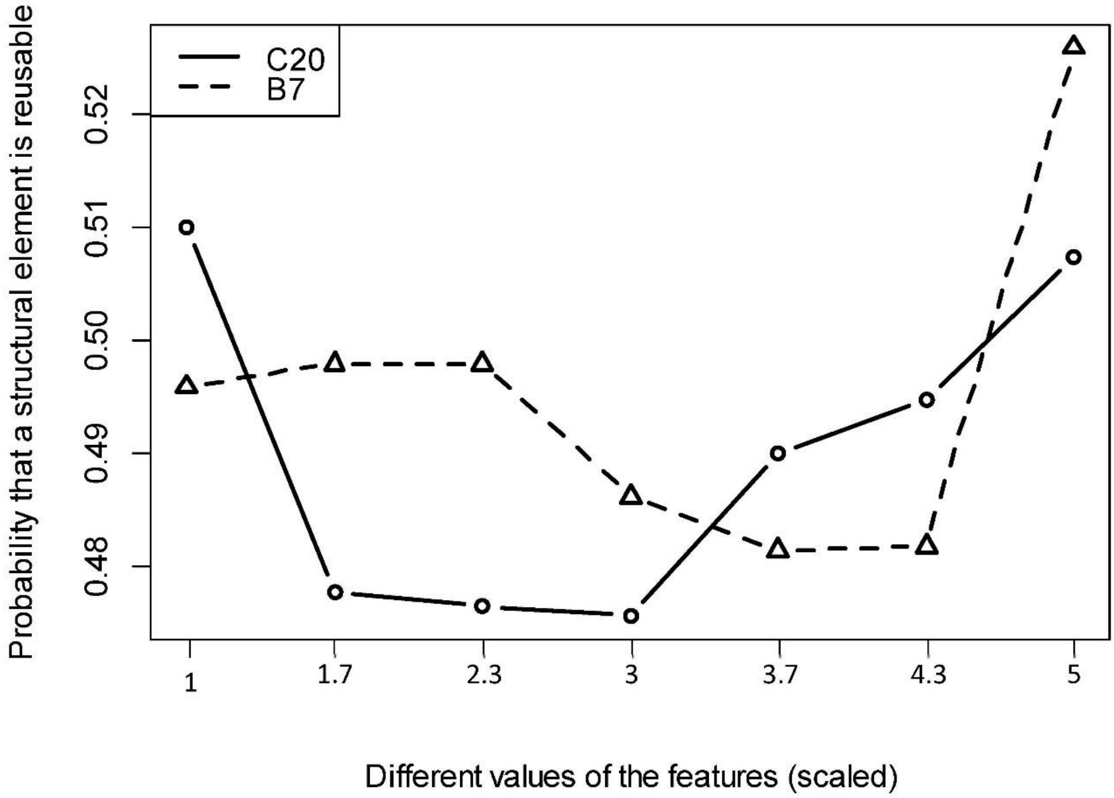

Figure 6.6 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for C20 and B7 (TEC-RF BSE-RPM)

Figures 6.5 and 6.6 follow the same trend observed in Figure 6.3. However, as discussed earlier, none of these features should be considered independently for estimating the technical reusability of building structural elements. This effect can be shown by drawing the VEC curves while showing the range of the sensitivity at each point. For this purpose, the most suitable feature (D23, the mechanical properties of the component) and the least significant variable (B7, the future deployment of the element, identified based on a minimum 2% threshold for the relative importance) are plotted in Figure 6.7. According to this figure, the average VEC curve for B7 is nearly flat (the diamonds on the curve). Moreover, while there is a leap from three to four for D23, the rest of the curve remains almost flat. However, the range of the sensitivity is high for both variables, as shown by the box plots in Figure 6.7. The above observation acknowledges that the technical reusability of the structural elements of a building depends on the interactions between the predictors, as well (Cortez and Embrechts 2013).

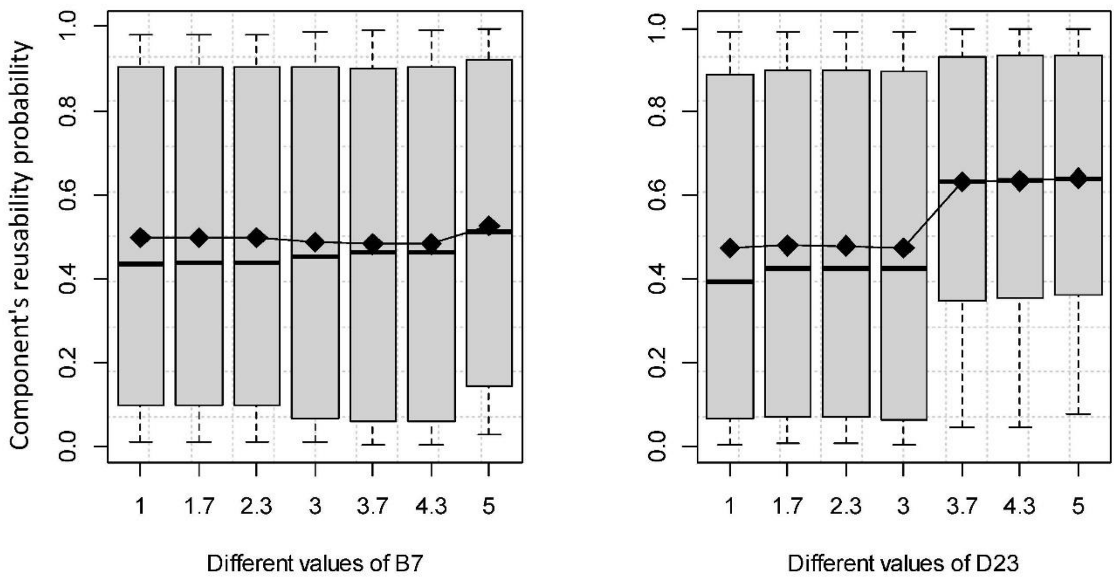

Figure 6.7 The VEC curves with box plots (to show the range of sensitivity at each point) to compare the impact of different values of B7 (left) and D23 (right) on the reusability probabilities of the elements (TEC-RF BSE-RPM)

In the next stage, and to further promote the clarity of the results of the selected TEC-RF BSE-RPM, a set of easy-to-understand rules (presented in Table 6.7) are developed based on the method suggested by Deng (2014). The steps followed for developing these rules are available in Figure 6.1 (Section 6.4).

The first column of Table 6.7 contains the sequence of the rules that need to be followed strictly. It means that checking should start with rule number one, and if its conditions are not satisfied, the next rule should be checked. This sequential process continues until a rule's conditions are satisfied. At this point, checking stops, and the rule number (the first column of Table 6.7) and prediction result (the sixth column of Table 6.7) should be recorded against the observation. It should be noted that the next rules should not be checked even if the collected data satisfy them.

The second column shows the length of a condition, which is the count of variable-value pairs (such as $C12 \leq 3$ in rule number one) in a rule (Deng 2014). For example, rule number 7 has three circumstances to be satisfied; hence, the length of its condition is equal to 3.

The third column is the frequency of a rule, which is defined as the proportion of the observations in the training dataset that satisfy the rule condition(s) (Deng 2014). For instance, the total number of observations in the training set is equal to 134, out of which twenty-one fall under the first rule. Therefore, the frequency of the first rule becomes 0.157 (the sum of frequency values is equal to one).

The fourth column is the error of a rule, which is equal to the number of misclassifications decided by the rule divided by the number of observations satisfying the rule condition(s) in the training dataset (Deng 2014). According to Table 6.7, out of 15 rules, only one (rule number 9) makes misclassifications on the training set. Rule number 9 covers 16 observations in the training set, out of which only one is wrongly classified as non-reusable, resulting in a misclassification error rate equal to 6.25%.

Column five of Table 6.7 shows the conditions of the rules. And the last column contains the predicted responses by the rules, which is equal to zero (0) for non-reusable elements and to one (1) for reusable components. As an example, rule number one states that if C12 (damage caused by living organisms), C20 (lack of earlier certificates), and D23 (the process of matching the design of the new building with the strength of the recovered element) are less than or equal to 3, then the component is not reusable.

In Table 6.7, there are both nominal variables (B3 and B5) and categorical factors (features of groups C and D). While the categorical variables are dealt with like numbers (because they are ordered) (Sauro and Lewis 2016) (see Section 4.10), the nominal factors have no correct order; hence, they appear in the learner presented in Table 6.7 as a vector. For instance, in rule number five, $B3 = c$ ('4') means only those observations where the age of a component (or a building) is between 81 to 100 years. For further details about the variables, please refer to Section 4.4.2 and Appendix C.2.

Table 6.7 The learner (rules set) developed based on the TEC-RF BSE-RPM

Rule No.	Length	Frequency	Error	Condition	Prediction
1	3	0.157	0	$\mathrm{C12} \leq 3\ \&\ \mathrm{C20} \leq 3\ \&\ \mathrm{D23} \leq 3$	0
2	2	0.134	0	$\mathrm{C16} > 4\ \&\ \mathrm{D24} > 2$	1
3	2	0.112	0	$\mathrm{B8} \leq 3\ \&\ \mathrm{C12} > 4$	0
4	3	0.075	0	$\mathrm{C20} \leq 2\ \&\ \mathrm{C28} > 3\ \&\ \mathrm{D24} > 2$	1
5	3	0.067	0	$\mathrm{B3} = \mathrm{c}\,('4')\ \&\ \mathrm{C27} > 3\ \&\ \mathrm{C28} > 2$	1
6	1	0.045	0	$\mathrm{D24} > 3$	1
7	3	0.030	0	$\mathrm{B3} = \mathrm{c}\,('1','2','3','5')\ \&\ \mathrm{B5} = \mathrm{c}\,('3','4')\ \&\ \mathrm{C12} > 4$	0
8	4	0.022	0	$\mathrm{B5} = \mathrm{c}\,('1','3','4','5')\ \&\ \mathrm{C6} > 3\ \&\ \mathrm{C15} \leq 4\ \&\ \mathrm{C28} \leq 3$	1
9	2	0.119	0.0625	$\mathrm{C28} \leq 4\ \&\ \mathrm{D23} > 2$	0
10	4	0.119	0	$\mathrm{B5} = \mathrm{c}\,('1','2','5')\ \&\ \mathrm{C6} > 3\ \&\ \mathrm{C20} > 3\ \&\ \mathrm{C28} > 2$	1
11	5	0.060	0	$\mathrm{B3} = \mathrm{c}\,('1','2')\ \&\ \mathrm{B5} = \mathrm{c}\,('1','2','3','5')\ \&\ \mathrm{C20} > 1\ \&\ \mathrm{C28} \leq 3\ \&\ \mathrm{D23} \leq 3$	0

Rule No.	Length	Frequency	Error	Condition	Prediction
12	3	0.015	0	$B7 > 4 \& C27 > 2 \& D23 > 1$	1
13	3	0.022	0	$B5 = c('1','5') \& C28 \leq 4 \& D24 \leq 3$	0
14	3	0.015	0	$B5 = c('1','2','3','5') \& B8 > 3 \& C28 > 3$	1
15	1	0.007	0	Else	0

Table 6.7 is developed based on the training dataset defined in Section 5.3. While the above set of rules provides an easy-to-understand and implement collections of conditions, it is essential to make sure that the resulting predictions on the unseen data satisfy the minimum requirements set in Section 6.3.1. Therefore, the corresponding testing dataset (unseen observations by the learner) was used to evaluate the performance of the learner presented in Table 6.7. For this purpose, the researcher followed the rules sequentially (from 1 to 15), identified the applicable set of conditions to each observation, and recorded the resulting prediction for each element. Next, the prediction results were compared with the correct responses, and the errors were recorded to evaluate the performance of the learner. Table 6.8 shows the results of the classifications made by this learner on the testing dataset (see Section 5.5.2.1).

Table 6.8 The confusion matrix of the learner presented in Table 6.7

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	2
Actual reusable (1)	8	21

As a result, the classifier misclassified two (2) non-reusable elements as reusable (Type-I errors, Section 5.5.2.2) and eight (8) reusable components as non-reusable (Type-II errors, Section 5.5.2.3). Based on Table 6.8, the Type-I error rate is equal to 6.9%, and the overall accuracy is equal to 85.3%. Therefore, this learner satisfies the minimum performance requirements defined in sections 6.3.1.1 and 6.3.1.2. Moreover, the learner in Table 6.7 is transparent and easy-to-understand and can be easily implemented in practice.

In Table 6.7, the rules are ordered, and the rules should be checked sequentially to find a condition that satisfies the predictor values of the component to determine the technical reusability of a structural element. According to Table 6.7, C25 is not available in any of the rules. Hence, a practitioner may not need to collect data on this variable. Table 6.9 summarises the survey that the practitioners need to perform before being able to use the learner in Table

6.7. In Table 6.9, the variable codes (Code) are kept equal to the original survey (Appendix C, Section C.2) to maintain uniformity.

Table 6.9 The required survey for assessing the technical reusability of a structural element using the learner in Table 6.7

Seq.	Code	Question / Options					Selected answer
1	B3	What is the approximate age of the building from which the element is recovered?					
		1	2	3	4	5	
		0 to 40	41 to 60	61 to 80	81 to 100	Above 100	
2	B5	What is the number of existing connections fixed to the element when purchased/acquired (plates or angles fixed to a beam, etc.)?					
		1	2	3	4	5	
		1 to 2	3 to 4	5 to 7	8 to 10	Above 10	
3	B7	The structural element is intended to be used for the same purpose (i.e. as a beam, slab, column, etc.) in its new installation.					
		1	2	3	4	5	
		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	
4	B8	The cross-section/thickness dimensions of the structural element in its new installation are expected to be equal or nearly equal to the cross-section/thickness dimensions of the element in its previous installation.					
		1	2	3	4	5	
		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	
5	C6	Estimated level of damage to the element due to the type of joints.					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
6	C12	Estimated level of damage to the element caused by living organisms (termite, bacterial attack, etc.)					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
7	C15	Estimated level of damage to the element due to post-production modifications (e.g. holes for ductwork, etc.)					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
8	C16	The negative impact of the lack of certificates of quality for the structural element.					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
9	C20	The negative impact of the lack of earlier certificates (inspection, material, etc.)					
		1	2	3	4	5	

Seq.	Code	Question / Options					Selected answer
		Very high	High	Moderate	Low	Very low	
10	C27	The negative impact of a potential problem with collateral warranties.					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
11	C28	The negative impact of the presence of hazardous, banned or contaminating coatings.					
		1	2	3	4	5	
		Very high	High	Moderate	Low	Very low	
12	D23	How do you expect that matching the design of the new building with the strength of the recovered element affects its reusability?					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
13	D24	How do you expect that challenges in designing with the reused element affect its reusability?					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	

For further details about how different steps of this method perform, please refer to (Deng 2014). Script E.24 (Appendix E) is used to extract the rules from the TEC-RF BSE-RPM.

6.4.2 Improving the transparency of the ECO-RF BSE-RPM

Figure 6.8 shows the results of the feature importance for the ECO-RF BSE-RPM. In this figure, the X-axis shows the relative importance of the variables, and the Y-axis shows the features. Based on Figure 6.8, all the variables are relevant and have relative importance above 0.02. This observation is in line with the results of the variable selection for the ECO dataset (Section 5.4). It results in a total number of twelve independent variables (Figure 6.8).

In the next stage, and to present how different values of a feature affects the economic reusability of building structural elements on average, a set of variable effect characteristic (VEC) curves are plotted for all predictors.

Figures 6.9 to 6.11 show the sensitivity analysis of the reusability factors based on Figure 6.8. According to these figures, in most cases, the economic reusability probabilities of a building's structural elements improves when the values of these variables increase. However, as discussed in Section 6.4.1, none of these features should be considered independently for estimating the economic reusability of building structural elements.

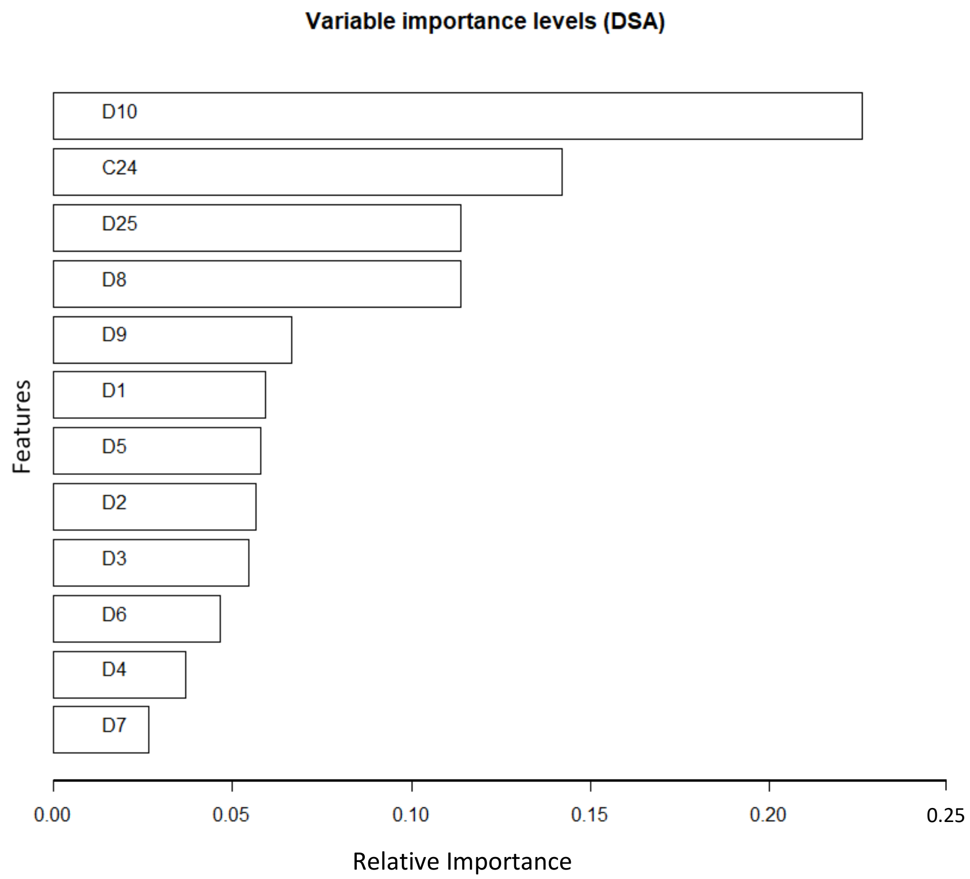

Figure 6.8 Bar plot with DSA and AAD relative feature importance for the ECO dataset based on the ECO-RF BSE-RPM

Figure 6.9 shows the sensitivity analysis of the top-four reusability factors (D10, C24, D25, D8) based on Figure 6.8. According to this figure, the economic reusability probabilities of a building's structural elements improves when the values of these variables increase from one (the highest negative impact) to five (the most positive effect). For the cash flow (D10), Figure 6.9 reveals that if it is necessary to purchase the required recovered elements early on and as soon as they are available, it could negatively affect the project due to additional costs such as the need to store the components for an extended period. Regarding C24, if the reuse of load-bearing building components reveals considerable financial risks as the result of extra efforts to find the required elements, changes in the original design to match with the properties of the recovered components, and other possible additional costs, reuse become economically unattractive. While a strict financial risk assessment at the beginning of any project is essential, the availability of financial incentives to recover and reuse building structural elements could overcome this barrier. Regarding the process to allocate and purchase the required components (D25), Figure 6.9 reveals that the increased difficulty in this process harms the economic reusability of the components. Eventually, Figure 6.9 shows that the increased cost of labour

(D8) could negatively affect the reuse rates because it could increase the overall project expenses.

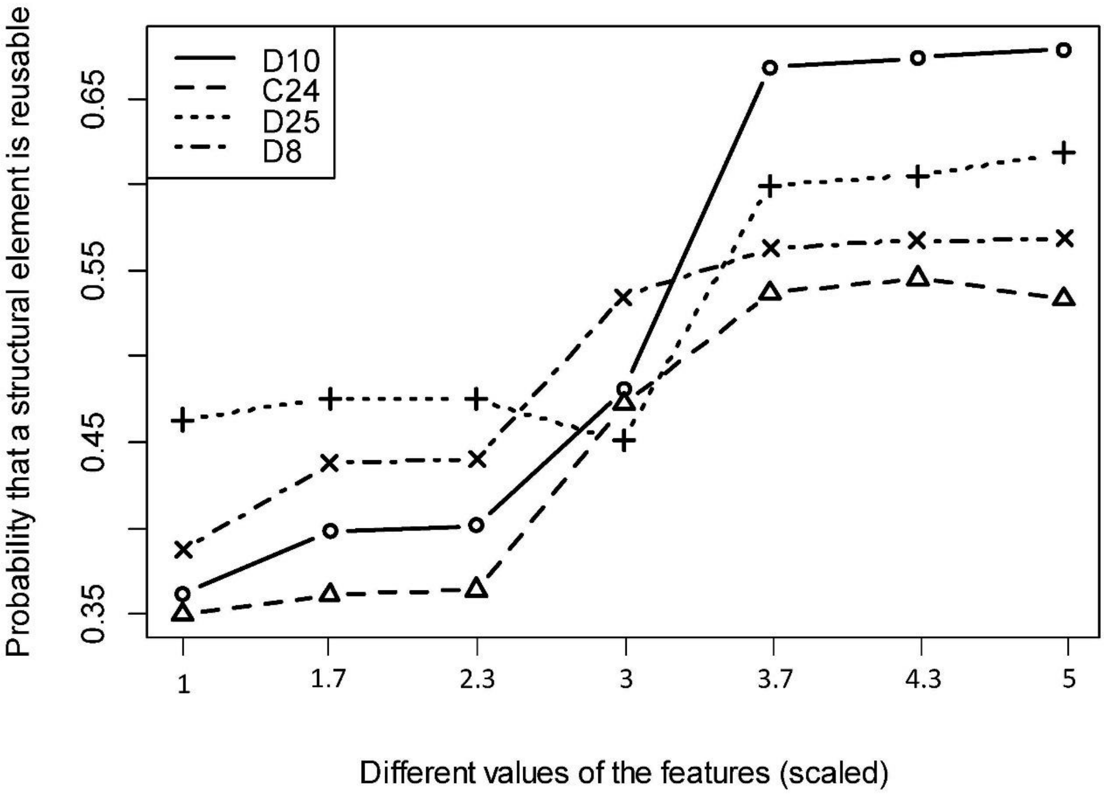

Figure 6.9 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D10, C24, D25, and D8 (the top-four variables in ECO-RF BSE-RPM)

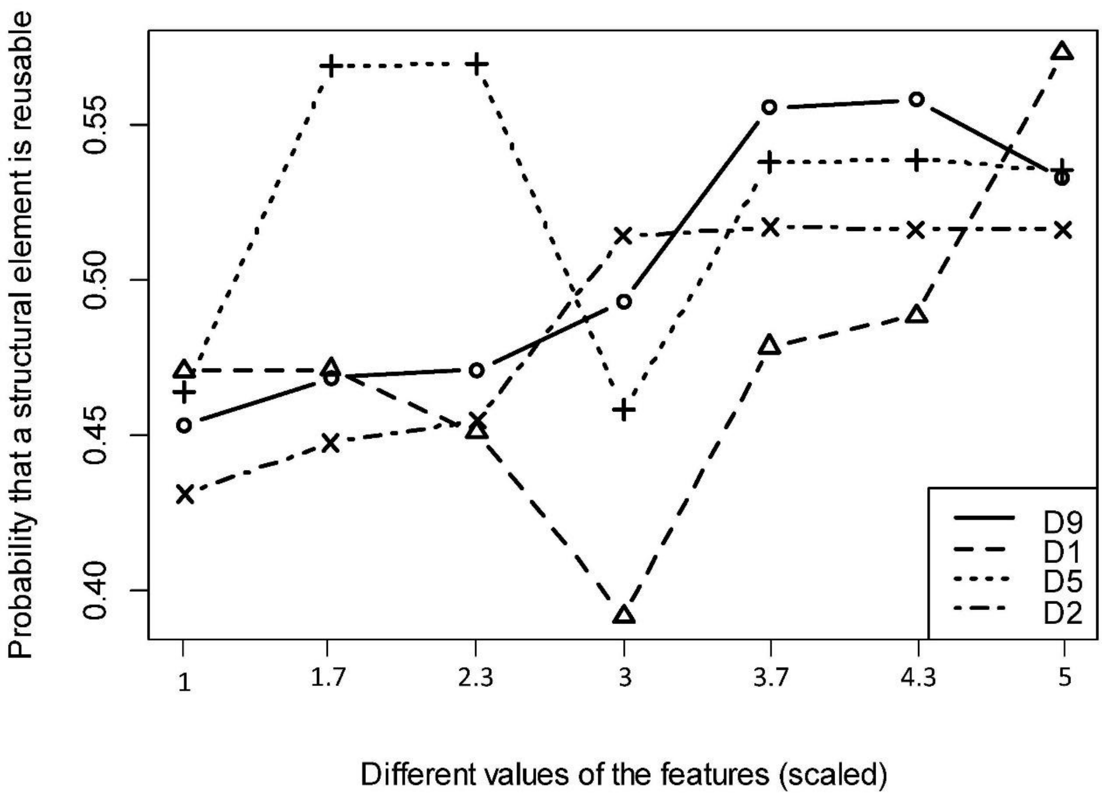

Figure 6.10 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D9, D1, D5, and D2 (ECO-RF BSE-RPM)

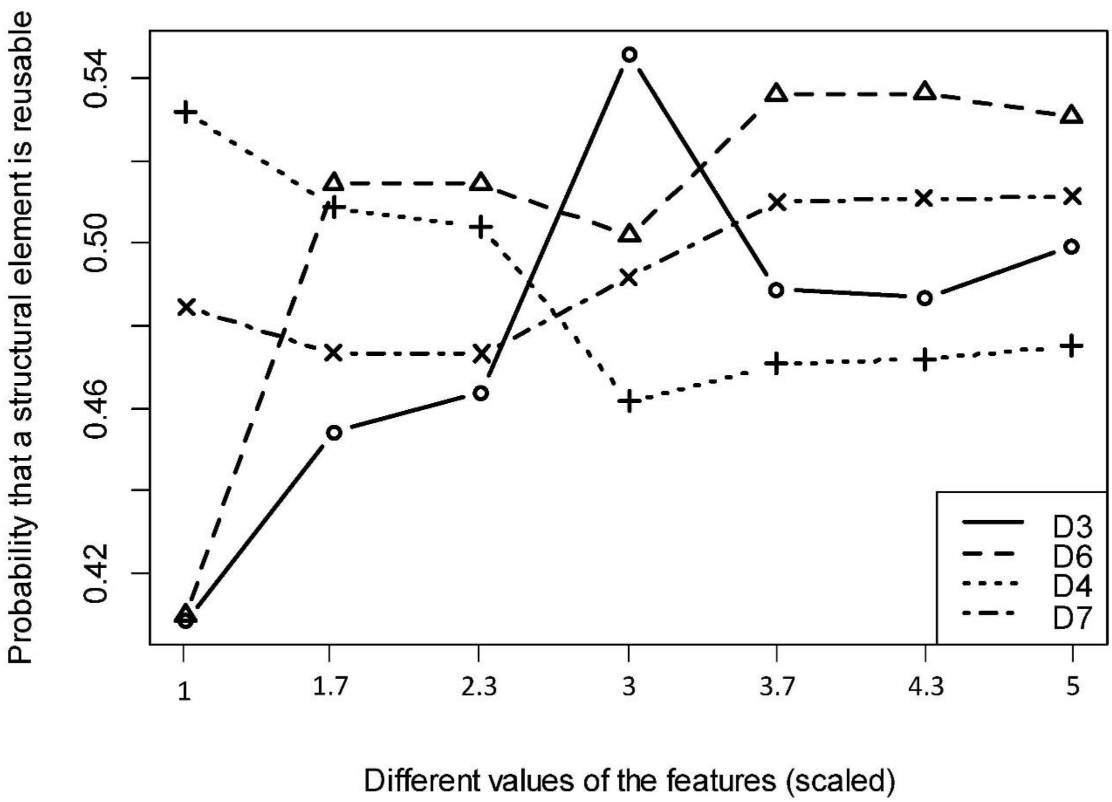

Figure 6.11 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D3, D6, D4, and D7 (ECO-RF BSE-RPM)

According to Figure 6.11, the higher values of D4 (cost of refurbishment) is associated with a decrease in the economic reusability of the structural elements of a building. However, this variable has the least importance among all other variables (Figure 6.8). Moreover, as discussed in Section 6.4.1, the interactions between variables should be considered for interpreting the results. Hence, to show the interdependency of the economic reusability factors, the most suitable feature (D10) and the least significant variable (D4) are plotted in Figure 6.12. According to this figure, the range of sensitivity for D4 is higher than D10 at all values. It acknowledges that D4 has much higher interdependency with other variables than D10. Consequently, Figure 6.12 shows that the economic reusability of the structural elements of a building depends on the interactions between the predictors, as well (Cortez and Embrechts 2013).

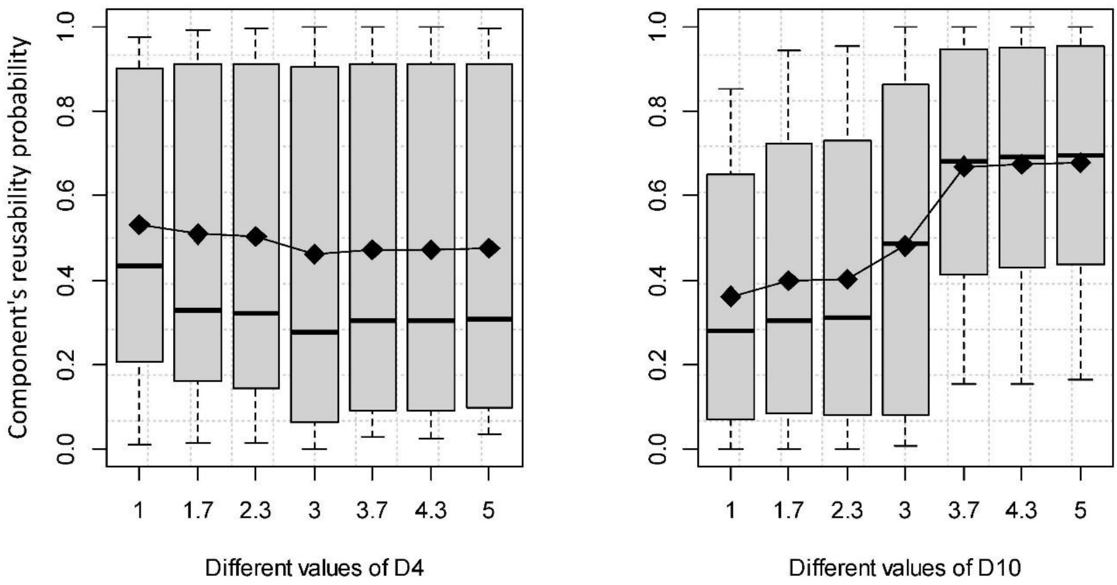

Figure 6.12 The VEC curves with box plots (to show the range of sensitivity at each point) to compare the impact of different values of D4 (left) and D10 (right) on the reusability probabilities of the elements (ECO-RF BSE-RPM)

While the above sensitivity analysis helps in improving the transparency of the ECO-RF BSE-RPM, as discussed in Section 6.4, a set of easy-to-understand rules are also developed to encourage the stakeholders to use the results of this research. These sets of rules are presented in Table 6.10. The details of this table are the same as Table 6.7 (Section 6.4.1) and are not repeated in this section. It is noteworthy that the first column of this table contains the sequence of the rules that need to be followed strictly.

Table 6.10 The learner (rules set) developed based on the ECO-RF BSE-RPM

Rule No.	Length	Frequency	Error	Condition	Prediction
1	3	0.225	0	$C24 > 4$ & $D8 > 2$ & $D25 > 2$	1
2	2	0.14	0	$D10 \leq 2$ & $D25 > 2$	0
3	5	0.101	0	$C24 > 3$ & $D1 = 3$ & $D5 \leq 3$ & $D10 \leq 3$ & $D25 > 1$	0
4	1	0.093	0	$D5 = 2$	1
5	4	0.093	0	$C24 \leq 3$ & $D4 > 2$ & $D5 \leq 3$ & $D25 \leq 3$	0
6	4	0.101	0	$C24 > 2$ & $D1 > 3$ & $D3 > 2$ & $D10 > 2$	1
7	3	0.047	0	$C24 \leq 3$ & $D6 > 3$ & $D10 \leq 4$	0
8	2	0.031	0	$C24 = 4$ & $D3 \leq 1$	1
9	2	0.031	0	$D1 > 3$ & $D3 \leq 1$	0
10	3	0.031	0	$D5 > 1$ & $D10 \leq 2$ & $D25 \leq 2$	1
11	2	0.023	0	$D5 \leq 1$ & $D6 > 2$	0
12	2	0.016	0	$D5 \leq 1$ & $D25 > 1$	0
13	2	0.062	0.125	$D3 \leq 2$ & $D6 \leq 2$	0
14	1	0.008	0	Else	1

Table 6.10 is developed based on the training dataset defined in Section 5.3. While the above set of rules provides an easy-to-understand and implement collections of conditions, it is essential to make sure that the resulting predictions on the unseen data satisfy the minimum requirements set in Section 6.3.1. Therefore, the corresponding testing dataset (unseen observations by the learner) was used to evaluate the performance of the learner presented in Table 6.10. Table 6.11 shows the results of the classifications made by this learner on the testing dataset (see Section 5.5.2.1).

Table 6.11 The confusion matrix of the learner presented in Table 6.10

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	26	2
Actual reusable (1)	8	19

Based on Table 6.11, the Type-I error rate is equal to 7.1%, and the overall accuracy is equal to 82%. Therefore, while this learner satisfies the minimum performance requirements defined in Section 6.3.1.1, its accuracy is slightly lower than 85%, which means it may classify an economically reusable component as non-reusable. Nonetheless, the learner in Table 6.10 is transparent and easy-to-understand and can be easily implemented in practice.

In Table 6.10, the rules are ordered, and they should be followed sequentially to find a condition that matches the predictor values to determine the economic reusability of a structural element. According to Table 6.10, D2, D7, and D9 are not available in any of the rules. Hence, a practitioner may not need to collect data on these variables to use the learner. Table 6.12 summarises the survey that the practitioners need to perform before being able to use the learner in Table 6.10. In Table 6.12, the variable codes (Code) are kept equal to the original survey (Appendix C, Section C.2) to maintain uniformity.

Table 6.12 The required survey for assessing the economic reusability of a structural element using the learner in Table 6.10

<table>
<tr><td>Seq.</td><td>Code</td><td colspan="5">Question / Options</td><td>Selected answer</td></tr>
<tr><td rowspan="3">1</td><td rowspan="3">C24</td><td colspan="5">The negative impact of the potential financial risks.</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very high</td><td>High</td><td>Moderate</td><td>Low</td><td>Very low</td></tr>
<tr><td colspan="8">For questions 2 to 9, please assess how do the following factors might affect the economic reusability of the structural element?</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2">D1</td><td colspan="5">The purchasing price / the analysis cost of an existing structure for reuse</td><td rowspan="2"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
</table>

Seq.	Code	Question / Options					Selected answer
		Very negatively	Negatively	No real effect	Positively	Very positively	
3	D3	Cost of testing					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
4	D4	Cost of refurbishment (sandblasting, treatment, etc.)					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
5	D5	Cost of design with the reused element					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
6	D6	Storage cost					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
7	D8	Cost of labour					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
8	D10	Cash flow (need to purchase the element early, etc.)					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
9	D25	Sourcing/procurement process					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	

For further details about how different steps of this method perform, please refer to (Deng 2014). Script E.24 (Appendix E) is used to extract the rules from the ECO-RF BSE-RPM.

6.4.3 Improving the transparency of the SOC-RF BSE-RPM

Figure 6.13 shows the results of the feature importance for the SOC-RF BSE-RPM. In this figure, the X-axis shows the relative importance of the variables, and the Y-axis shows the features. Based on Figure 6.13, all the variables are relevant and have relative importance above 0.02. This observation is in line with the results of the variable selection for the SOC dataset (Section 5.4). It results in a total number of ten independent variables (Figure 6.13).

In the next stage, and to present how different values of a feature affects the social reusability of building structural elements on average, a set of variable effect characteristic (VEC) curves are plotted for all predictors (Figures 6.14 to 6.16).

Figure 6.14 shows the sensitivity analysis of the top-four features in the SOC dataset. For D16, C22, and D15, the higher values of the variables are associated with an improvement in social reusability. Whereas for C23, this increase has a counter effect. Notwithstanding, as discussed in Section 6.4.1, this variable cannot determine the social reusability of a component on its own, and the interactions with other variables should be considered, as well. For instance, according to Table 6.13, C23 is positively correlated with C22 and has a negative correlation with all other variables. While Table 6.13 clearly shows the linear interdependencies among the variables, it does not mean that the real relationship between predictors is linear. The result of the parametric models (Table 6.5) shows that the non-linear classifiers outperform the linear methods, an indication that the actual relationship between the predictors and the outcome is non-linear.

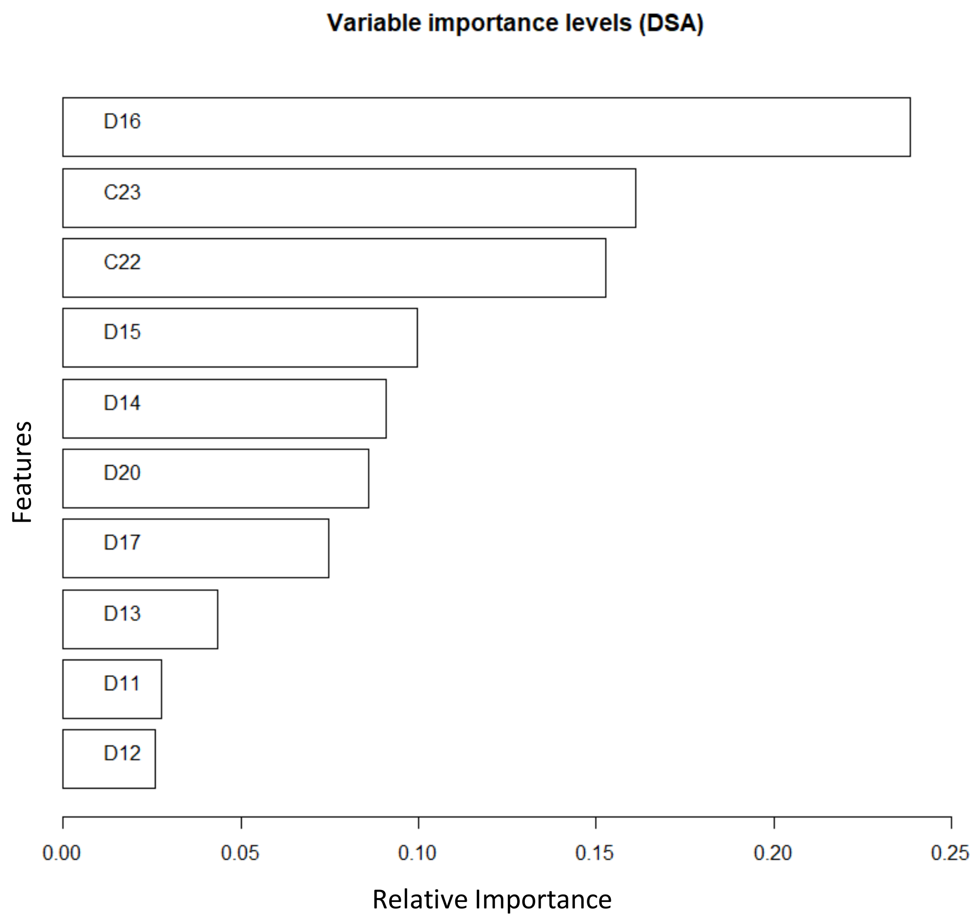

Figure 6.13 Bar plot with DSA and AAD relative feature importance for the ECO dataset based on the SOC-RF BSE-RPM

Table 6.13 Correlation between features in the SOC dataset (Pearson's)

	C23	D11	D12	D13	D14	D15	D16	D17	D20
C22	0.563**	-0.052	-0.044	-0.015	-0.091	-0.014	-0.013	0.048	-0.157
C23	-	-0.184	-0.161	-0.091	-0.21*	-0.097	-0.221*	-0.044	-0.15**
D11		-	0.816**	0.645**	0.685**	0.291**	0.328**	0.520**	0.386**
D12			-	0.704**	0.722**	0.434**	0.392**	0.535**	0.398**
D13				-	0.649**	0.443**	0.59**	0.484**	0.534**
D14					-	0.432**	0.334**	0.573**	0.501**
D15						-	0.379**	0.445**	0.492**
D16							-	0.327**	0.518**
D17								-	0.398**
**. Correlation is significant at the 0.01 level (2-tailed).									
*. Correlation is significant at the 0.05 level (2-tailed).									

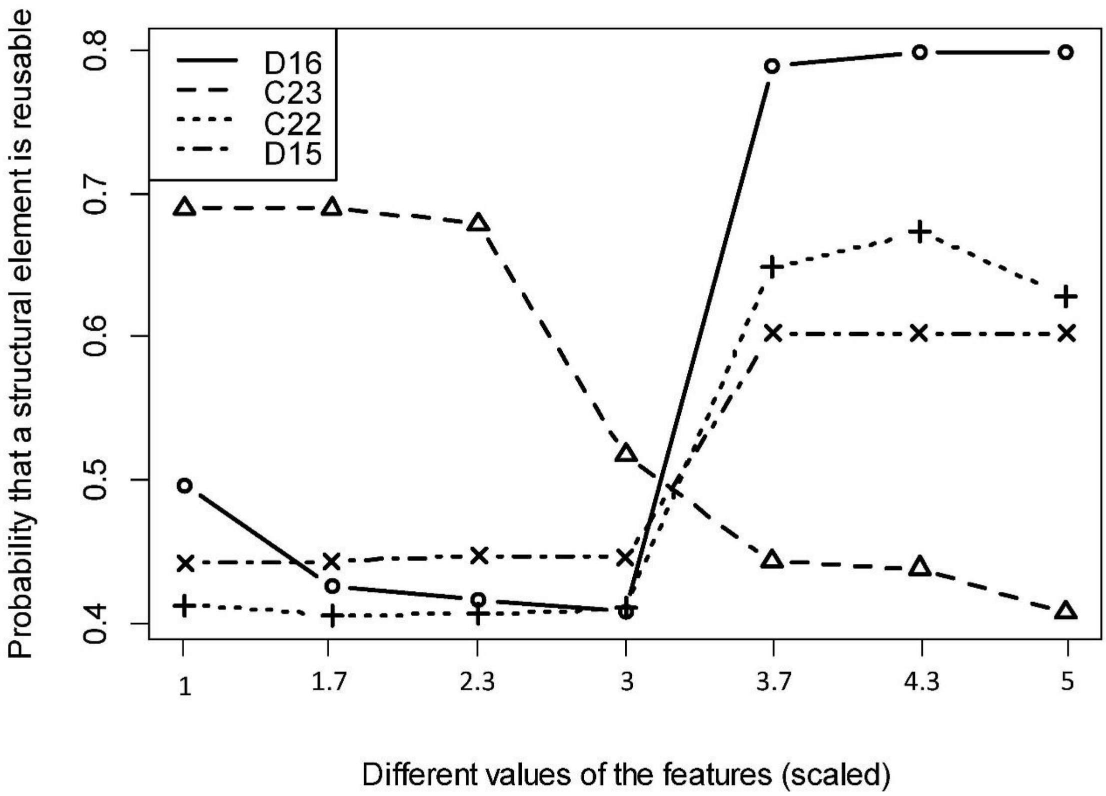

Figure 6.14 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D16, C23, C22, and D15 (the top-four variables in SOC-RF BSE-RPM)

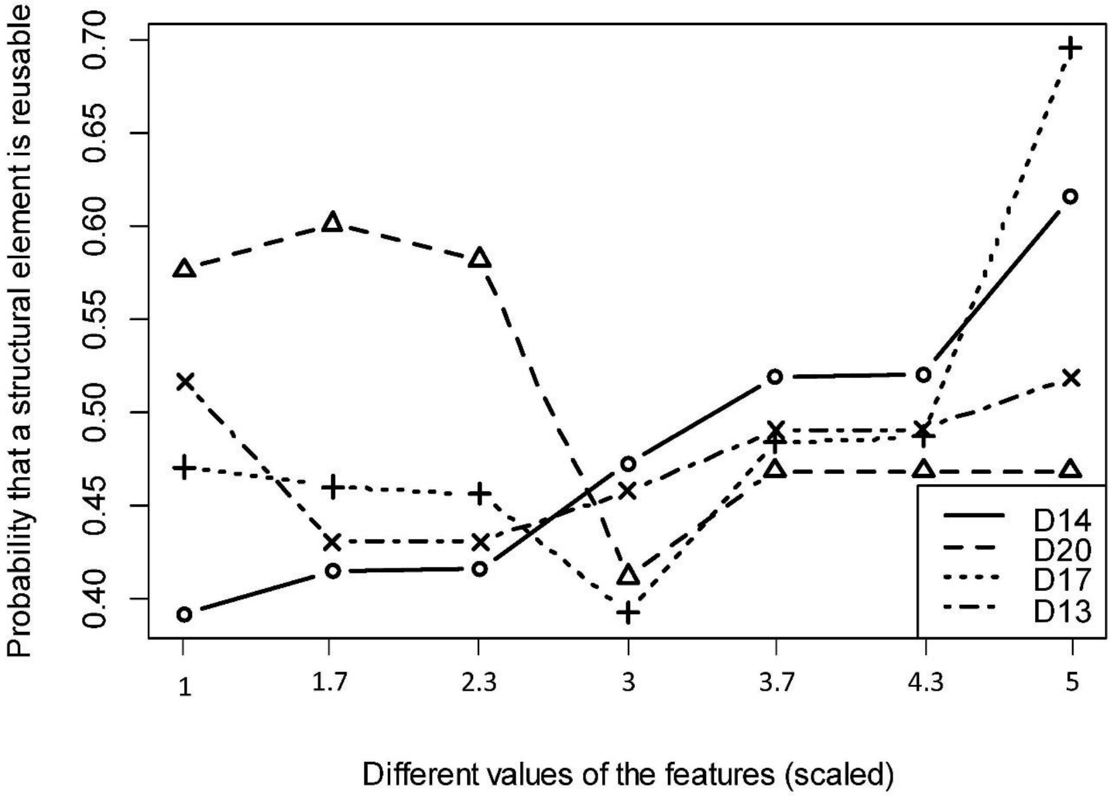

Figure 6.15 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D14, D20, D17, and D13 (SOC-RF BSE-RPM)

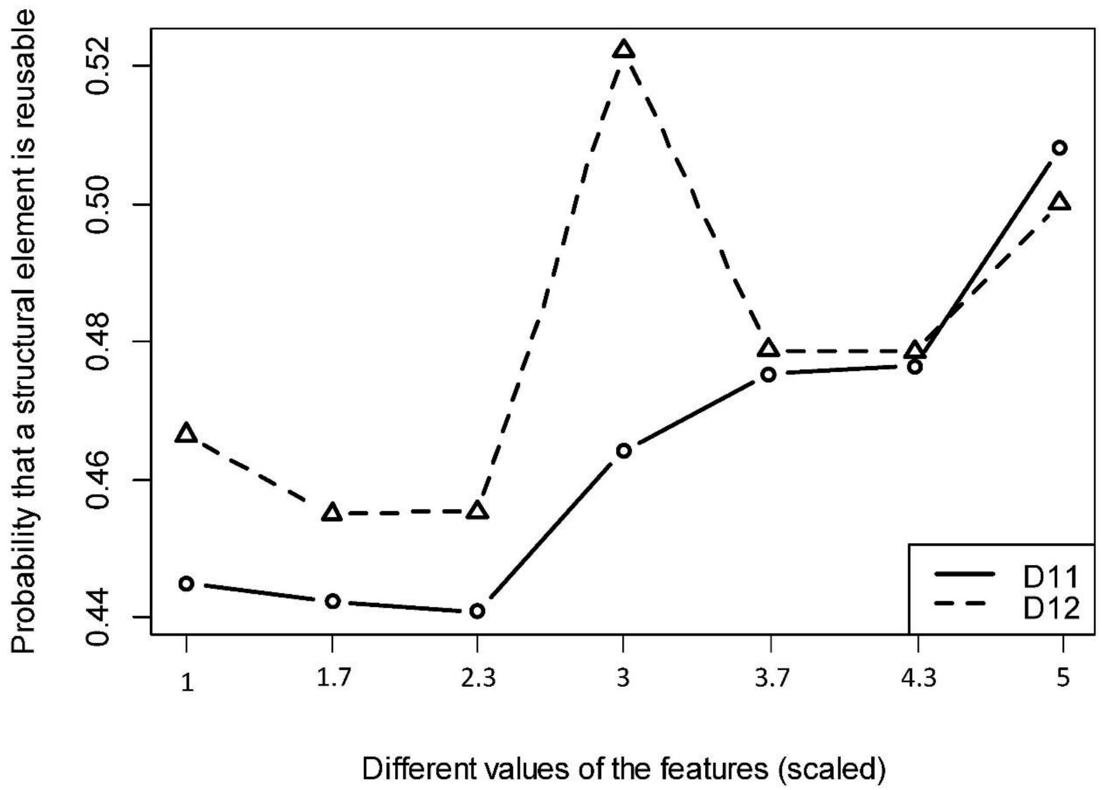

Figure 6.16 The impact of different values the features on the reusability probabilities of the elements (sensitivity analysis) for D11 and D12 (SOC-RF BSE-RPM)

Figures 6.15 and 6.16 show that, in most cases, the reusability probabilities of a building's structural elements improve when the values of these variables increase. However, as discussed earlier, because of the interdependencies between the predictors (Table 6.13), none of these features should be considered independently for estimating the social reusability of a building's structural elements.

In the next stage, and following the approach adopted in Sections 6.4.1 and 6.4.2, a set of easy-to-understand rules are also developed to further clarify the outcome of the selected SOC-RF BSE-RPM. The resulting set of rules is presented in Table 6.14.

The details of Table 6.14 are the same as Table 6.7 (Section 6.4.1) and are not repeated in this section. It is noteworthy that the first column of this table contains the sequence of the rules that need to be followed strictly.

Table 6.14 The learner (rules set) developed based on the SOC-RF BSE-RPM

Rule No.	Length	Frequency	Error	Condition	Prediction
1	3	0.244	0	$C23 > 3 \& D13 = 2 \& D17 \leq 4$	0
2	3	0.179	0	$C22 > 3 \& D11 > 2 \& D12 > 1$	1
3	5	0.218	0	$C22 \leq 3 \& D15 \leq 3 \& D16 \leq 3 \& D17 \leq 4 \& D20 > 2$	0
4	2	0.064	0	$C23 \leq 3 \& D16 \leq 2$	1
5	2	0.064	0	$D14 \leq 2 \& D17 = 3$	0
6	2	0.051	0	$C23 \leq 4 \& D12 = 3$	1
7	1	0.179	0	Else	1

Table 6.14 is developed based on the training dataset defined in Section 5.3. While the above set of rules provides an easy-to-understand and implement collections of conditions, it is essential to make sure that the resulting predictions on the unseen data satisfy the minimum requirements set in Section 6.3.1. Therefore, the corresponding testing dataset (unseen observations by the learner) was used to evaluate the performance of the learner presented in Table 6.14. Table 6.15 shows the results of the classifications made by this learner on the testing dataset (see Section 5.5.2.1).

Table 6.15 The confusion matrix of the learner presented in Table 6.14

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	18	0
Actual reusable (1)	3	13

Based on Table 6.15, there is no Type-I error, and the overall accuracy is equal to 91%. Therefore, this learner satisfies the minimum performance requirements defined in sections 6.3.1.1 and 6.3.1.2. Moreover, the learner in Table 6.14 is transparent and easy-to-understand and can be easily implemented in practice.

In Table 6.14, the rules are ordered, and they should be followed sequentially to find a condition that matches the predictor values to determine the social reusability of a structural element. Table 6.16 summarises the survey that the practitioners need to perform before being able to use the learner in Table 6.14. In Table 6.16, the variable codes (Code) are kept equal to the original survey (Appendix C, Section C.2) to maintain uniformity.

Table 6.16 The required survey for assessing the social reusability of a structural element using the learner in Table 6.14

<table>
<tr><th>Seq.</th><th>Code</th><th colspan="5">Question / Options</th><th>Selected answer</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">C22</td><td colspan="5">The potential liability risks related to reusing the recovered structural elements.</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very high</td><td>High</td><td>Moderate</td><td>Low</td><td>Very low</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">C23</td><td colspan="5">The potential health and safety risks related to reusing the recovered structural elements.</td><td></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td></td></tr>
<tr><td>Very high</td><td>High</td><td>Moderate</td><td>Low</td><td>Very high</td><td></td></tr>
<tr><td colspan="8">For questions 3 to 10, please assess how do the following factors might affect the social reusability of the structural element?</td></tr>
<tr><td rowspan="3">3</td><td rowspan="3">D11</td><td colspan="5">Perception of the client/top management team about the element</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very negatively</td><td>Negatively</td><td>No real effect</td><td>Positively</td><td>Very positively</td></tr>
<tr><td rowspan="3">4</td><td rowspan="3">D12</td><td colspan="5">Perception of the designers about the element</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very negatively</td><td>Negatively</td><td>No real effect</td><td>Positively</td><td>Very positively</td></tr>
<tr><td rowspan="3">5</td><td rowspan="3">D13</td><td colspan="5">Perception of the builders/contractors about the element</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very negatively</td><td>Negatively</td><td>No real effect</td><td>Positively</td><td>Very positively</td></tr>
<tr><td rowspan="3">6</td><td rowspan="3">D14</td><td colspan="5">Perception of the end-users (when it is not the client) about the element</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very negatively</td><td>Negatively</td><td>No real effect</td><td>Positively</td><td>Very positively</td></tr>
<tr><td rowspan="3">7</td><td rowspan="3">D15</td><td colspan="5">Perception of the stockist about the element</td><td rowspan="3"></td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Very negatively</td><td>Negatively</td><td>No real effect</td><td>Positively</td><td>Very positively</td></tr>
</table>

Seq.	Code	Question / Options					Selected answer
8	D16	Perception of the regulatory authorities about the element					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
9	D17	Visual appearance					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	
10	D20	Changes in the health and safety regulations (fire, etc.)					
		1	2	3	4	5	
		Very negatively	Negatively	No real effect	Positively	Very positively	

For further details about how different steps of this method perform, please refer to (Deng 2014). Script E.24 (Appendix E) is used to extract the rules from the SOC-RF BSE-RPM.

6.5 Instructions for using the developed learners

In Section 6.4, three learners were developed and presented in Tables 6.7, 6.10, and 6.14 using the results of the best-practice random forest models for the TEC, ECO, and SOC datasets, respectively. In Section 6.5, a flow chart is developed to help the practitioners in the building sector to use these learners effectively. Figure 6.17 shows this flow chart.

Before using these learners, it is essential to consider the following. If the below conditions are not satisfied, the learners in Tables 6.7, 6.10, and 6.14 cannot be used.

- The learners presented in Tables 6.7, 6.10, and 6.14 are designed to assist the construction professionals in their decision-making process for reusing load-bearing building components from technical, economic, and social aspects. First, an item should be confirmed reusable using the learner presented in Table 6.7. Next, if the learner is technically reusable, using the learner in Table 6.10, its economic reusability should be assessed. Eventually, the item should be assessed from a social perspective using the learner in Table 6.14.
- It is assumed that the elements are/would be recovered through deconstruction. If demolition is considered, reuse of the load-bearing building components is not practical due to the damages during this process (Section 1.10).

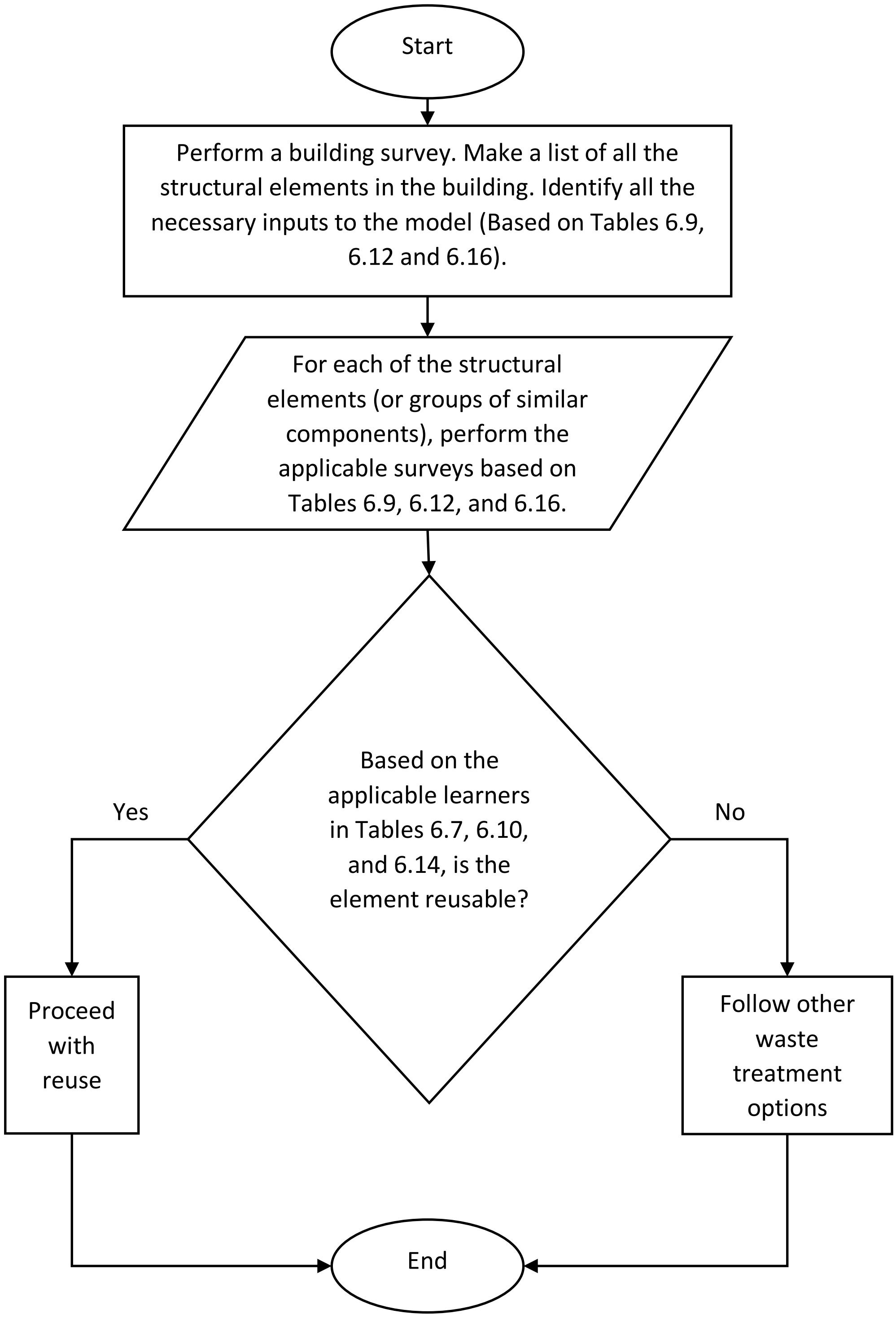

Figure 6.17 Instructions for using the learners developed in Tables 6.7, 6.10, and 6.14

Figure 6.18 presents an example of using these learners for predicting the technical reusability of the structural elements at the end-of-life of a building. It is noteworthy that the structural component presented in this example is the result of a real survey that was received after the development of the predictive models in this study. Hence, it was not used for training or performance evaluation of the predictive models. This component was technically reusable based on the confirmation of the respondent. According to Figure 6.18, the learner predicts that the element is reusable, which agrees with the real status of the component.

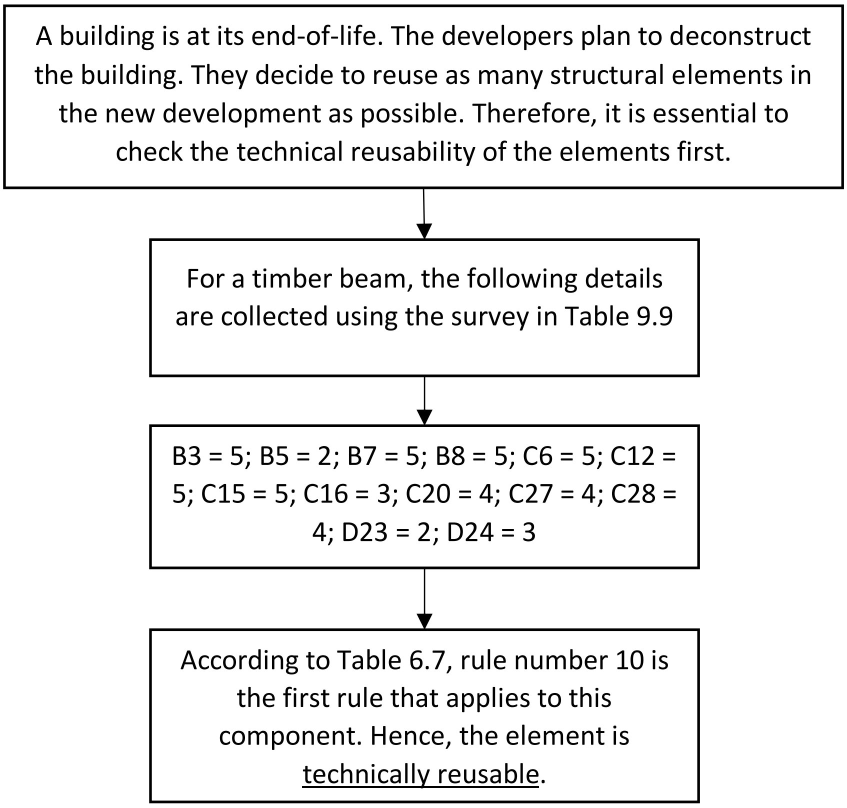

Figure 6.18 An example of using the learner presented in Table 6.7 for predicting the technical reusability of a timber beam

6.6 Technical reusability factors

Based on Figure 6.2, the most important factor affecting the reusability of the building structural elements is the mechanical properties of the component (D23). This observation is in line with the attempts of some researchers in estimating the mechanical properties of the load-bearing components as an indicator of reusability (Fujita and Masuda 2014, Fujita and Kuki 2016, Cavalli et al. 2016).

The next important variable is the other design challenges observed by the stakeholders (D24). In the literature, these challenges are identified as integrating reused and new components into the new building (Gorgolewski 2008), need for flexibility in the design (Gorgolewski 2008), and overdesigned structures due to the available supply (Brütting et al. 2019).

The third variable affecting the reusability of building structural elements is the presence of hazardous, banned or contaminating coatings (C28). This variable has been reported in various articles in the literature including (Rameezdeen et al. 2016, Tatiya et al. 2017, Tingley et al.

2017). If such coatings are present on the structural elements, the chance for recovery and reuse decreases drastically. As a solution, and to overcome this barrier in new buildings, Basta et al. (2020) proposed a reusable fireproofing system to promote the reusability of the building structure.

According to Figure 6.2, the fourth most important barrier is a potential problem with collateral warranties. Surprisingly, this barrier was not observed by other researchers. However, according to Addis (2006), issues related to the performance of the recovered structural element should be resolved early to avoid a problem with collateral warranties.

6.7 Economic reusability factors

According to Figure 6.8, the most important economic factors affecting the reusability of the structural components of a building is the need to purchase reused elements early in the project, which can have cash flow implications. This observation is in line with (Gorgolewski 2008, Gorgolewski et al. 2008). According to Gorgolewski et al. (2008), the need to purchase early on requires the client to allocate resources and can increase the cost of storage.

The second most important factor, based on Figure 6.8, is the potential financial risks. According to the literature (Rameezdeen et al. 2016, Pun, Liu, and Langston 2006), these potential financial risks might be the result of other variables such as deconstruction, transportation, and storage costs (Dantata, Touran, and Wang 2005, Chileshe, Rameezdeen, and Hosseini 2015, Yeung, Walbridge, and Haas 2015, Tingley et al. 2017, Rose and Stegemann 2018, Dunant et al. 2018, Tatiya et al. 2017). As discussed in Section 2.4.2, a strict financial risk assessment at the beginning of any project with reused structural elements is then necessary. As shown in Figure 6.9, if these risks are low, there is a higher chance for reuse.

The third most important economic factor is the sourcing/procurement process. This factor has been continuously reported in the literature as one of the main factors affecting reuse (Section 2.3.2). According to Section 2.4.2, this factor is categorised under the supply chain level, and it is observed that there is a significant correlation between the market and cost. If an established market for the reused structural elements is not available (Shaurette 2006, Gorgolewski 2008, Gorgolewski et al. 2008, Dunant et al. 2018), the design team need to put extra efforts to allocate the desired element, which in turn can increase the overall cost of the project (Gorgolewski et al. 2008). According to Figure 6.9, the reusability of building components increases if the difficulty in sourcing decreases.

Based on Figure 6.8, the fourth most affecting variable is the cost of labour. Dantata et al. (2005) observed that deconstruction and recovery of the structural elements are time-consuming and can decrease the economic viability of reuse. According to Figure 6.9, the lower cost of labour is associated with the higher reusability of building components.

6.8 Social reusability factors

According to Figure 6.13, the most important social factor affecting reuse is the perception of the regulatory authorities about a recovered structural component. This factor was observed by Chileshe et al. (2015) in the context of South Australian construction, as well. According to this study, to improve the perception of the building regulators, it is essential to increase the awareness of the stakeholders about the advantages of reuse (Chileshe, Rameezdeen, and Hosseini 2015). Figure 6.14 reveals that when this perception is in favour of reuse, the reusability of the structural components increases.

The second and third ranks among the social reusability factors belong to risks. These factors were reported as reuse barriers by several authors in the literature (Huuhka and Hakanen 2015, Rameezdeen et al. 2016, Klang, Vikman, and Brattebø 2003, Gorgolewski 2008, Tingley et al. 2017). According to Section 2.4.1, there is a strong correlation between perception and risk. As discussed in Section 2.4.1, the potential risks associated with reusing structural elements affect the stakeholders' perception about reuse.

The fourth most important social factor is the perception of the stockist about the element (D15). This factor has been reported by Dunant et al. (2017). According to Dunant et al. (2017), the stockists are sensitive to the visual appearance of the recovered structural elements, which could affect their perception of the reusability of these components. Based on Figure 6.14, the positive perception of the stakeholders towards recovered load-bearing structural elements of a building could potentially improve their reusability.

6.9 Chapter summary

Chapter 6 was focused on fulfilling the fourth objective of this research by developing best-practice BSE-RPMs using advanced supervised machine learning methods, which provide reliable predictions. Initially, this chapter assessed the performance of the developed BSE-RPMs in Chapter 5 using a k-fold Cross-validation method with $k = 10$ (Section 6.3). While both performance and interpretability are essential in the selection of the best-practice models, the results revealed that the understandable models were having poor performance. Hence, only the Type-I error rate, overall accuracy, and the AUC were used to select the best-practice models. The result was choosing random forest models for the technical, economic, and social aspects of this research (Section 6.3).

The selected TEC-RF BSE-RPM, ECO-RF BSE-RPM, and SOC-RF BSE-RPM outperform all other models. However, since they are known as black boxes, they lack transparency. Therefore, to improve clarity, this research opened the selected black-box models in two ways (Section 6.4). First, the models were opened using advanced sensitivity analysis and visualisation techniques. Using these methods, the author identified the relative importance of the features and demonstrated the effect of different values of the features on the reusability of the structural components. Next, the author used the results of the previous stage and developed a set of easy-to-understand rules so that the stakeholders could use them as a guideline to identify the technical, economic, and social reusability of these elements. The researcher then evaluated the performance of the developed learners (Tables 6.7, 6.10, and 6.14) and concluded that they produce reliable predictions. Eventually, the author revised the original survey (Appendix C.2) and produced three new questionnaires that stakeholders can use to gather information for using the developed learners (Tables 6.9, 6.12, and 6.16 for the technical, economic, and social aspects of this research).

Chapter 7 – Conclusion and recommendations

7.1 Chapter introduction

This chapter concludes this research, which aimed to develop a set of tools to predict the reuse potential of the load-bearing building components based on professional experience from technical, economic, and social perspectives. This research considered four objectives to fulfil its aim, as specified in Section 1.6. Section 7.2 presents a summary of the findings concerning the identified objectives of the research. Next, Section 7.3 highlights the contributions of this research from academic and industrial perspectives (Sections 7.3.1 and 7.3.2, respectively). Section 7.4 reintroduces the limitations of the research, and Section 7.5 discusses the future research opportunities in this field. Eventually, this chapter concludes by summarising the results in Section 7.6.

7.2 Summary of the findings

Four objectives were considered to answer the research questions (Section 1.5) and fulfil the aim of this research (Section 1.6). Achieving each of these objectives helped to uncover unknown dimensions of a new paradigm in the field of reuse in the construction sector, which is determining the reusability of the load-bearing building components. Subsections 7.2.1 to 7.2.4 present a summary of these findings.

7.2.1 Objective One: To identify and assess factors affecting the reusability of a building's structural elements (reusability factors) through a literature review.

The identification of the reusability factors was performed through a systematic literature review targeting peer-reviewed journal articles (Chapter 2). After a careful study of the top-tier construction journals (Chapter 2, Section 2.2), 76 peer-reviewed journal articles were reviewed to identify factors affecting the reuse of load-bearing building components. In total, 57 drivers and 130 barriers affecting the reuse of these components were identified. These factors were then categorised into economic, environmental, social, technical, regulatory, and organisational groups.

The review of the categories of the variables showed that the top-three groups of the identified drivers were economic, organisational, and environmental. Also, reviewing the frequency of the reported barriers in the literature revealed that the economic factors were playing a significant role in the successful implementation of reuse in the building sector, followed by technical, social, regulatory, and organisational barriers. As discussed in Section 1.4, identifying the technical reusability of the load-bearing building components has introduced a new paradigm in the field of reuse and has been the focus of research recently. Moreover, the analysis of the inter-relationship between the sub-categories of barriers in Chapter 2 revealed that social and economic factors are having a significant impact on the widespread of reusing recovered load-bearing building components. Therefore, this research focused on estimating the technical, economic, and social reusability of the recovered structural elements of a building.

It should be noted that since the focus of this research was to develop tools to estimate the technical, economic, and social reusability of the load-bearing building components, only factors under these categories were used to prepare the questionnaire survey to achieve the second objective of this research.

7.2.2 Objective Two: To quantify the weightage and impact of the reusability factors based on the experience of the professionals using questionnaires.

According to the collected questionnaires, it was observed that among different structural components, 62.5% of respondents referred to beams (of various materials) to complete the survey. Therefore, to evaluate if the type of the element (question B1) affects the scores provided for the factors affecting the reusability of the structural components, a non-parametric test (Kruskal-Wallis H test) was performed at a 5% significance level. The results revealed that there was no statistical difference between the groups of the structural elements at 95% confidence level, which means that the type of the component (i.e., beam, column, truss, etc. based on question B1) does not affect the scores given to the reusability factors.

The results of the descriptive statistics for the technical (TEC), economic (ECO), and social (SOC) datasets are as follows.

From a technical perspective, the following factors were identified as the most significant barriers ahead of the reuse of load-bearing building components (Table 4.5 and Appendix D.4).

- Matching the original design with the dimensions of the reused element (D22)
- Changes in the design codes (BS codes to Eurocodes, etc.) (D19)
- CE marking (D21)
- Matching the original design with the strength of the reused element (D23)

From these variables, it can be observed that the design-related factors are the most significant variables affecting the reusability of the load-bearing building components.

The following factors were identified as the most significant barriers against the economic reusability of the structural elements of a building (Table 4.6 and Appendix D.4).

- Cost of testing (D3)
- Cost of insurance (D2)
- Storage cost (D6)
- Cost of refurbishment (sandblasting, treatment, etc.) (D4)

The above observations reflect the fact that the reuse of the load-bearing building components is associated with additional costs that could negatively affect the successful integration of the reused elements in the new buildings.

The results of the descriptive statistics of the received questionnaires revealed that the following barriers significantly affect the social reusability of the load-bearing building components (Table 4.7 and Appendix D.4).

- Changes in the health and safety regulations (fire, etc.) (D20)
- Perception of the stockist about the element (D15)
- Perception of the regulatory authorities about the element (D16)
- Perception of the builders/contractors about the element (D13)

According to these variables, the perception of the stakeholders has the highest impact on the social reusability of the load-bearing building components.

While the results of the descriptive statistics provide an overview of the barriers to reuse from different perspectives, it should be noted that these variables cannot be directly used to determine if a structural component is reusable or not. For instance, considering the technical dataset (72 valid responses, see Section 4.7), matching the original design with the dimensions of the reused element (D22) is the most significant barrier with a mean of 2.53 and a standard deviation of 1.14 (Table 4.5 and Appendix D.4). Considering $D22$ to decide if an element is technically reusable or not, a model predicts reusable if $D22 \geq 3$, which results in predicting 37 reusable and 35 non-reusable components (Table 7.1). The following confusion matrix (Table 7.1) is developed based on this classifier for the entire TEC dataset (72 responses). As can be observed, the model's overall accuracy (Section 5.5.2.6) is equal to 46%.

Whereas, since the number of reusable components in the received dataset is 48, a baseline model (Section 6.3.1.2) always predicts reusable for all elements, which results in an overall accuracy of 67%. It shows that using the $D22 \geq 3$ rule as an indication of reusability results in a prediction worse than the baseline model, which is not acceptable.

Table 7.1 Technical reusability of the elements in the original dataset using $D22 \geq 3$ rule only

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	10	14
Actual reusable (1)	25	23

Moreover, the results of the descriptive statistics do not reveal which combination of variables could provide the most accurate estimate of the reusability of the components. In this research, these shortcomings were addressed through the third and fourth objectives of this research.

7.2.3 Objective Three: To determine the best combination of the identified factors to develop the BSE-RPMs.

This research aimed to predict the technical, economic, and social reusability of the load-bearing building components based on the experts' opinions. Therefore, this research used a combination of filter and wrapper techniques to select the best combination of variables that could result in reliable predictions on unseen observations.

In this research, ten different filter methods were used to rank the importance of the variables in all three datasets. While the outcome of the filter methods provided an overview of the importance of the variables, it was decided to use wrappers to compare the results and choose the best combination of variables in all datasets. For this purpose, the Boruta method was used for variable selection in all three datasets as well. In the case of economic and social datasets, the Boruta method identified that all variables were suitable for the development of the BSE-RPMs. Moreover, there was a good agreement between the rankings made by the filter techniques and the Boruta method in these two datasets. However, in the case of the technical dataset, some of the variables were rejected, and the Boruta method could not determine the suitability of one feature. Moreover, the rankings made by the filter methods were different from the feature selection of the Boruta technique. Therefore, it was decided to employ the RFE method to add a new layer to the process of feature selection. In this research, four different supervised machine learning methods were used to perform the RFE technique. It was observed that the RFE results are in good agreement with the Boruta technique. Therefore, twenty-six variables that were not rejected by the RFE and Boruta methods were selected for the development of the technical BSE-RPMs.

7.2.4 Objective Four: To develop a best-practice BSE-RPM using advanced supervised machine learning techniques, which provides reliable predictions.

In this research, thirteen different methods were used to develop 13 predictive models for each dataset. These methods cover both parametric and non-parametric techniques and range from interpretable Logistic Regression to complex Gaussian Processes and Support Vector Machines. While it is desirable to have accurate and interpretable models to encourage the building experts to use the results of this research, analysing the results revealed that such models were inefficient in terms of accuracy. Hence, the selection of the best-practice models was performed based on the predictive performance of the models only. From the above discussion, it can be concluded that the relationship between features and responses in these datasets are not linear.

For all three datasets, the random forest (RF) models were selected as the best-practice BSE-RPMs because they outperformed all other models. However, these models are known as black boxes because they cannot be interpreted easily. Therefore, this study used advanced sensitivity analysis and visualisation techniques to open the best-practice BSE-RPMs.

Opening the TEC-RF BSE-RPM using the sensitivity analysis and visualisation techniques revealed that the following factors are the most important variables affecting the technical reusability of the load-bearing building components.

- Matching the original design with the strength of the reused element (D23)
- Other design challenges with the reused element (D24)
- Presence of hazardous, banned or contaminating coatings (C28)
- A potential problem with collateral warranties (C27)

The results of sensitivity analysis and visualisation techniques to improve the transparency of the ECO-RF BSE-RPM revealed that the following variables significantly affect the economic reusability of the load-bearing building components.

- Cash flow (need to purchase the element early, etc.) (D10)
- Potential financial risks (C24)
- Sourcing/procurement process (D25)
- Cost of labour (D8)

Improving the transparency of the SOC-RF BSE-RPM using sensitivity analysis and visualisation techniques revealed that the following factors are having the highest effect on the social reusability of the structural elements of a building at its end-of-life.

- Perception of the regulatory authorities about the element (D16)
- Potential health and safety risks (C23)
- Potential liability risks (C22)
- Perception of the stockist about the element (D15)

The above findings are different from the results of the descriptive statistics on the collected data. As discussed earlier, measures such as mean, or median do not consider the possible interdependency of the features and the response. Also, from the developed predictive models, it was observed that this relationship is non-linear. Therefore, the outcome of the descriptive statistics presented in Section 7.2.2 is not reliable.

While the findings presented in Section 7.2.4 help in opening the selected best-practice models, they cannot be used directly to determine the reusability of the load-bearing building components. Hence, this research used advanced rule-extraction techniques to develop three easy-to-understand models that can be used by practitioners in the building sector to assess if a recovered structural element is reusable or not from technical, economic, and social perspectives. The resulting tools are easily interpretable and produce reliable predictions on unseen observations, hence, fulfilling the aim of this research.

7.3 Contributions of the research

7.3.1 Contributions of the research to the body of knowledge

This research contributes to the body of knowledge in different ways. First, this research shows how advanced supervised machine learning techniques such as random forests, K-Nearest Neighbours algorithm, Gaussian processes, support vector machines, adaptive boosting, BART machine, etc., (Section 5.5) can be used to promote the circular economy in the building sector. Second, this research showed that the relationship between factors affecting reuse is not linear and that the results of the ordinary statistics have significant restrictions. Also, this research successfully ranked the factors affecting the reusability of load-bearing building components. This achievement assists other researchers to take progressive steps towards the circularity of materials in this sector by prioritising their research. Likewise, this research showcased how complex supervised machine learning techniques could be handled to produce practical tools that can be used by practitioners who have no prior knowledge about these complex data analysis techniques. While this research is focused on the building sector, the techniques used could be employed to perform similar studies in different divisions of the construction industry, and on a larger scale, in other economic sectors.

7.3.2 Contributions of the research in practice

This research contributes to the industry in different ways. This research developed three easy-to-understand models that can be used by professionals in the building sector for estimating the technical, economic, and social reusability of the structural components effectively. The easy-to-understand predictive tools developed during this research have several advantages, as follows.

- They can be used by any practitioner in the building sector, and they do not need a machine learning background.
- They give a first-hand idea about the feasibility of reusing a structural component from technical, economic, and social dimensions by collecting the necessary data.
- They have the potential to promoting reuse by increasing the reuse rate, which, in turn, can accelerate the growth of reuse markets.

Considering the UK economy post-Brexit and the impact of the COVID-19 outbreak on the employment rate, the results of this project can provide new job opportunities in the building sector in the UK.

7.4 Limitations of the research

In contrast to the mentioned contributions, this study has some limitations. The most important constraint in this research is the low rate of reuse in the building sector that restricts access to more experts with such experience. Likewise, while the researcher tried to decrease error by employing a wide range of machine learning methods, there still might be some errors due to a missing key factor that has not been integrated into the questionnaire.

Moreover, this research limits itself to the reuse of load-bearing building components in the superstructure of buildings; hence, the findings may not be generalised to the substructure of buildings. Also, this research is limited to the building sector, and the findings should not be expanded to other sub-sectors of the construction industry. Besides, while the questionnaire was not limited to any material, the responses provided were restricted to timber, steel, and concrete. Hence, the developed predictive tools in Chapter 6 can only be used to determine the reusability of timber, steel, and concrete load-bearing building components. Furthermore, as discussed in Section 5.4.4, the results of this research would be limited to load-bearing building components recovered using deconstruction technique or its variations such as component-specific recovery.

7.5 Future research

The resulting outcome of this research is three easy-to-understand predictive tools that can estimate the technical, economic, and social reusability of the load-bearing building components. While these tools are developed based on real reused components, because of the time constraints, it was not possible to employ them in real projects. Therefore, one possible future research would be utilising these tools to predict the reusability of the structural elements in relevant buildings' case studies and evaluate the impacts of this research on promoting reuse in the building sector.

Another potential future research would be using the developed learners in case study buildings with different structural materials and comparing their effectiveness in correctly classifying the reusable and non-reusable components based on their material (i.e., steel, timber, and concrete). Since the embodied energy and CO_2 of construction of similar structural elements with different materials are not equal, the learners' accuracy could be associated with the amount of CO_2 saved as the result of reusing the structural element. This way, a new metric (accuracy plus the percentage of saved embodied CO_2 of construction) could be developed to give a broader indication of the tools' effectiveness.

As discussed earlier, this research is limited to the superstructure of buildings. Therefore, it is advised to perform such investigation in other sub-divisions of the construction industry, such as foundations, roads, bridges, and infrastructures. While this research is limited to the building sector, the researcher strongly believes that similar studies can be performed in other sub-divisions of the construction industry to develop tools that can assess the reusability of the structures.

This research is focused on the technical, economic, and social reusability of the load-bearing building components. However, as discussed in Chapter 2, the reusability factors extend to a broader domain including, the environment, organisations, and regulations. As observed in this research, the relationship between variables is non-linear, which requires advanced tools to analyse the reusability factors under these domains. Therefore, one other potential future research is using the developed methodology in this research to identify the key factors affecting the reuse of load-bearing building components from organisational and regulatory perspectives. Such investigations would have considerable impacts on the existing policies and promote the circular economy in all aspects of the construction industry.

7.6 Chapter summary

Determining the reusability of load-bearing building components has introduced a new paradigm in the field of reuse. The focus of the existing body of knowledge in recent years has been limited to estimating the physical properties of recovered building structural elements to evaluate their reusability. However, these studies are not comprehensive because they only consider physical properties and ignore the impact of the multitude of variables including, economic and social factors affecting the reusability of these elements. Moreover, the few studies that have tried to consider the impact of other variables are too simplistic, consider a linear relationship between variables, are not based on real reuse projects, and are restricted to a very particular type of building and material.

This research performed a systematic literature review to identify the factors affecting reuse. Then, it developed an online questionnaire to quantify the reusability factors based on the experts' opinions. Next, this research used the results of the survey and showed the effectiveness of employing advanced supervised machine learning techniques such as random forests, K-Nearest Neighbours algorithm, Gaussian processes, support vector machines, adaptive boosting, BART machine, etc., in determining the reusability of the load-bearing building components. The results of this research revealed that the relationship between variables is far from being linear, which is evident by reviewing the performance of linear regression (LR), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and decision trees (DT) models. Moreover, this research revealed that not all variables play a significant role in the reusability of the load-bearing building components. It should be noted that this research is the first study to use advanced feature selection techniques to identify the most important variables affecting the reuse of structural elements in the building sector.

The results of sensitivity analysis and visualisation techniques to open the RF BSE-RPMs showed that design-related variables are having the highest impact on the technical reusability of the building components. Moreover, they showed that cost-related barriers have a significant effect on economic reusability. They eventually revealed that perception plays a significant role in the success of a project that intends to integrate recovered structural components, regardless of being technically and economically reusable.

This research took a further step and, for the first time, developed a series of tools that can be used by building experts to evaluate if a structural element is reusable from technical, economic, and social perspectives. While these tools perform effectively on the unseen observations, it is essential to utilise them in real case-study construction projects to evaluate their accuracy in an attempt to fine-tune them as future research work.

This research concludes that the complex interdependencies of factors affecting reuse cause a high level of uncertainty about the feasibility of reusing load-bearing building structural components, which hampers the widespread adoption of reuse. Notwithstanding, this research unveils that by using the probability theory foundations and combining it with advanced supervised machine learning methods, it is possible to develop tools that could reliably estimate the reusability of these elements based on affecting variables.

References

Abergel, T., Dean, B., and Dulac, J. (2017) *Towards a Zero-Emission, Efficient, and Resilient Buildings and Construction Sector, Global Status Report* [online] available from <https://www.worldgbc.org/sites/default/files/UNEP 188_GABC_en %28web%29.pdf>

Adams, J., Hillier-Brown, F.C., Moore, H.J., Lake, A.A., Araujo-Soares, V., White, M., and Summerbell, C. (2016) 'Searching and Synthesising "grey Literature" and "Grey Information" in Public Health: Critical Reflections on Three Case Studies'. *Systematic Reviews* 5 (1)

Addelman, S. (1970) 'Variability of Treatments and Experimental Units in the Design and Analysis of Experiments'. *Journal of the American Statistical Association* 65 (331), 1095–1108

Addis, B. (2006) *Building with Reclaimed Components and Materials, A Design Handbook for Reuse and Recycling*. UK: Earthscan

Agrawal, K., Baweja, Y., Dwivedi, D., Saha, R., Prasad, P., Agrawal, S., Kapoor, S., Chaturvedi, P., Mali, N., Kala, V.U., and Dutt, V. (2018) *A Comparison of Class Imbalance Techniques for Real-World Landslide Predictions*. held 2018. 1–8

Ajayi, S.O., Oyedele, L.O., Bilal, M., Akinade, O.O., Alaka, H.A., and Owolabi, H.A. (2017) 'Critical Management Practices Influencing On-Site Waste Minimization in Construction Projects'. *Waste Management* 59, 330–339

Ajayi, S.O., Oyedele, L.O., Bilal, M., Akinade, O.O., Alaka, H.A., Owolabi, H.A., and Kadiri, K.O. (2015) 'Waste Effectiveness of the Construction Industry: Understanding the Impediments and Requisites for Improvements'. *Resources, Conservation and Recycling* 102, 101–112

Ajayi, S.O., Oyedele, L.O., Kadiri, K.O., Akinade, O.O., Bilal, M., Owolabi, H.A., and Alaka, H.A. (2016) 'Competency-Based Measures for Designing out Construction Waste: Task and Contextual Attributes'. *Engineering, Construction and Architectural Management* 23 (4), 464–490

Akinade, O.O., Oyedele, L.O., Ajayi, S.O., Bilal, M., Alaka, H.A., Owolabi, H.A., Bello, S.A., Jaiyeoba, B.E., and Kadiri, K.O. (2017) 'Design for Deconstruction (DfD): Critical Success Factors for Diverting End-of-Life Waste from Landfills'. *Waste Management* 60, 3–13

Akinade, O.O., Oyedele, L.O., Munir, K., Bilal, M., Ajayi, S.O., Owolabi, H.A., Alaka, H.A., and Bello, S.A. (2016) 'Evaluation Criteria for Construction Waste Management Tools: Towards a Holistic BIM Framework'. *International Journal of Sustainable Building Technology and Urban Development* 7 (1), 3–21

Al-Bahrani, R., Agrawal, A., and Choudhary, A. (2013) *Colon Cancer Survival Prediction Using Ensemble Data Mining on SEER Data*. held 2013. 9–16

Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Ajayi, S.O., Bilal, M., and Akinade, O.O. (2016) 'Methodological Approach of Construction Business Failure Prediction Studies: A Review'. *Construction Management and Economics* 34 (11), 808–842

Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O., and Bilal, M. (2018) 'Systematic Review of Bankruptcy Prediction Models: Towards a Framework for Tool Selection'. *Expert Systems with Applications* 94, 164–184

de Alencar, M.S. and de Alencar, R.T. (2016) *Probability Theory*. New York: Momentum Press

Allison, P.D. (2001) *Missing Data*. 1st edn. Sage Publications, Inc

Anink, D., Boonstra, C., Mak, J., and Morris, A. (1996) *Handbook of Sustainable Building : An Environmental Preference Method for Selection of Materials in Construction and Refurbishment*. London: London : James & James Science

Arif, M., Bendi, D., Toma-Sabbagh, T., and Sutrisna, M. (2012) 'Construction Waste Management in India: An Exploratory Study'. *Construction Innovation* 12 (2), 133–155

Atlas.ti (2014) *DATA ANALYSIS AND THEORY-BUILDING TOOLS* [online] available from <https://atlasti.com/wp-content/uploads/2014/05/atlas-ti-7-co-occurrence-tools1.pdf> [26 August 2018]

Audigier, V., Husson, F., and Josse, J. (2017) 'MIMCA: Multiple Imputation for Categorical Variables with Multiple Correspondence Analysis'. *Statistics and Computing* 27 (2), 501–518

Aye, L., Ngo, T., Crawford, R.H., Gammampila, R., and Mendis, P. (2012) 'Life Cycle Greenhouse Gas Emissions and Energy Analysis of Prefabricated Reusable Building Modules'. *Energy and Buildings* 47, 159–168

Barbosa, F., Woetzel, J., Mischke, J., Ribeirinho, M.J., Sridhar, M., Parsons, M., Bertram, N., and Brown, S. (2017) *Reinventing Construction: A Route to Higher Productivity* [online] available from <https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/reinventing-construction-through-a-productivity-revolution#>

Baruch, Y. and Holtom, B.C. (2008) 'Survey Response Rate Levels and Trends in Organizational Research'. *Human Relations* 61 (8), 1139–1160

Basta, A., Serror, M.H., and Marzouk, M. (2020) 'A BIM-Based Framework for Quantitative Assessment of Steel Structure Deconstructability'. *Automation in Construction* 111, 189–200

BCIS (2012) *Elemental Standard Form of Cost Analysis: Principles, Instructions, Elements and Definitions*. 4th edn. London: The Building Cost Information Service (BCIS)

Bell, J. and Waters, S. (2014) *Doing Your Research Project*. 6th edn. Maidenhead: Open University Press

Benyus, J.M. (1997) *Biomimicry : Innovation Inspired by Nature*. New York: New York : Morrow

Bettany-Saltikov, J. (2016) *How to Do a Systematic Literature Review in Nursing : A Step-by-Step Guide.* Second edi.

Bhushan, B. (2009) 'Biomimetics: Lessons from Nature - an Overview'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1893), 1445–1486

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., and Jones, Z.M. (2016) 'Mlr: Machine Learning in R'. *Journal of Machine Learning Research* 17

Boulding, K.E. (2013) 'The Economics of the Coming Spaceship Earth'. in *Environmental Quality in A Growing Economy: Essays from the Sixth RFF Forum*. vol. 3. 1–20

Boyd, S., Stevenson, C., and Augenbraun, J.J. (2012) 'Deconstructing Deconstruction: Is a Ton of Material Worth a Ton of Work?' *Sustainability* 5 (6), 391–400

Brace, I. (2013) *Questionnaire Design : How to Plan, Structure and Write Survey Material for Effective Market Research*. 3rd editio. ed. by Society, M.R. London: London : KoganPage

Brambilla, G., Lavagna, M., Vasdravellis, G., and Castiglioni, C.A. (2019) 'Environmental Benefits Arising from Demountable Steel-Concrete Composite Floor Systems in Buildings'. *Resources, Conservation and Recycling* 141, 133–142

Breiman, L. (2001) 'Random Forests'. *Machine Learning* 45 (1), 5–32

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (2017) *Classification and Regression Trees*.

Brewer, G. and Mooney, J. (2008) 'A Best Practice Policy for Recycling and Reuse in Building'. *Proceedings of the Institution of Civil Engineers: Engineering Sustainability* 161 (3), 173–180

Brütting, J., Desruelle, J., Senatore, G., and Fivet, C. (2019) 'Design of Truss Structures Through Reuse'. *Structures* 18, 128–137

Bryman, A. and Cramer, D. (2005) *Quantitative Data Analysis for SPSS 12 and 13: A Guide for Social Scientists*. 1st edn. London: Routledge

Bullen, P.A. (2007) 'Adaptive Reuse and Sustainability of Commercial Buildings'. *Facilities* [online] 25 (1–2), 20–31. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33846665856&doi=10.1108%2F02632770710716911&partnerID=40&md5=e7447bad58a0088a7d91ec7ed37da6db>

Burrell, G. and Morgan, G. (2016) *Sociological Paradigms and Organisational Analysis : Elements of the Sociology of Corporate Life*. Routledge

van Buuren, S. and Groothuis-Oudshoorn, K. (2011) 'Mice: Multivariate Imputation by Chained Equations in R'. *Journal of Statistical Software* [online] 45 (3), 1–67. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-79953732420&partnerID=40&md5=2ca1881ca1e36a4bf6093577312336d7>

Cash, P. editor, Stanković, T. editor, and Štorga, M. editor (2016) *Experimental Design Research Approaches, Perspectives, Applications*. 1st ed. 20.

Cavalli, A., Bevilacqua, L., Capecchi, G., Cibecchini, D., Fioravanti, M., Goli, G., Togni, M., and Uzielli, L. (2016) 'MOE and MOR Assessment of in Service and Dismantled Old Structural Timber'. *Engineering Structures* [online] 125, 294–299. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84990060548&doi=10.1016%2Fj.engstruct.2016.06.054&partnerID=40&md5=a04b52ee00d6d9320c907f596489fa32>

Charef, R., Alaka, H., and Emmitt, S. (2018) 'Beyond the Third Dimension of BIM: A Systematic Review of Literature and Assessment of Professional Views'. *Journal of Building Engineering* 19, 242–257

Chau, C.K., Hui, W.K., Ng, W.Y., and Powell, G. (2012) 'Assessment of CO 2 Emissions Reduction in High-Rise Concrete Office Buildings Using Different Material Use Options'. *Resources, Conservation and Recycling* 61, 22–34

Chau, C.K., Xu, J.M., Leung, T.M., and Ng, W.Y. (2017) 'Evaluation of the Impacts of End-of-Life Management Strategies for Deconstruction of a High-Rise Concrete Framed Office Building'. *Applied Energy* 185, 1595–1603

Chawla, N. V, Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002) ‘SMOTE: Synthetic Minority over-Sampling Technique’. *Journal of Artificial Intelligence Research* 16, 321–357

Check, J. and Schutt, R.K. (2011) *Research Methods in Education*. 1st edn. SAGE Publishing

Chen, C., Habert, G., Bouzidi, Y., and Jullien, A. (2010) ‘Environmental Impact of Cement Production: Detail of the Different Processes and Cement Plant Variability Evaluation’. *Journal of Cleaner Production* 18 (5), 478–485

Cheng, J.C.P. and Ma, L.Y.H. (2013) ‘A BIM-Based System for Demolition and Renovation Waste Estimation and Planning’. *Waste Management* 33 (6), 1539–1551

Chileshe, N., Rameezdeen, R., and Hosseini, M.R. (2016) ‘Drivers for Adopting Reverse Logistics in the Construction Industry: A Qualitative Study’. *Engineering, Construction and Architectural Management* 23 (2), 134–157

Chileshe, N., Rameezdeen, R., and Hosseini, M.R. (2015) ‘Barriers to Implementing Reverse Logistics in South Australian Construction Organisations’. *Supply Chain Management* 20 (2), 179–204

Chileshe, N., Rameezdeen, R., Hosseini, M.R., Lehmann, S., and Udeaja, C. (2016) ‘Analysis of Reverse Logistics Implementation Practices by South Australian Construction Organisations’. *International Journal of Operations and Production Management* 36 (3), 332–356

Chileshe, N., Rameezdeen, R., Hosseini, M.R., Martek, I., Li, H.X., and Panjehbashi-Aghdam, P. (2018) ‘Factors Driving the Implementation of Reverse Logistics: A Quantified Model for the Construction Industry’. *Waste Management* 79, 48–57

Chilisa, B. and Kawulich, B. (2012) ‘Selecting a Research Approach: Paradigm, Methodology, and Methods’. *Doing Social Research: A Global Context* 51–61

Chinda, T. and Ammarapala, V. (2016) ‘Decision-Making on Reverse Logistics in the Construction Industry’. *Songklanakarin Journal of Science and Technology* 38 (1), 7–14

Chini, A.R. and Acquaye, L. (2001) ‘Grading and Strength of Salvaged Lumber from Residential Buildings’. *Environmental Practice* 3 (4), 247–256

Chipman, H.A., George, E.I., and McCulloch, R.E. (2010) ‘BART: Bayesian Additive Regression Trees’. *Annals of Applied Statistics* 4 (1), 266–298

Chung, N.C., Miasojedow, B., Startek, M., and Gambin, A. (2018) *Jaccard: Test Similarity Between Binary Data Using Jaccard/Tanimoto Coefficients. R Package Version 0.1.0.* available from <https://cran.r-project.org/package=jaccard>

Ciaburro, G. and Venkateswaran, B. (2017) *Neural Networks with R*. 1st edn. Birmingham: Packt Publishing Ltd.

Clark, C., Jambeck, J., and Townsend, T. (2006) ‘A Review of Construction and Demolition Debris Regulations in the United States’. *Critical Reviews in Environmental Science and Technology* 36 (2), 141–186

Coelho, A., de Brito, J., and Brito, J. De (2012) ‘Influence of Construction and Demolition Waste Management on the Environmental Impact of Buildings’. *Waste Management (New York, N.Y.)* 32 (3), 532–541

Cohen, L., Manion, L., and Morrison, K. (2018) *Research Methods in Education*. Eighth edi. ed.

by Morrison, K. (Keith R.B.. author, Manion author, L., and Morrison author, K. (Keith R.B.. Routledge

Cohen, W.W. (1995) *Fast Effective Rule Induction*. ed. by Prieditis, A. and Russell, S.B.T.-M.L.P. 1995. San Francisco (CA): Morgan Kaufmann, 115–123

Collier, A. (1994) *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. 1st editio. London: Verso

Cooper, S., Skelton, A.C.H., Owen, A., Densley-Tingley, D., and Allwood, J.M. (2016) 'A Multi-Method Approach for Analysing the Potential Employment Impacts of Material Efficiency'. *Resources, Conservation and Recycling* 109, 54–66

Cortez, P. (2020) 'Rminer: Data Mining Classification and Regression Methods'. *R Package Version 1.4.5*

Cortez, P. and Embrechts, M.J. (2013) 'Using Sensitivity Analysis and Visualization Techniques to Open Black Box Data Mining Models'. *Information Sciences* 225, 1–17

Cover, T.M. and Thomas, J.A. (2005) *Elements of Information Theory*.

Creswell, J.W. and Creswell, J.D. (2018) *Research Design : Qualitative, Quantitative & Mixed Methods Approaches*. 5th editio. ed. by Creswell author, J.D. SAGE Publications Inc.

Crotty, M. (1998) *The Foundations of Social Research : Meaning and Perspective in the Research Process.* London: London : SAGE

Daneshkhah, A. and Bedford, T. (2008) 'Sensitivity Analysis of a Reliability System Using Gaussian Processes'. in *Advances in Mathematical Modeling for Reliability*. 46–62

Daneshkhah, A., Hosseinian-Far, A., and Chatrabgoun, O. (2017) 'Sustainable Maintenance Strategy under Uncertainty in the Lifetime Distribution of Deteriorating Assets'. in *Strategic Engineering for Cloud Computing and Big Data Analytics*. 29–50

Dantata, N., Touran, A., and Wang, J. (2005) 'An Analysis of Cost and Duration for Deconstruction and Demolition of Residential Buildings in Massachusetts'. *Resources, Conservation and Recycling* 44 (1), 1–15

Decarlo, M. (2018) *Scientific Inquiry in Social Work* [online] ROANOKE, VA: OPEN SOCIAL WORK EDUCATION. available from <https://scientificinquiryinsocialwork.pressbooks.com/front-matter/welcome/>

Defra (2019) *UK Statistics on Waste - February 2019 Update* [online] available from <https://www.gov.uk/government/statistics/uk-waste-data> [7 May 2019]

Defra (2007) *Waste Strategy for England 2007*. The Stationery Office, 123. available from <https://www.gov.uk/government/publications/waste-strategy-for-england-2007> [29 July 2018]

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977) 'Maximum Likelihood from Incomplete Data via the EM Algorithm'. *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1), 1–38

Deng, H. (2014) 'Interpreting Tree Ensembles with InTrees'. *ArXiv:1408.5456*

Denhart, H. (2010) 'Deconstructing Disaster: Economic and Environmental Impacts of Deconstruction in Post-Katrina New Orleans'. *Resources, Conservation and Recycling* 54 (3), 194–204

Densley Tingley, D., Davison, B., Tingley, D.D., and Davison, B. (2012) 'Developing an LCA Methodology to Account for the Environmental Benefits of Design for Deconstruction'. *Building and Environment* 57, 387–395

Denyer, D. and Tranfield, D. (2009) 'Producing a Systematic Review'. *The Sage Handbook of Organizational Research Methods* 671–689

Denzin editor, N.K. and Lincoln editor, Y.S. (2018) *The SAGE Handbook of Qualitative Research*. Fifth edit. Los Angeles: Los Angeles : SAGE

Designing Buildings Wiki (2021) *Global Construction Market Projections from 2020 to 2030* [online] available from <https://www.designingbuildings.co.uk/wiki/Global_construction_market_projections_from_2020_to_2030> [27 May 2021]

Dewulf, J., Van der Vorst, G., Versele, N., Janssens, A., and Van Langenhove, H. (2009) 'Quantification of the Impact of the End-of-Life Scenario on the Overall Resource Consumption for a Dwelling House'. *Resources, Conservation and Recycling* 53 (4), 231–236

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. (2019) 'E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien'. *R Package Version 1.7-3.*

Ding, C. and Peng, H. (2003) *Minimum Redundancy Feature Selection from Microarray Gene Expression Data*. held 2003. 523–528

Ding, Z., Yi, G., Tam, V.W.Y., and Huang, T. (2016) 'A System Dynamics-Based Environmental Performance Simulation of Construction Waste Reduction Management in China'. *Waste Management* 51, 130–141

Diyamandoglu, V. and Fortuna, L.M. (2015) 'Deconstruction of Wood-Framed Houses: Material Recovery and Environmental Impact'. *Resources, Conservation and Recycling* 100, 21–30

Duffy, J. (2015) *LinkedIn* [online] available from <https://uk.pcmag.com/internet-3/8297/linkedin> [8 August 2019]

Dunant, C.F., Drewniok, M.P., Sansom, M., Corbey, S., Allwood, J.M., and Cullen, J.M. (2017) 'Real and Perceived Barriers to Steel Reuse across the UK Construction Value Chain'. *Resources, Conservation and Recycling* 126, 118–131

Dunant, C.F., Drewniok, M.P., Sansom, M., Corbey, S., Cullen, J.M., and Allwood, J.M. (2018) 'Options to Make Steel Reuse Profitable: An Analysis of Cost and Risk Distribution across the UK Construction Value Chain'. *Journal of Cleaner Production* 183, 102–111

Dunant, C.F., Skelton, A.C.H., Drewniok, M.P., Cullen, J.M., and Allwood, J.M. (2019) 'A Marginal Abatement Cost Curve for Material Efficiency Accounting for Uncertainty'. *Resources, Conservation and Recycling* 144, 39–47

Durão, V., Caixinhas, J., Osório-Peters, S., Den Boer, E., Williams, I.D., Curran, T., and Pertl, A. (2014) 'Zero-Waste Networks in Construction and Demolition in Portugal'. *Proceedings of Institution of Civil Engineers: Waste and Resource Management* 167 (4), 153–168

Dusek, G.A., Yurova, Y. V, and Ruppel, C.P. (2015) 'Using Social Media and Targeted Snowball Sampling to Survey a Hard-to-Reach Population: A Case Study'. *International Journal of Doctoral Studies* 10, 279–299

Eberhardt, L.C.M., Birgisdóttir, H., and Birkved, M. (2019) 'Life Cycle Assessment of a Danish Office Building Designed for Disassembly'. *Building Research and Information* 47 (6), 666–680

Eck, N.J. van and Waltman, L. (2009) 'How to Normalize Cooccurrence Data? An Analysis of Some Well-Known Similarity Measures'. *Journal of the American Society for Information Science and Technology* 60 (8), 1635–1651

Elias-Ozkan, S.T. (2014) 'Demolition versus Deconstruction: Impacts of Fenestration Disposal in Building Renovation Projects'. *Open House International* 39 (1), 48–55

Elias Özkan, S.T. (2012) 'Selective Demolition of Redundant and Earthquake Damaged Buildings in Turkey'. *Metu Journal of the Faculty of Architecture* 29 (1), 139–152

Ellen MacArthur Foundation (2015) *Delivering the Circular Economy: A Toolkit for Policymakers* [online] available from <https://www.ellenmacarthurfoundation.org/assets/downloads/publications/EllenMacArthurFoundation_PolicymakerToolkit.pdf>

Ellen MacArthur Foundation (2013) *Towards the Circular Economy: Economic and Business Rationale for an Accelerated Transition* [online] available from <https://www.ellenmacarthurfoundation.org/assets/downloads/publications/Ellen-MacArthur-Foundation-Towards-the-Circular-Economy-vol.1.pdf> [20 August 2018]

Ellen MacArthur Foundation (n.d.) *What Is the Circular Economy?* [online] available from <https://www.ellenmacarthurfoundation.org/circular-economy/what-is-the-circular-economy> [28 July 2020]

European Union (2008) *DIRECTIVE 2008/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL* [online] available from <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1521196549736&uri=CELEX:32008L0098> [16 March 2018]

Eurostat (2020) *Energy, Transport and Environment Statistics — 2020 Edition* [online] available from <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-DK-20-001>

Eurostat (2016) *Waste Statistics - Statistics Explained* [online] available from <http://ec.europa.eu/eurostat/statistics-explained/index.php/Waste_statistics> [9 September 2020]

Faleschini, F., Zanini, M.A., Hofer, L., Zampieri, P., and Pellegrino, C. (2017) 'Sustainable Management of Demolition Waste in Post-Quake Recovery Processes: The Italian Experience'. *International Journal of Disaster Risk Reduction* 24, 172–182

Farsi, M., Daneshkhah, A., Hosseinian-Far, A., and Jahankhani, H. (eds.) (2020) *Digital Twin Technologies and Smart Cities*. 1st ed. 20.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., and Herrera, F. (2018) *Learning from Imbalanced Data Sets*. 1st edn. Cham, Switzerland: Springer Nature Switzerland AG

Ferreira, J., Duarte Pinheiro, M., and De Brito, J. (2015) 'Economic and Environmental Savings of Structural Buildings Refurbishment with Demolition and Reconstruction - A Portuguese Benchmarking'. *Journal of Building Engineering* 3, 114–126

Field, A.P. (2009) *Discovering Statistics Using SPSS : (And Sex and Drugs and Rock 'n' Roll)*. 3rd ed.. Los Angeles: Los Angeles

Forsythe, P. (2011) 'Drivers of Housing Demolition Decision Making and the Impact on Timber Waste Management'. *Australasian Journal of Construction Economics and Building* 11

(1), 1–14

Fowler, F.J. (2014) *Survey Research Methods*. 5th edn. Thousand Oaks, Calif.: SAGE Publications Inc.

Freund, Y. and Schapire, R.E. (1997) 'A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting'. *Journal of Computer and System Sciences* 55 (1), 119–139

Frisch, S.A., Harris, D.B., Kelly, S.Q., and Parker, D.C.W. (eds.) (2012) *Doing Archival Research in Political Science*. New York: Cambria Press

Frosch, R.A. and Gallopoulos, N.E. (1989) 'Strategies for Manufacturing'. *Scientific American* 261 (3), 144–152

Fujita, M. and Iwata, M. (2008) 'Reuse System of Building Steel Structures'. *Structure and Infrastructure Engineering* 4 (3), 207–220

Fujita, M. and Kuki, K. (2016) 'An Evaluation of Mechanical Properties with the Hardness of Building Steel Structural Members for Reuse by NDT'. *Metals* 6 (10)

Fujita, M. and Masuda, T. (2014) 'Application of Various NDT Methods for the Evaluation of Building Steel Structures for Reuse'. in *Materials*. vol. 7 (10)

Fürnkranz, J., Gamberger, D., and Lavrač, N. (2012) *Foundations of Rule Learning*. 1st edn. Springer-Verlag Berlin Heidelberg

Gálvez-Martos, J.-L., Styles, D., Schoenberger, H., and Zeschmar-Lahl, B. (2018) 'Construction and Demolition Waste Best Management Practice in Europe'. *Resources, Conservation and Recycling* 136, 166–178

Gangolells, M., Casals, M., Forcada, N., and Macarulla, M. (2014) 'Analysis of the Implementation of Effective Waste Management Practices in Construction Projects and Sites'. *Resources, Conservation and Recycling* 93, 99–111

Geman, S., Bienenstock, E., and Doursat, R. (1992) 'Neural Networks and the Bias/Variance Dilemma'. *Neural Computation* 4 (1), 1–58

GET-IT Glossary (n.d.) *Systematic Review* [online] available from <http://getitglossary.org/term/systematic+review> [31 July 2018]

Geyer, R., Jackson, T., and Clift, R. (2002) 'Economic and Environmental Comparison between Recycling and Reuse of Structural Steel Sections'. in *36th International Iron and Steel Institute World Conference*. held 2002 at Rome, Italy. 13–18

Given, L.M. (ed.) (2008) *The Sage Encyclopedia of Qualitative Research Methods*. Los Angeles: SAGE Publications Inc.

Gorgolewski, M. (2008) 'Designing with Reused Building Components: Some Challenges'. *Building Research & Information* 36 (2), 175–188

Gorgolewski, M., Straka, V., Edmonds, J., and Sergio-Dzoutzidis, C. (2008) 'Designing Buildings Using Reclaimed Steel Components'. *Journal of Green Building* 3 (3), 97–107

Gottsche, J. and Kelly, M. (2018) 'Assessing the Impact of Construction Waste Reduction on Selected Projects in Ireland'. *Proceedings of Institution of Civil Engineers: Waste and Resource Management* 171 (3), 71–81

Government Construction Strategy (2011) available from

<https://www.gov.uk/government/publications/government-construction-strategy>

Graham, J.W. (2012) *Missing Data: Analysis and Design*. 1st edn. Springer-Verlag New York

Graham, J.W. (2009) 'Missing Data Analysis: Making It Work in the Real World'. in *Annual Review of Psychology*. vol. 60. 549–576

Graham, J.W. and Schafer, J.L. (1999) 'On the Performance of Multiple Imputation for Multivariate Data with Small Sample Size'. in *Statistical Strategies for Small Sample Research*. Thousand Oaks, Calif. : SAGE Publications, 1–29

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018) 'A Survey of Methods for Explaining Black Box Models'. *ACM Computing Surveys* 51 (5)

Guo, H. (2016) 'A System Dynamics Model for Construction Waste Resource Recovery Management in China'. *Revista de La Facultad de Ingeniería* 31 (6), 186–203

Guo, S., Zheng, S., Hu, Y., Hong, J., Wu, X., and Tang, M. (2019) 'Embodied Energy Use in the Global Construction Industry'. *Applied Energy* 256

Guy, B. (2006) 'The Optimization of Building Deconstruction for Department of Defense Facilities: FT. Mcclellan Deconstruction Project'. *Journal of Green Building* 1 (1), 102–122

Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L.A. (eds.) (2006) *Feature Extraction Foundations and Applications*. 1st ed. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg

Hammond, G.P. and Jones, C.I. (2008) 'Embodied Energy and Carbon in Construction Materials'. *Proceedings of Institution of Civil Engineers: Energy* [online] 161 (2), 87–98. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-58149144250&doi=10.1680%2Fener.2008.161.2.87&partnerID=40&md5=64dad57d99704a91b0133555e0a42f20>

Harzing, A.W. (1997) 'Response Rates in International Mail Surveys: Results of a 22-Country Study'. *International Business Review* 6 (6), 641–665

Hastie, T., Tibshirani, R., and Friedman, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edn. New York: Springer

He, H. and Ma, Y. (eds.) (2013) *Imbalanced Learning: Foundations, Algorithms, and Applications*. Hoboken, New Jersey.: John Wiley & Sons, Inc.

Healy, M. and Perry, C. (2000) 'Comprehensive Criteria to Judge Validity and Reliability of Qualitative Research within the Realism Paradigm'. *Qualitative Market Research: An International Journal* 3 (3), 118–126

Hechenbichler, K. and Schliep, K. (2004) 'Weighted K-Nearest-Neighbor Techniques and Ordinal Classification'. *LMU*

Herczeg, M., McKinnon, D., Milios, L., Bakas, I., Klaassens, E., Svatikova, K., and Wilderberg, O. (2014) *Resource Efficiency in the Building Sector* [online] available from <http://ec.europa.eu/environment/eussd/pdf/Resource efficiency in the building sector.pdf>

Heron, J. (1996) *Co-Operative Inquiry Research into the Human Condition*. London: London

Hglmeier, K., Weber-Blaschke, G., Richter, K., Höglmeier, K., Weber-Blaschke, G., and Richter, K. (2013) 'Potentials for Cascading of Recovered Wood from Building Deconstruction - A

Case Study for South-East Germany'. *Resources, Conservation and Recycling* 78, 81–91

Holdnack, J.A., Millis, S., Larrabee, G.J., and Iverson, G.L. (2013) 'Assessing Performance Validity with the ACS'. in *WAIS-IV, WMS-IV, and ACS: Advanced Clinical Interpretation*. ed. by Holdnack, J.A., Drozdick, L.W., Weiss, L.G., and Iverson, G.L. San Diego: Academic Press, 331–365

Hoornweg, D. and Bhada-Tata, P. (2012) *What a Waste: A Global Review of Solid Waste Management* [online] available from <https://www.ifc.org/wps/wcm/connect/1e5ca7004c07698db58eb7d8bd2c3114/What-A-Waste-Report.pdf?MOD=AJPERES> [22 January 2019]

Hornik, K., Buchta, C., and Zeileis, A. (2009) 'Open-Source Machine Learning: R Meets Weka'. *Computational Statistics* 24 (2), 225–232

Hosseini, M.R., Rameezdeen, R., Chileshe, N., and Lehmann, S. (2015) 'Reverse Logistics in the Construction Industry'. *Waste Management and Research* 33 (6), 499–514

Hothorn, T., Hornik, K., and Zeileis, A. (2006) 'Party: A Laboratory for Recursive Partytioning'. *R Package Version 0.9-0, URL Http://CRAN. R-Project. Org*

Hradil, P., Fülöp, L., and Ungureanu, V. (2019) 'Reusability of Components from Single-Storey Steel-Framed Buildings'. *Steel Construction* 12 (2), 91–97

Hradil, P., Talja, A., Ungureanu, V., Koukkari, H., and Fülöp, L. (2017) 'Reusability Indicator for Steel-Framed Buildings and Application for an Industrial Hall'. *Special Issue: Proceedings of Eurosteel 2017* 1 (2–3), 4512–4521

Huang, B., Wang, X., Kua, H., Geng, Y., Bleischwitz, R., and Ren, J. (2018) 'Construction and Demolition Waste Management in China through the 3R Principle'. *Resources, Conservation and Recycling* 129, 36–44

Huuhka, S. and Hakanen, J.H. (2015) 'Potential and Barriers for Reusing Load-Bearing Building Components in Finland'. *International Journal for Housing Science and Its Applications* 39 (4), 215–224

Huuhka, S., Kaasalainen, T., Hakanen, J.H., and Lahdensivu, J. (2015) 'Reusing Concrete Panels from Buildings for Building: Potential in Finnish 1970s Mass Housing'. *Resources, Conservation and Recycling* 101, 105–121

Iacovidou, E. and Purnell, P. (2016) 'Mining the Physical Infrastructure: Opportunities, Barriers and Interventions in Promoting Structural Components Reuse'. *Science of the Total Environment* 557–558 (Supplement C), 791–807

ICE (2008) *Demolition Protocol* [online] available from <https://apps2.staffordshire.gov.uk/scc/TrimDocProvider/?ID=13/174>

Iguyon, I. and Elisseeff, A. (2003) 'An Introduction to Variable and Feature Selection'. *Journal of Machine Learning Research* 3, 1157–1182

International Energy Agency (2020) *Energy Technology Perspectives 2020* [online] available from <https://www.iea.org/reports/energy-technology-perspectives-2020>

International Finance Corporation (2018) *Construction Industry Value Chain: How Companies Are Using Carbon Pricing to Address Climate Risk and Find New Opportunities* [online] Washington, D.C. available from <www.ifc.org>

International Finance Corporation (2016) *Climate Investment Opportunities in Emerging*

Markets [online] Washington, D.C. available from <https://www.ifc.org/wps/wcm/connect/59260145-ec2e-40de-97e6-3aa78b82b3c9/3503-IFC-Climate_Investment_Opportunity-Report-Dec-FINAL.pdf?MOD=AJPERES&CVID=lBLd6Xq>

Jaillon, L. and Poon, C.-S. (2010) 'Design Issues of Using Prefabrication in Hong Kong Building Construction'. *Construction Management and Economics* 28 (10), 1025–1042

Jaillon, L. and Poon, C.S. (2014) 'Life Cycle Design and Prefabrication in Buildings: A Review and Case Studies in Hong Kong'. *Automation in Construction* 39, 195–202

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017) *An Introduction to Statistical Learning: With Applications in R*. Springer

Jamjoom, M. (2020) 'The Pertinent Single-Attribute-Based Classifier for Small Datasets Classification'. *International Journal of Electrical and Computer Engineering* 10 (3), 3227–3234

Jang, W., Lee, J.K., Lee, J., and Han, S.H. (2015) 'Naive Bayesian Classifier for Selecting Good/Bad Projects during the Early Stage of International Construction Bidding Decisions'. *Mathematical Problems in Engineering* 2015

Japkowicz, N. (2000) 'The Class Imbalance Problem: Significance and Strategies'. in *Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI'2000): Special Track on Inductive Learning*. held 2000 at Las Vegas, Nevada. 111–117

Jin, R., Li, B., Zhou, T., Wanatowski, D., and Piroozfar, P. (2017) 'An Empirical Study of Perceptions towards Construction and Demolition Waste Recycling and Reuse in China'. *Resources, Conservation and Recycling* 126, 86–98

Jisc (2019) *Online Surveys (Formerly BOS)* [online] available from <https://www.onlinesurveys.ac.uk/>

Johnson, P. and Clark, M. (2006) 'Mapping the Terrain: An Overview of Business and Management Research Methodologies'. *Business and Management Research Methodologies*

Johnson, T. and Owens, L. (2003) 'Survey Response Rate Reporting in the Professional Literature'. in *58th Annual Meeting of the American Association for Public Opinion Research*. held 2003. 127–133

Josse, J. and Husson, F. (2016) 'MissMDA: A Package for Handling Missing Values in Multivariate Data Analysis'. *Journal of Statistical Software* 70

Kaethner, S.C. and Burridge, J.A. (2012) 'Embodied CO2 of Structural Frames'. *Structural Engineer* 90 (5), 33–40

Kajikawa, Y., Tacoa, F., and Yamaguchi, K. (2014) 'Sustainability Science: The Changing Landscape of Sustainability Research'. *Sustainability Science* 9 (4), 431–438

Kalyun, M. and Wodajo, T. (2012) *Application of a Design Method for Manufacture and Assembly: Flexible Assembly Methods and Their Evaluation for the Construction of Bridges* [online] CHALMERS UNIVERSITY OF TECHNOLOGY. available from <http://publications.lib.chalmers.se/records/fulltext/164233.pdf>

Kang, H. (2013) 'The Prevention and Handling of the Missing Data'. *Korean Journal of Anesthesiology* 64 (5), 402–406

Kapelner, A. and Bleich, J. (2016) 'BartMachine: Machine Learning with Bayesian Additive Regression Trees'. *Journal of Statistical Software* 70

Keller, P., McConnell, J., Schumacher, T., and Thostenson, E.T. (2019) 'Construction Stress Monitoring Using a Wireless Sensor Network to Evaluate Reuse Potential of Structural Steel'. *Journal of Structural Engineering (United States)* 145 (12)

KHL Group (2018) 'Top 100 World Demolition Contractors'. *Demolition and Recycling International* [online] 20 (3). available from <https://www.khl.com/magazines/demolition-and-recycling-international/>

Kibert, C.J. and Chini, A.R. (eds.) (2000) 'Overview of Deconstruction in Selected Countries (CIB Report, Publication 252)'. in *Overview of Deconstruction in Selected Countries*. held 2000 at Rotterdam. CIB

Kira, K. and Rendell, L.A. (1992) 'A Practical Approach to Feature Selection'. in *Proceedings of the Ninth International Workshop on Machine Learning*. held 1992 at Aberdeen, Scotland. Morgan Kaufmann Publishers Inc., 249–256

Klang, A., Vikman, P.-A., and Brattebø, H. (2003) 'Sustainable Management of Demolition Waste - An Integrated Model for the Evaluation of Environmental, Economic and Social Aspects'. *Resources, Conservation and Recycling* 38 (4), 317–334

Kruskal, W.H. (1952) 'A Nonparametric Test for the Several Sample Problem'. *Annals of Mathematical Statistics* 23 (4), 525–540

Kruskal, W.H. and Wallis, W.A. (1952) 'Use of Ranks in One-Criterion Variance Analysis'. *Journal of the American Statistical Association* 47 (260), 583–621

Kuehlen, A., Thompson, N., and Schultmann, F. (2014) 'Barriers for Deconstruction and Reuse/Recycling of Construction Materials in Germany'. *Barriers for Deconstruction and Reuse/Recycling of Construction Materials* 38–52

Kuhn, M. (2008) 'Building Predictive Models in R Using the Caret Package'. *Journal of Statistical Software* 28 (5), 1–26

Kuhn, M. and Johnson, K. (2020) *Feature Engineering and Selection: A Practical Approach for Predictive Models*. 1st edn. New York: Chapman and Hall/CRC

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*. [2nd ed.,. Chicago: Chicago : University of Chicago Press

Kursa, M.B. and Rudnicki, W.R. (2010) 'Feature Selection with the Boruta Package'. *Journal of Statistical Software* 36 (11), 1–13

Labovitz, S. (1970) 'The Assignment of Numbers to Rank Order Categories'. *American Sociological Review* 35 (3), 515–524

Lachimpadi, S.K., Pereira, J.J., Taha, M.R., and Mokhtar, M. (2012) 'Construction Waste Minimisation Comparing Conventional and Precast Construction (Mixed System and IBS) Methods in High-Rise Buildings: A Malaysia Case Study'. *Resources, Conservation and Recycling* 68, 96–103

Lavrakas, P.J. (ed.) (2008) *Encyclopedia of Survey Research Methods*. Thousand Oaks, Calif.: SAGE Publications Inc.

Lazarus, N. (2003) *Beddington Zero (Fossil) Energy Development Construction Materials Report Toolkit for Carbon Neutral Developments-Part 1* [online] available from

<www.bioregional.com> [23 January 2019]

Leigh, N.G. and Patterson, L.M. (2006) 'Deconstructing to Redevelop: A Sustainable Alternative to Mechanical Demolition'. *Journal of the American Planning Association* [online] 72 (2), 217–225. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33745393101&doi=10.1080%2F01944360608976740&partnerID=40&md5=92548c28cc2f06aaa4ad3489f9e2417f>

Li, H., Chen, Z., and Wong, C.T.C. (2003) 'Barcode Technology for an Incentive Reward Program to Reduce Construction Wastes'. *Computer-Aided Civil and Infrastructure Engineering* 18 (4), 313–324

Liaw, A. and Wiener, M. (2002) 'Classification and Regression by RandomForest'. *R News* 2 (3), 18–22

Likert, R. (1932) 'A Technique for the Measurement of Attitudes'. *Archives of Psychology* 22 (140), 5–55

Ling, Y.Y. and Leo, K.C. (2000) 'Reusing Timber Formwork: Importance of Workmen's Efficiency and Attitude'. *Building and Environment* 35 (2), 135–143

LinkedIn (2019) *About LinkedIn* [online] available from <https://about.linkedin.com/> [8 August 2019]

Little, R.J.A. and Rubin, D.B. (2019) *Statistical Analysis with Missing Data*. 3rd edn. John Wiley & Sons

Lord, F.M. (1953) 'On the Statistical Treatment of Football Numbers'. *American Psychologist* 8 (12), 750–751

Machado, R.C., de Souza, H.A., and Veríssimo, G.S. (2018) 'Analysis of Guidelines and Identification of Characteristics Influencing the Deconstruction Potential of Buildings'. *Sustainability (Switzerland)* 10 (8)

MacKinnon, C.T. (2000) 'School Construction and Waste Diversion: An Interorganizational Perspective'. *Environment and Planning B: Planning and Design* 27 (6), 877–891

Mahpour, A. (2018) 'Prioritizing Barriers to Adopt Circular Economy in Construction and Demolition Waste Management'. *Resources, Conservation and Recycling* 134, 216–227

Mahpour, A. and Mortaheb, M.M. (2018) 'Financial-Based Incentive Plan to Reduce Construction Waste'. *Journal of Construction Engineering and Management* 144 (5), art. no. 04018029

McGill, W.J. (1954) 'Multivariate Information Transmission'. *Psychometrika* 19 (2), 97–116

Merriam-Webster Dictionary (n.d.) available from <https://www.merriam-webster.com/>

Migliore, M., Campioli, A., Lavagna, M., Oberti, I., Paganin, G., and Talamo, C. (2015) 'Intersectorial Reuse of Waste and Scraps for the Production of Building Products: Strategies and Valorization of Waste'. *Environmental Engineering and Management Journal* 14 (7), 1675–1681

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., and Group, T.P. (2009) 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement'. *PLOS Medicine* 6 (7), e1000097

Molnar, C. (2020) *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable* [online] Github. available from <https://christophm.github.io/interpretable-ml-book/>

Munroe, T., Hatamiya, L., and Westwind, M. (2006) 'Deconstruction of Structures: An Overview of Economic Issues'. *International Journal of Environmental Technology and Management* 6 (3–4), 374–385

Murphy, K.P. (2012) *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: Cambridge, MA : MIT Press

Naseriparsa, M. and Kashani, M.M.R. (2013) 'Combination of PCA with SMOTE Resampling to Boost the Prediction Rate in Lung Cancer Dataset'. *International Journal of Computer Applications* 77 (3), 33–38

Nembrini, S., König, I.R., and Wright, M.N. (2018) 'The Revival of the Gini Importance?' *Bioinformatics* 34 (21), 3711–3718

Nesselroade, J.R. and Cattell, R.B. (eds.) (1988) *Handbook of Multivariate Experimental Psychology*. 2nd edn. New York: Plenum Press

Nevzat, M.Z. and Atakara, C. (2015) 'A Contemporary Connection to Historic Buildings through Transparency and Reusability'. *Open House International* 40 (3), 52–60

Nordby, A.S., Berge, B., Hakonsen, F., and Hestnes, A.G. (2009) 'Criteria for Salvageability: The Reuse of Bricks'. *Building Research and Information* 37 (1), 55–67

Nußholz, J.L.K., Nygaard Rasmussen, F., and Milios, L. (2019) 'Circular Building Materials: Carbon Saving Potential and the Role of Business Model Innovation and Public Policy'. *Resources, Conservation and Recycling* 141, 308–316

OECD (2019) *Global Material Resources Outlook to 2060: Economic Drivers and Environmental Consequences*.

Oxford English Dictionary (n.d.) available from <https://www.oed.com/>

Parker, D. and Deegan, K. (2007) *An Analysis of the Spectrum of Re-Use A Component of the Remanufacturing Pilot for Defra, BREW Programme* [online] available from <http://www.remanufacturing.org.uk/pdf/story/1p297.pdf> [21 August 2018]

Patton, M.Q. (2002) *Qualitative Research & Evaluation Methods*. 3rd ed.. ed. by Patton, M.Q. Thousand Oaks, Calif. : Thousand Oaks, Calif.

Peters, A. and Hothorn, T. (2019) 'Ipred: Improved Predictors.' *R Package Version 0.9-9.* [online] available from <https://cran.r-project.org/package=ipred>

Petrosino, A. and Lavenberg, J. (2007) 'Systematic Reviews and Meta-Analyses: Best Evidence on "What Works" for Criminal Justice Decision Makers'. *Western Criminology Review* 8 (1), 1–15

Pomponi, F. and Moncaster, A. (2017) 'A Theoretical Framework for Circular Economy Research in the Built Environment'. in *Building Information Modelling, Building Performance, Design and Smart Construction*. ed. by Dastbaz, M., Gorse, C., and Moncaster, A. Springer International Publishing AG, 31–44

Pongiglione, M. and Calderini, C. (2014) 'Material Savings through Structural Steel Reuse: A Case Study in Genoa'. *Resources, Conservation and Recycling* 86, 87–92

Ponto, J. (2015) ‘Understanding and Evaluating Survey Research’. *Journal of the Advanced Practitioner in Oncology* 6 (2), 168–171

PRISMA (2018) *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)* [online] available from <http://prisma-statement.org/> [1 August 2018]

Pun, S.K. and Liu, C. (2006) ‘A Framework for Material Management in the Building Demolition Industry’. *Architectural Science Review* 49 (4), 391–398

Pun, S.K., Liu, C., and Langston, C. (2006) ‘Case Study of Demolition Costs of Residential Buildings’. *Construction Management and Economics* 24 (9), 967–976

R Core Team (2020) *R: A Language and Environment for Statistical Computing.* 4.0.2. R Foundation for Statistical Computing, Vienna, Austria. available from <https://www.r-project.org/>

Rakhshan, K., Morel, J.-C., Alaka, H., and Charef, R. (2020) ‘Components Reuse in the Building Sector – A Systematic Review’. *Waste Management & Research* [online] 38 (4), 347–370. available from <https://doi.org/10.1177/0734242X20910463>

Rameezdeen, R., Chileshe, N., Hosseini, M.R., and Lehmann, S. (2016) ‘A Qualitative Examination of Major Barriers in Implementation of Reverse Logistics within the South Australian Construction Sector’. *International Journal of Construction Management* 16 (3), 185–196

Rasmussen, C.E. and Williams, C.K.I. (2006) *Gaussian Processes for Machine Learning*. MIT Press

Reed, M. (2005) ‘Reflections on the “realist Turn” in Organization and Management Studies’. *Journal of Management Studies* 42 (8), 1621–1644

da Rocha, C.G. and Sattler, M.A. (2009) ‘A Discussion on the Reuse of Building Components in Brazil: An Analysis of Major Social, Economical and Legal Factors’. *Resources, Conservation and Recycling* 54 (2), 104–112

Rogers, S. (2011) ‘Battling Construction Waste and Winning: Lessons from Uae’. *Proceedings of the Institution of Civil Engineers: Civil Engineering* 164 (1), 41–48

Romanski, P. and Kotthoff, L. (2018) ‘FSelector: Selecting Attributes.’ *R Package Version 0.31.* [online] available from <https://cran.r-project.org/package=FSelector>

Rose, C.M. and Stegemann, J.A. (2018) ‘From Waste Management to Component Management in the Construction Industry’. *Sustainability (Switzerland)* 10 (1), art. no. 229

Ruxton, G.D. and Beauchamp, G. (2008) ‘Some Suggestions about Appropriate Use of the Kruskal–Wallis Test’. *Animal Behaviour* 76 (3), 1083–1087

Saeys, Y., Inza, I., and Larrañaga, P. (2007) ‘A Review of Feature Selection Techniques in Bioinformatics’. *Bioinformatics* 23 (19), 2507–2517

Salkind, N.J. (ed.) (2010) *Encyclopedia of Research Design*. Thousand Oaks, Calif.: Thousand Oaks, Calif. : SAGE Publications

Sanchez, B. and Haas, C. (2018) ‘A Novel Selective Disassembly Sequence Planning Method for Adaptive Reuse of Buildings’. *Journal of Cleaner Production* 183, 998–1010

Sansom, M. and Avery, N. (2014) ‘Briefing: Reuse and Recycling Rates of UK Steel Demolition

Arisings'. *Proceedings of the Institution of Civil Engineers: Engineering Sustainability* 167 (3), 89–94

Sapsford, R.J. (2007) *Survey Research*. 2nd ed. London: Sage

Sára, B., Antonini, E., and Tarantini, M. (2001) 'Application of Life Cycle Assessment (LCA) Methodology for Valorization of Building Demolition Materials and Products'. *Proceedings of SPIE - The International Society for Optical Engineering* 4193, 382–390

Sarhrouni, E., Hammouch, A., and Aboutajdine, D. (2012) 'Application of Symmetric Uncertainty and Mutual Information to Dimensionality Reduction and Classification of Hyperspectral Images'. *International Journal of Engineering and Technology* 4 (5), 268–276

Sarkar, S., Pramanik, A., Maiti, J., and Reniers, G. (2020) 'Predicting and Analyzing Injury Severity: A Machine Learning-Based Approach Using Class-Imbalanced Proactive and Reactive Data'. *Safety Science* 125

Saunders, M., Lewis, P., and Thornhill, A. (2016) *Research Methods for Business Students*. Seventh ed. Harlow, Essex: Pearson Education Limited

Sauro, J. and Lewis, J.R. (2016) *Quantifying the User Experience. Practical Statistics for User Research*. 2nd edn. Cambridge, MA: Morgan Kaufmann

Schliep, K., Hechenbichler, K., and Lizee, A. (2016) 'Kknn: Weighted k-Nearest Neighbors'. *R Package Version 1.3.1.*

Schlosser, R.W. (2007) 'Appraising the Quality of Systematic Reviews'. in *Focus: Technical Briefs* [online] held 2007. Southwest Educational Development Laboratory, 8. available from <http://www.ncddr.org/kt/products/focus/focus17/>

Schultmann, F. and Sunke, N. (2007) 'Energy-Oriented Deconstruction and Recovery Planning'. *Building Research and Information* 35 (6), 602–615

Scutari, M. and Denis, J.-B. (2015) *Bayesian Networks: With Examples in R*. Boca Raton, FL: CRC Press: Taylor & Francis Group

Sea-Lim, K., Plianpho, C., Sukmake, P., Pongcharoenkiat, W., and Chinda, T. (2018) 'Feasibility Study of Reverse Logistic of Steel Waste in the Construction Industry'. *Songklanakarin Journal of Science and Technology* 40 (2), 271–277

Sfakianaki, E. and Moutsatsou, K. (2015) 'A Decision Support Tool for the Adaptive Reuse or Demolition and Reconstruction of Existing Buildings'. *International Journal of Environment and Sustainable Development* [online] 14 (1), 1–19. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84921364286&doi=10.1504%2FIJESD.2015.066893&partnerID=40&md5=c6b026e7a2b1a91ea16b7686059f6494>

Shaurette, M. (2006) 'Demolition Contractors' Perceptions of Impediments to Salvage and Reuse of Wood Structural Components'. *Journal of Green Building* 1 (2), 145–163

Sirisawat, P. and Kiatcharoenpol, T. (2019) 'A Study of Correlation for Reverse Logistics Barriers to Solutions Using Structural Equation Modelling'. *International Journal of Productivity and Quality Management* 26 (2), 139–159

Smith, M., Bull, J.W., Morris, T., and Underwood, C. (2013) 'Strength Prediction of Recovered Softwood Beams for Structural Reuse'. *14th International Conference on Civil, Structural and Environmental Engineering Computing, CC 2013* [online] 102. available from

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84893935153&partnerID=40&md5=f6cc040ccfc0e65fc420e31e82b423e8>

Srinagesh, K. (2006) *The Principles of Experimental Research*. Amsterdam : Amsterdam

Statista (2021a) *Construction Industry in the UK - Statistics & Facts* [online] available from <https://www.statista.com/topics/3797/construction-industry-in-the-uk/> [27 May 2021]

Statista (2021b) *Gross Domestic Product of the United Kingdom from 1948 to 2020* [online] available from <https://www.statista.com/statistics/281744/gdp-of-the-united-kingdom/> [27 May 2021]

Statista (2021c) *Distribution of Steel End-Usage Worldwide in 2019, by Sector* [online] available from <https://www.statista.com/statistics/1107721/steel-usage-global-segment/> [27 May 2021]

Statista (2021d) *World Crude Steel Production from 2012 to 2020* [online] available from <https://www.statista.com/statistics/267264/world-crude-steel-production/> [27 May 2021]

Statista (2019) *Statista Industry Report – Global* [online] available from <https://www.statista.com/study/57203/construction---global/>

Stevens, S.S. (1946) 'On the Theory of Scales of Measurement'. *Science* 103 (2684), 677–680

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008) 'Conditional Variable Importance for Random Forests'. *BMC Bioinformatics* 9

Strobl, C., Boulesteix, A.-N., Zeileis, A., and Hothorn, T. (2007) 'Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution'. *BMC Bioinformatics* 8

Surahman, U., Higashi, O., and Kubota, T. (2017) 'Evaluation of Current Material Stock and Future Demolition Waste for Urban Residential Buildings in Jakarta and Bandung, Indonesia: Embodied Energy and CO2 Emission Analysis'. *Journal of Material Cycles and Waste Management* 19 (2), 657–675

Szolnoki, G. and Hoffmann, D. (2013) 'Online, Face-to-Face and Telephone Surveys - Comparing Different Sampling Methods in Wine Consumer Research'. *Wine Economics and Policy* 2 (2), 57–66

Taft, L.M., Evans, R.S., Shyu, C.R., Egger, M.J., Chawla, N., Mitchell, J.A., Thornton, S.N., Bray, B., and Varner, M. (2009) 'Countering Imbalanced Datasets to Improve Adverse Drug Event Predictive Models in Labor and Delivery'. *Journal of Biomedical Informatics* 42 (2), 356–364

Tam, V.W.Y. and Tam, C.M. (2008) 'Waste Reduction through Incentives: A Case Study'. *Building Research and Information* 36 (1), 37–43

Tan, Y., Shen, L.-Y., and Langston, C. (2014) 'A Fuzzy Approach for Adaptive Reuse Selection of Industrial Buildings in Hong Kong'. *International Journal of Strategic Property Management* [online] 18 (1), 66–76. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84897054148&doi=10.3846%2F1648715X.2013.864718&partnerID=40&md5=0865290b7b6bb46b787279f209bbdd3c>

Tatiya, A., Zhao, D., Syal, M., Berghorn, G.H., and LaMore, R. (2017) 'Cost Prediction Model for Building Deconstruction in Urban Areas'. *Journal of Cleaner Production*

Tavakol, M. and Dennick, R. (2011) ‘Making Sense of Cronbach’s Alpha’. *International Journal of Medical Education* 2, 53–55

Thai, H.-T., Ngo, T., and Uy, B. (2020) ‘A Review on Modular Construction for High-Rise Buildings’. *Structures* 28, 1265–1290

The Product-Life Institute (n.d.) *Cradle to Cradle* [online] available from <http://www.product-life.org/en/cradle-to-cradle> [29 July 2020]

The World Bank (2019) *Labor Force, Total* [online] available from <https://data.worldbank.org/indicator/SL.TLF.TOTL.IN> [27 May 2021]

Tingley, D.D., Cooper, S., Cullen, J., Densley Tingley, D., Cooper, S., and Cullen, J. (2017) ‘Understanding and Overcoming the Barriers to Structural Steel Reuse, a UK Perspective’. *Journal of Cleaner Production* 148, 642–652

Tingley, D.D. and Davison, B. (2011) ‘Design for Deconstruction and Material Reuse’. *Proceedings of Institution of Civil Engineers: Energy* 164 (4), 195–204

Torgo, L. (2016) *Data Mining with R: Learning with Case Studies*. 2nd edn. ed. by Kumar, V. Boca Raton, FL: Taylor & Francis; Chapman & Hall/CRC

Tranfield, D., Denyer, D., and Smart, P. (2003) ‘Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review’. *British Journal of Management* 14 (3), 207–222

Tuszynski, J. (2020) ‘CaTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, Etc.’ *R Package Version 1.18.0.*

UK Government (2020) *HMRC Exchange Rates for 2020: Monthly* [online] available from <https://www.gov.uk/government/publications/hmrc-exchange-rates-for-2020-monthly> [27 May 2021]

UN (2015) *The Paris Agreement* [online] available from <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement> [19 November 2020]

UNECE (2020) *Share of Construction in GDP* [online] available from <https://w3.unece.org/PXWeb/en/CountryRanking?IndicatorCode=8> [27 May 2021]

UNEP (2020) *2020 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION: Towards a Zero-Emissions, Efficient and Resilient Buildings and Construction Sector*.

UNEP (2015) *Global Waste Management Outlook*.

UNFCCC (2015) *Climate Get the Big Picture* [online] available from <http://bigpicture.unfccc.int/#content-the-paris-agreemen> [19 June 2018]

Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., and Moore, J.H. (2018) ‘Relief-Based Feature Selection: Introduction and Review’. *Journal of Biomedical Informatics* 85, 189–203

De Vaus, D.A. (2014) *Surveys in Social Research*. Sixth edit. London: Routledge

Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S*. 4th edn. New York: Springer

Vitale, P. and Arena, U. (2017) ‘An Attributional Life Cycle Assessment for an Italian Residential Multifamily Building’. in *Environmental Technology (United Kingdom)* [online] Department of Environmental, Biological, Pharmaceutical Sciences and Technologies, University of Campania 'Luigi Vanvitelli', Caserta, Italy: Taylor and Francis Ltd., 1–13. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029420703&doi=10.1080%2F09593330.2017.1371252&partnerID=40&md5=c157a6d37101e13220b38ab5788885cd>

Weihs, C., Ligges, U., Luebke, K., and Raabe, N. (2005) 'KlaR Analyzing German Business Cycles'. *Data Analysis and Decision Support* 335–343

Weinberger, K.Q. (2018) *Lecture 5: Bayes Classifier and Naive Bayes* [online] available from <https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote05.html> [28 June 2020]

Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J. (2017) *Data Mining: Practical Machine Learning Tools and Techniques*. ed. by 4th. Cambridge, MA: Morgan Kaufmann

Wolpert, D.H. and Macready, W.G. (1997) 'No Free Lunch Theorems for Optimization'. *IEEE Transactions on Evolutionary Computation* [online] 1 (1), 67–82. available from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031118203&doi=10.1109%2F4235.585893&partnerID=40&md5=fac8c56be911367d556066800e863066>

World Economic Forum (2016) *Shaping the Future of Construction: A Breakthrough in Mindset and Technology* [online] available from <http://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction_full_report__.pdf>

WRAP (2008) *Practical Solutions for Sustainable Construction Reclaimed Building Products Guide* [online] available from <www.aggregain.org.uk> [29 July 2018]

WSA (2012) *Sustainable Steel: At the Core of a Green Economy* [online] available from <https://www.worldsteel.org/publications/bookshop.html> [13 February 2018]

Wu, Z., Shen, L., Yu, A.T.W., and Zhang, X. (2016) 'A Comparative Analysis of Waste Management Requirements between Five Green Building Rating Systems for New Residential Buildings'. *Journal of Cleaner Production* 112, 895–902

Yeung, J., Walbridge, S., and Haas, C. (2015) 'The Role of Geometric Characterization in Supporting Structural Steel Reuse Decisions'. *Resources, Conservation and Recycling* 104, 120–130

Yeung, J., Walbridge, S., Haas, C., and Saari, R. (2017) 'Understanding the Total Life Cycle Cost Implications of Reusing Structural Steel'. *Environment Systems and Decisions* 37 (1), 101–120

Yi, W. and Chan, A.P.C. (2014) 'Critical Review of Labor Productivity Research in Construction Journals'. *Journal of Management in Engineering* 30 (2), 214–225

Yin, R.K. (2018) *Case Study Research and Applications : Design and Methods*. Sixth edit.

Zaman, A.U., Arnott, J., Mclntyre, K., and Hannon, J. (2018) 'Resource Harvesting through a Systematic Deconstruction of the Residential House: A Case Study of the "Whole House Reuse" Project in Christchurch, New Zealand'. *Sustainability (Switzerland)* 10 (10), art. no. 3430

Appendix A The complete list of reusability factors

A.1 Reuse drivers identified during the systematic literature review (Chapter 2)

Table A-1 The complete list of identified reuse drivers (Table 2.1)

Category	Sub-category	Driver	Reference (sequence number in Table 2.1)
Economic	Cost	Access to finance to offset additional costs	12, 49
Economic	Cost	Deconstruction costs less than demolition	43
Economic	Cost	Increased cost of landfilling	5, 12, 41, 42
Economic	Cost	In-situ reuse of the reused elements	4, 11, 12, 21, 59
Economic	Cost	Low labour cost due to reusing the modules of the structural systems	54
Economic	Cost	Low labour cost due to using custom plates and reusing the existing bolt holes	54
Economic	Cost	Low price of new steel and scrap	59
Economic	Cost	Lower cost of deconstruction compared to demolition due to low cost of manual labour and high demand for demolition products	14
Economic	Cost	Lower cost of reused elements	1, 4, 11, 14, 33, 47, 55, 62
Economic	Cost	Savings due to the purchase of fewer new steel sections	36
Economic	Cost	Sourcing reused material from nearby locations	11, 31
Economic	Market	High demand for reused building components	4
Economic	Market	Supporting the growth of reuse market	49
Economic	Market	Well-established reuse market	14, 33
Economic	Value for money	Increased profit as the result of decreased CDW sent to landfill	7
Economic	Value for money	Increased revenue from reused elements resale	4, 5, 7, 14, 17, 19, 30, 33, 36, 47, 52, 55, 56, 62
Environmental	Energy and GHG	Decrease in embodied energy and carbon of construction	2, 4, 20, 23, 25, 26, 30, 33, 45, 46, 48, 49, 50, 51, 52, 54, 60, 61, 62, 63, 64
Environmental	Preservation of resources	Decrease in the amount of waste disposed in the landfills	20
Environmental	Preservation of resources	Decrease in the use of virgin materials	2, 16, 24, 26, 30, 49

Category	Sub-category	Driver	Reference (sequence number in Table 2.1)
Environmental	Preservation of resources	Decrease in water consumption	50
Environmental	Preservation of resources	Scarcity of the landfilling sites	41, 62
Organisational	Contracts	Legal contractual requirement to use reused elements	1, 11, 12
Organisational	Experience	Knowledge and experience in using reused elements makes firms more competitive	49
Organisational	Experience	Training the operators for effective deconstruction	5, 8, 27
Organisational	Infrastructure	Availability of space for storage of reusable materials after deconstruction	18
Organisational	Infrastructure	Proper separation and storage of the reusable materials after deconstruction	18, 27, 38, 44
Organisational	Management	Companies' entrepreneurial activities to integrate circular principles	62
Organisational	Management	Existence of a reclaimed components management coordinator	12, 49
Organisational	Management	Integrating reuse in the design process of the new projects	11, 12, 18, 25, 44, 49
Organisational	Management	Knowledge of a known list of structural elements to reuse early on in the design process	12, 58
Organisational	Sustainability	Corporate social responsibility	58
Organisational	Sustainability	Improving the overall sustainability of the building sector	18, 49
Organisational	Sustainability	Promoting the green image of the firms to improve competitiveness	18, 32, 39, 41, 42, 55
Organisational	Sustainability	Reducing the CDW generation by the firms	7, 9, 10, 24, 26, 30, 39, 45, 52, 53, 55
Regulatory	Compliance	Availability of standards to certify the quality of reused elements	62
Regulatory	Compliance	Compliance to regulations enhances deconstruction	55
Regulatory	Compliance	Compliance to regulations enhances reuse	55
Regulatory	Incentive	Availability of regulatory/financial incentives to promote deconstruction	55
Regulatory	Incentive	Availability of regulatory/financial incentives to promote reuse	3, 13, 55, 57
Regulatory	Sustainability	Impact of building rating systems such as BREEAM, LEED, etc.	8, 12, 35, 62
Regulatory	Sustainability	Impact of environmental policies	55
Regulatory	Sustainability	Legislative pressure for resource preservation	41
Social	Awareness	Increased awareness by recognition of reuse in the public debate	62

Category	Sub-category	Driver	Reference (sequence number in Table 2.1)
Social	Awareness	Increased awareness of the full benefits of reuse among the stakeholders	1, 13
Social	Perception	Positive perception of contractors about reuse	18
Social	Sustainability	Impact of society's environmental concerns	42
Social	Trust	Informality and good relationship among the stakeholders can enhance reuse	8, 14, 39
Social	Willingness	Client willingness to integrate reused elements	8, 11, 12, 21, 29, 47, 59
Social	Willingness	Contractor willingness to integrate reused elements	11, 18, 39, 47, 55
Social	Willingness	Design team willingness to integrate reused elements	11, 12, 37, 47, 49, 59
Social	Willingness	Unique appearance of reused elements	62
Technical	Deconstruction	Deconstruction technique can enhance the chance for reuse	6, 7, 10, 11, 17, 19, 24, 28, 31, 34
Technical	Deconstruction	Use of advanced construction techniques (e.g. pre-fabrications for installation) increases the reuse rate	22, 40
Technical	Design challenges	Durability of the recovered building component	51
Technical	Design challenges	Proper estimation of the required size and lengths at the beginning of the design phase	11
Technical	Design challenges	Use of the reused structural elements to support similar loads	11, 12, 31
Technical	Information	Availability of information about characteristics, details, certificates and drawings of the reused structural elements	11, 12, 15, 31, 51

A.2 Reuse barriers identified during the systematic literature review (Chapter 2)

Table A-2 The complete list of identified reuse barriers (Table 2.2)

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Economic	Cost	Cost of insurance for reused materials	38
Economic	Cost	Cost of marketing for reused elements	4
Economic	Cost	Cost of sorting for reused elements	12, 32
Economic	Cost	Cost of testing for the reused elements	10, 29, 31, 32, 38, 46
Economic	Cost	Deconstruction costs more than demolition	4, 8, 25, 29, 38, 45, 46
Economic	Cost	Extra effort by design team to find reused components	9, 31
Economic	Cost	Extra effort required for deconstruction/reuse	9, 25, 32
Economic	Cost	Extra time required for treatment and fabrication of the salvaged components	37, 46
Economic	Cost	Higher cost of reused elements	7, 25, 32, 35, 38, 46
Economic	Cost	Impact of access to the building on deconstruction cost	36
Economic	Cost	Impact of complexity of the building design on deconstruction cost	36
Economic	Cost	Impact of location of the building on deconstruction cost	36
Economic	Cost	Increased cost due to the need for treatment/modification of the salvaged components	2, 12, 15, 27, 31, 46
Economic	Cost	Increased cost of design with the reused elements	9, 10, 37
Economic	Cost	Increased cost of fabrication of the reused materials	37, 38, 46
Economic	Cost	Increased labour cost	3, 4, 6, 7, 9, 32, 34, 38, 39, 44, 45, 46, 48
Economic	Cost	Increased storage cost	6, 8, 9, 10, 11, 29, 31, 34, 37, 38, 46
Economic	Cost	Increased transportation cost	8, 9, 10, 11, 23, 29, 31, 32, 34, 35, 44, 46
Economic	Cost	Lower cost of landfilling	8, 15, 32, 35, 42
Economic	Cost	Need to purchase reused elements early in the project	9, 10
Economic	Cost	Potential financial risks	6, 32
Economic	Cost	Recycling is preferred to reuse due to market conditions	21, 27, 29, 38, 39, 49

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Economic	Cost	Time required for deconstruction and project scheduling	1, 4, 6, 7, 9, 10, 15, 21, 25, 27, 29, 32, 34, 35, 36, 37, 38, 45
Economic	Cost	Wrong estimation of deconstruction cost hinders its application	36
Economic	Market	Lack of an established market for reused structural elements	5, 6, 7, 9, 10, 32, 33, 34, 35, 36, 37, 38, 41, 42, 45, 46, 49
Economic	Market	Lack of demand for reused structural elements	5, 6, 14, 27, 33, 38, 41, 49
Economic	Market	Lack of information sharing in the supply chain (e.g. disconnection between supply and demand)	5, 9, 32, 45
Economic	Market	Lack of sufficient supply for the reused elements with desired characteristics (dimension, quality, etc.)	10, 11, 37, 38, 43, 45, 49
Economic	Market	Lack of supply and demand for reused structural elements	7, 27
Economic	Market	Uncertainty in the demand for reused materials	6, 45
Economic	Value for money	Economic benefits of reuse not defined properly	38
Economic	Value for money	Uncertainty about revenue from reused elements resale	15, 29, 35, 45, 46, 48
Environmental	Energy and GHG	Emissions due to the higher operating time of the heavy machinery during deconstruction	50
Environmental	Energy and GHG	Emissions due to transportation	27, 49, 50
Organisational	Contracts	Proprietary lock-in	38
Organisational	Experience	Lack of companies' expert in deconstruction	33
Organisational	Experience	Lack of skills, experience and knowledge in deconstruction, salvage, and using reused elements	7, 10, 17, 20, 25, 29, 33, 48
Organisational	Experience	Uncommon practice	37, 46
Organisational	Infrastructure	Lack of facilities to recover the used products	33, 49
Organisational	Infrastructure	Lack of space for storage of reusable materials after deconstruction	7, 10, 18, 34, 37, 45, 46, 48
Organisational	Infrastructure	Need for infrastructure and equipment to perform deconstruction	7, 18, 40, 44
Organisational	Infrastructure	Need for specific technology	34
Organisational	Management	Inconsistency in waste management practices	16

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Organisational	Management	Lack of a decision-making framework for reuse	29
Organisational	Management	Lack of cooperation with demolition contractors to jointly recover materials from construction sites	49
Organisational	Management	Lack of coordination between the owners of the demolition site and the new building	46
Organisational	Management	Lack of integration of reuse in the design process of the new projects	45
Organisational	Management	Lack of ownership due to too many players	16
Organisational	Management	Lack of systems thinking	45
Organisational	Management	Uncertainty about the timely availability of desired reused elements	9, 25
Regulatory	Compliance	Change in the applicable design norms (e.g. room height, fire, stress, etc.)	8, 27
Regulatory	Compliance	Existing codes, standards, and procedures do not consider component reuse	10, 27, 32, 38
Regulatory	Compliance	Existing codes, standards, and procedures do not mandate component reuse	49
Regulatory	Compliance	Existing codes, standards, and procedures do not mandate deconstruction	49
Regulatory	Compliance	Existing regulations do not support deconstruction	32
Regulatory	Compliance	Existing regulations do not support reuse	10, 18, 19, 25, 27, 28, 32, 33, 37, 38, 45, 48
Regulatory	Compliance	Inconsistency and lack of coordination among the regulatory bodies	32, 35
Regulatory	Compliance	Lack of government control for effective implementation of existing regulations	11
Regulatory	Compliance	Lack of government support	33, 35
Regulatory	Compliance	Lack of guidance, knowledge and information sharing about C&DW management	11, 42
Regulatory	Compliance	Lack of insurance for reused elements	37
Regulatory	Compliance	Lack of quality certificates for the reused element	2, 6, 8, 27, 33
Regulatory	Compliance	Lack of standardisation for reused components	27
Regulatory	Compliance	Lack of standards to certify the quality of reused elements	2, 7, 12, 25, 37, 38, 42
Regulatory	Compliance	Lack of traceability and certification for reused elements	37, 38
Regulatory	Compliance	Need for CE marking	37, 38
Regulatory	Compliance	PI insurance in case of using reused elements	37, 38

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Regulatory	Compliance	Regulations do not allow storage of salvaged material and consider them as waste	32
Regulatory	Incentive	Lack of incentives for component reuse in environmental assessment methods	38
Regulatory	Incentive	Lack of incentives for waste minimisation	16
Regulatory	Incentive	Lack of regulatory/financial incentives to promote deconstruction	33, 35
Regulatory	Incentive	Lack of regulatory/financial incentives to promote reuse	29, 33, 38, 45, 49
Social	Awareness	Lack of awareness about reused elements across the supply chain	6, 38
Social	Awareness	Lack of awareness about the deconstruction risks and challenges	32, 33
Social	Awareness	Lack of awareness of the full benefits of deconstruction among the stakeholders	25, 32, 33
Social	Awareness	Lack of awareness of the full benefits of reuse among the stakeholders	10, 27, 33
Social	Perception	Demolition is preferred to deconstruction due to the perceived economic and scheduling reasons.	10
Social	Perception	Negative perception of contractors about reused elements	7, 10
Social	Perception	Negative perception of the clients about reused elements	11, 25, 37, 45
Social	Perception	Negative perception of the designers about reuse	10
Social	Perception	Negative perception of the stakeholders about reused elements	1, 3, 25, 27, 32, 35, 37, 38, 47
Social	Perception	Negative perception of the supervisors about reused elements	25
Social	Perception	Reused structural elements are not visually attractive	24, 38
Social	Risk	Inequality in the distribution of risk among the stakeholders	46
Social	Risk	Lack of confidence in the quality of reused components	30, 35, 46
Social	Risk	Liability risk due to informality and trust	11
Social	Risk	Potential health and safety risks	3, 27, 32
Social	Risk	Risks associated with reuse (liability, fear, etc.)	38, 37, 25, 10, 7, 32
Social	Sustainability	Unsatisfactory working environment during the treatment of the reused elements	3, 12
Social	Trust	Lack of trust to the supplier of reused elements	37, 46
Social	Willingness	Construction sector inertia/resistance against reuse	10, 24, 32, 38, 49

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Social	Willingness	Contractors unwillingness to work with the reused element	10
Social	Willingness	Design team unwillingness to integrate reused elements	25, 32
Social	Willingness	Lack of client demand/ support	27, 32, 38, 45
Social	Willingness	Lack of interest to integrate reused materials in the projects	32, 33
Social	Willingness	Regulatory authority unwillingness to integrate reused elements	25
Technical	Deconstruction	Composite structural elements	38
Technical	Deconstruction	Existing building not designed for deconstruction	22, 25, 27, 30, 33, 36, 37, 39, 45
Technical	Deconstruction	Hard to access joints	38, 40
Technical	Deconstruction	Permanent jointing techniques	10, 12, 13, 18, 22, 23, 38, 50, 51
Technical	Deconstruction	Presence of fire permanent protection on the reused elements	51
Technical	Deconstruction	Type of connection can affect deconstruction	12, 28
Technical	Design challenges	Damage caused by living organisms (termite, bacterial attack, etc.)	2, 8
Technical	Design challenges	Damage during refurbishment (nail removal, etc.)	2
Technical	Design challenges	Damage to the structural elements	27
Technical	Design challenges	Damage to the structural elements due to corrosion	2, 28, 29
Technical	Design challenges	Damage to the structural elements due to deconstruction	2, 8, 10, 15, 23, 40
Technical	Design challenges	Damage to the structural elements due to degradation	24, 26
Technical	Design challenges	Damage to the structural elements due to fatigue	29
Technical	Design challenges	Damage to the structural elements due to frost	28
Technical	Design challenges	Damage to the structural elements due to impact	29
Technical	Design challenges	Damage to the structural elements due to post-production modifications (e.g. holes for duct work, etc.)	2, 8, 29
Technical	Design challenges	Damage to the structural elements due to type of joints	10, 12, 40, 51
Technical	Design challenges	Damage to the structural elements due to water/fire/holes	8, 26, 29, 36
Technical	Design challenges	Damage to the structural elements during storage	10, 40

Category	Sub-category	Driver	Reference (sequence number in Table 2.2)
Technical	Design challenges	Design with long spans	9
Technical	Design challenges	Difference in the loading requirements of the old and the new buildings	9
Technical	Design challenges	Difficulty in designing with reused elements	9, 23, 38, 43
Technical	Design challenges	Difficulty in reusing the elements due to the short length	43
Technical	Design challenges	Integration of the reused and new elements in the new structure	10, 23
Technical	Design challenges	Lower quality of reclaimed products compared to new	11, 24, 36
Technical	Design challenges	Need for the flexibility in the design	9, 10
Technical	Design challenges	Old spans do not match new design features	27
Technical	Design challenges	Overdesigned structures due to the available supply	9, 43
Technical	Design challenges	Reused elements exposed to weather conditions	8, 27
Technical	Health and safety	Additional health and safety precautions necessary for deconstruction and element recovery & reuse	15, 21, 25, 27, 29, 32, 35, 38
Technical	Health and safety	Presence of fire protection on the reused elements	38
Technical	Health and safety	Presence of hazardous, banned or contaminating coatings on the reused elements	6, 8, 15, 32, 36, 38
Technical	Information	Lack of information about characteristics, details, certificates and drawings of the reused structural elements	9, 10, 27, 29, 38, 45
Technical	Information	Lack of information about the remaining capacity of the reused structural elements	27, 29

Appendix B Checklists used to improve the quality of the questionnaire before pilot study

Seq	Questionnaire layout	Status
1	(For self-completed questionnaires) Do questions appear well spaced on the page or screen? A cramped design will put the respondent off reading it and reduce the response rate. Unfortunately, a thick questionnaire is equally off-putting!	Yes
2	(For paper-based self-completed questionnaires) Is the questionnaire going to be printed on good quality paper? Poor-quality paper implies that the survey is not important.	Yes
3	(For self-completed questionnaires) Is the questionnaire going to be printed or displayed on a warm pastel colour? Warm pastel shades, such as yellow and pink, generate slightly more responses than white (Edwards et al. 2002) or cool colours, such as green or blue. White is a good neutral colour but bright or fluorescent colours should be avoided.	No (neutral white)
4	(For structured interviews) Will the questions and instructions be printed on one side of the paper only? An interviewer will find it difficult to read the questions on the back of pages if you are using a questionnaire attached to a clipboard!	N/A
5	Is your questionnaire easy to read? Questionnaires should be typed in 12 point or 10 point using a plain font. Excessively long and unduly short lines reduce legibility. Similarly, respondents find CAPITALS, italics and shaded backgrounds more difficult to read. However, if used consistently, they can make completing the questionnaire easier.	Yes (you should remove italics)
6	Have you ensured that the use of shading, colour, font sizes, spacing and the formatting of questions is consistent throughout the questionnaire?	Yes
7	Is your questionnaire laid out in a format that respondents are accustomed to reading? Research has shown that many people skim-read questionnaires (Dillman et al. 2014). Instructions that can be read one line at a time from left to right moving down the page are, therefore, more likely to be followed correctly.	Yes

Actions taken (6th online revision):

- Changes incorporated
- Italics changed to normal
- Page 7, definition changed to bold & black

Seq	Question order	Status
1	Are questions at the beginning of your questionnaire more straightforward and ones the respondent will enjoy answering? Questions about attributes and behaviours are usually more straightforward to answer than those collecting data on opinions.	No
2	Are questions at the beginning of your questionnaire obviously relevant to the stated purpose of your questionnaire? For example, questions requesting contextual information may appear irrelevant.	Yes
3	Are questions and topics that are more complex placed towards the middle of your questionnaire? By this stage most respondents should	Yes
4	Are personal and sensitive questions towards the end of your questionnaire, and is their purpose clearly explained? On being asked these a respondent may refuse to answer; however, if they are at the end of an interviewer-completed questionnaire you will still have the rest of the data!	Yes
5	Are filter questions and routing instructions easy to follow so that there is a clear route through the questionnaire?	No filter questions
6	(For interviewer-completed questionnaires) Are instructions to the interviewer easy to follow?	N/A
7	Are questions grouped into obvious sections that will make sense to the respondent?	Yes
8	Have you re-examined the wording of each question and ensured it is consistent with its position in the questionnaire as well as with the data you require?	Yes

Note: I developed the 6th online revision to fulfil the requirements of the 4th question.

Seq	Question wording	Status
1	Does your question collect data at the right level of detail to answer your investigative question as specified in your data requirements table?	Yes
2	Will respondents have the necessary knowledge to answer your question? A question on the implications of a piece of European Union legislation would yield meaningless answers from those who were unaware of that legislation.	Yes
3	Does your question appear to talk down to respondents? It should not!	No
4	Does your question challenge respondents' mental or technical abilities? Questions that do this are less likely to be answered.	No
5	Are the words used in your question familiar to all respondents, and will all respondents understand them in the same way? In particular, you should use simple words and avoid jargon, abbreviations and colloquialisms.	Yes
6	Are there any words that sound similar and might be confused with those used in your question? This is a particular problem with interviewer-completed questionnaires.	No
7	Are there any words that look similar and might be confused if your question is read quickly? This is particularly important for self-completed questionnaires.	No
8	Are there any words in your question that might cause offence? These might result in biased responses or a lower response rate.	No
9	Can your question be shortened? Long questions are often difficult to understand, especially in interviewer-completed questionnaires, as the respondent needs to remember the whole question. Consequently, they often result in no response at all.	Yes
10	Are you asking more than one question at the same time? The question 'How often do you visit your mother and father?' contains two separate questions, one about each parent, so responses would probably be impossible to interpret.	No
11	Does your question include a negative or double negative? Questions that include the word 'not' are sometimes difficult to understand. The question 'Would you rather not use a	No
12	Is your question unambiguous? This can arise from poor sentence structure, using words with several different meanings or having an unclear investigative question. If you ask 'When did you leave school?' some respondents might state the year, others might give their age, while those still in education might give the time of day! Ambiguity can also occur in category questions. If you ask employers how many employees they have on their payroll and categorise their answers into three groups (up to 100, 100– 250, 250 plus), they will not be clear which group to choose if they have 100 or 250 employees.	Checked. Questions are not ambiguous.
13	Does your question imply that a certain answer is correct? If it does, the question is biased and will need to be reworded, such as with the question 'Many people believe that too little money is spent on our public Health Service. Do you believe this to be the case?' For this question, respondents are more likely to answer 'yes' to agree with and please the interviewer.	No
14	Does your question prevent certain answers from being given? If it does, the question is biased and will need to be reworded. The	No

Seq	Question wording	Status
	question 'Is this the first time you have pretended to be sick?' implies that the respondent has pretended to be sick whether they answer yes or no!	
15	Is your question likely to embarrass the respondent? If it is, then you need either to reword it or to place it towards the end of the survey when you will, it is to be hoped, have gained the respondent's confidence. Questions on income can be asked as either precise amounts (more embarrassing), using a quantity question, or income bands (less embarrassing), using a category question. Questions on self-perceived shortcomings are unlikely to be answered.	No
16	Have you incorporated advice appropriate for your type of questionnaire (such as the maximum number of categories) outlined in the earlier discussion of question types?	Yes
17	Are answers to closed questions written so that at least one will apply to every respondent and so that each of the responses listed is mutually exclusive?	Yes
18	Are the instructions on how to record each answer clear?	Yes

Notes:

- Performed on the 5th online revision

Appendix C Data collection tool

C.1 Example of the email sent to the professionals with experience in reuse in buildings

Content removed on data protection grounds

C.2 Example of the questionnaire survey

<u>Section A: Respondent's details:</u>

Please answer the following questions by choosing the applicable boxes or filling in the blank spaces.

1. Where is the geographic location of your organisation (Country name)?

   ……………………………………………………………………

2. What is the type of organisation you work in?

☐Client ☐Consultancy (architectural, structural, etc.) ☐Contractor ☐Deconstruction/Demolition ☐Supplier/Stockist ☐ University/Academic institution ☐other (please specify):

………

3. How many years of experience do you have in the construction sector?

☐1-5 ☐6-10 ☐11-15 ☐16-20 ☐21-25 ☐26-30 ☐31-35 ☐36-40 ☐over 40 ☐other (please specify):

……………………………………………………………………………………

4. What is your position/job title (Architect, CEO, etc.)?

   ………………………………………………………………………………………

5. Do you or your company have any experience with the reuse of the building structural elements? ☐Yes☐No

<u>Section B: Details about the reused structural element</u>

Based on your experience, **please select only one structural element** that you reused in the past and complete the rest of the questionnaire based on that.

1. Which structural element that you reused before are you basing your answers?

☐Beam ☐Brace ☐Column ☐Slab ☐Truss ☐other (please specify):

………

2. What is the material of construction (MoC) of the structural element that you reused?

☐Concrete ☐Steel ☐Timber ☐Cast Iron ☐Wrought Iron ☐ Composite ☐other (please specify):

………

3. What is the approximate age of the building from which the element is recovered?

☐0 to 40 ☐41 to 60 ☐61 to 80 ☐81 to 100 ☐100 years and older ☐other (please specify):

………

4. What is the recovery technique used to recover the particular element?

☐Demolition ☐Component-specific recovery ☐ Deconstruction ☐other (please specify):

....

5. What is the number of existing connections fixed to the element when purchased/acquired (plates or angles fixed to a beam, etc.)?

☐1 to 2 ☐3 to 4 ☐5 to 7 ☐8 to 10 ☐11 and above ☐other (please specify):

....

6. What are the types of the end connections (joints) of the element when purchased/acquired?

☐ Reversible (bolts, screws, etc.) ☐ Permanent (welding, cast in-situ concrete, etc.) ☐ Mixed ☐other (please specify):

....

Instructions for questions 7 to 11:

You may ignore any question if not applicable or the details are/were not available.

Questions 7 to 11 compare the current use (or use after deconstruction) of the structural element with its previous use before it was removed/deconstructed from a building.

7. The structural element is serving the same purpose (i.e. as a beam, slab, column, etc.) in its new installation as in its previous installation.

☐ Strongly agree ☐ Agree ☐ Neither agree nor disagree ☐ Disagree ☐ Strongly disagree

8. The cross-section/thickness dimensions of the structural element in its new installation are equal or nearly equal to the cross-section/thickness dimensions of the element in its previous installation.

☐ Strongly agree ☐ Agree ☐ Neither agree nor disagree ☐ Disagree ☐ Strongly disagree

9. The length dimensions of the structural element in its new installation are equal or nearly equal to the length dimensions of the element in its previous installation.

☐ Strongly agree ☐ Agree ☐ Neither agree nor disagree ☐ Disagree ☐ Strongly disagree

10. The amount of load supported by the structural element in its new installation compared to the amount of load supported by the element in its previous installation.

☐ Much lower ☐ Lower ☐ Equal ☐ Higher ☐ Much Higher

11. The life expectancy of the structural element in its new installation compared to the life expectancy of the element in its previous installation.

☐ Much lower ☐ Lower ☐ Equal ☐ Higher
☐ Much Higher

Section C: Factors affecting the reusability of the structural element

You may ignore any question if not applicable or the details are/were not available.

Please rate the followings on the scale of 1 to 5 where:

5 = Very low ***4 = Low*** ***3 = Moderate*** ***2 = High*** ***1 = Very High***

	What was the negative impact of the following factors on the reusability of the structural element?	Scale				
		1	2	3	4	5
C1	Damage during deconstruction/demolition					
C2	Damage due to fatigue					
C3	Damage due to fire					
C4	Damage during transportation					
C5	Damage during storage					
C6	Damage due to the type of joints					
C7	Damage due to corrosion					
C8	Damage due to frost					
C9	Damage due to water penetration/presence					
C10	Damage during refurbishment (nail removal, etc.)					
C11	Damage due to exposure to wind, acidic rain, etc.					
C12	Damage caused by living organisms (termite, bacterial attack, etc.)					
C13	Damage due to earthquake					
C14	Damage due to impact					
C15	Damage due to post-production modifications (e.g. holes, etc.)					
C16	Lack of certificates of quality for the element when acquired					
C17	Lack of standards to certify the element					
C18	Lack of the original drawings					
C19	Lack of the original design calculations					
C20	Lack of earlier certificates (inspection, material, etc.)					
C21	Lack of traceability of the element					
C22	Potential liability risks					
C23	Potential health and safety risks					
C24	Potential financial risks					
C25	The potential risk associated with the structural integrity					
C26	The potential risk of damage to the machinery (nails in timber, etc.)					
C27	A potential problem with collateral warranties					
C28	Presence of hazardous, banned or contaminating coatings					

Section D: Other factors affecting the reusability of the structural element

You may ignore any question if not applicable or the details are/were not available.

Please rate the followings on the scale of 1 to 5 where:

1 = Very negatively 2 = Negatively 3 = No real effect 4 = Positively 5 = Very Positively

	How did the following factors affect the reusability of the structural element?	Scale				
		1	2	3	4	5
D1	The purchasing price					
D2	Cost of insurance					
D3	Cost of testing					
D4	Cost of refurbishment (sandblasting, treatment, etc.)					
D5	Cost of design with the reused element					
D6	Storage cost					
D7	Transportation cost					
D8	Cost of labour					
D9	Cost of fabrication					
D10	Cash flow (need to purchase the element early, etc.)					
D11	Perception of the client/top management team about the element					
D12	Perception of the designers about the element					
D13	Perception of the builders/contractors about the element					
D14	Perception of the end users (when it is not the client) about the element					
D15	Perception of the stockist about the element					
D16	Perception of the regulatory authorities about the element					
D17	Visual appearance					
D18	Presence of fire protection on the element					
D19	Changes in the design codes (BS codes to Eurocodes, etc.)					
D20	Changes in the health and safety regulations (fire, etc.)					
D21	CE marking					
D22	Matching the original design with the dimensions of the reused element					
D23	Matching the original design with the strength of the reused element					
D24	Other design challenges with the reused element					
D25	Sourcing/procurement process					

Section E: The overall reusability of the structural element

Definitions:

Technical reusability:

- The extent to which the reused structural element in its new life could perform similarly to its earlier life.

Economic reusability:

- The cost savings in the project as the result of using the reused structural element when compared to a similar project using a new structural element with the same performance.

Social reusability:

- The acceptance level of the stakeholders (clients, CEO, designers, construction team, occupants, etc.) about using the reused structural element in the new building.

Please refer to the **definitions** section (above) for further clarity. Please rate the followings on the scale of 1 to 5 where:

1 = Very low 2 = Low 3 = Moderate 4 = High 5 = Very High

	Please rate the relative level of reusability of the structural element by providing the actual or approximate answers.	Scale				
		1	2	3	4	5
E1	The technical reusability					
E2	The economic reusability					
E3	The social reusability					

Please feel free to write any additional comments in the space provided below.

....
....
....
....

If you are willing to know the results of this study, please provide your contact details in the space provided below. Kindly note that this is totally optional.

....
....
....

If you have an experience with another reused structural element, please feel free to fill this survey again based on that other structural element.

Thank you for taking the time to complete this questionnaire.

If you have any queries, please do not hesitate to contact me (Kambiz Rakhshanbabanari) by telephoning (+44)7443-305756 or emailing rakhshak@uni.coventry.ac.uk.

Appendix D Record of the statistical tests and descriptive statistics

D.1 Little's MCAR test (Technical dataset)

Table D-1 Little's MCAR test (Technical dataset)

Univariate Statistics

Variables	N	Mean	Std. Deviation	Missing Count	Missing Percent	No. of Extremes[a] Low	No. of Extremes[a] High	EM Means[b]
B7	69	3.70	1.204	3	4.2	0	0	3.70
B8	69	3.93	1.019	3	4.2	9	0	3.89
B9	69	3.23	1.214	3	4.2	0	0	3.22
B10	68	2.31	.778	4	5.6	0	0	2.30
B11	65	2.83	.802	7	9.7	0	2	2.69
C1	71	3.01	1.282	1	1.4	0	0	2.98
C2	72	3.85	1.070	0	.0	0	0	3.85
C3	71	4.27	1.253	1	1.4	10	0	4.22
C4	71	4.35	.864	1	1.4	3	0	4.36
C5	72	4.21	1.061	0	.0	7	0	4.21
C6	72	3.78	1.178	0	.0	0	0	3.78
C7	70	4.19	1.133	2	2.8	8	0	4.16
C8	71	4.58	.710	1	1.4	1	0	4.56
C9	72	3.53	1.267	0	.0	0	0	3.53
C10	71	3.85	1.023	1	1.4	0	0	3.81
C11	72	4.42	.946	0	.0	5	0	4.42
C12	72	3.87	1.310	0	.0	0	0	3.88
C13	71	4.85	.497	1	1.4	.	.	4.85
C14	70	4.39	.997	2	2.8	5	0	4.34
C15	72	3.76	1.081	0	.0	0	0	3.76
C16	72	2.97	1.472	0	.0	0	0	2.97
C17	71	3.06	1.511	1	1.4	0	0	3.04
C18	71	3.75	1.481	1	1.4	0	0	3.73
C19	71	3.80	1.480	1	1.4	0	0	3.79
C20	71	3.70	1.468	1	1.4	0	0	3.69
C21	71	3.86	1.437	1	1.4	0	0	3.85
C25	72	3.43	1.276	0	.0	0	0	3.43
C26	71	3.80	1.116	1	1.4	0	0	3.80
C27	72	3.97	1.162	0	.0	0	0	3.97
C28	72	3.60	1.241	0	.0	0	0	3.60

Univariate Statistics

Variables	N	Mean	Std. Deviation	Missing Count	Missing Percent	No. of Extremes[a] Low	No. of Extremes[a] High	EM Means[b]
D18	68	2.69	1.069	4	5.6	0	5	2.69
D19	67	2.58	1.103	5	6.9	0	4	2.60
D21	65	2.65	1.138	7	9.7	0	5	2.74
D22	69	2.51	1.146	3	4.2	0	4	2.55
D23	69	2.71	1.238	3	4.2	0	8	2.76
D24	68	2.75	1.098	4	5.6	0	7	2.72
E1	72	3.76	1.216	0	.0	0	0	3.76
B1	72			0	.0			
B2	72			0	.0			
B3	72			0	.0			
B4	72			0	.0			
B5	71			1	1.4			
B6	71			1	1.4			

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

b. Little's MCAR test: Chi-Square = 662.391, DF = 611, Sig. = .074

D.2 Little's MCAR test (Economic dataset)

Table D-2 Little's MCAR test (Economic dataset)

Univariate Statistics

Variables	N	Mean	Std. Deviation	Missing Count	Missing Percent	No. of Extremes[a] Low	No. of Extremes[a] High	EM Means[b]
C24	72	4.01	1.132	0	.0	0	0	4.00
D1	72	3.68	1.265	0	.0	0	0	3.68
D2	68	2.66	.940	4	5.6	0	3	2.67
D3	68	2.57	1.083	4	5.6	0	4	2.55
D4	72	2.81	1.043	0	.0	0	4	2.79
D5	72	2.82	1.079	0	.0	0	5	2.82
D6	72	2.78	1.213	0	.0	0	9	2.77
D7	72	2.82	1.179	0	.0	0	0	2.82

Univariate Statistics

Variables	N	Mean	Std. Deviation	Missing		No. of Extremes[a]		EM Means[b]
				Count	Percent	Low	High	
D8	72	2.94	1.112	0	.0	0	10	2.94
D9	71	2.89	1.090	1	1.4	0	6	2.91
D10	71	2.83	1.134	1	1.4	0	6	2.81
D25	69	2.83	1.200	3	4.2	0	0	2.82
E2	72	3.93	.969	0	.0	0	0	3.92
B1	72			0	.0			
B2	72			0	.0			
B3	72			0	.0			
B4	72			0	.0			
B5	71			1	1.4			
B6	71			1	1.4			

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

b. Little's MCAR test: Chi-Square = 55.684, DF = 44, Sig. = .111

D.3 Little's MCAR test (Social dataset)

Table D-3 Little's MCAR test (Social dataset)

Univariate Statistics

Variables		Mean	Std. Deviation	Missing		No. of Extremes[a]		EM Means[b]
				Count	Percent	Low	High	
C22	72	3.24	1.399	0	.0	0	0	3.23
C23	72	3.81	1.274	0	.0	0	0	3.79
D11	70	3.46	1.151	2	2.8	5	0	3.44
D12	71	3.35	1.255	1	1.4	0	0	3.36
D13	71	3.11	1.304	1	1.4	0	0	3.10

Univariate Statistics

Variables		Mean	Std. Deviation	Missing		No. of Extremes[a]		EM Means[b]
				Count	Percent	Low	High	
D14	70	3.46	1.441	2	2.8	0	0	3.46
D15	61	2.70	1.038	11	15.3	0	2	2.81
D16	70	3.00	1.155	2	2.8	0	0	3.02
D17	72	3.32	1.309	0	.0	0	0	3.34
D20	68	2.63	1.006	4	5.6	0	3	2.64
E3	72	4.29	.956	0	.0	3	0	4.28
B1	72			0	.0			
B2	72			0	.0			
B3	72			0	.0			
B4	72			0	.0			
B5	71			1	1.4			
B6	71			1	1.4			

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

b. Little's MCAR test: Chi-Square = 60.730, DF = 59, Sig. = .413

D.4 Descriptive statistics of the received questionnaire (after estimating the missing values)

Ranking of the variables is performed based on the values of the Mean. Since questions B7 to B11 intend to compare the current use of the element with its previous deployment, they were not included in the ranking. Questions in Section C act as barrier and questions in Section D can act as drivers or barriers depending on the value of the Mean. For instance, for values of Mean above 3 in Section D, the variable acts as a reuse driver. The comparison has been made in two stages. In Stage 1 variables are compared within their respective group (e.g., ranking is based on being in Section B or C or D and being a driver or a barrier). In Stage 2, which includes variables in Sections C & D, the variables are ranked from 1 (the lowest Mean) to the highest Mean. Hence, D22 in the TEC dataset with Mean equal to 2.53 has the worst impact on the reusability of an element and is ranked 1.

Table D-4 Descriptive statistics for TEC dataset (Number of observations = 72)

Variables	Mean	Standard Error	Median	Standard Deviation	Variance	Role	In group rank (Stage 1)	Overall rank (Stage 2)
B7	3.71	0.14	4.00	1.19	1.42	NA	2	N/A
B8	3.93	0.12	4.00	1.01	1.02	NA	1	N/A
B9	3.25	0.14	3.00	1.21	1.46	NA	3	N/A
B10	2.31	0.09	2.00	0.76	0.58	NA	5	N/A
B11	2.88	0.10	3.00	0.82	0.67	NA	4	N/A
C1	2.99	0.15	3.00	1.29	1.68	Barrier	2	8
C2	3.85	0.13	4.00	1.07	1.15	Barrier	14	19
C3	4.26	0.15	5.00	1.24	1.55	Barrier	20	26
C4	4.31	0.11	5.00	0.94	0.89	Barrier	21	27
C5	4.21	0.13	5.00	1.06	1.13	Barrier	19	25
C6	3.78	0.14	4.00	1.18	1.39	Barrier	10	16
C7	4.19	0.13	5.00	1.12	1.26	Barrier	18	24
C8	4.58	0.08	5.00	0.71	0.50	Barrier	24	30
C9	3.53	0.15	4.00	1.27	1.60	Barrier	5	11
C10	3.85	0.12	4.00	1.02	1.03	Barrier	13	20
C11	4.42	0.11	5.00	0.95	0.89	Barrier	23	29
C12	3.88	0.15	4.00	1.31	1.72	Barrier	16	22
C13	4.85	0.06	5.00	0.49	0.24	Barrier	25	31
C14	4.35	0.13	5.00	1.06	1.13	Barrier	22	28
C15	3.76	0.13	4.00	1.08	1.17	Barrier	9	15
C16	2.97	0.17	3.00	1.47	2.17	Barrier	1	7
C17	3.04	0.18	3.00	1.51	2.27	Barrier	3	9
C18	3.76	0.17	5.00	1.48	2.18	Barrier	8	14
C19	3.81	0.17	5.00	1.47	2.16	Barrier	12	18
C20	3.72	0.17	4.50	1.47	2.15	Barrier	7	13
C21	3.88	0.17	5.00	1.43	2.05	Barrier	15	21
C25	3.43	0.15	4.00	1.28	1.63	Barrier	4	10
C26	3.79	0.13	4.00	1.11	1.24	Barrier	11	17
C27	3.97	0.14	4.00	1.16	1.35	Barrier	17	23
C28	3.60	0.15	4.00	1.24	1.54	Barrier	6	12
D18	2.72	0.13	3.00	1.13	1.27	Barrier	5	5
D19	2.63	0.13	3.00	1.11	1.22	Barrier	2	2
D21	2.64	0.13	3.00	1.10	1.22	Barrier	3	3
D22	2.53	0.13	3.00	1.14	1.29	Barrier	1	1
D23	2.71	0.15	3.00	1.26	1.59	Barrier	4	4
D24	2.79	0.13	3.00	1.10	1.21	Barrier	6	6

Table D-5 Descriptive statistics for ECO dataset (Number of observations = 72)

Variable	Mean	Standard Error	Median	Standard Deviation	Variance	Role	In group rank (Stage 1)	Overall rank (Stage 2)
C24	4.01	0.13	4.00	1.13	1.28	Barrier	1	12
D1	3.68	0.15	4.00	1.27	1.60	Driver	1	11
D2	2.69	0.11	3.00	0.96	0.92	Barrier	2	2
D3	2.58	0.12	3.00	1.06	1.12	Barrier	1	1
D4	2.81	0.12	3.00	1.04	1.09	Barrier	4	4
D5	2.82	0.13	3.00	1.08	1.16	Barrier	5	5
D6	2.78	0.14	3.00	1.21	1.47	Barrier	3	3
D7	2.82	0.14	3.00	1.18	1.39	Barrier	6	6
D8	2.94	0.13	3.00	1.11	1.24	Barrier	10	10
D9	2.89	0.13	3.00	1.08	1.17	Barrier	9	9
D10	2.86	0.14	3.00	1.15	1.33	Barrier	8	8
D25	2.83	0.14	3.00	1.19	1.41	Barrier	7	7

Table D-6 Descriptive statistics for SOC dataset (Number of observations = 72)

Variable	Mean	Standard Error	Median	Standard Deviation	Variance	Role	In group rank (Stage 1)	Overall rank (Stage 2)
C22	3.24	0.16	3.00	1.40	1.96	Barrier	1	5
C23	3.81	0.15	4.00	1.27	1.62	Barrier	2	10
D11	3.46	0.14	4.00	1.16	1.35	Driver	1	9
D12	3.36	0.15	4.00	1.25	1.56	Driver	3	7
D13	3.10	0.15	3.00	1.30	1.69	Driver	5	4
D14	3.44	0.17	4.00	1.43	2.05	Driver	2	8
D15	2.71	0.12	3.00	0.98	0.97	Barrier	2	2
D16	2.97	0.14	3.00	1.15	1.32	Barrier	3	3
D17	3.32	0.15	3.00	1.31	1.71	Driver	4	6
D20	2.65	0.12	3.00	0.98	0.96	Barrier	1	1

Appendix E The R codes

E.1 The code to estimate the missing data

Depending on the dataset, `dataset1` represents the original TEC, ECO, and SOC dataset.

```
>library(missMDA)

>library(mice)

>library(readxl)

>dataset=read_xlsx("dataset1.xlsx")

>col_names = names(dataset1)

>dataset1[,col_names] <- lapply(dataset1[,col_names] , factor)

>res.ncp = estim_ncpMCA(dataset1,method.cv="loo")#optionally use “kfold”

>plot(names(res.ncp$criterion),res.ncp$criterion,xlab="number of
dimensions",ylab="cv error")

>res.MIMCA = MIMCA(dataset1,ncp = res.ncp$ncp)

>imp=prelim(res.MIMCA,dataset1)

>dataset2 = complete(imp, action = "long", include = TRUE)

>dataset2_imputed=complete(mice(dataset2))
```

E.2 The code to balance the datasets

```
>library(mlr)

>task1 = makeClassifTask(data = as.data.frame(dataset1), target =
"response")

>dataset2= smote(task1, rate, nn = 5L, standardize = TRUE, alt.logic = TRUE)
```

E.3 The code to randomly split the datasets to training and testing sets

```
>library(caTools)

>set.seed(88)

>split=sample.split(dataset2$response,SplitRatio = 0.70)

>train=subset(dataset2,split1==TRUE)

>test=subset(dataset2,split1==FALSE)
```

E.4 The code to install packages required for feature selection using filter methods

```
>install.packages("mlr")
>install.packages("FSelector")
>install.packages("randomForest")
>install.packages("party")
>install.packages("praznik")
>install.packages("xtable")
>install.packages("xlsx")
>library(mlr)
>library(FSelector)
>library(randomForest)
>library(party)
>library(praznik)
>library(xtable)
>library(xlsx)
```

E.5 The code to rank features using filter methods

```
>task = makeClassifTask(data = as.data.frame(train), target = "response")
>gfvd=generateFilterValuesData(task, method =
c("party_cforest.importance","FSelector_chi.squared","FSelector_information
.gain","FSelector_gain.ratio","kruskal.test","praznik_MRMR","FSelector_oneR
","randomForest_importance","FSelector_relief","FSelector_symmetrical.uncer
tainty"))
>featureScors=xtable(gfvd$data)
>write.xlsx(featureScors, file = "featureScors.xlsx")
```

E.6 The code to select features using the Boruta method

```
>library(Boruta)
>feature=Boruta(response~., data=train, maxRuns = 10000)
```

E.7 The code to select features using RFE technique

```
library(caret)
library(randomForest)
library(ipred)
library(klaR)
# x=train1 (excluding response)
# y=response
#List of functions used: rfFuncs, nbFuncs, treebagFuncs, caretFuncs

ctrl <- rfeControl(functions = (see list of functions used)
                   method = "repeatedcv",
                   repeats = 5,
                   verbose = FALSE)

Profile <- rfe(x, y,
                 sizes = c(1:20),
                 rfeControl = ctrl)
```

E.8 The code to install packages required for the development of the predictive models

Below is the list of all required packages to develop the predictive models. After installing these packages, it is necessary to call the package using the library() function in R. Script E.8 (Appendix E) is used to call the required packages.

Scripts E.9 to E.21 were used to develop the predictive models for the list of machine learning methods in Table 5.10. In these scripts, using the `makeLearner()` function, the class of learner and the type of prediction is specified. Next, using the `makeClassifTask()` function, the training and testing datasets (Section 5.3), as well as the targeting response vector are defined for use to fit the models and perform predictions. Then, using the `train()` function, the predictive model is developed by fitting the learner to the training dataset. Finally, using the `predict()` function, the predictions of the models on the unseen data are performed.

```
library(ada)
library(bartMachine)
library(caret)
library(clusterGeneration)
library(devtools)
library(e1071)
library(kernlab)
library(kknn)
library(mlr)
library(neuralnet)
library(nnet)
library(randomForest)
library(ROCR)
library(rpart)
library(rpart.plot)
library(RWeka)
library(rJava)
```

E.9 The code to develop the KNN models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
#Estimating the number of neighbours using the rminer package.

>s=list(smethod="grid",search=mparheuristic("kknn",n=10),convex=0,metric="A
UC",method=c("holdout",2/3,123))

>model1=fit(E1C ~ ., data = train, model="kknn",task="prob",search=s)

>print(model1@mpar)

#Using mlr package to develop the model

>obj_mlr_knn = makeLearner("classif.kknn", predict.type = "prob")

obj_mlr_knn$par.set$pars$k=s #s is equal to 6, 5, or 8 for the TEC, ECO,
and SOC, respectively

>train1_task = makeClassifTask(data = as.data.frame(train), target)

>model1_knn = train(obj_mlr_knn, train1_task)

>test1_task = makeClassifTask(data = as.data.frame(test), target)

>predTest1_knn = predict(model1_knn, test1_task, predict.type = "prob")

>calculateConfusionMatrix(predTest1_knn)

>ROCRpredTest1 = asROCRPrediction(predTest1_knn)

>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")

>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))

>knn_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.10 The code to develop the LR models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_lr = makeLearner("classif.logreg", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_lr = train(obj_mlr_lr, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_lr = predict(model1_lr, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_lr)
>ROCRpredTest1 = asROCRPrediction(predTest1_lr)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>lr_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.11 The code to develop the LDA models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_lda = makeLearner("classif.lda", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_lda = train(obj_mlr_lda, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_lda = predict(model1_lda, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_lda)
>ROCRpredTest1 = asROCRPrediction(predTest1_lda)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>lda_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.12 The code to develop the QDA models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_qda = makeLearner("classif.qda", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_qda = train(obj_mlr_qda, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_qda = predict(model1_qda, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_qda)
>ROCRpredTest1 = asROCRPrediction(predTest1_qda)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>qda_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.13 The code to develop the NB models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_nb = makeLearner("classif.naiveBayes", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_nb = train(obj_mlr_nb, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_nb = predict(model1_nb, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_nb)
>ROCRpredTest1 = asROCRPrediction(predTest1_nb)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>nb_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.14 The code to develop the DT models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_dt = makeLearner("classif.rpart", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_dt = train(obj_mlr_dt, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_dt = predict(model1_dt, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_dt)
>ROCRpredTest1 = asROCRPrediction(predTest1_dt)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>dt_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
>rpart.plot.version1(model1_dt$learner.model)
```

E.15 The code to develop the RF models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_rf = makeLearner("classif.randomForest", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_rf = train(obj_mlr_rf, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_rf = predict(model1_rf, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_rf)
>ROCRpredTest1 = asROCRPrediction(predTest1_rf)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>rf_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.16 The code to develop the AB models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_ab = makeLearner("classif.ada", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_ab = train(obj_mlr_ab, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_ab = predict(model1_ab, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_ab)
>ROCRpredTest1 = asROCRPrediction(predTest1_ab)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>ab_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.17 The code to develop the BM models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_bm = makeLearner("classif.bartMachine", predict.type = "prob")
>train1_task = makeClassifTask(data = as.data.frame(train), target)
>model1_bm = train(obj_mlr_bm, train1_task)
>test1_task = makeClassifTask(data = as.data.frame(test), target)
>predTest1_bm = predict(model1_bm, test1_task, predict.type = "prob")
>calculateConfusionMatrix(predTest1_bm)
>ROCRpredTest1 = asROCRPrediction(predTest1_bm)
>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")
>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))
>bm_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.18 The code to develop the ANN models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset. The hyperparameters (size and decay) for each of the datasets are as follows: TEC (size=9, decay=0.09), ECO (size=9, decay=0.08), SOC (size=8, decay=0.04). For k-fold cross-validation (Tables 6.1 to 6.3), `dataset` represents the entire TEC, ECO, and SOC observations (Section 5.2), and not the training set.

```
#Estimating the hyperparameters using the caret package.

>fitControl = trainControl(method = "repeatedcv",number = 10,repeats =
5,classProbs = TRUE, summaryFunction = twoClassSummary)

>nnetGrid <= expand.grid(size = seq(from = 1, to = 10, by = 1),decay =
seq(from = 0, to = 0.5, by = 0.01))

>nnetFit = train(target ~ .,data = train,method = "nnet",metric =
"ROC",trControl = fitControl,tuneGrid = nnetGrid,verbose = FALSE)

#Using mlr package to develop the model

>model1=fit(target ~ ., data = train ,model="mlpe",task="prob",size ,decay)

>predTest1=predict(model1,newdata = test)

>print(mmetric(test$target,predTest1,"CONF",TC=2))

>print(mmetric(test$target,predTest1,metric =
c("ACC","AUC","TPR","TNR"),TC=2))

>print(mmetric(test$target,predTest1,"ROC",TC=2))

>ROCRpredTest1 = ROCR::prediction(predTest1[,2], test$target)

>ROCRperfTest1 = ROCR::performance(ROCRpredTest1,"tpr","fpr")

>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7), main="ANN
(TEC) Testing dataset ROC curve")

>aucTest1 = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)

>aucTest1

#10-fold Cross-validation for Tables 9.1, 9.2, and 9.3

>M=crossvaldata(target ~
.,data=dataset,fit,predict,seed=88,model="mlpe",task="prob",size,decay)

>print(mmetric(dataset$TEC,M$cv.fit,metric =
c("ACC","AUC","TPR","TNR"),TC=2))
```

E.19 The code to develop the GP models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_gp = makeLearner("classif.gausspr", predict.type = "prob")

>train1_task = makeClassifTask(data = as.data.frame(train), target)

>model1_gp = train(obj_mlr_gp, train1_task)

>test1_task = makeClassifTask(data = as.data.frame(test), target)

>predTest1_gp = predict(model1_gp, test1_task, predict.type = "prob")

>calculateConfusionMatrix(predTest1_gp)

>ROCRpredTest1 = asROCRPrediction(predTest1_gp)

>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")

>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))

>gp_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.20 The code to develop the PRL models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
>obj_mlr_prl = makeLearner("classif.JRip", predict.type = "prob")

>train1_task = makeClassifTask(data = as.data.frame(train), target)

>model1_prl = train(obj_mlr_prl, train1_task)

>test1_task = makeClassifTask(data = as.data.frame(test), target)

>predTest1_prl = predict(model1_prl, test1_task, predict.type = "prob")

>calculateConfusionMatrix(predTest1_prl)

>ROCRpredTest1 = asROCRPrediction(predTest1_prl)

>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")

>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))

>prl_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)

>as.matrix(scan(text=.jcall(model1_prl$learner.model$classifier,"S",
"toString") ,sep="\n", what="") )[-c(1:2, 20), ,drop=FALSE]
```

E.21 The code to develop the SVM models

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset.

```
> train1_task = makeClassifTask(data = as.data.frame(train), target)

>num_ps_Tuning = makeParamSet(

  makeNumericParam("C", lower = -10, upper = 10, trafo = function(x) 10^x),

  makeNumericParam("sigma", lower = -10, upper = 10, trafo = function(x)
10^x))

>ctrlTuning = makeTuneControlRandom(maxit = 100L)

>rdescTuning = makeResampleDesc("CV", iters = 10L)

>resTuning = tuneParams("classif.ksvm", task = train1_task, resampling =
rdescTuning, par.set = mum_ps_Tuning, control = ctrlTuning, measures =
list(acc, setAggregation(acc, test.sd)))

>obj_mlr_svm = setHyperPars(makeLearner("classif.ksvm", predict.type =
"prob"), C = resTuning$x$C, sigma = resTuning$x$sigma)

>test1_task = makeClassifTask(data = as.data.frame(test), target)

>model1_svm = train(obj_mlr_svm, train1_task)

>predTest1_svm = predict(model1_svm, test1_task, predict.type = "prob")

>calculateConfusionMatrix(predTest1_svm)

>ROCRpredTest1 = asROCRPrediction(predTest1_svm)

>ROCRperfTest1 = ROCR::performance(ROCRpredTest1, "tpr", "fpr")

>plot(ROCRperfTest1,
colorize=TRUE,print.cutoffs.at=seq(0,1,0.1),text.adj=c(-0.2,1.7))

>svm_auc = as.numeric(ROCR::performance(ROCRpredTest1, "auc")@y.values)
```

E.22 The code to assess the performance of the BSE-RPMs using the kfCV method

In `makeClassifTask()` function, `target` is replaced with `target = "E1C"` for the TEC dataset, `target = "E2C"` for the ECO dataset, and `target = "E3C"` for the SOC dataset. Moreover, in `resample()` function, learner is replaced with `obj_mlr_knn` (Script E.9), `obj_mlr_lr` (Script E.10), `obj_mlr_lda` (Script E.11), `obj_mlr_qda` (Script E.12), `obj_mlr_nb` (Script E.13), `obj_mlr_dt` (Script E.14), `obj_mlr_rf` (Script E.15), `obj_mlr_bm` (Script E.16), `obj_mlr_ab` (Script E.17), `obj_mlr_ann` (Script E.18), `obj_mlr_gp` (Script E.19), `obj_mlr_prl` (Script E.20), `obj_mlr_svm` (Script E.21).

```
>task = makeClassifTask(data = as.data.frame(reuse2), target)

>rdesc = makeResampleDesc("CV", iters = 10, predict = "both")

>r = resample(learner, task, rdesc, measures = list(mmce, acc, fpr, fnr,
tnr, tpr, auc))
```

E.23 The code to perform Sensitivity Analysis and open the best-practice RF BSER-RPMs

`dataset2` is defined in Script E.1

```
>library(rminer)

>model=fit(response ~ ., data=dataset2,model = "randomForest",task="prob")

>dsa_imp=Importance(model,dataset2,method = "DSA",measure = "AAD")

>list_dsa=list(runs=1,sen=t(dsa_imp$imp),sresponses=dsa_imp$sresponses)

>mgraph(list_dsa,graph="IMP",leg=names(dataset2),col="white",

       xval=0.0105,main = "TEC-RF BSE-RPM Variable importance levels
(DSA)",metric="ALL",

       axis=c(1),cex = 0.75)
```

E.24 The code to extract rules from the best-practice RF BSER-RPMs

```
>library(inTrees)
>library(randomForest)
>y=train$response
>nonvars=c("response")
>x=train[ , !(names(train) %in% nonvars) ]
>y=as.factor(y)
>x=as.data.frame(x)
>train=as.data.frame(train)
>test=as.data.frame(test)
>model=randomForest(x,y)
>treeList = RF2List(model)
>ruleExec = extractRules(treeList,x)
>ruleExec = unique(ruleExec)
>ruleMetric = getRuleMetric(ruleExec,x,y)
>ruleMetric = pruneRule(ruleMetric,x,y)
>ruleMetric = selectRuleRRF(ruleMetric,x,y)
>learner = buildLearner(ruleMetric,x,y)
>pred = applyLearner(learner,x)
>read = presentRules(learner,colnames(x))
```

Appendix F Outcome of the predictive models

Results of the predictive models (Chapter 5, Section 5.5.3, Table 5.10)

F.1 Predictive models on the TEC dataset

In this section, the results of the models used to predict the technical reusability of the structural elements are presented.

Table F-1 Summary of the results of the TEC BSE-RPMs developed (the validation set approach method).

Predictive model	Type-I error	Type-II error	Specificity	Sensitivity	Overall accuracy	Overall error rate	AUC
KNN	0.03	0.28	0.97	0.72	0.85	0.15	0.95
LR*	0.28	0.17	0.72	0.83	0.78	0.22	0.81
LDA	0.14	0.21	0.86	0.79	0.83	0.17	0.86
QDA	0.07	0.17	0.93	0.83	0.88	0.12	0.96
NB	0.24	0.35	0.76	0.65	0.71	0.29	0.82
DT	0.10	0.41	0.90	0.59	0.74	0.26	0.76
RF	0.00	0.17	1.00	0.83	0.91	0.09	0.98
AB	0.07	0.31	0.93	0.69	0.81	0.19	0.93
BM	0.07	0.38	0.93	0.62	0.78	0.22	0.91
ANN	0.14	0.14	0.86	0.86	0.86	0.14	0.90
GP	0.14	0.31	0.86	0.69	0.78	0.22	0.91
PRL	0.21	0.17	0.79	0.83	0.81	0.19	0.84
SVM	0.07	0.14	0.93	0.86	0.90	0.10	0.97
* The LR BSE-RPM did not converge. Hence, this model is excluded from further analysis.							

F.1.1 TEC dataset K-Nearest Neighbours (KNN) BSE-RPM

Script E.9 (Appendix E) is used to develop the TEC K-Nearest Neighbours (KNN) BSE-RPM.

Table F-2 The confusion matrix of the KNN BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	28	1
Actual reusable (1)	8	21

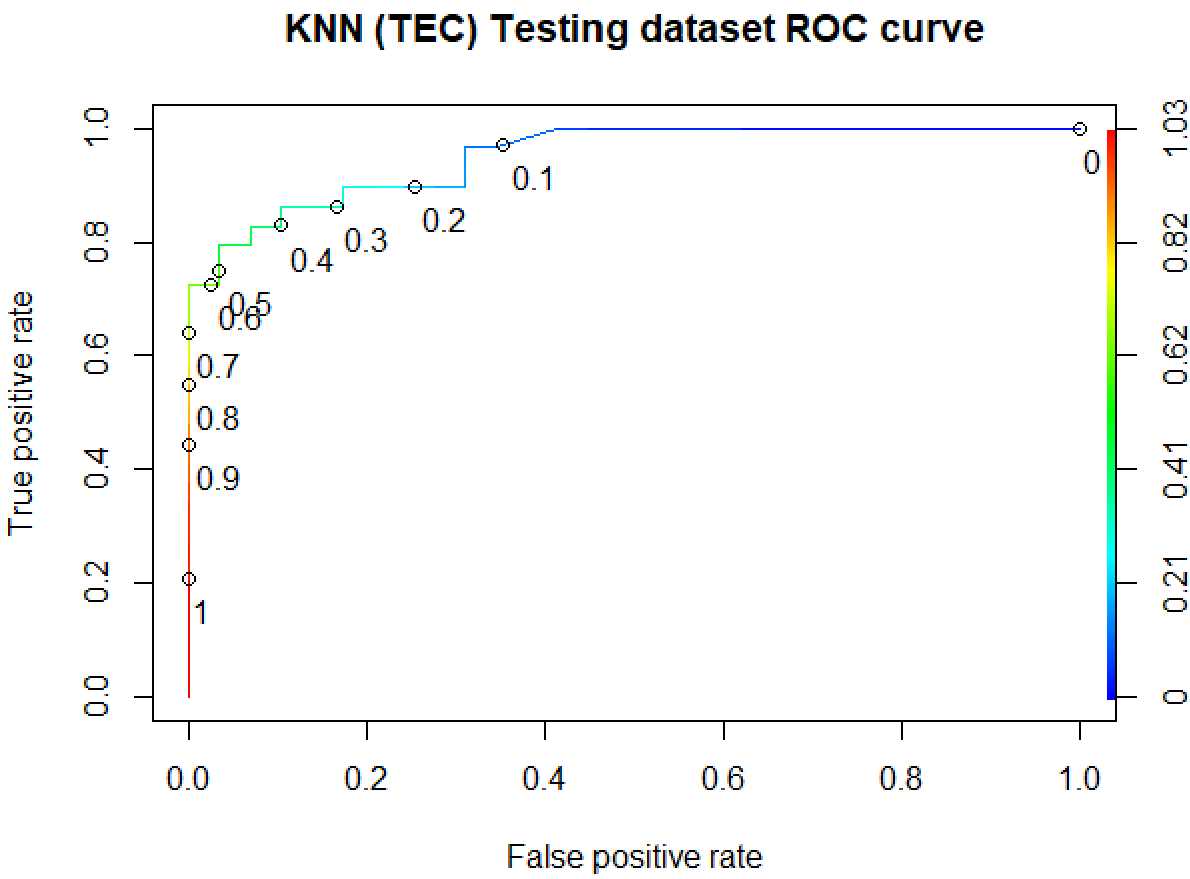

Figure F-1 The ROC curve of the KNN BSE-RPM (TEC dataset)

The **AUC value** for the TEC K-Nearest Neighbours (KNN) BSE-RPM is equal to 0.95.

F.1.2 TEC dataset Logistic Regression (LR) BSE-RPM

Script E.10 (Appendix E) is used to develop the TEC Logistic Regression (LR) BSE-RPM.

Table F-3 The confusion matrix of the LR BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	21	8
Actual reusable (1)	5	24

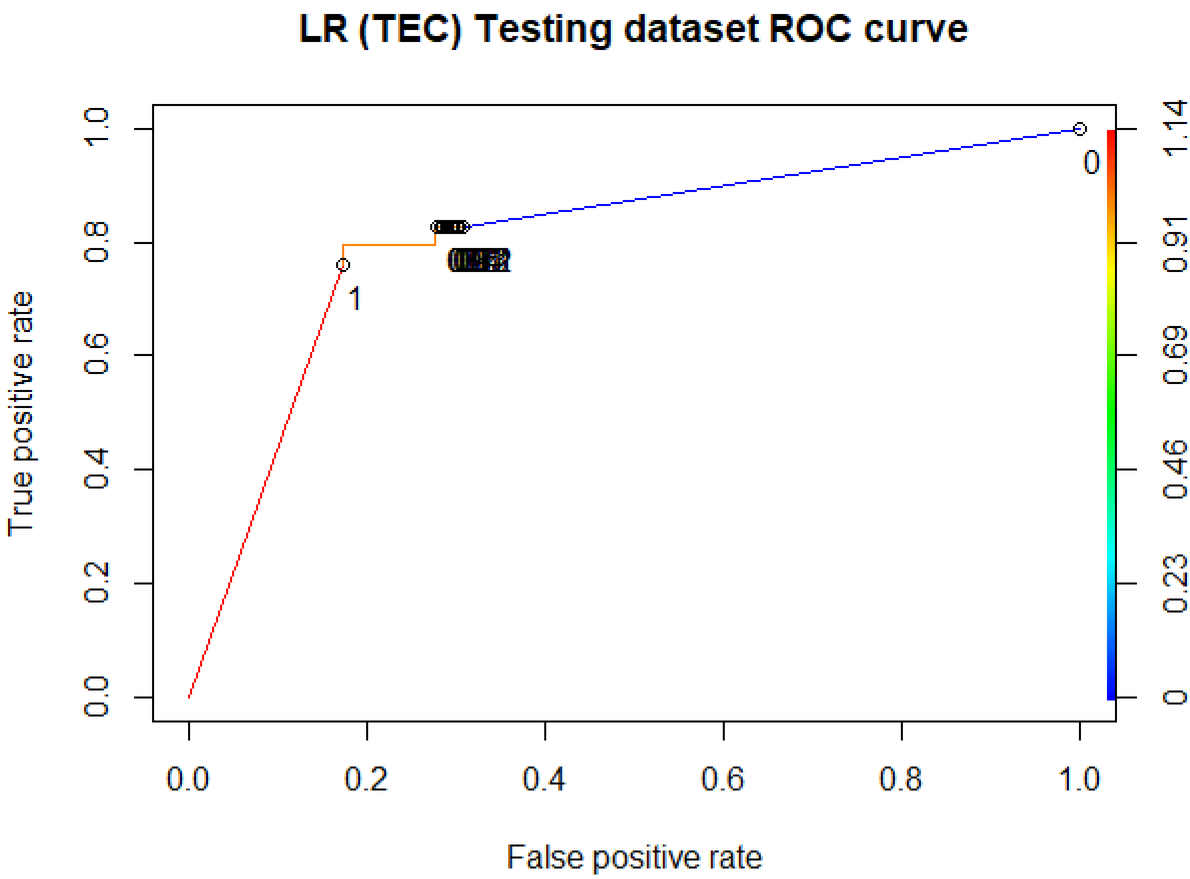

Figure F-2 The ROC curve of the LR BSE-RPM (TEC dataset)

The **AUC value** for the TEC Logistic Regression (LR) BSE-RPM is equal to 0.81. However, the model for the TEC Logistic Regression (LR) BSE-RPM did not converge. Hence, this model is not considered during the final evaluation (Chapter 6).

F.1.3 TEC dataset Linear Discriminant Analysis (LDA) BSE-RPM

Script E.11 (Appendix E) is used to develop the TEC Linear Discriminant Analysis (LDA) BSE-RPM.

Table F-4 The confusion matrix of the LDA BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	25	4
Actual reusable (1)	6	23

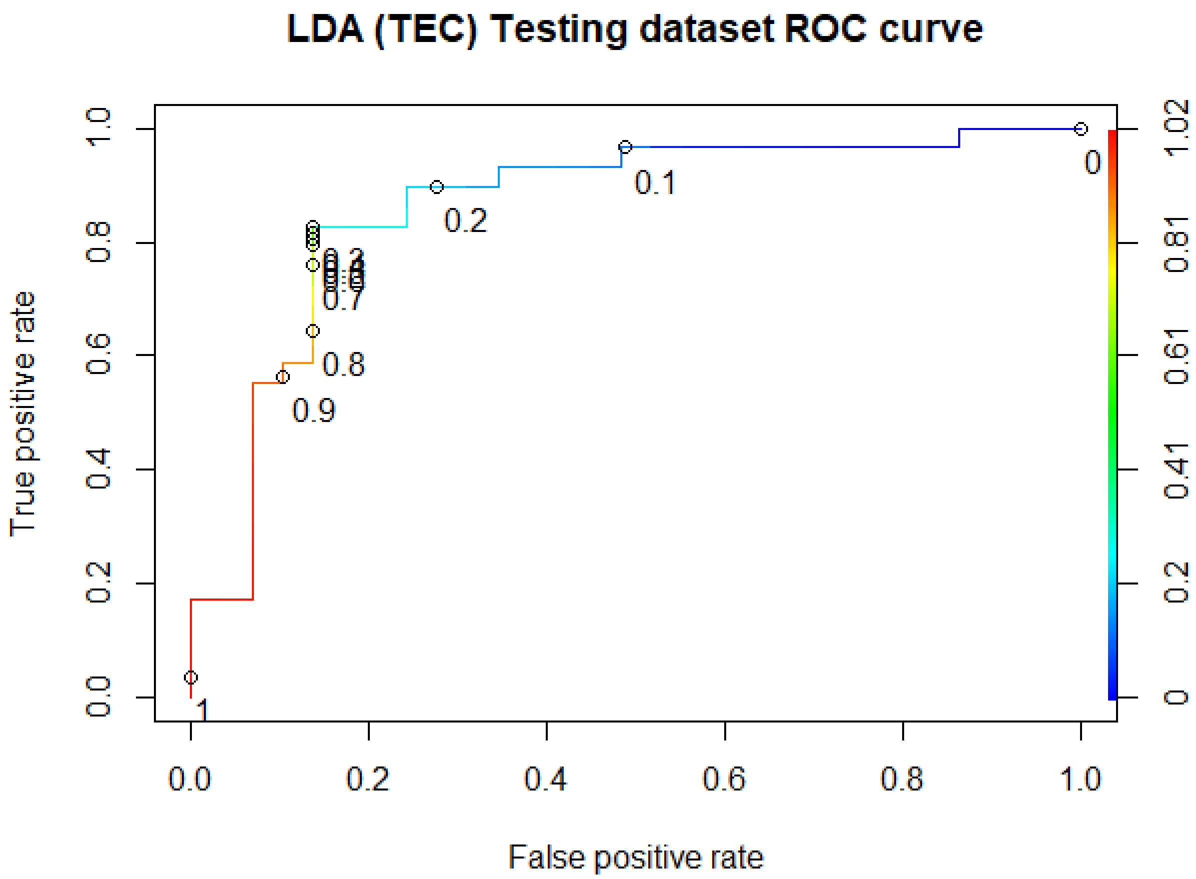

Figure F-3 The ROC curve of the LDA BSE-RPM (TEC dataset)

The **AUC value** for the TEC Linear Discriminant Analysis (LDA) BSE-RPM is equal to 0.862069.

F.1.4 TEC dataset Quadratic Discriminant Analysis (QDA) BSE-RPM

Script E.12 (Appendix E) is used to develop the TEC Quadratic Discriminant Analysis (QDA) BSE-RPM.

Table F-5 The confusion matrix of the QDA BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	2
Actual reusable (1)	5	24

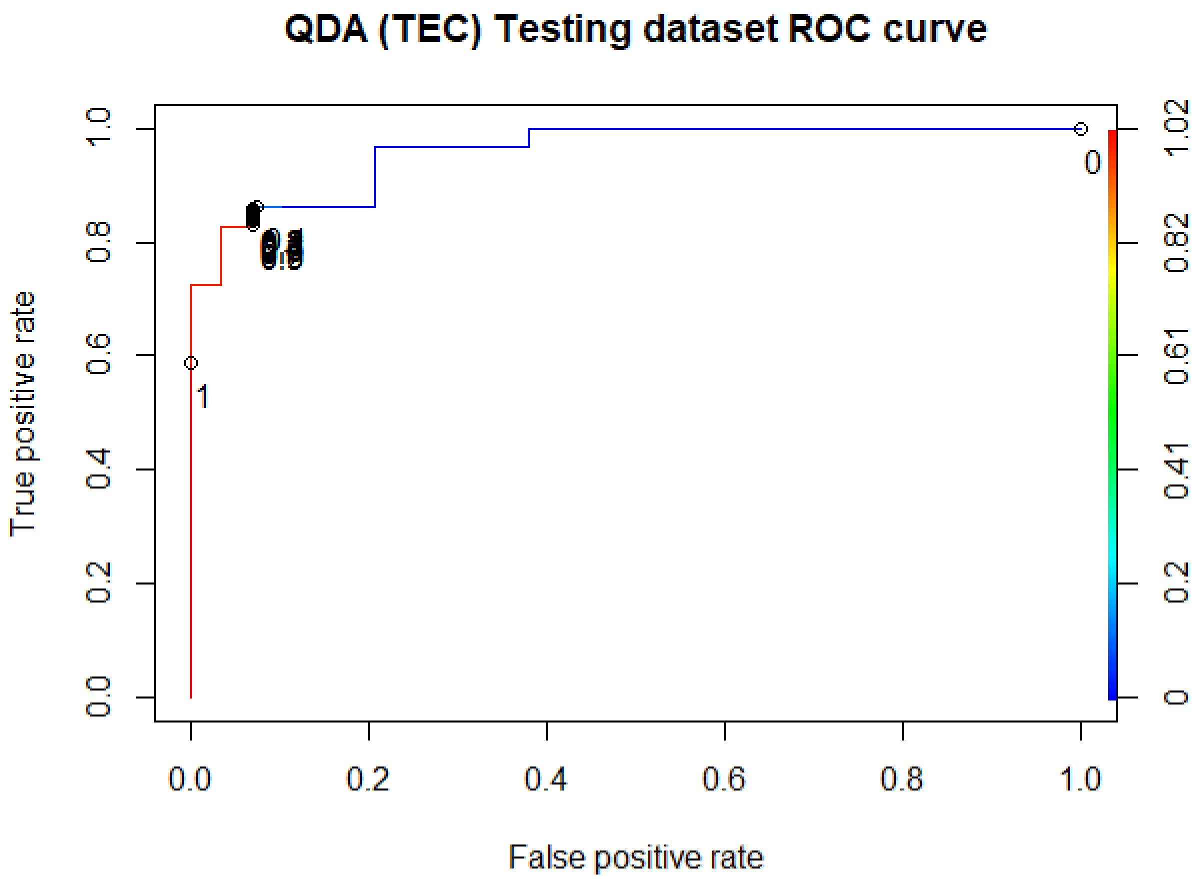

Figure F-4 The ROC curve of the QDA BSE-RPM (TEC dataset)

The **AUC value** for the TEC Quadratic Discriminant Analysis (QDA) BSE-RPM is equal to 0.96.

F.1.5 TEC dataset Naïve Bayes (NB) BSE-RPM

Script E.13 (Appendix E) is used to develop the TEC Naïve Bayes (NB) BSE-RPM.

Table F-6 The confusion matrix of the NB BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	22	7
Actual reusable (1)	10	19

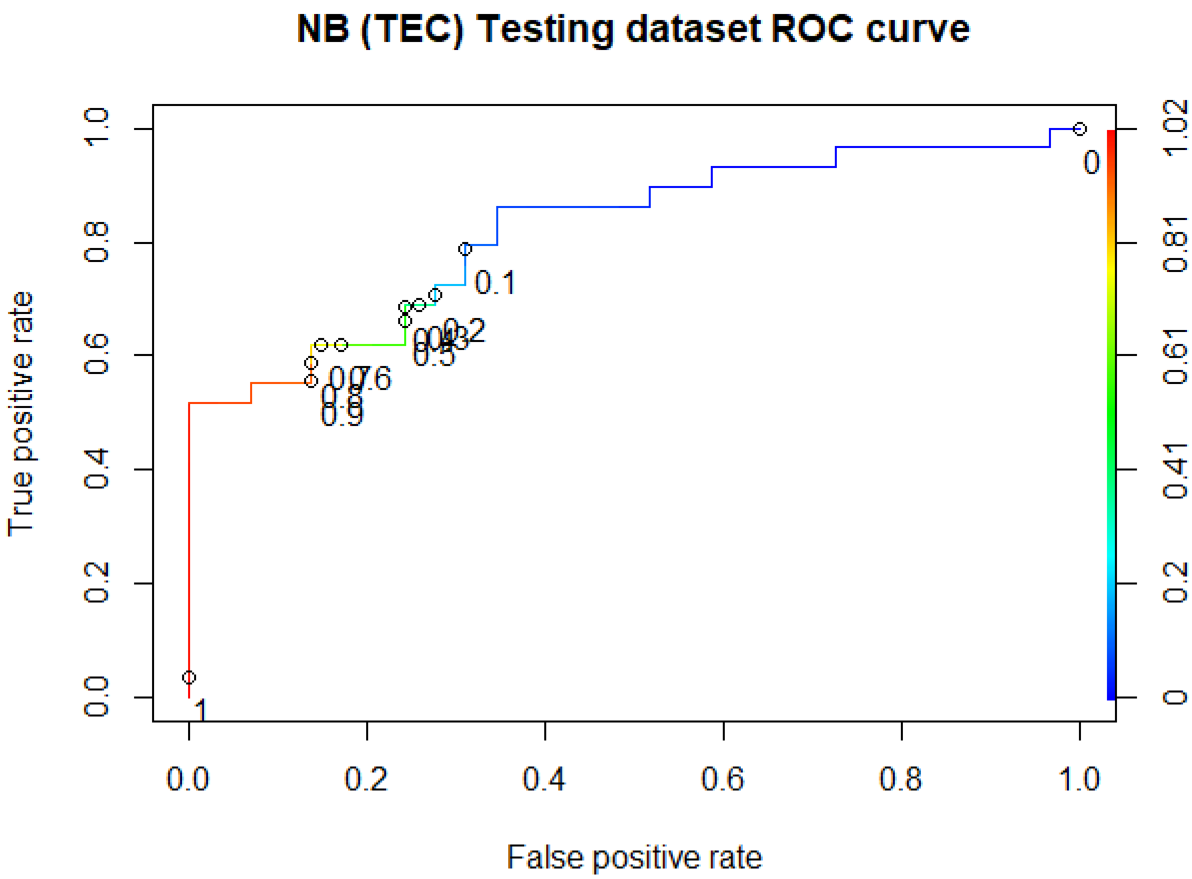

Figure F-5 The ROC curve of the NB BSE-RPM (TEC dataset)

The **AUC value** for the TEC Naïve Bayes (NB) BSE-RPM is equal to 0.82.

F.1.6 TEC dataset Decision Trees (DT) BSE-RPM

Script E.14 (Appendix E) is used to develop the TEC Decision Trees (DT) BSE-RPM.

Table F-7 The confusion matrix of the DT BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	26	3
Actual reusable (1)	12	17

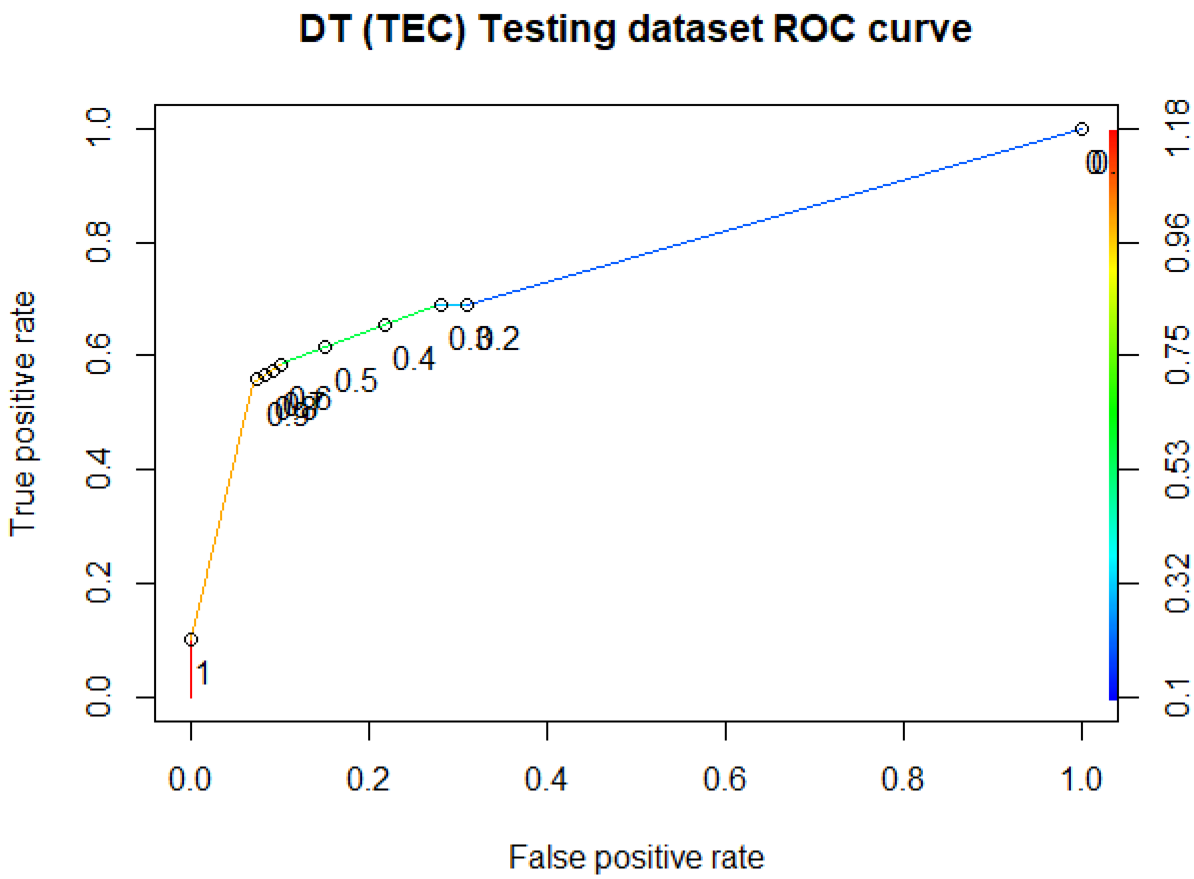

Figure F-6 The ROC curve of the DT BSE-RPM (TEC dataset)

The **AUC value** for the TEC Decision Trees (DT) BSE-RPM is equal to 0.76.

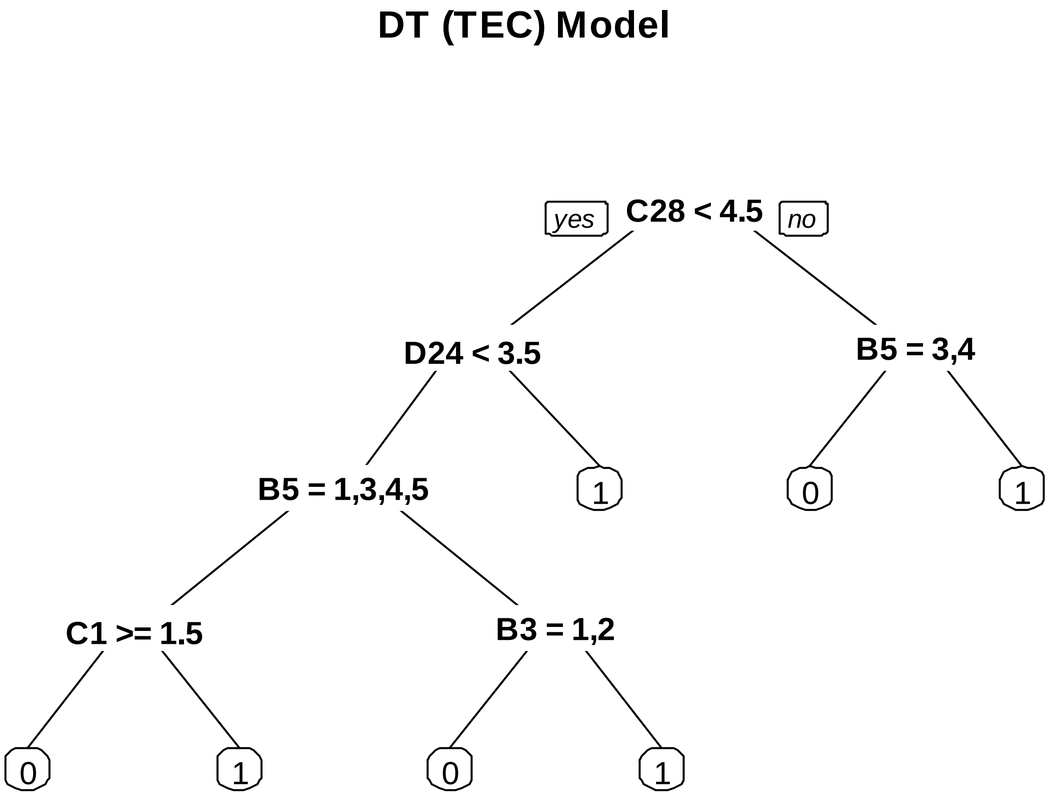

Figure F-7 The DT BSE-RPM Model (TEC dataset)

F.1.7 TEC dataset Random Forests (RF) BSE-RPM

Script E.15 (Appendix E) is used to develop the TEC Random Forests (RF) BSE-RPM.

Table F-8 The confusion matrix of the RF BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	29	0
Actual reusable (1)	5	24

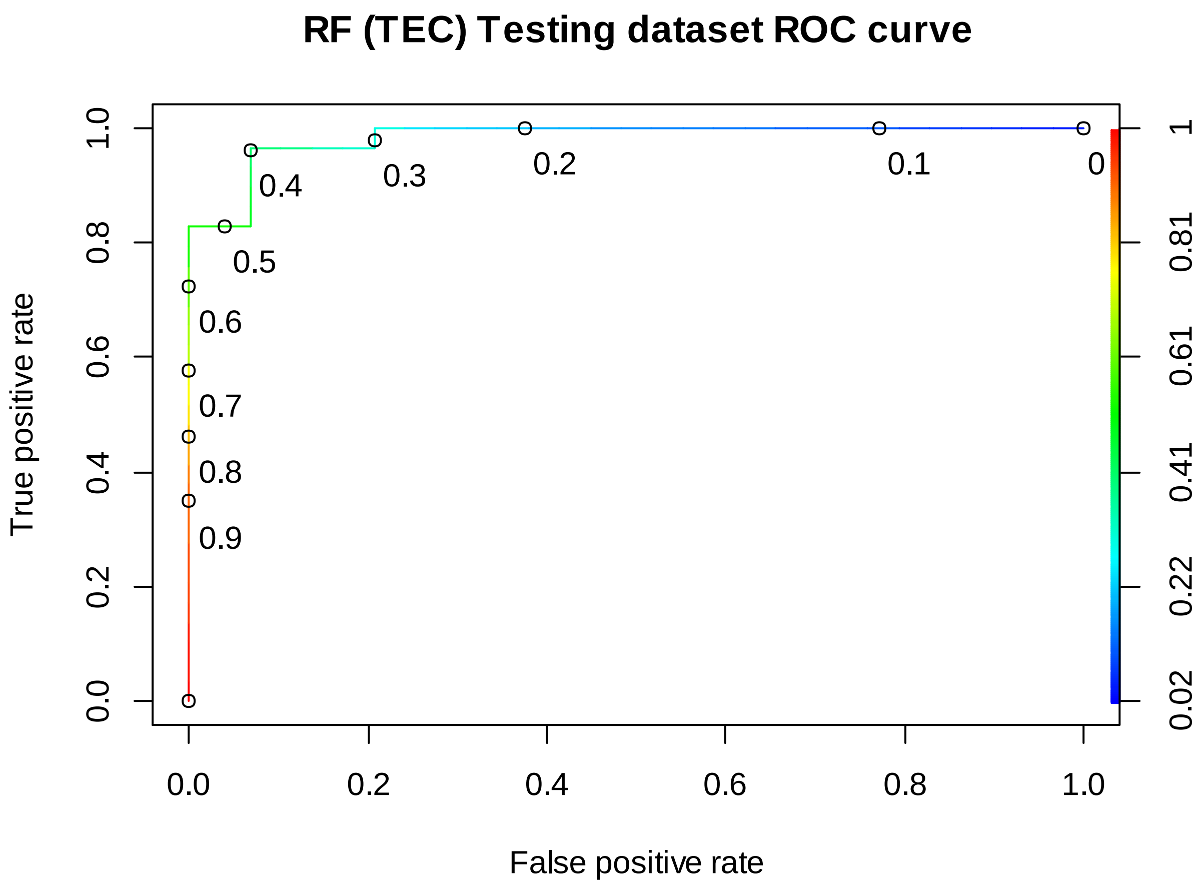

Figure F-8 The ROC curve of the RF BSE-RPM (TEC dataset)

The **AUC value** for the TEC Random Forests (RF) BSE-RPM is equal to 0.98.

F.1.8 TEC dataset Adaptive Boosting (AB) BSE-RPM

Script E.17 (Appendix E) is used to develop the TEC Adaptive Boosting (AB) BSE-RPM.

Table F-9 The confusion matrix of the AB BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	2
Actual reusable (1)	11	18

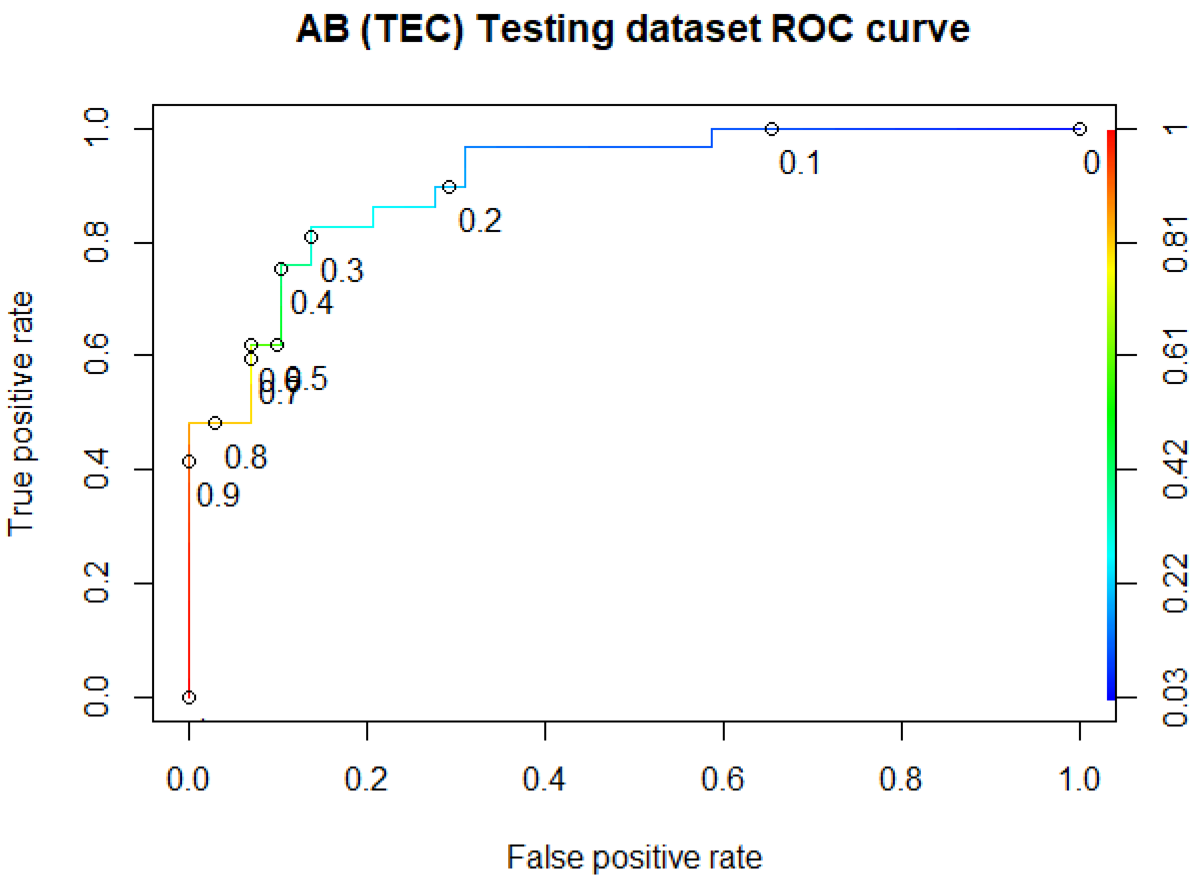

Figure F-9 The ROC curve of the AB BSE-RPM (TEC dataset)

The **AUC value** for the TEC Adaptive Boosting (AB) BSE-RPM is equal to 0.91.

F.1.9 TEC dataset Bart Machine (BM) BSE-RPM

Script E.16 (Appendix E) is used to develop the TEC Bart Machine (BM) BSE-RPM.

Table F-10 The confusion matrix of the BM BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	2
Actual reusable (1)	9	20

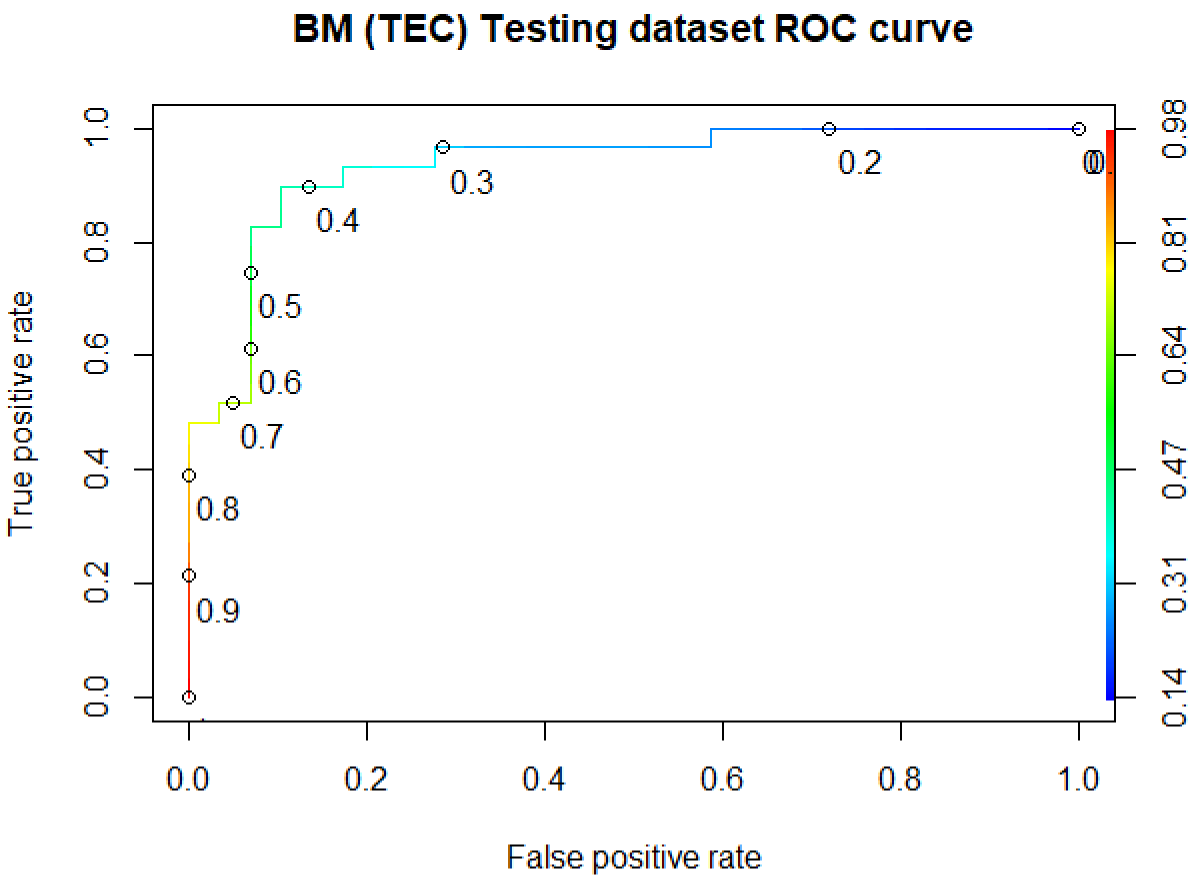

Figure F-10 The ROC curve of the BM BSE-RPM (TEC dataset)

The **AUC value** for the TEC Bart Machine (BM) BSE-RPM is equal to 0.93.

F.1.10 TEC dataset Artificial Neural Networks (ANN) BSE-RPM

Script E.18 (Appendix E) is used to develop the TEC Artificial Neural Networks (ANN) BSE-RPM.

Table F-11 The confusion matrix of the ANN BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	25	4
Actual reusable (1)	4	25

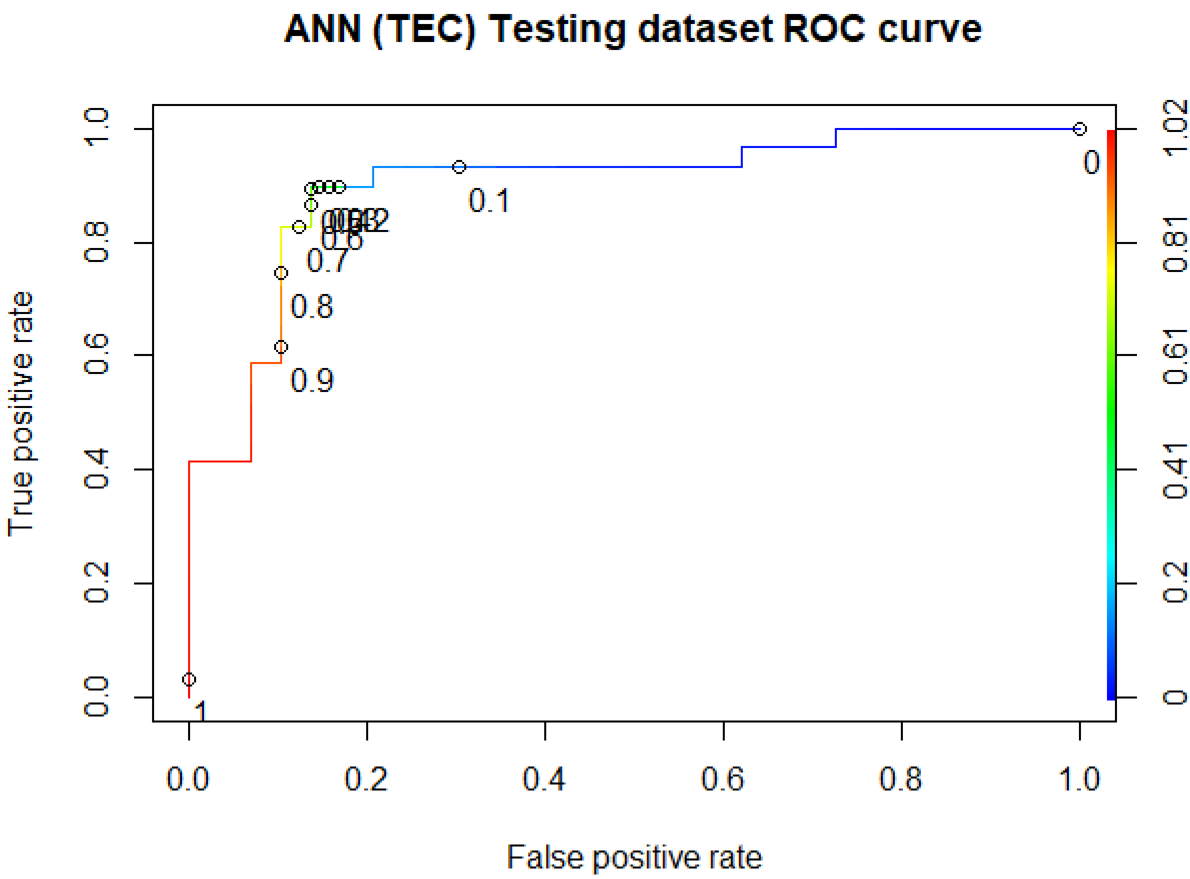

Figure F-11 The ROC curve of the ANN BSE-RPM (TEC dataset)

The **AUC value** for the TEC Artificial Neural Networks (ANN) BSE-RPM is equal to 0.90.

F.1.11 TEC dataset Gaussian Processes (GP) BSE-RPM

Script E.19 (Appendix E) is used to develop the TEC Gaussian Processes (GP) BSE-RPM.

Table F-12 The confusion matrix of the GP BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	25	4
Actual reusable (1)	9	20

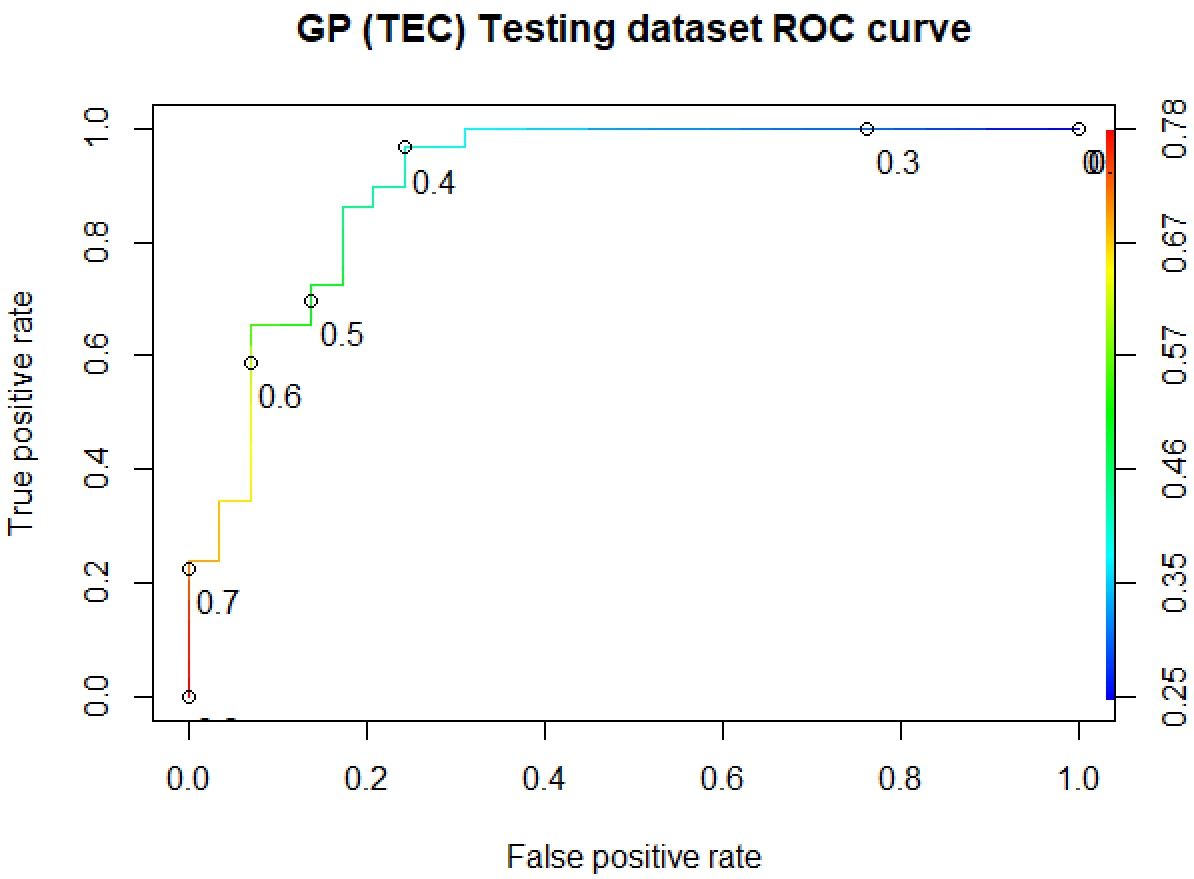

Figure F-12 The ROC curve of the GP BSE-RPM (TEC dataset)

The **AUC value** for the TEC Gaussian Processes (GP) BSE-RPM is equal to 0.91.

F.1.12 TEC dataset Propositional Rule Learner (PRL) BSE-RPM

Script E.20 (Appendix E) is used to develop the TEC Propositional Rule Learner (PRL) BSE-RPM.

Table F-13 The confusion matrix of the PRL BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	23	6
Actual reusable (1)	5	24

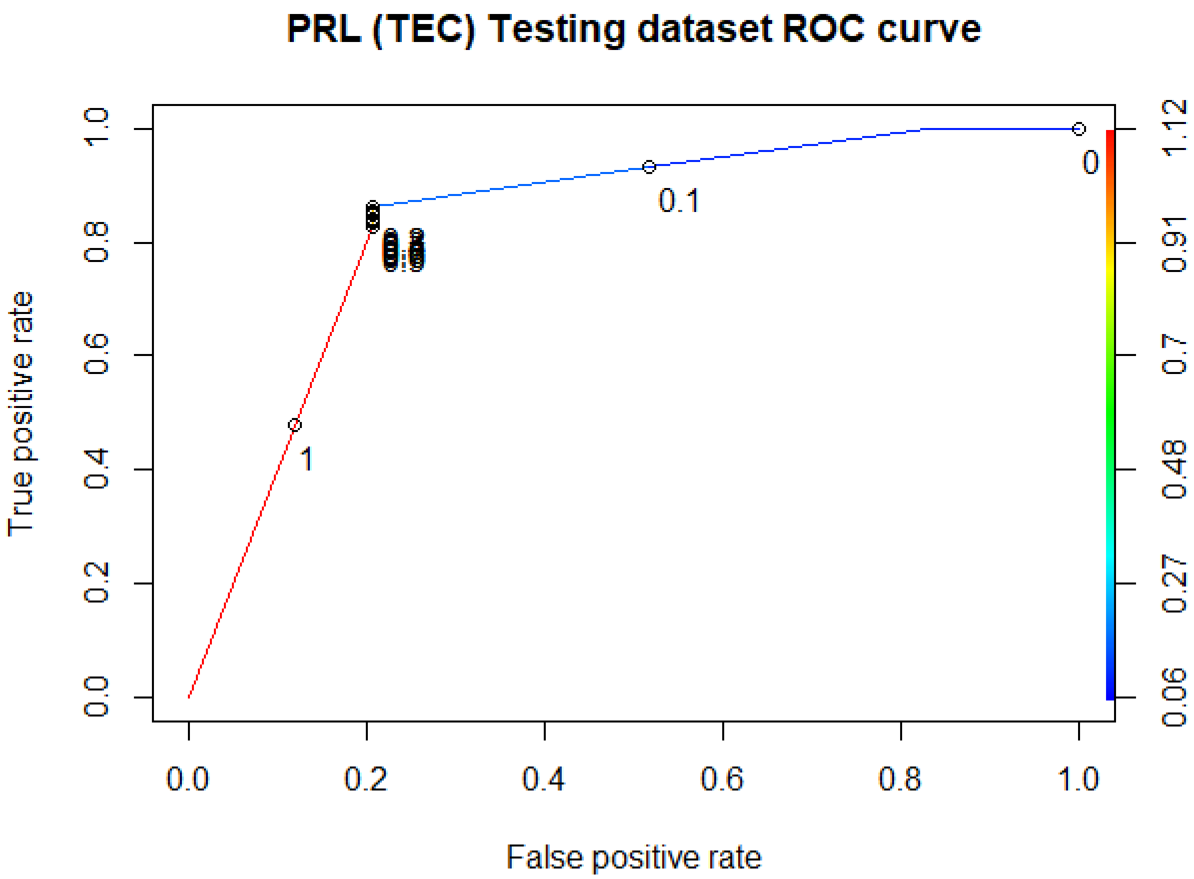

Figure F-13 The ROC curve of the PRL BSE-RPM (TEC dataset)

The **AUC value** for the TEC Propositional Rule Learner (PRL) BSE-RPM is equal to 0.84.

Table F-14 The rules set of the PRL BSE-RPM (TEC dataset)

Rule number (to be considered in order)	Rule	Result
1st	If: (C25 <= 3) and (C9 >= 3) and (C20 >= 2) and (C17 <= 4)	Then: E1C=0 (34.0/3.0)

Rule number (to be considered in order)	Rule	Result
2nd	Else if: (C28 <= 3) and (C15 = 5)	Then: E1C=0 (17.0/1.0)
3rd	Else if: (C6 <= 3) and (B9 <= 3)	Then: E1C=0 (10.0/1.0)
4th	Else if: (B3 = 2) and (B7 <= 4)	Then: E1C=0 (6.0/1.0)
5th	Else if none	Then: E1C=1 (67.0/6.0)

F.1.13 TEC dataset Support Vector Machines (SVM) BSE-RPM

Script E.21 (Appendix E) is used to develop the TEC Support Vector Machines (SVM) BSE-RPM.

Table F-15 The confusion matrix of the SVM BSE-RPM (TEC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	2
Actual reusable (1)	4	25

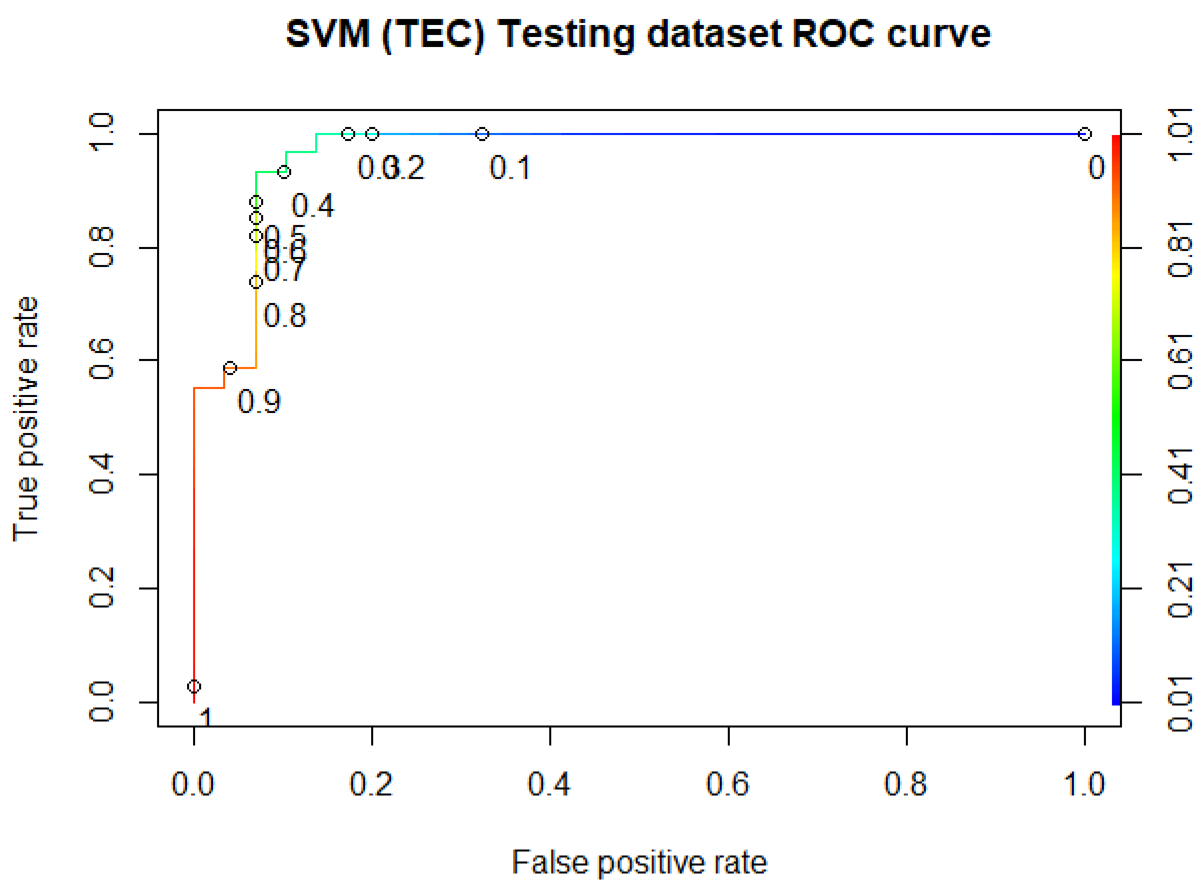

Figure F-14 The ROC curve of the SVM BSE-RPM (TEC dataset)

The **AUC value** for the TEC Support Vector Machines (SVM) BSE-RPM is equal to 0.97.

F.2 Predictive models on the ECO dataset

In this section, the results of the models used to predict the economic reusability of the structural elements are presented.

Table F-16 Summary of the results of the ECO BSE-RPMs developed (the validation set approach method).

Predictive model	Type-I error	Type-II error	Specificity	Sensitivity	Overall accuracy	Overall error rate	AUC
KNN	0.00	0.30	1.00	0.70	0.86	0.14	0.96
LR	0.21	0.30	0.79	0.70	0.75	0.25	0.81
LDA	0.25	0.37	0.75	0.63	0.69	0.31	0.79
QDA	0.21	0.26	0.79	0.74	0.76	0.24	0.83
NB	0.32	0.30	0.68	0.70	0.69	0.31	0.77
DT	0.25	0.19	0.75	0.81	0.78	0.22	0.80
RF	0.00	0.30	1.00	0.70	0.86	0.14	0.98
AB	0.04	0.33	0.96	0.67	0.82	0.18	0.94
BM	0.00	0.33	1.00	0.67	0.84	0.16	0.90
ANN	0.00	0.22	1.00	0.78	0.89	0.11	0.96
GP	0.07	0.37	0.93	0.63	0.78	0.22	0.86
PRL	0.25	0.33	0.75	0.67	0.71	0.29	0.72
SVM	0.07	0.15	0.93	0.85	0.89	0.11	0.95

F.2.1 ECO dataset K-Nearest Neighbours (KNN) BSE-RPM

Script E.9 (Appendix E) is used to develop the ECO K-Nearest Neighbours (KNN) BSE-RPM.

Table F-17 The confusion matrix of the KNN BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	28	0
Actual reusable (1)	8	19

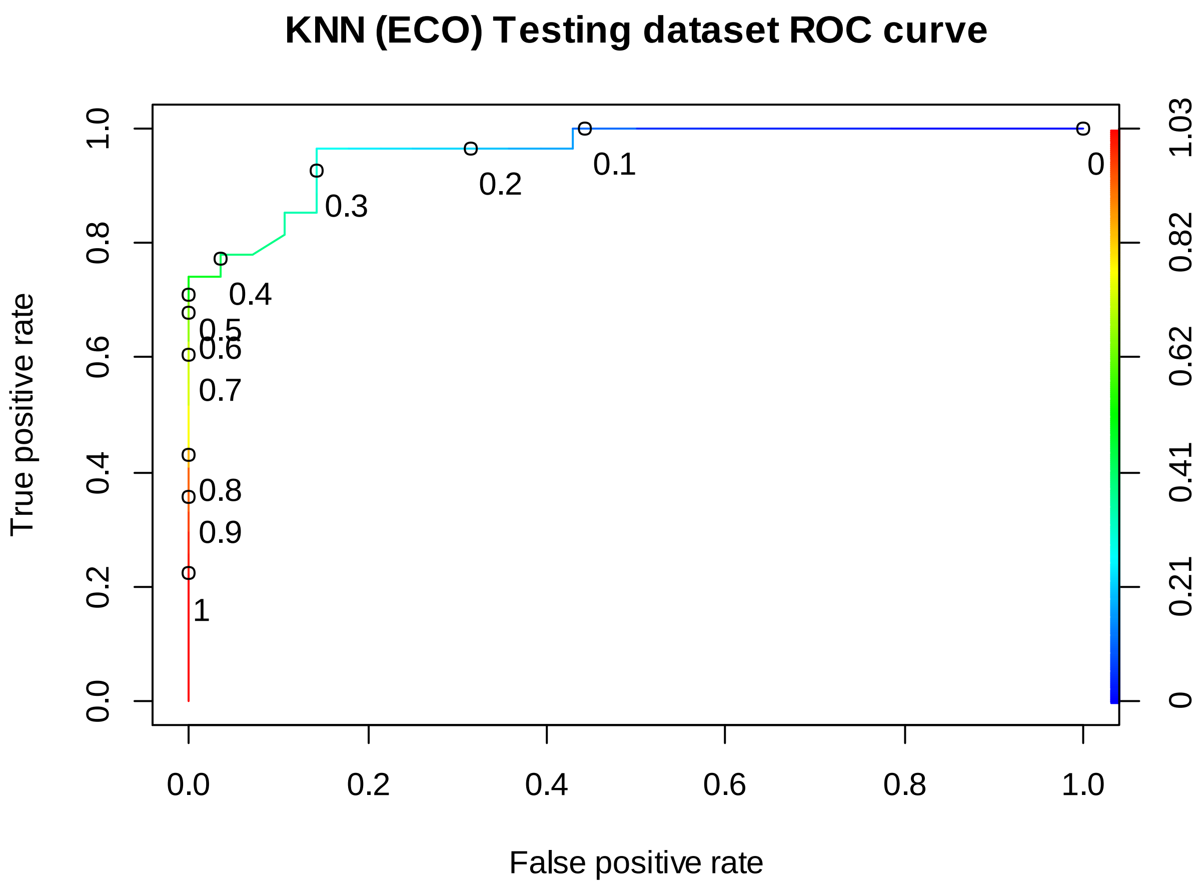

Figure F-15 The ROC curve of the KNN BSE-RPM (ECO dataset)

The **AUC value** for the ECO K-Nearest Neighbours (KNN) BSE-RPM is equal to 0.96.

F.2.2 ECO dataset Logistic Regression (LR) BSE-RPM

Script E.10 (Appendix E) is used to develop the ECO Logistic Regression (LR) BSE-RPM.

Table F-18 The confusion matrix of the LR BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	22	6
Actual reusable (1)	8	19

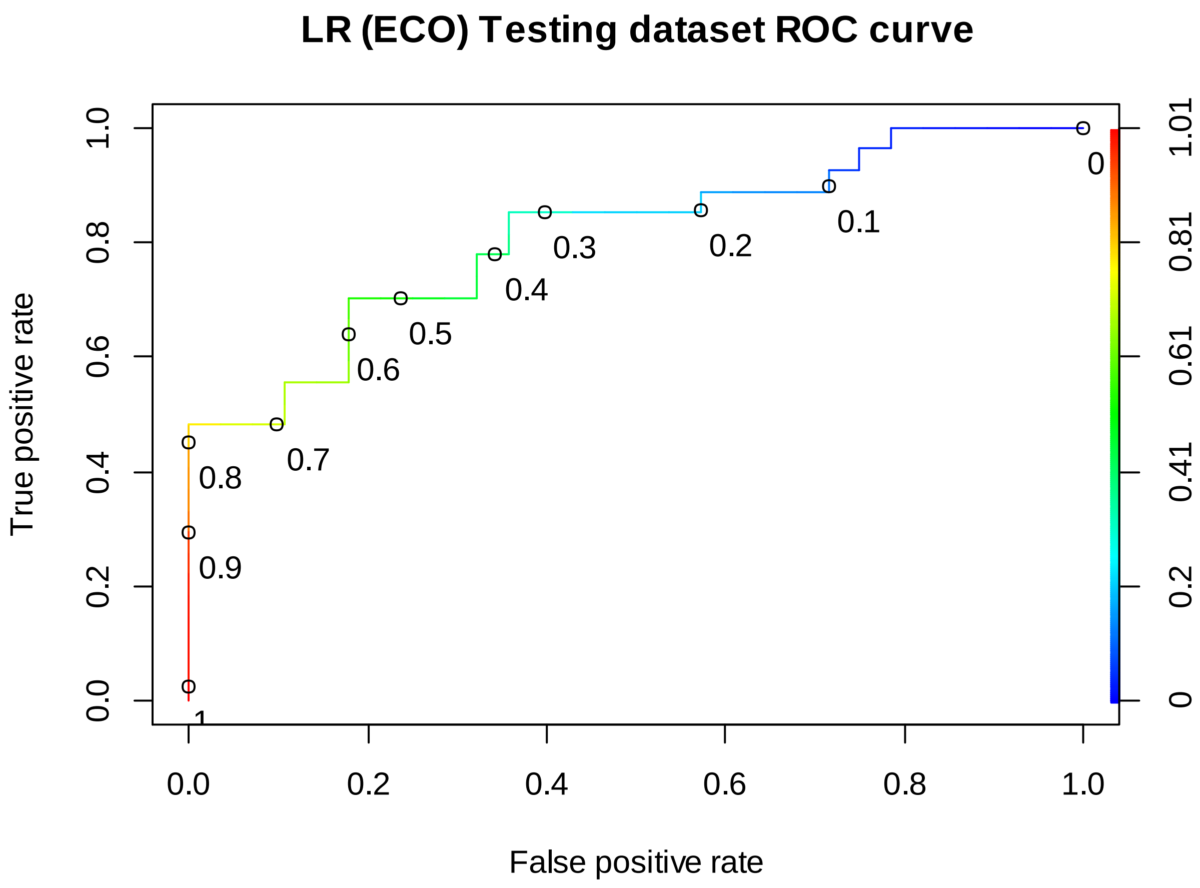

Figure F-16 The ROC curve of the LR BSE-RPM (ECO dataset)

The **AUC value** for the ECO Logistic Regression (LR) BSE-RPM is equal to 0.81.

F.2.3 ECO dataset Linear Discriminant Analysis (LDA) BSE-RPM

Script E.11 (Appendix E) is used to develop the ECO Linear Discriminant Analysis (LDA) BSE-RPM.

Table F-19 The confusion matrix of the LDA BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	21	7
Actual reusable (1)	10	17

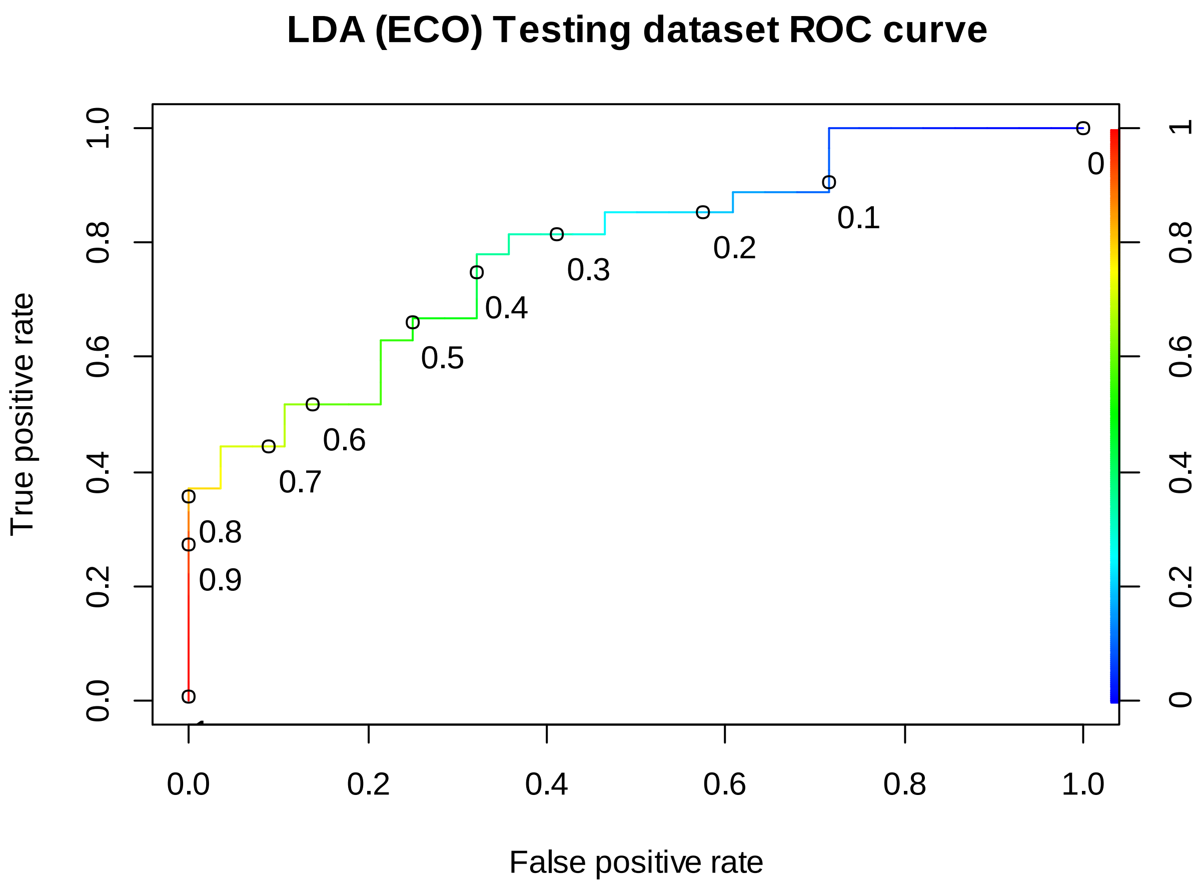

Figure F-17 The ROC curve of the LDA BSE-RPM (ECO dataset)

The **AUC value** for the ECO Linear Discriminant Analysis (LDA) BSE-RPM is equal to 0.79.

F.2.4 ECO dataset Quadratic Discriminant Analysis (QDA) BSE-RPM

Script E.12 (Appendix E) is used to develop the ECO Quadratic Discriminant Analysis (QDA) BSE-RPM.

Table F-20 The confusion matrix of the QDA BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	22	6
Actual reusable (1)	7	20

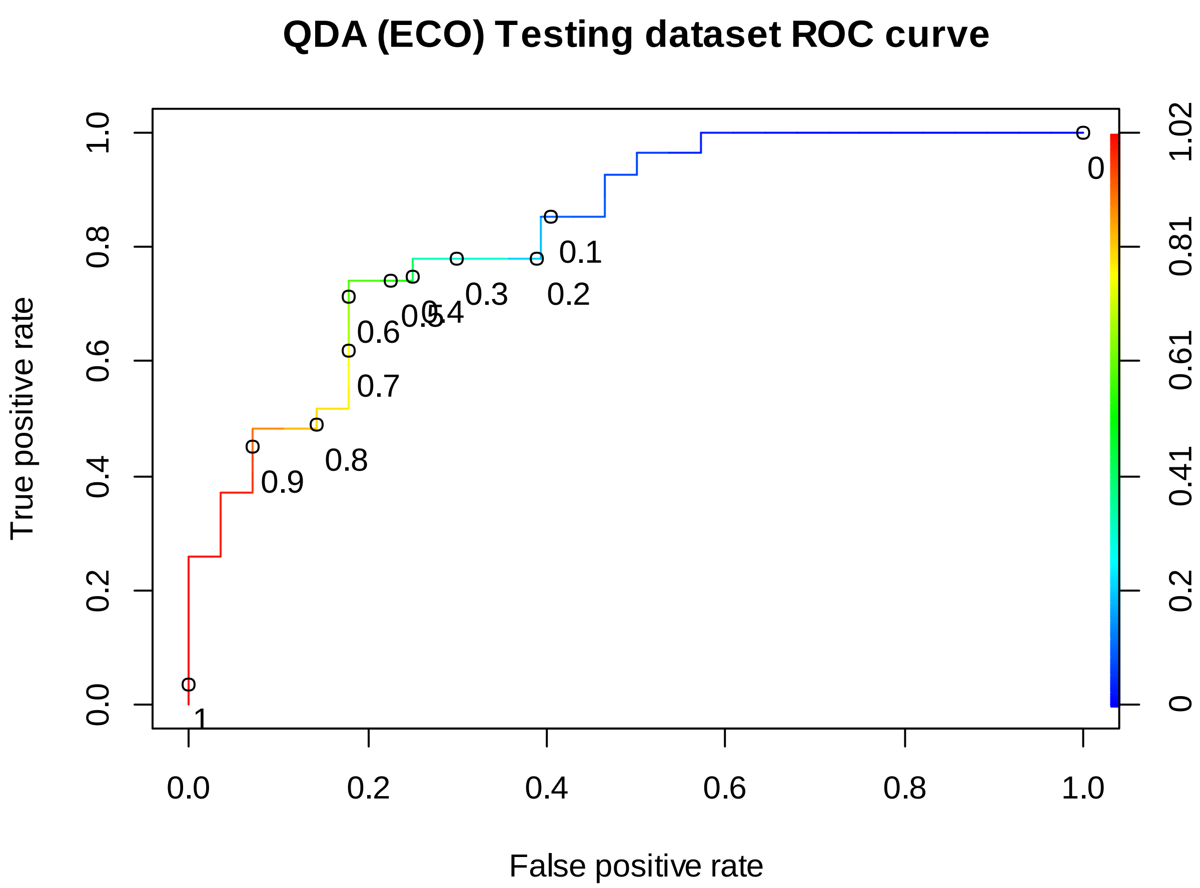

Figure F-18 The ROC curve of the QDA BSE-RPM (ECO dataset)

The **AUC value** for the ECO Quadratic Discriminant Analysis (QDA) BSE-RPM is equal to 0.83.

F.2.5 ECO dataset Naïve Bayes (NB) BSE-RPM

Script E.13 (Appendix E) is used to develop the ECO Naïve Bayes (NB) BSE-RPM.

Table F-21 The confusion matrix of the NB BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	19	9
Actual reusable (1)	8	19

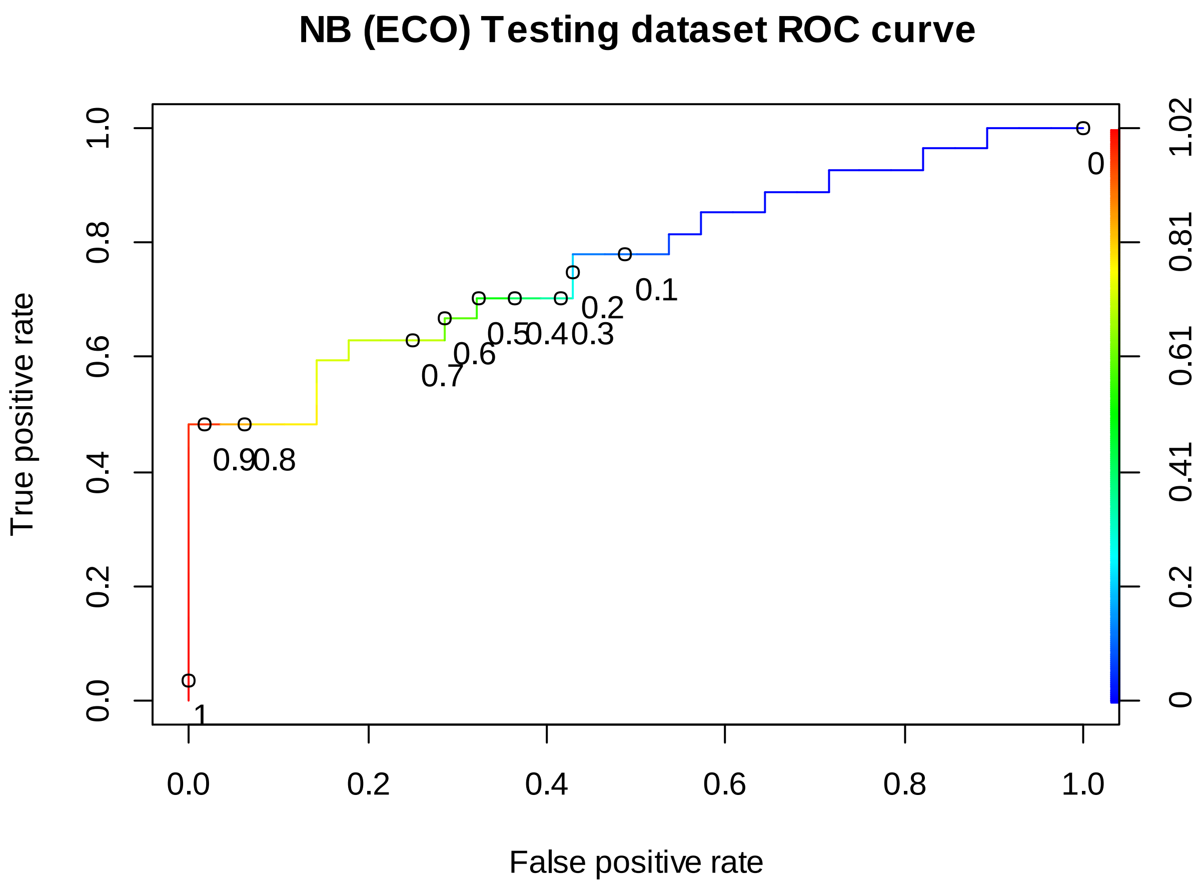

Figure F-19 The ROC curve of the NB BSE-RPM (ECO dataset)

The **AUC value** for the ECO Naïve Bayes (NB) BSE-RPM is equal to 0.77.

F.2.6 ECO dataset Decision Trees (DT) BSE-RPM

Script E.14 (Appendix E) is used to develop the ECO Decision Trees (DT) BSE-RPM.

Table F-22 The confusion matrix of the DT BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	21	7
Actual reusable (1)	5	22

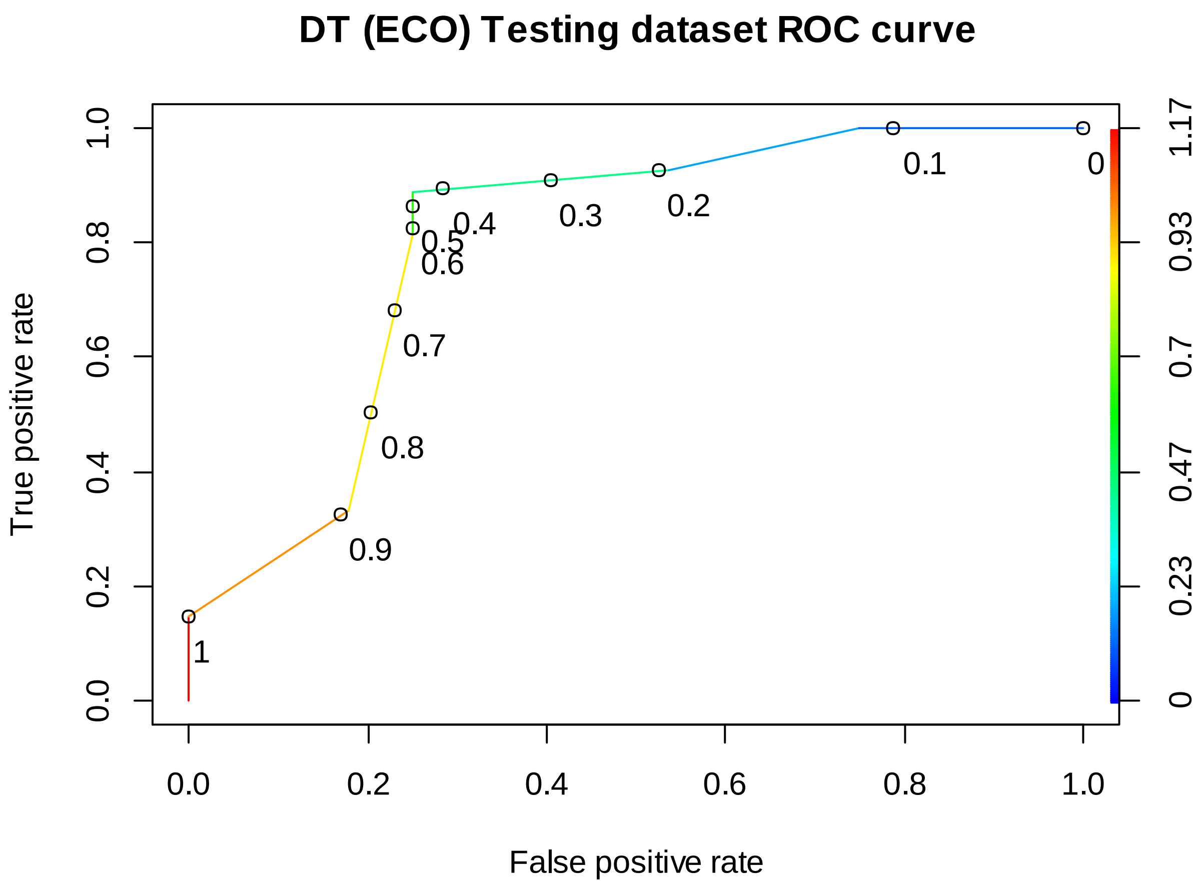

Figure F-20 The ROC curve of the DT BSE-RPM (ECO dataset)

The **AUC value** for the ECO Decision Trees (DT) BSE-RPM is equal to 0.80.

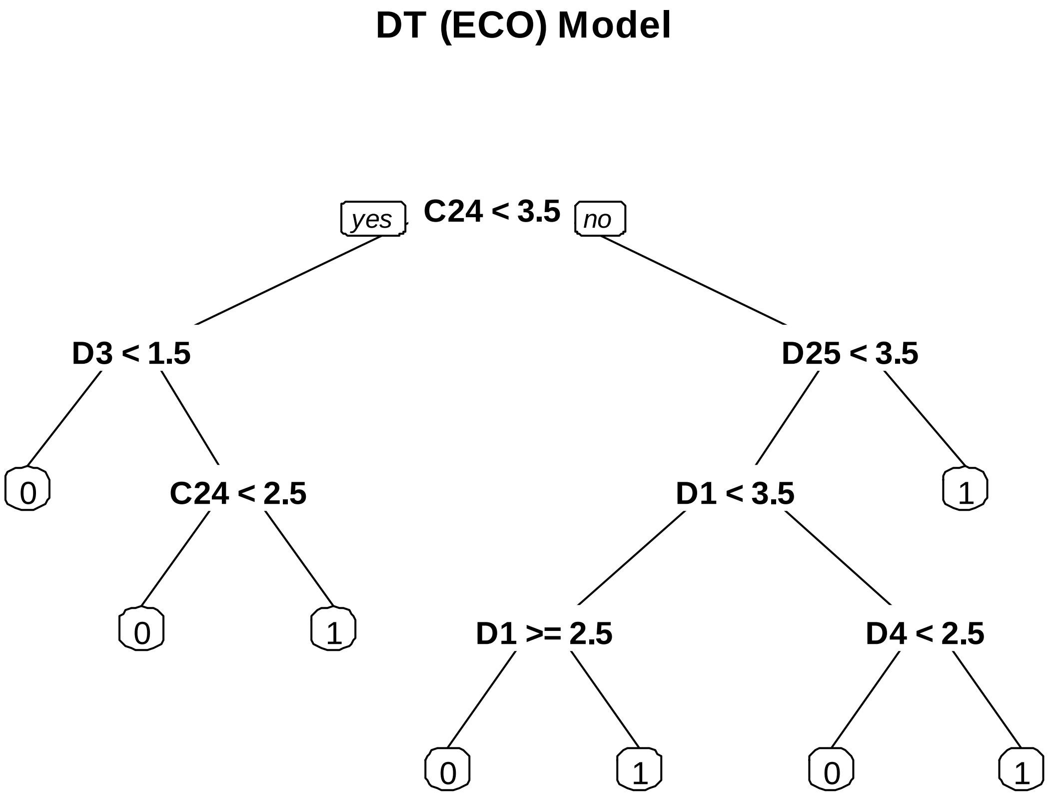

Figure F-21 The DT BSE-RPM Model (ECO dataset)

F.2.7 ECO dataset Random Forests (RF) BSE-RPM

Script E.15 (Appendix E) is used to develop the ECO Random Forests (RF) BSE-RPM.

Table F-23 The confusion matrix of the RF BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	28	0
Actual reusable (1)	8	19

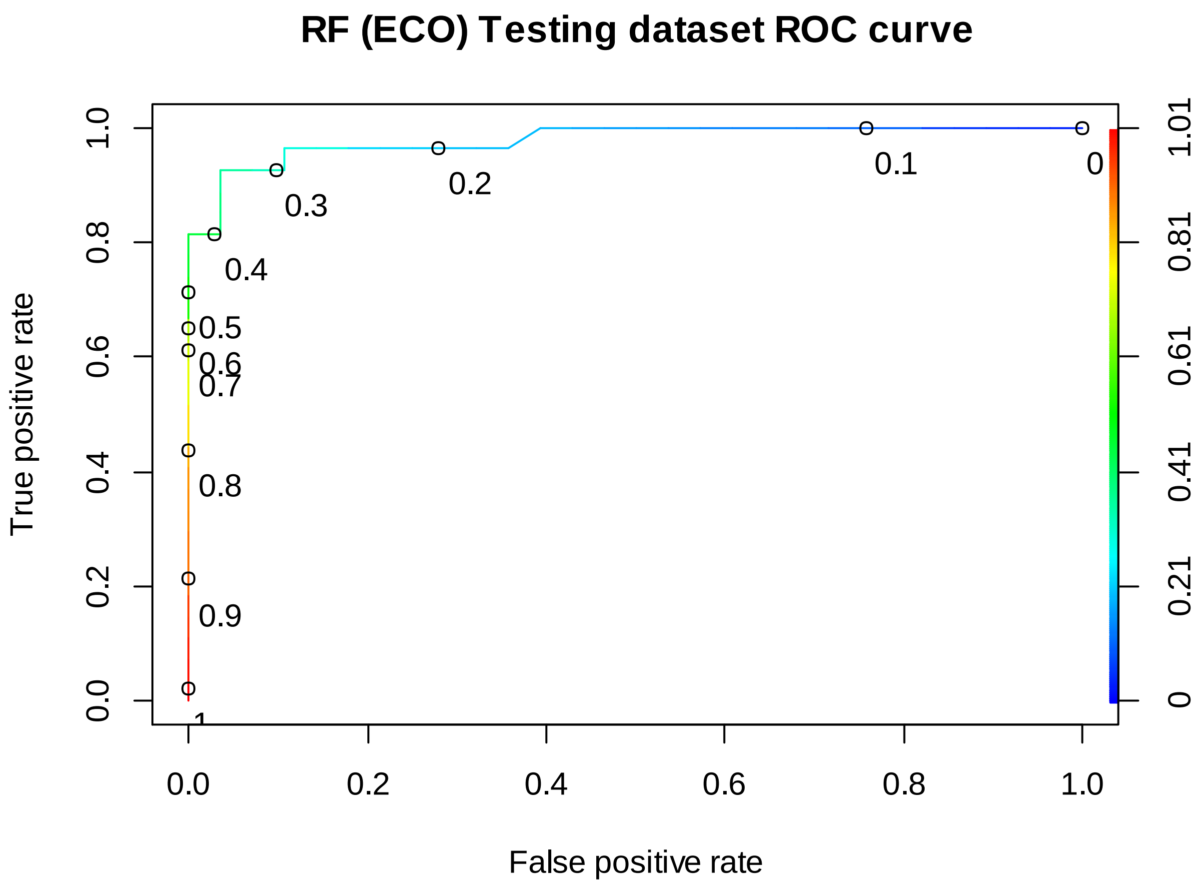

Figure F-22 The ROC curve of the RF BSE-RPM (ECO dataset)

The **AUC value** for the ECO Random Forests (RF) BSE-RPM is equal to 0.98.

F.2.8 ECO dataset Adaptive Boosting (AB) BSE-RPM

Script E.17 (Appendix E) is used to develop the ECO Adaptive Boosting (AB) BSE-RPM.

Table F-24 The confusion matrix of the AB BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	27	1
Actual reusable (1)	9	18

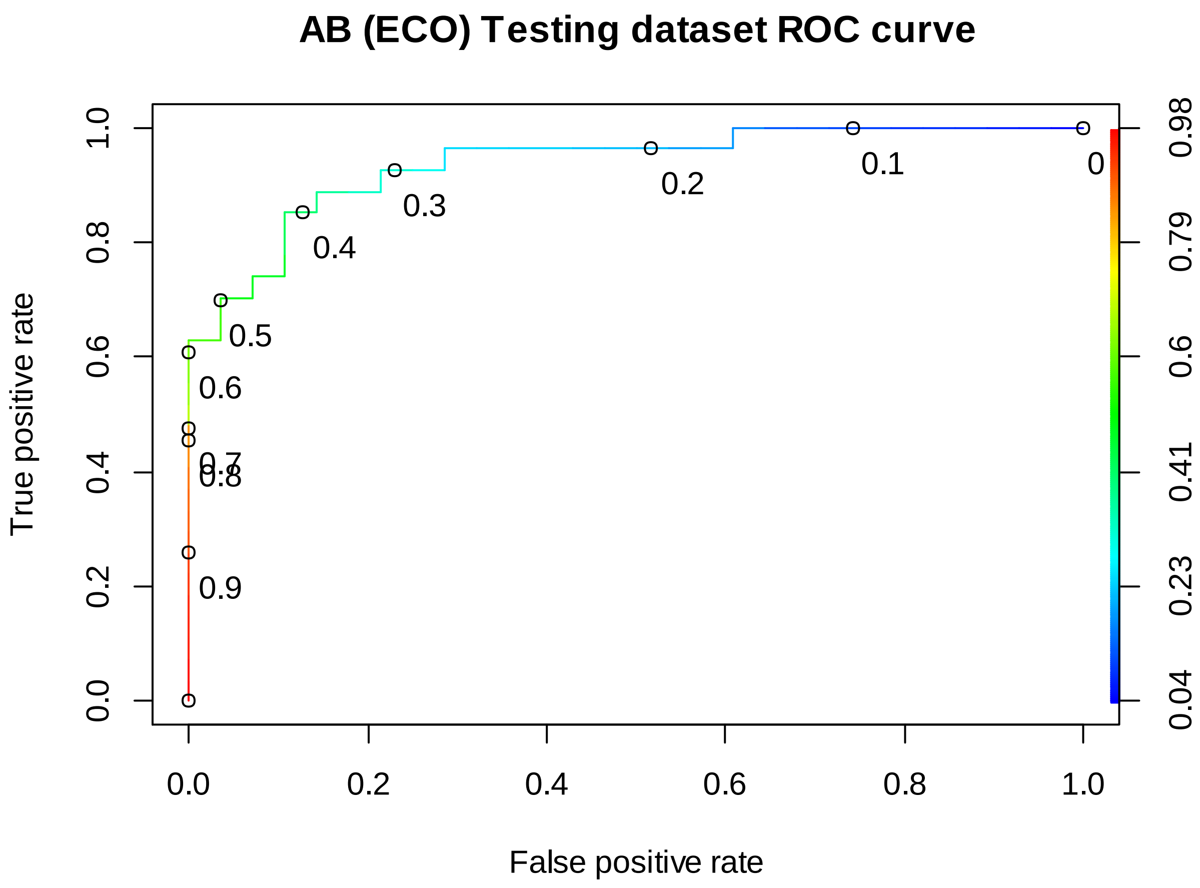

Figure F-23 The ROC curve of the AB BSE-RPM (ECO dataset)

The **AUC value** for the ECO Adaptive Boosting (AB) BSE-RPM is equal to 0.94.

F.2.9 ECO dataset Bart Machine (BM) BSE-RPM

Script E.16 (Appendix E) is used to develop the ECO Bart Machine (BM) BSE-RPM.

Table F-25 The confusion matrix of the BM BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	28	0
Actual reusable (1)	9	18

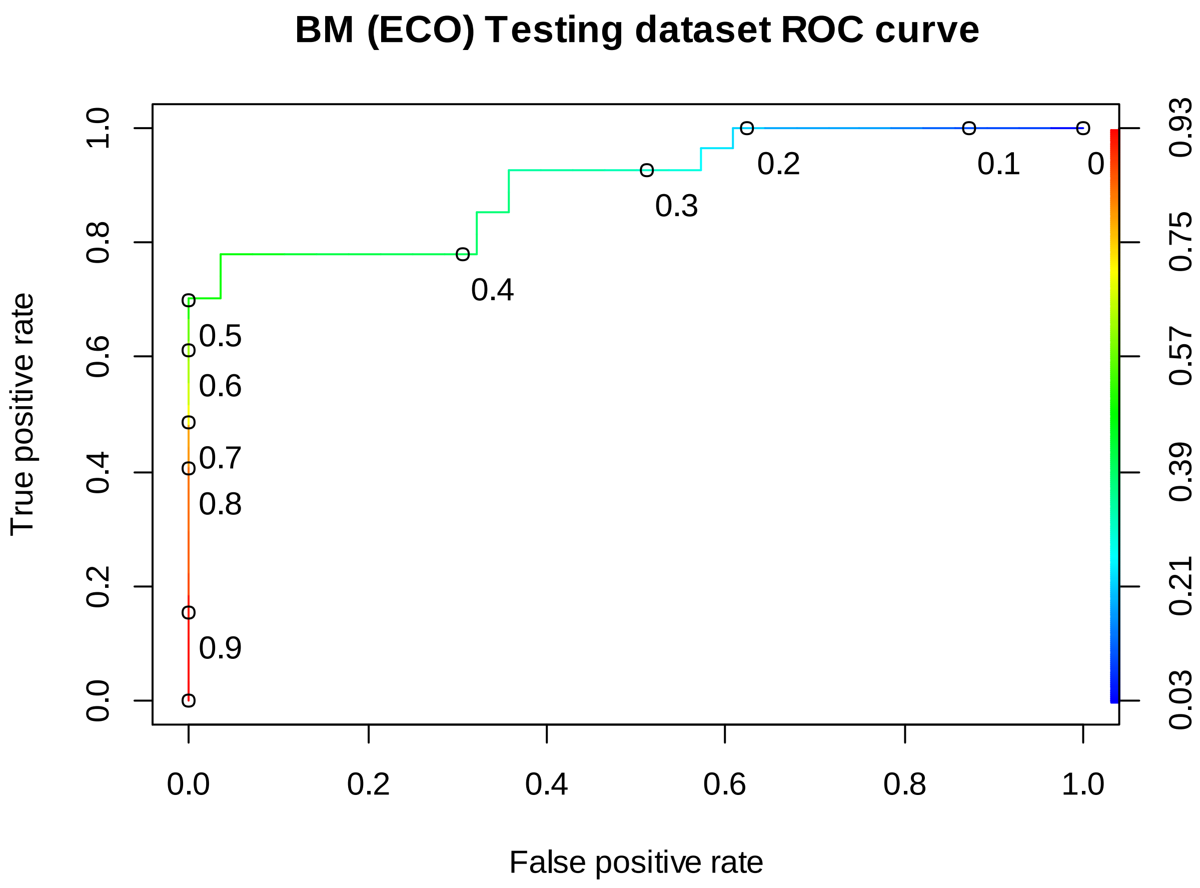

Figure F-24 The ROC curve of the BM BSE-RPM (ECO dataset)

The **AUC value** for the ECO Bart Machine (BM) BSE-RPM is equal to 0.90.

F.2.10 ECO dataset Artificial Neural Networks (ANN) BSE-RPM

Script E.18 (Appendix E) is used to develop the ECO Artificial Neural Networks (ANN) BSE-RPM.

Table F-26 The confusion matrix of the ANN BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	28	0
Actual reusable (1)	6	21

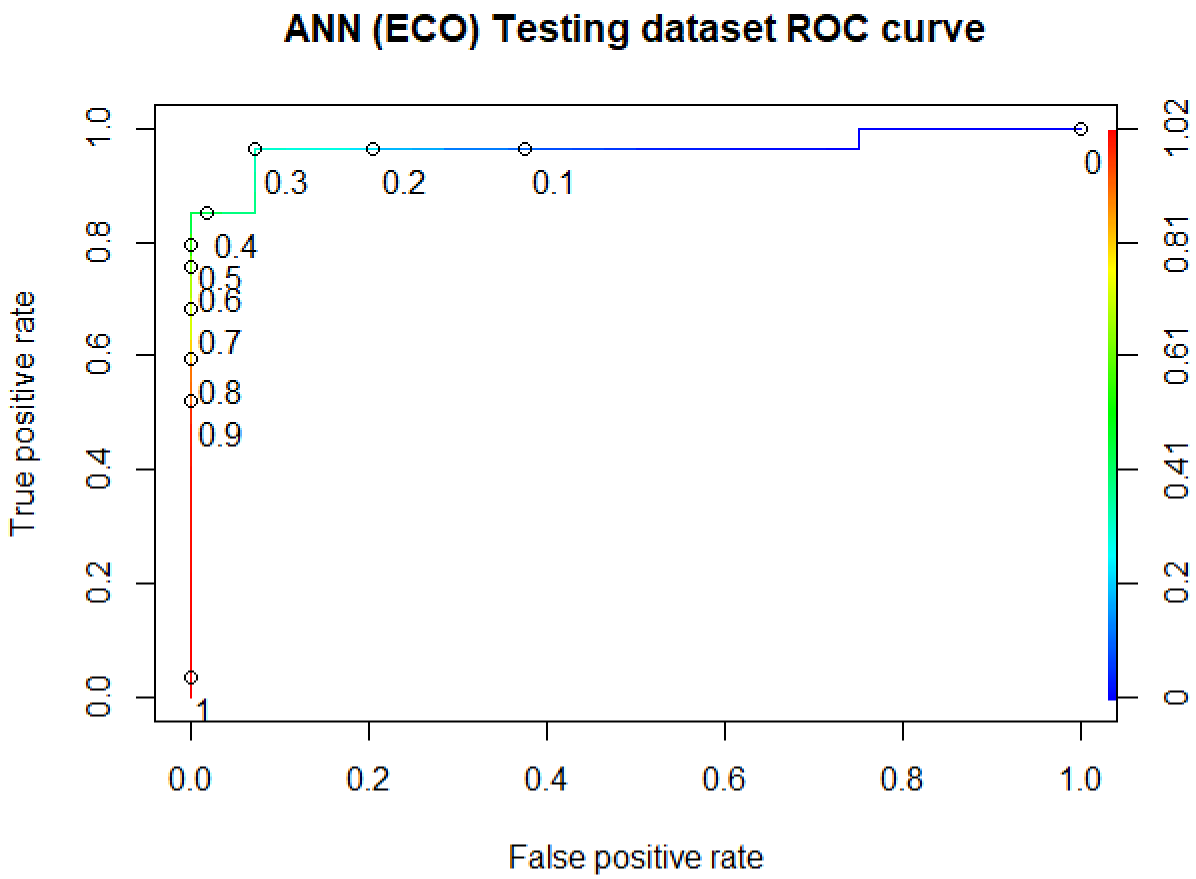

Figure F-25 The ROC curve of the ANN BSE-RPM (ECO dataset)

The **AUC value** for the ECO Artificial Neural Networks (ANN) BSE-RPM is equal to 0.96.

F.2.11 ECO dataset Gaussian Processes (GP) BSE-RPM

Script E.19 (Appendix E) is used to develop the ECO Gaussian Processes (GP) BSE-RPM.

Table F-27 The confusion matrix of the GP BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	26	2
Actual reusable (1)	10	17

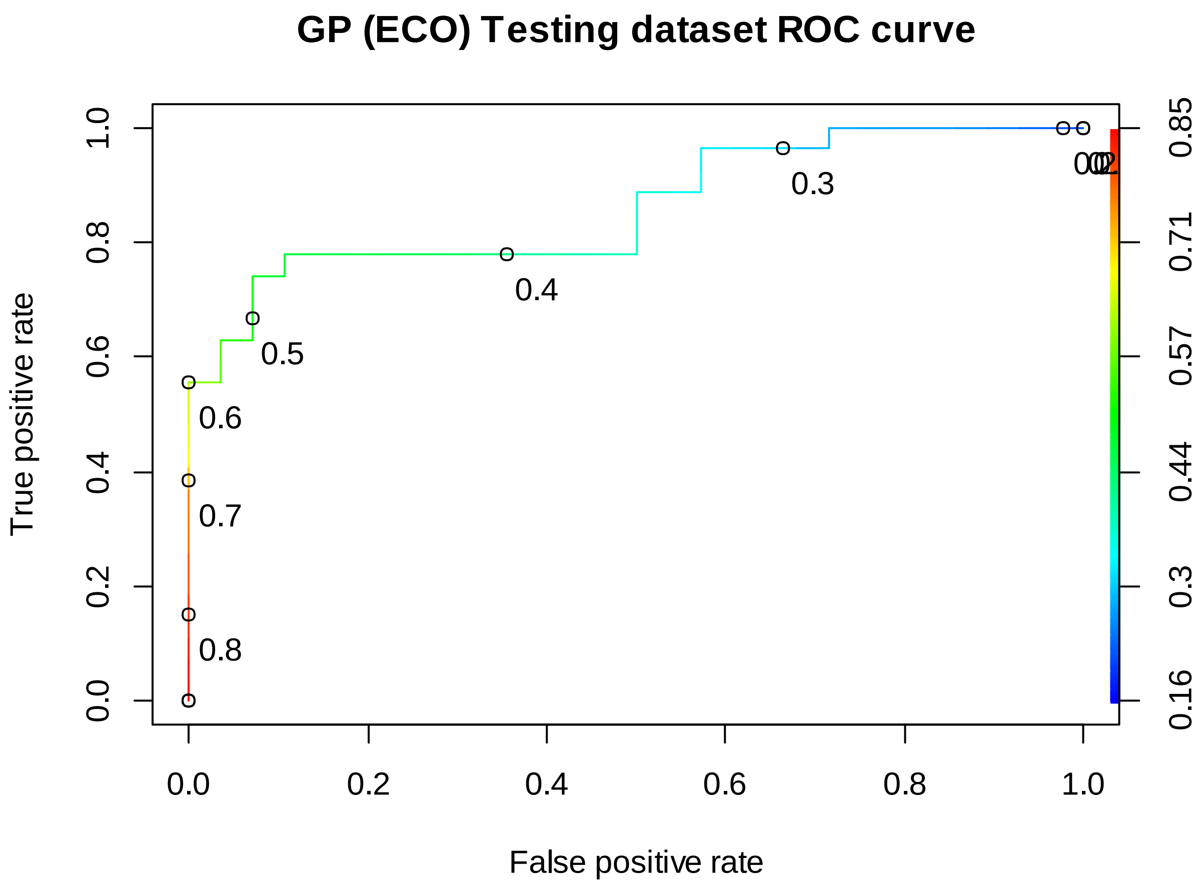

Figure F-26 The ROC curve of the GP BSE-RPM (ECO dataset)

The **AUC value** for the ECO Gaussian Processes (GP) BSE-RPM is equal to 0.86.

F.2.12 ECO dataset Propositional Rule Learner (PRL) BSE-RPM

Script E.20 (Appendix E) is used to develop the ECO Propositional Rule Learner (PRL) BSE-RPM.

Table F-28 The confusion matrix of the PRL BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	21	7
Actual reusable (1)	9	18

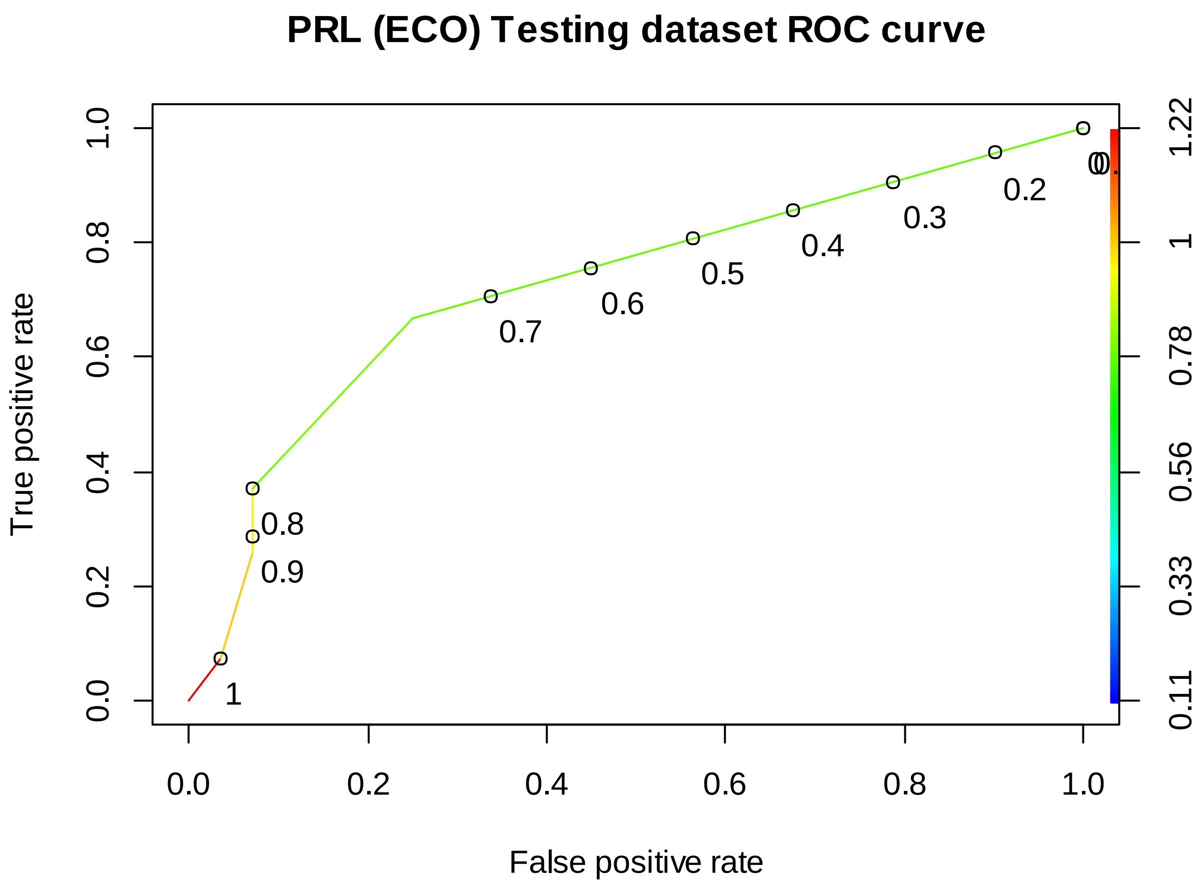

Figure F-27 The ROC curve of the PRL BSE-RPM (ECO dataset)

The **AUC value** for the ECO Propositional Rule Learner (PRL) BSE-RPM is equal to 0.72.

Table F-29 The rules set of the PRL BSE-RPM (ECO dataset)

Rule number (to be considered in order)	Rule	Result
1st	If: (D10 >= 3) and (D25 >= 4)	Then: E2C=1 (29.0/2.0)
2nd	Else if: (C24 >= 4) and (D1 >= 4)	Then: E2C=1 (27.0/6.0)
3rd	Else if: (D8 >= 3) and (D10 >= 4)	Then: E2C=1 (5.0/1.0)

Rule number (to be considered in order)	Rule	Result
4th	Else if: (D4 <= 2) and (D2 >= 3) and (C24 <= 4)	Then: E2C=1 (5.0/0.0)
5th	Else if none	Then: E2C=0 (63.0/7.0)

F.2.13 ECO dataset Support Vector Machines (SVM) BSE-RPM

Script E.21 (Appendix E) is used to develop the ECO Support Vector Machines (SVM) BSE-RPM.

Table F-30 The confusion matrix of the SVM BSE-RPM (ECO dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	26	2
Actual reusable (1)	4	23

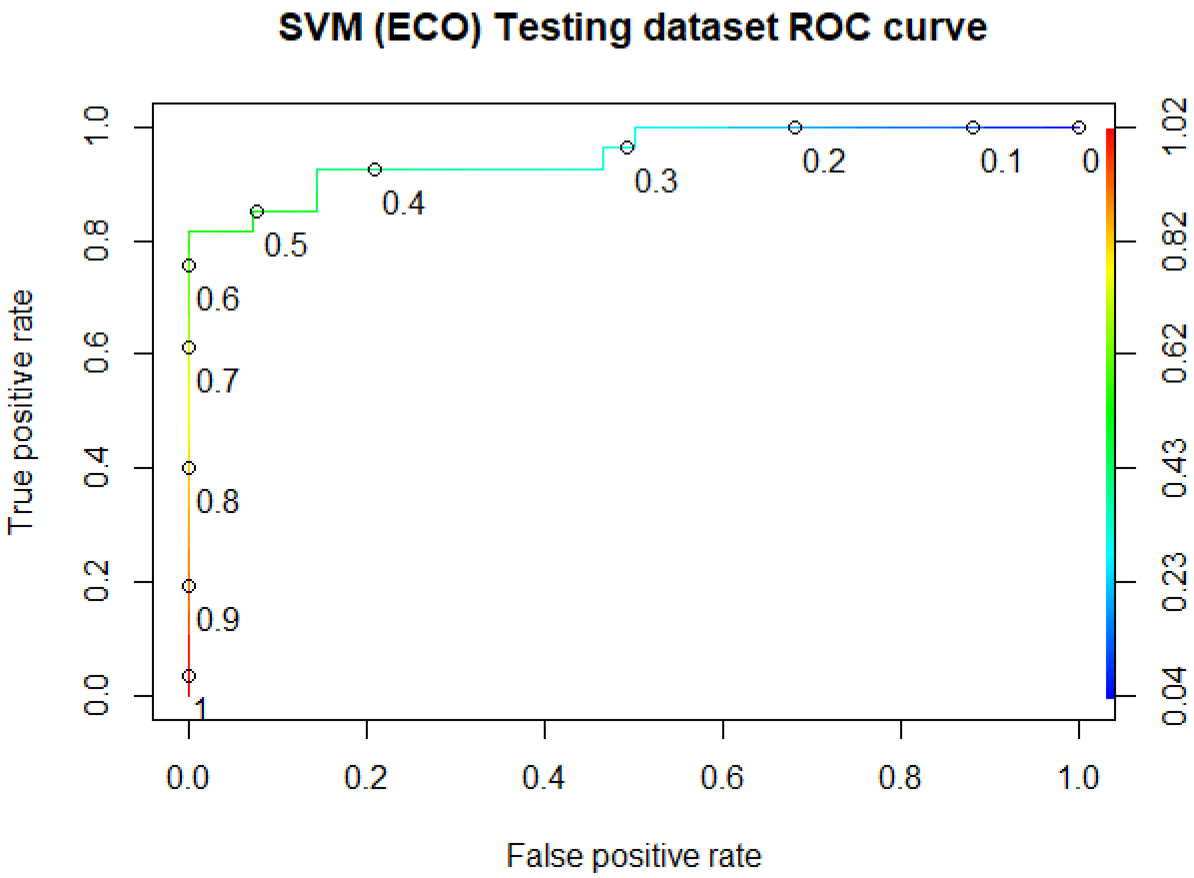

Figure F-28 The ROC curve of the SVM BSE-RPM (ECO dataset)

The **AUC value** for the ECO Support Vector Machines (SVM) BSE-RPM is equal to 0.95.

F.3 Predictive models on the SOC dataset

In this section, the results of the models used to predict the social reusability of the structural elements are presented.

Table F-31 Summary of the results of the SOC BSE-RPMs developed (the validation set approach method).

Predictive model	Type-I error	Type-II error	Specificity	Sensitivity	Overall accuracy	Overall error rate	AUC
KNN	0.06	0.38	0.93	0.62	0.79	0.21	0.95
LR	0.11	0.38	0.89	0.62	0.77	0.23	0.76
LDA	0.11	0.44	0.89	0.56	0.74	0.26	0.77
QDA	0.11	0.06	0.89	0.94	0.91	0.09	0.97
NB	0.22	0.06	0.78	0.94	0.85	0.15	0.97
DT	0.33	0.13	0.67	0.87	0.77	0.23	0.88
RF	0.00	0.19	1.00	0.81	0.91	0.09	0.99
AB	0.11	0.06	0.89	0.94	0.91	0.09	0.94
BM	0.06	0.19	0.94	0.81	0.88	0.12	0.98
ANN	0.11	0.13	0.89	0.87	0.88	0.12	0.92
GP	0.06	0.25	0.94	0.75	0.85	0.15	0.96
PRL	0.17	0.13	0.83	0.87	0.85	0.15	0.85
SVM	0.11	0.00	0.89	1.00	0.94	0.06	0.97

F.3.1 SOC dataset K-Nearest Neighbours (KNN) BSE-RPM

Script E.9 (Appendix E) is used to develop the SOC K-Nearest Neighbours (KNN) BSE-RPM.

Table F-32 The confusion matrix of the KNN BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	17	1
Actual reusable (1)	6	10

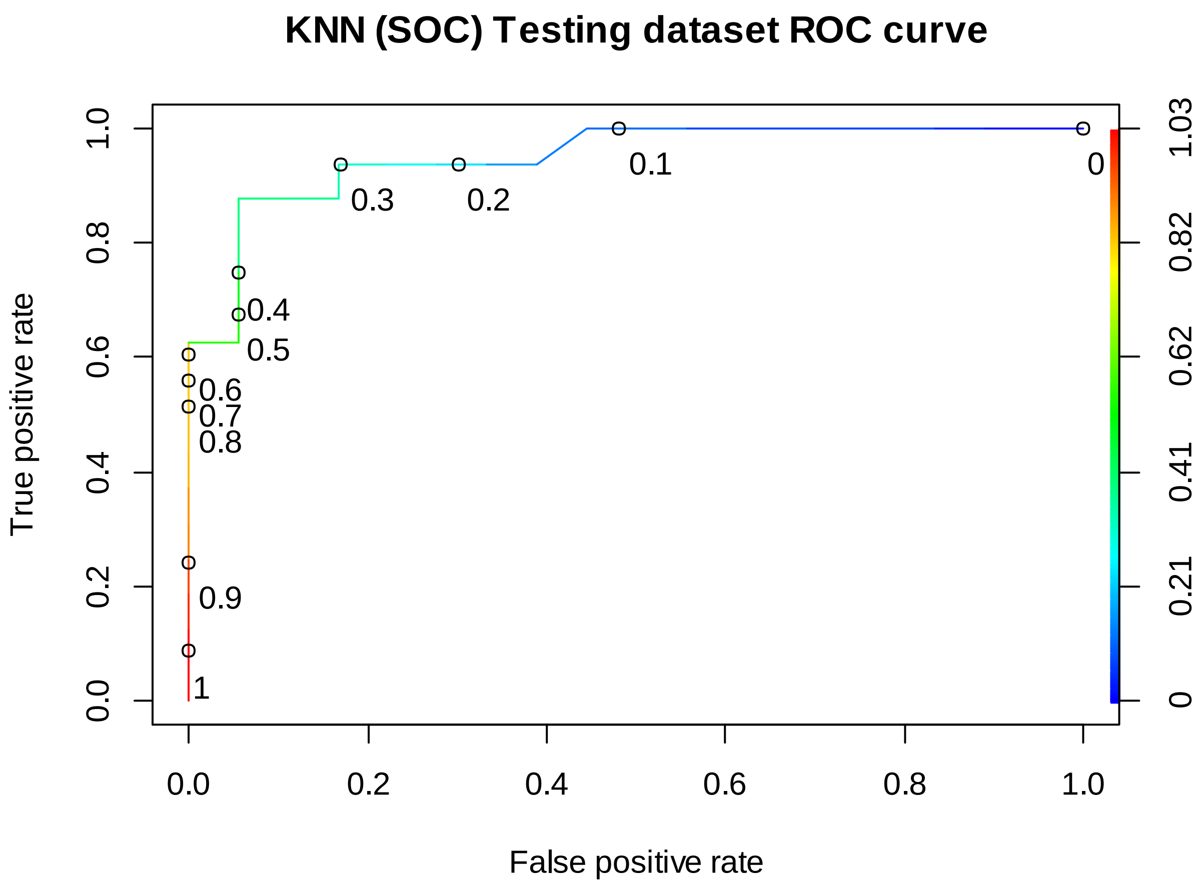

Figure F-29 The ROC curve of the KNN BSE-RPM (SOC dataset)

The **AUC value** for the SOC K-Nearest Neighbours (KNN) BSE-RPM is equal to 0.95.

F.3.2 SOC dataset Logistic Regression (LR) BSE-RPM

Script E.10 (Appendix E) is used to develop the SOC Logistic Regression (LR) BSE-RPM.

Table F-33 The confusion matrix of the LR BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	6	10

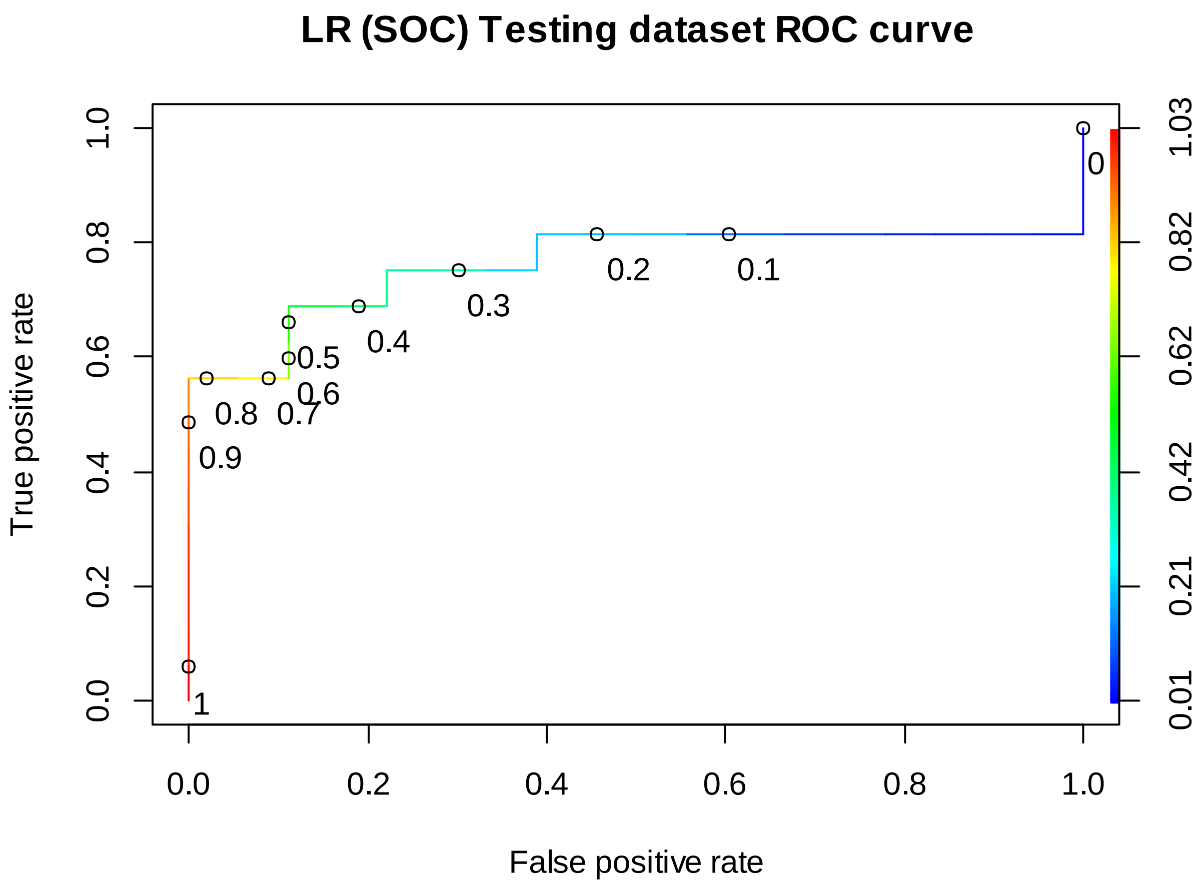

Figure F-30 The ROC curve of the LR BSE-RPM (SOC dataset)

The **AUC value** for the SOC Logistic Regression (LR) BSE-RPM is equal to 0.76.

F.3.3 SOC dataset Linear Discriminant Analysis (LDA) BSE-RPM

Script E.11 (Appendix E) is used to develop the SOC Linear Discriminant Analysis (LDA) BSE-RPM.

Table F-34 The confusion matrix of the LDA BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	7	9

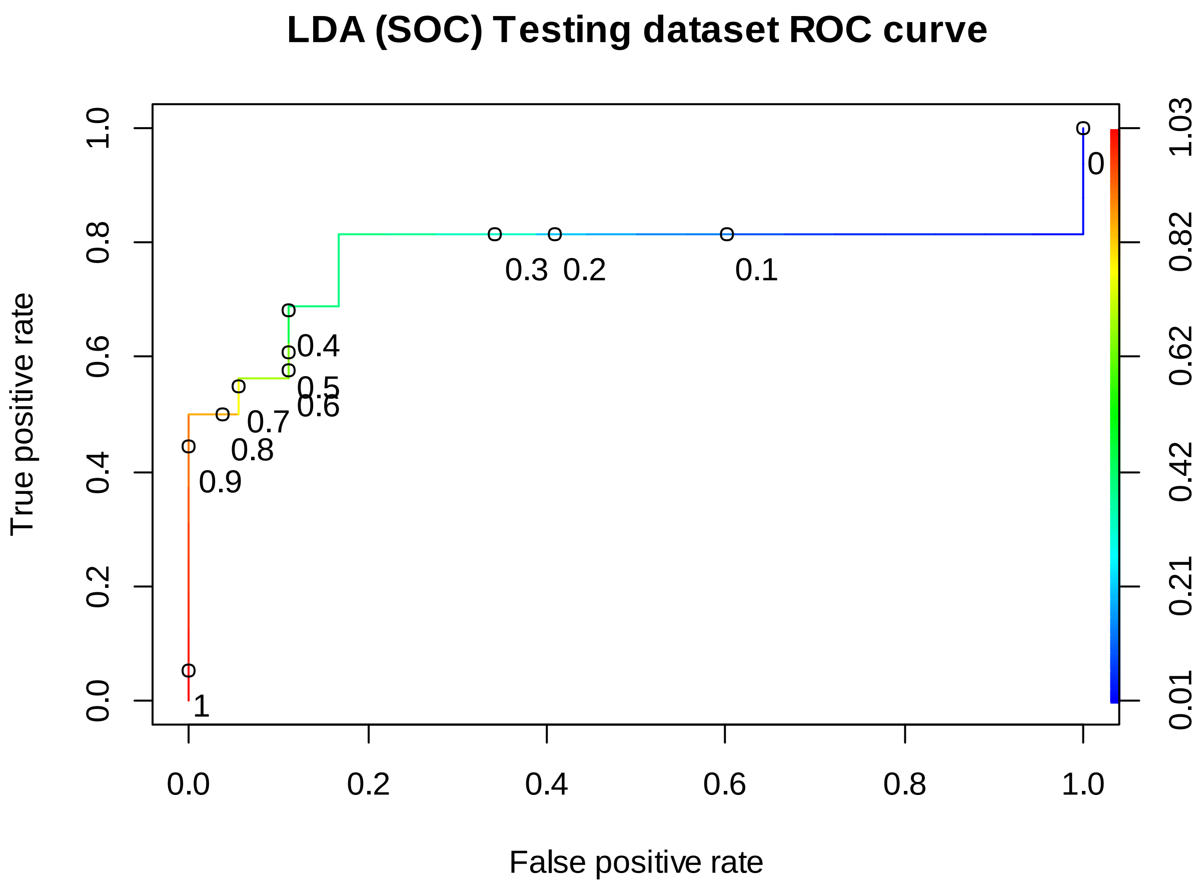

Figure F-31 The ROC curve of the LDA BSE-RPM (SOC dataset)

The **AUC value** for the SOC Linear Discriminant Analysis (LDA) BSE-RPM is equal to 0.77.

F.3.4 SOC dataset Quadratic Discriminant Analysis (QDA) BSE-RPM

Script E.12 (Appendix E) is used to develop the SOC Quadratic Discriminant Analysis (QDA) BSE-RPM.

Table F-35 The confusion matrix of the QDA BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	1	15

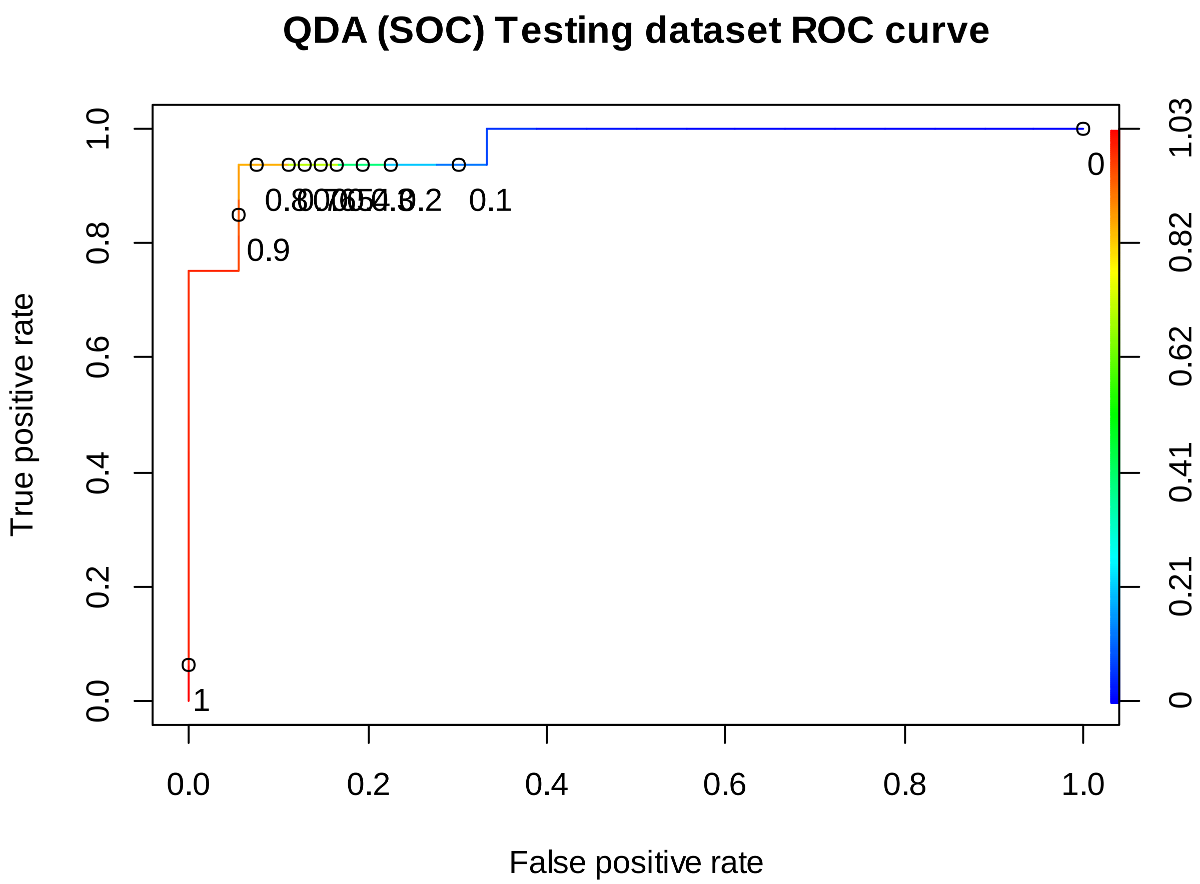

Figure F-32 The ROC curve of the QDA BSE-RPM (SOC dataset)

The **AUC value** for the SOC Quadratic Discriminant Analysis (QDA) BSE-RPM is equal to 0.97.

F.3.5 SOC dataset Naïve Bayes (NB) BSE-RPM

Script E.13 (Appendix E) is used to develop the SOC Naïve Bayes (NB) BSE-RPM.

Table F-36 The confusion matrix of the NB BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	14	4
Actual reusable (1)	1	15

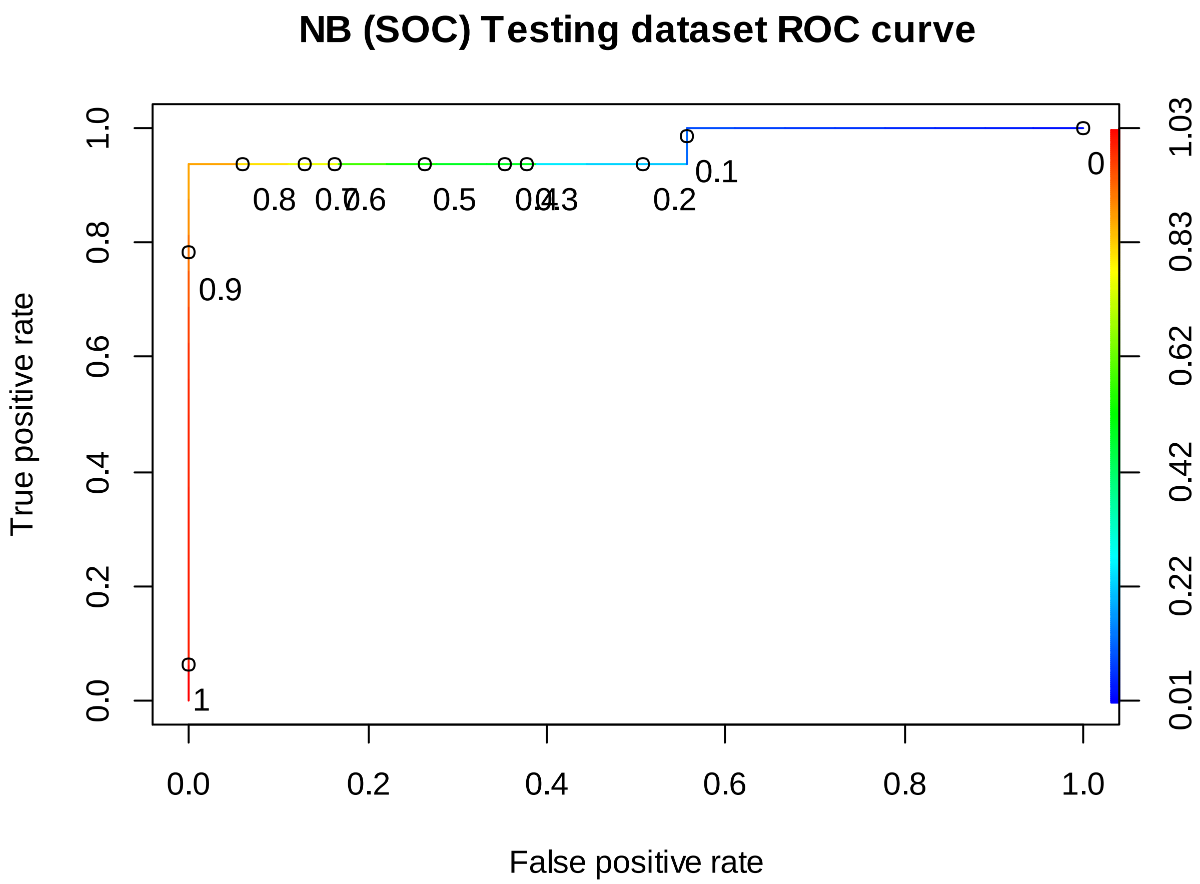

Figure F-33 The ROC curve of the NB BSE-RPM (SOC dataset)

The **AUC value** for the SOC Naïve Bayes (NB) BSE-RPM is equal to 0.97.

F.3.6 SOC dataset Decision Trees (DT) BSE-RPM

Script E.14 (Appendix E) is used to develop the SOC Decision Trees (DT) BSE-RPM.

Table F-37 The confusion matrix of the DT BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	12	6
Actual reusable (1)	2	14

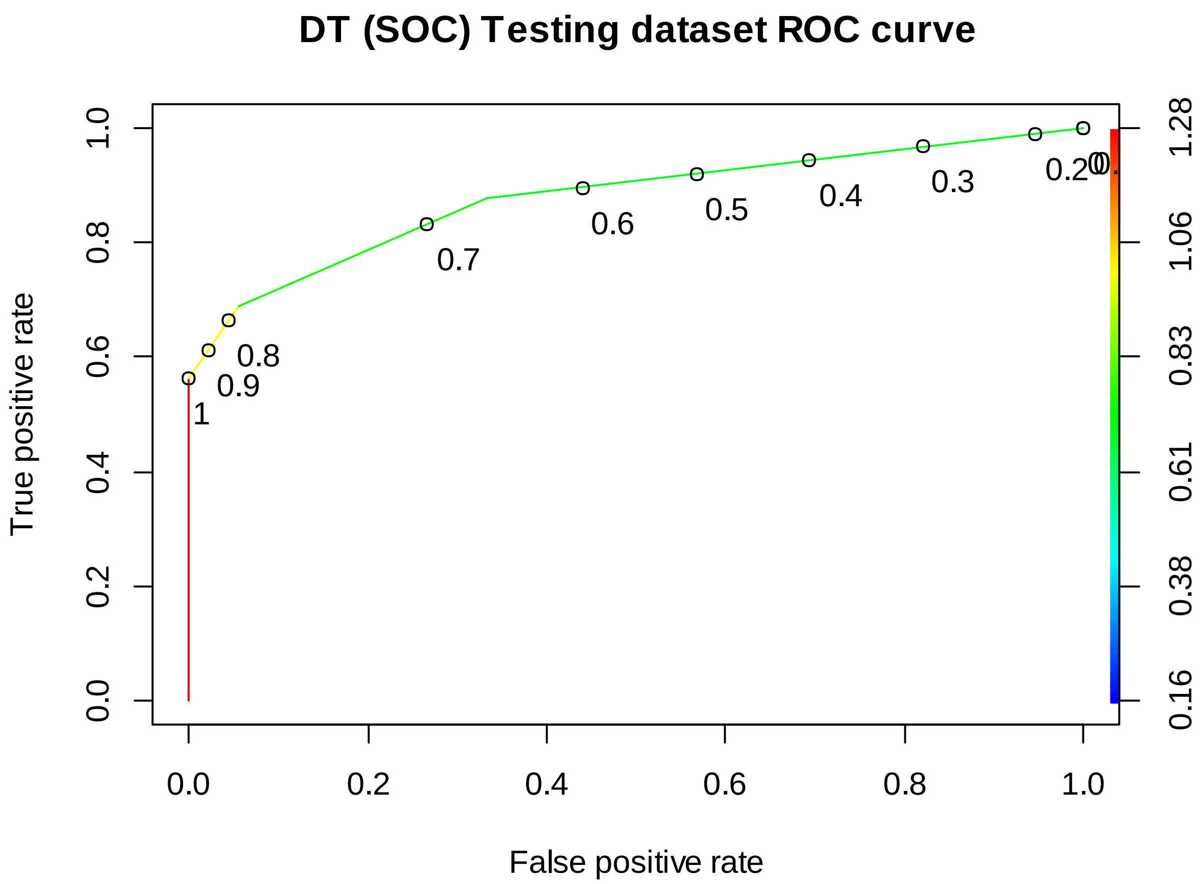

Figure F-34 The ROC curve of the DT BSE-RPM (SOC dataset)

The **AUC value** for the SOC Decision Trees (DT) BSE-RPM is equal to 0.88.

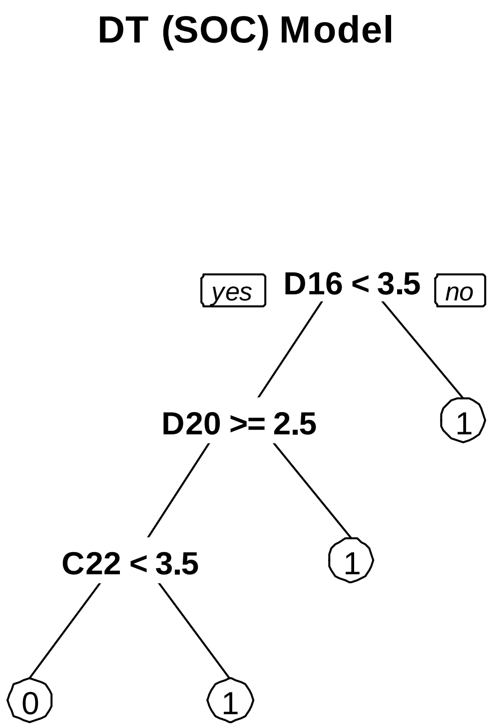

Figure F-35 The DT BSE-RPM Model (SOC dataset)

F.3.7 SOC dataset Random Forests (RF) BSE-RPM

Script E.15 (Appendix E) is used to develop the SOC Random Forests (RF) BSE-RPM.

Table F-38 The confusion matrix of the RF BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	18	0
Actual reusable (1)	3	13

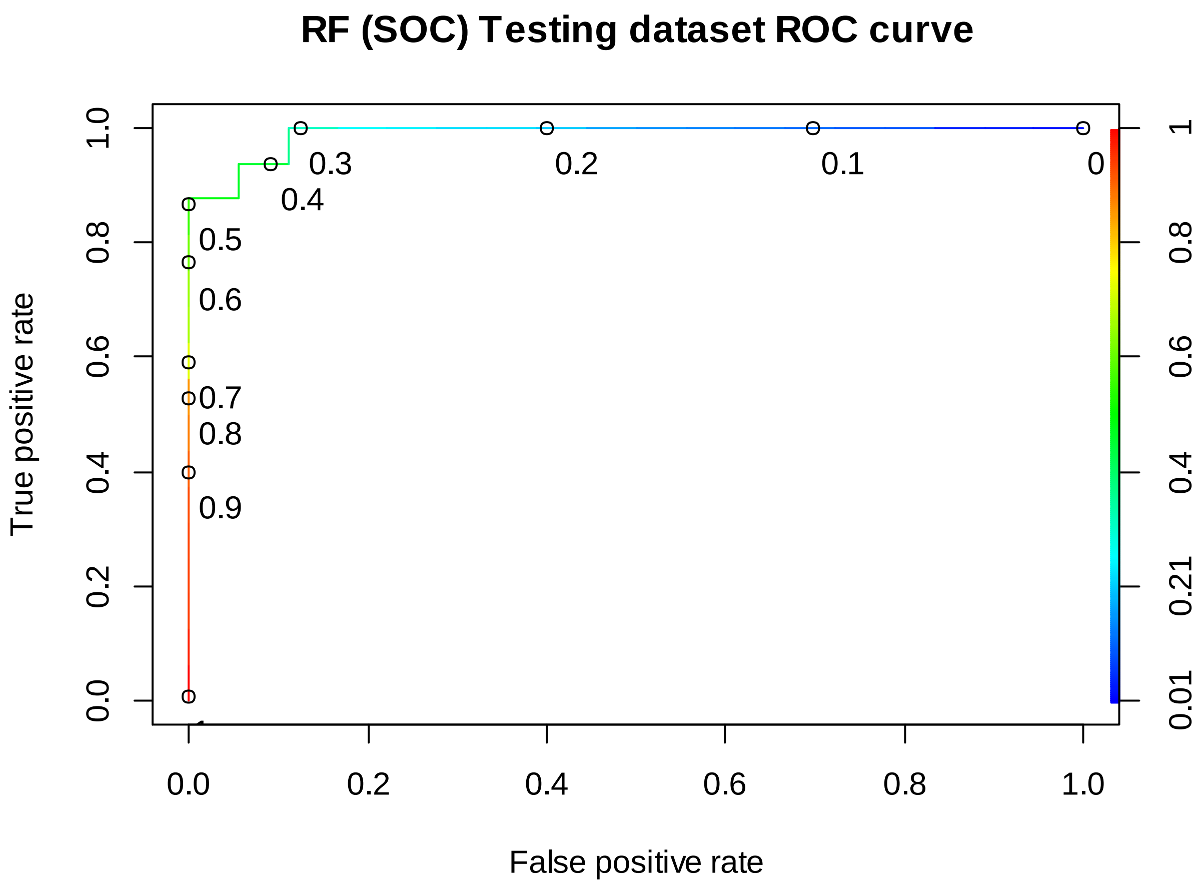

Figure F-36 The ROC curve of the RF BSE-RPM (SOC dataset)

The **AUC value** for the SOC Random Forests (RF) BSE-RPM is equal to 0.99.

F.3.8 SOC dataset Adaptive Boosting (AB) BSE-RPM

Script E.17 (Appendix E) is used to develop the SOC Adaptive Boosting (AB) BSE-RPM.

Table F-39 The confusion matrix of the AB BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	1	15

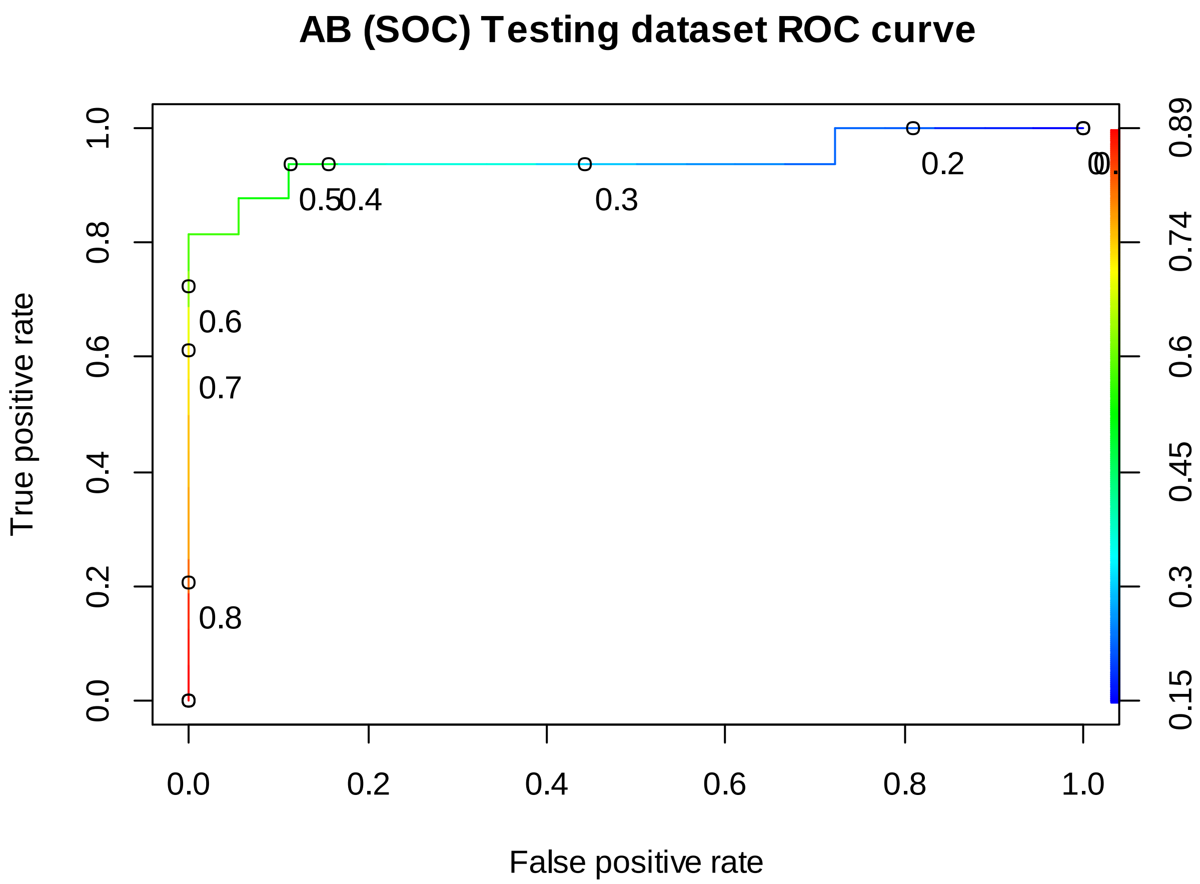

Figure F-37 The ROC curve of the AB BSE-RPM (SOC dataset)

The **AUC value** for the SOC Adaptive Boosting (AB) BSE-RPM is equal to 0.94.

F.3.9 SOC dataset Bart Machine (BM) BSE-RPM

Script E.16 (Appendix E) is used to develop the SOC Bart Machine (BM) BSE-RPM.

Table F-40 The confusion matrix of the BM BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	17	1
Actual reusable (1)	3	13

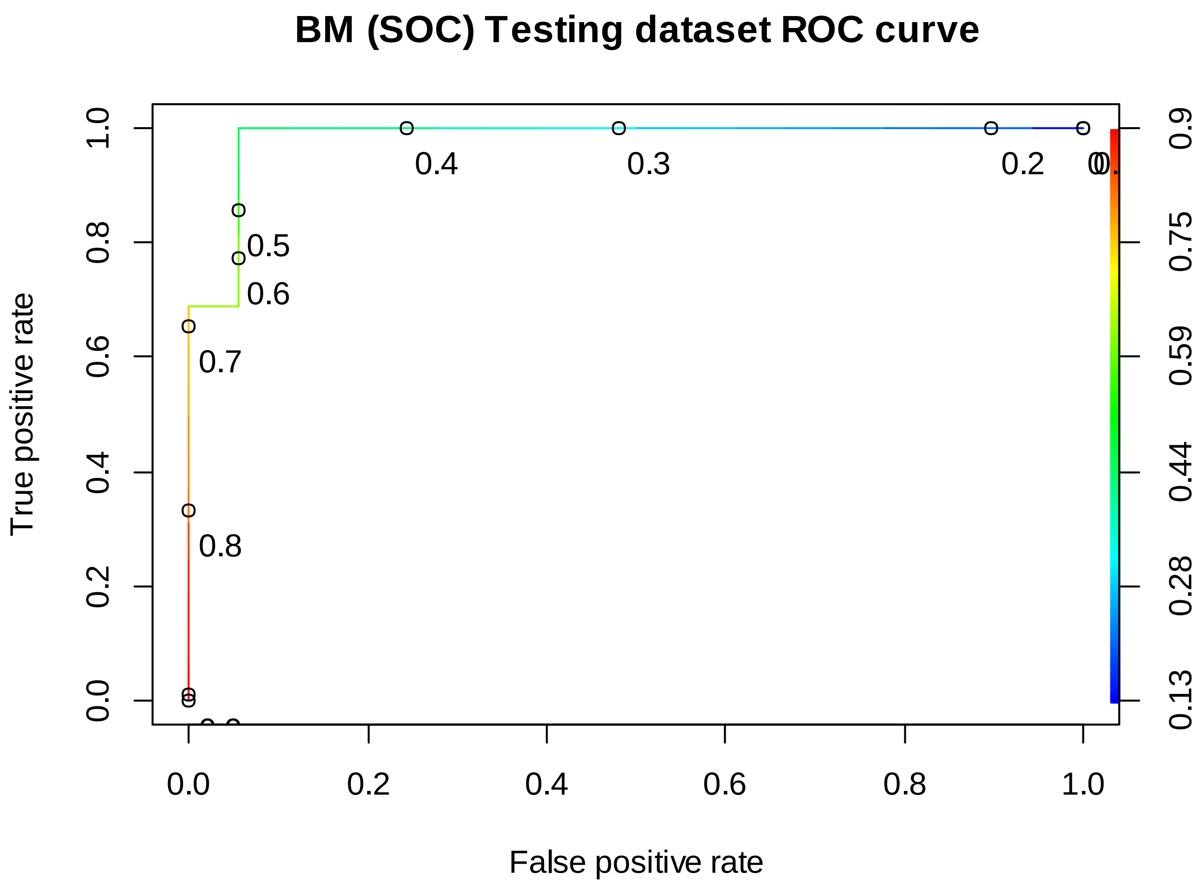

Figure F-38 The ROC curve of the BM BSE-RPM (SOC dataset)

The **AUC value** for the SOC Bart Machine (BM) BSE-RPM is equal to 0.98.

F.3.10 SOC dataset Artificial Neural Networks (ANN) BSE-RPM

Script E.18 (Appendix E) is used to develop the SOC Artificial Neural Networks (ANN) BSE-RPM.

Table F-41 The confusion matrix of the ANN BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	2	14

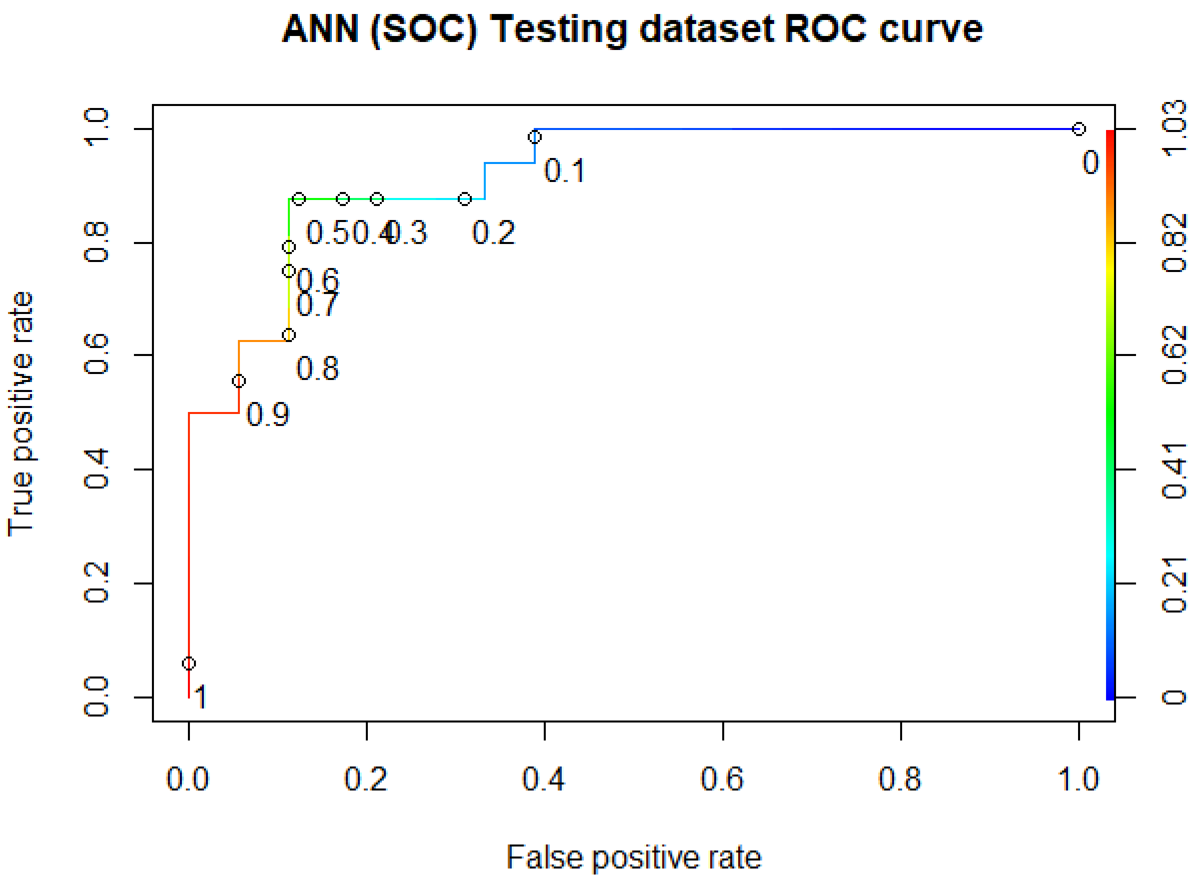

Figure F-39 The ROC curve of the ANN BSE-RPM (SOC dataset)

The **AUC value** for the SOC Artificial Neural Networks (ANN) BSE-RPM is equal to 0.92.

F.3.11 SOC dataset Gaussian Processes (GP) BSE-RPM

Script E.19 (Appendix E) is used to develop the SOC Gaussian Processes (GP) BSE-RPM.

Table F-42 The confusion matrix of the GP BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	17	1
Actual reusable (1)	4	12

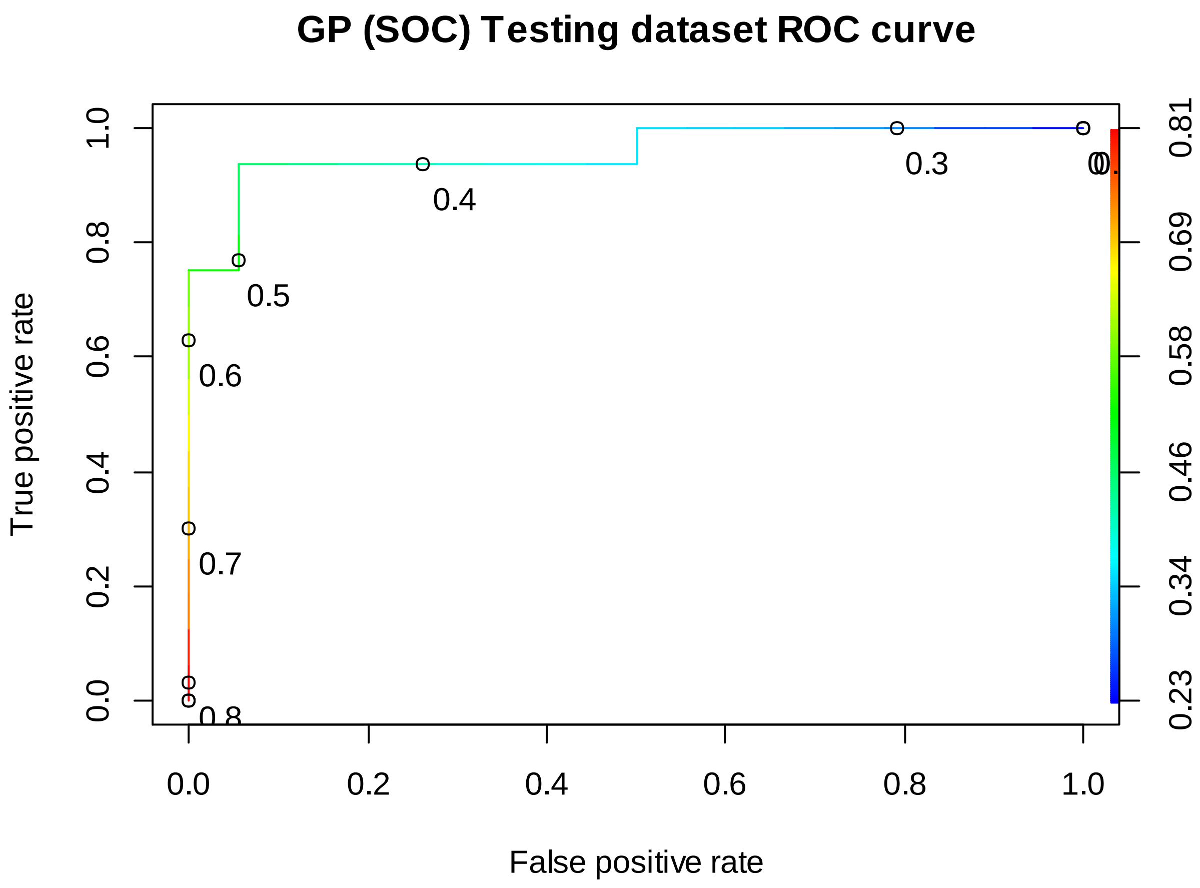

Figure F-40 The ROC curve of the GP BSE-RPM (SOC dataset)

The **AUC value** for the SOC Gaussian Processes (GP) BSE-RPM is equal to 0.96.

F.3.12 SOC dataset Propositional Rule Learner (PRL) BSE-RPM

Script E.20 (Appendix E) is used to develop the SOC Propositional Rule Learner (PRL) BSE-RPM.

Table F-43 The confusion matrix of the PRL BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	15	3
Actual reusable (1)	2	14

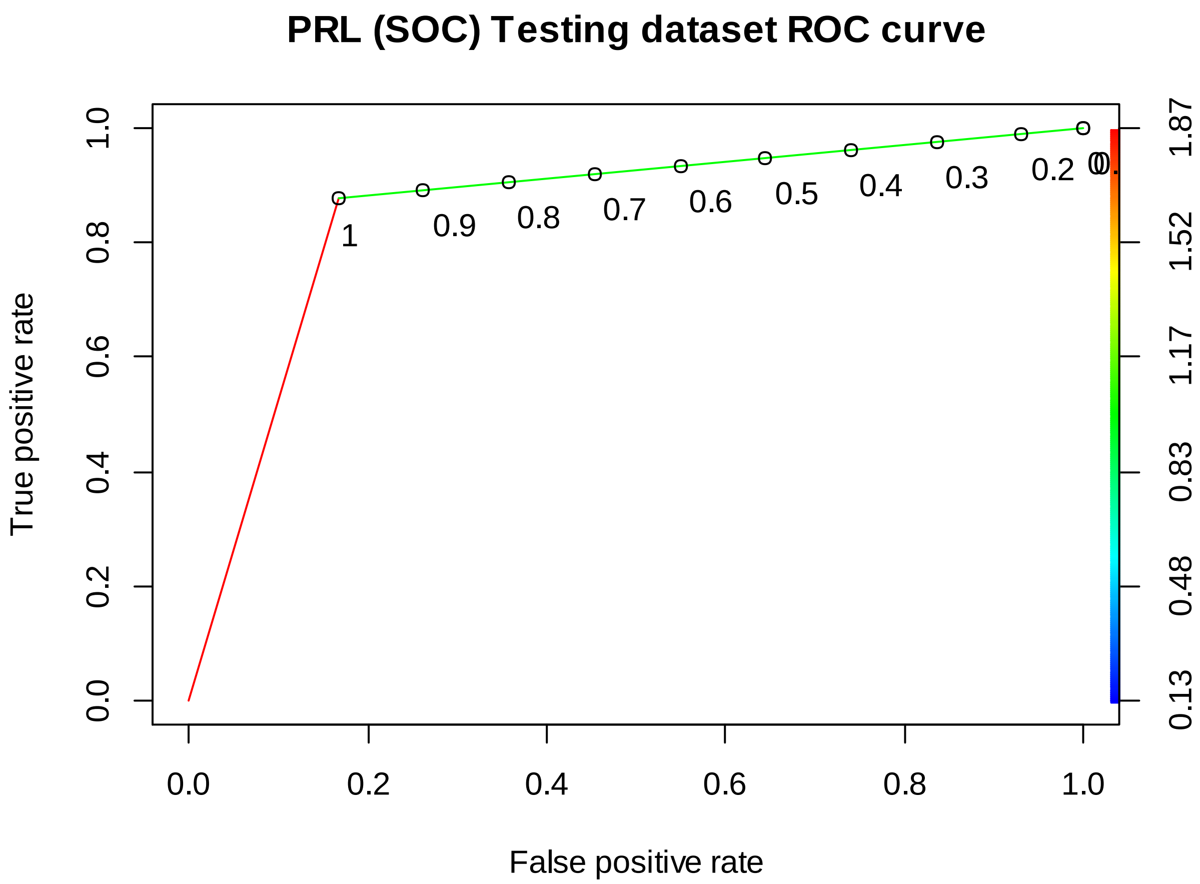

Figure F-41 The ROC curve of the PRL BSE-RPM (SOC dataset)

The **AUC value** for the SOC Propositional Rule Learner (PRL) BSE-RPM is equal to 0.85.

Table F-44 The rules set of the PRL BSE-RPM (SOC dataset)

Rule number (to be considered in order)	Rule	Result
1st	If: (D14 >= 5) and (D17 >= 4)	Then: E3C=1 (11.0/0.0)
2nd	Else if: (D20 <= 2) and (C22 <= 4)	Then: E3C=1 (8.0/0.0)
3rd	Else if: (D16 >= 4)	Then: E3C=1 (4.0/0.0)

Rule number (to be considered in order)	Rule	Result
4th	Else if: (C22 >= 4) and (D12 >= 3)	Then: E3C=1 (8.0/0.0)
5th	Else if none	Then: E3C=0 (47.0/6.0)

F.3.13 SOC dataset Support Vector Machines (SVM) BSE-RPM

Script E.21 (Appendix E) is used to develop the SOC Support Vector Machines (SVM) BSE-RPM.

Table F-45 The confusion matrix of the SVM BSE-RPM (SOC dataset)

	Predicted response values	
	Non-reusable (0)	**Reusable (1)**
Actual non-reusable (0)	16	2
Actual reusable (1)	0	16

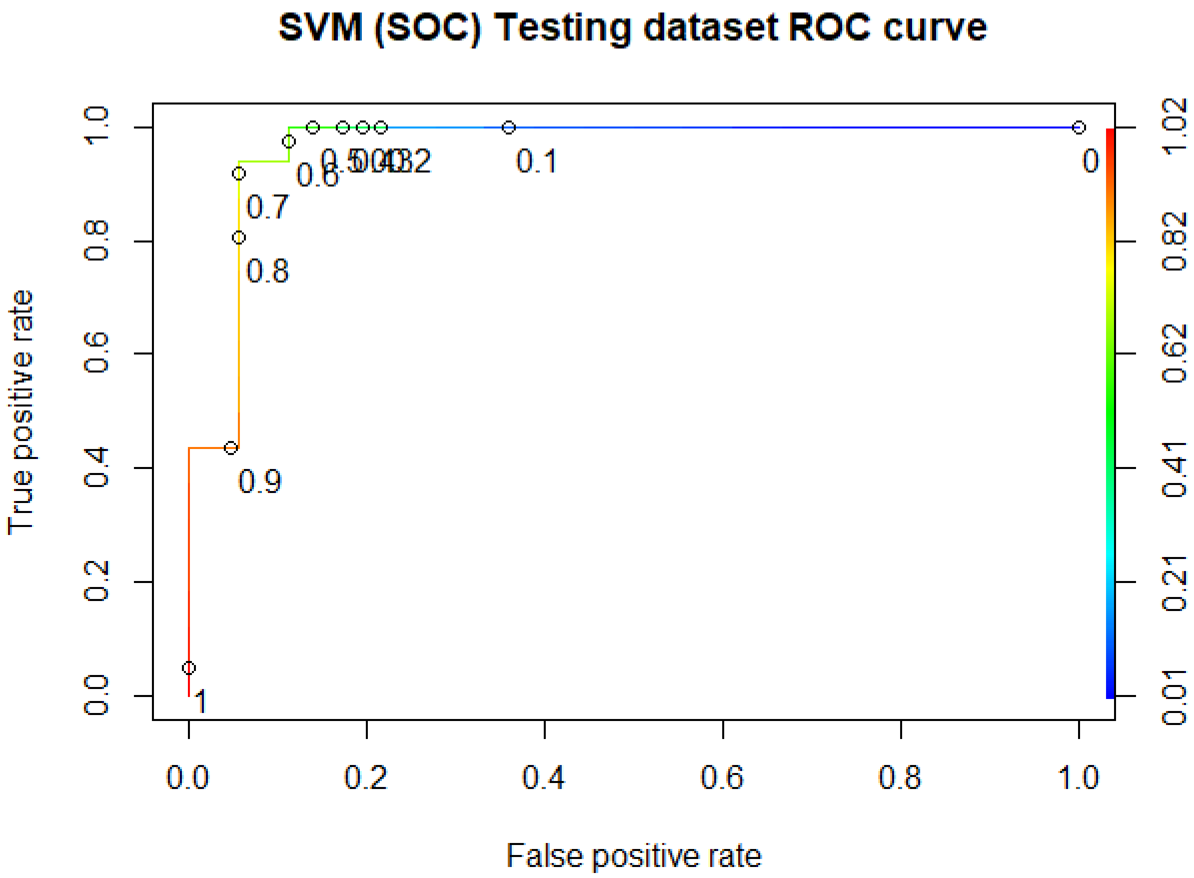

Figure F-42 The ROC curve of the SVM BSE-RPM (SOC dataset)

The **AUC value** for the SOC Support Vector Machines (SVM) BSE-RPM is equal to 0.97.

Appendix G Certificate of Ethical Approvals

Ethical Approval P105548 is presented at the beginning of the thesis

G.1 Ethical Approval P88781

Coventry University

Certificate of Ethical Approval

Applicant:

Kambiz Rakhshanbabanari

Project Title:

A predictive model for assessing the reuse potential of structural elements at the end-of-life of a building based on professional experience - Towards achieving a circular economy

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

24 March 2019

Project Reference Number:

P88781

G.2 Ethical Approval P67530

Certificate of Ethical Approval

Applicant:

Kambiz Rakhshanbabanari

Project Title:

A predictive model for assessing reuse potential of structural elements at the end-of-life of a building - Towards achieving a circular economy

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

04 July 2018

Project Reference Number:

P67530