Room Occupancy Detection Based on Random Forest with Timestamp Features and ANOVA Feature Selection Method

Alam, S., Sari, R. M., Alfian, G. & Farooq, U.

Original citation:
Alam, S, Sari, RM, Alfian, G & Farooq, U 2024, 'Room Occupancy Detection Based on Random Forest with Timestamp Features and ANOVA Feature Selection Method', Journal of Computing Science and Engineering (JCSE), vol. 18, no. 1, pp. 10-18.
https://dx.doi.org/10.5626/JCSE.2024.18.1.00

DOI 10.5626/JCSE.2024.18.1.00
ISSN 1976-4677
ESSN 2093-8020

Publisher: Korean Institute of Information Scientists and Engineers (KIISE)

JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 18, No. 1, March 2024, pp. 10-18

Room Occupancy Detection Based on Random Forest with Timestamp Features and ANOVA Feature Selection Method

Sahirul Alam*, **Risa Mahardika Sari, and Ganjar Alfian**
Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta, Indonesia
sahirul.alam@ugm.ac.id, risamahardika89@mail.ugm.ac.id, ganjar.alfian@ugm.ac.id

Umar Farooq
Faculty of Business and Law, Coventry University, Coventry, United Kingdom
ad9061@coventry.ac.uk

Abstract

To improve energy efficiency, understanding occupant behavior is crucial for adaptive temperature control and optimal electronic device usage. Our study introduces a room occupancy detection system using machine learning and Internet-of-Things sensors to predict occupant behavior patterns. Initially, indoor IoT sensor devices are installed to observe occupant behavior, and datasets are generated from sensor data, including temperature, humidity, light, and CO_2 levels, in both occupied and vacant rooms. The collected dataset undergoes analysis through a machine learning-based model designed to classify room occupancy. First, the timestamp features, extracted from date-time data, such as time of day and part of the day, are extracted. ANOVA feature selection is applied to identify five crucial features. Ultimately, the random forest model is employed to classify room occupancy based on the selected features. Results indicate that our proposed model significantly outperforms other models—achieving improvements of up to 99.713%, 99.467%, 99.676%, 99.676%, and 99.571% in accuracy, precision, recall, specificity, and F1-score, respectively. The trained model holds potential for integration into web-based systems for real-time applications. This predictive model is poised to contribute to the optimization of electronic device efficiency within a room or building by continuously monitoring real-time room conditions.

Category: Information Retrieval / Web

Keywords: Occupancy detection; Machine learning; Feature selection; IoT; Web-based system

I. INTRODUCTION

Effective management of room occupancy has become a pivotal aspect in energy consumption management while ensuring optimal comfort within indoor spaces. Room occupancy detection, a burgeoning field in smart building technology, plays a crucial role in achieving this delicate balance [1]. By accurately identifying whether a room is occupied or vacant, intelligent systems can dynamically adjust lighting, heating, ventilation, and air conditioning (HVAC) settings to minimize energy wastage without compromising on comfort. This indicates the significance of room occupancy detection as a sustainable solution for energy-efficient building management. Occupancy detection

Open Access http://dx.doi.org/10.5626/JCSE.2024.18.1.00 http://jcse.kiise.org

Received 15 February 2024; **Accepted** 12 March 2024
*Corresponding Author

pISSN: 1976-4677 eISSN: 2093-8020

without the use of cameras is preferred due to heightened privacy concerns associated with visual surveillance [2]. This preference has led to the exploration and implementation of alternative methods that utilize various sensors. The utilization of various sensors, including those measuring temperature, humidity, light, and CO_2 levels, forms the backbone of occupancy detection systems, enabling a comprehensive understanding of the environmental conditions that indicate occupant presence.

The advent of the Internet-of-Things (IoT) emerges as a pivotal technological paradigm that seamlessly integrates with non-intrusive sensor technologies to enable efficient room occupancy detection [3]. By harnessing data from an array of sensors, IoT transforms static spaces into dynamic, responsive environments even in industrial spaces [4]. This interconnected ecosystem allows for real-time monitoring and analysis of occupancy patterns, facilitating precise adjustments to lighting, HVAC, and other environmental factors to optimize energy consumption. The fusion of IoT and sensor technologies not only ensures privacy preservation but also underscores a sophisticated approach toward creating intelligent, energy-efficient spaces capable of adapting to the dynamic needs of occupants while maintaining a commitment to sustainability.

Earlier studies demonstrated the successful application of IoT in indoor occupancy detection systems. Utilizing sensor data within the room, encompassing light, temperature, humidity, and CO_2 levels, is employed for discerning the presence or absence of occupants through the implementation of a classification model based on machine learning [2]. Past studies also indicate that enhanced efficiency of an occupancy detection system is attainable through the utilization of more accurate sensor equipment. The careful choice of prediction models and features holds significant importance in the endeavor to optimize the system's overall performance [5].

Addressing these challenges underscores the paramount need to enhance the efficiency of the prediction model for room occupancy detection through deliberate consideration of feature extraction and selection. In this study, we incorporate two features derived from the timestamps: the hour of the day and the specific part of the day. Additionally, we deploy the analysis of variance (ANOVA) feature selection method to identify and retain crucial features, ultimately utilizing five selected features in our proposed method.

In machine learning, a timestamp serves as a feature that denotes the time of occurrence for each data point. This temporal information is crucial for tasks such as time series analysis, forecasting, and event prediction, allowing models to capture and leverage chronological patterns in the data to improve predictive accuracy in time-dependent scenarios [6]. Timestamps enable algorithms to adapt and make informed decisions based on the temporal order of events. Timestamps have demonstrated success in sequential recommendation [7], and given the relevance of occupancy behavior to time-dependent scenarios, there is potential for the implementation of timestamps in occupancy detection. However, as of now, no research has explored the application of timestamps for occupancy detection.

The structure of this paper is outlined as follows. Section II initiates the discourse on related works, delineating the contributions of this paper. Section III provides details on the dataset utilized in this research, introduces new features proposed as novel contributions, outlines the feature selection method employed, and offers a concise explanation of random forest (RF) as the proposed methodology. Section IV unfolds the results and discussion, encompassing the performance evaluation of machine learning models, the influence of the feature selection method, and the practical applications of the system. Finally, Section V concludes the paper, summarizing key findings, and outlining avenues for future research within this domain.

II. LITERATURE REVIEW

The focus on occupancy data is driven by its potential to optimize the trade-off between energy consumption and occupant comfort [1]. The identification of occupancy within buildings or rooms can be achieved by utilizing data gathered from various sensors. Analyzing this data enables the implementation of occupancy detection systems, allowing for intelligent and adaptive control of building systems such as lighting, HVAC to optimize energy consumption while ensuring occupant's comfort. The existing studies employ a range of sensors for occupancy detection, including climate sensors (measuring humidity, temperature, and CO_2 levels) [8-10], mobility sensors (such as passive infrared, Bluetooth low energy [11], GPS, and Wi-Fi sensor [12]), smart meters (for energy and water) [13], and cameras [14]. Nevertheless, owing to privacy apprehensions, camera sensors are less preferred [10, 15]. Consequently, this research concentrates on harnessing climate sensors, specifically those measuring temperature, humidity, light, and CO_2.

Several studies utilizing climate sensors have employed machine learning models for detecting room occupancy. Table 1 presents a summary of prior research that utilized machine learning for occupancy detection. The studies listed in Table 1 demonstrate promising performance, indicating the effectiveness of climate sensors in occupancy detection [8, 16-21]. However, many of these studies lack real-life experimentation. Therefore, our research not only proposes a method to enhance the performance of occupancy detection systems but also includes practical implementation to address this gap.

No prior research has explored the combination of timestamp features, ANOVA feature selection, and RF classifiers for occupancy detection. This study addresses

Table 1. Studies utilizing machine learning for occupancy detection using climate sensor

ML model	Year	Sensors used	Accuracy (%)
RBF [16]	2012	CO_2, L, H, T, S, PIR	87.62
AE [17]	2017	CO_2, L, H, T	98.88
LSTM and RNN [18]	2019	CO_2, H, T, P, PIR, camera	70.00
CNN and LSTM [19]	2020	CO_2, H, T, AP	95.42
LSTM [20]	2021	CO_2, H, T, L	96.80
AE [21]	2021	CO_2, H, T, L	94.50
PI-PRM [8]	2021	CO_2, T	97.00

RBF: radial basis function, AE: autoencoder, LSTM: long short-term memory, RNN: recurrent neural network, CNN: convolutional neural network, PI-PRM: physics-informed pattern-recognition machine, CO_2: carbon dioxide, H: humidity, L: light, P: power, PIR: passive infrared S: sound, T: temperature, AP: air pressure.

this gap by proposing the integration of these three methods to enhance the accuracy of occupancy detection. The contributions of this study can be summarized as follows:

- Introduction of the timestamp as a promising feature for occupancy detection.
- Integration of the timestamp feature, ANOVA feature selection, and RF for improved occupancy detection.
- Executing the feasibility of our proposed model for the web-based occupancy detection system.

III. METHODOLOGY

The illustrated model in Fig. 1 employs data from IoT sensors, encompassing variables such as temperature, humidity, light, and CO_2. Subsequent to data collection, a data preprocessing step is implemented to eliminate inconsistent data and substitute missing values with averages or medians through data imputation. In the feature extraction phase, supplementary timestamp features are derived. ANOVA feature selection is applied to identify crucial features, instrumental in predicting room occupancy. The RF algorithm is employed for occupancy prediction, and model performance is assessed by comparing it with alternative machine learning models. To conduct model evaluation, a hold-out method is employed, with 70% of the dataset allocated for training and the remaining portion for testing. Following training, the trained model is seamlessly integrated into a web-based application, facilitating straightforward utilization by administrators and staff.

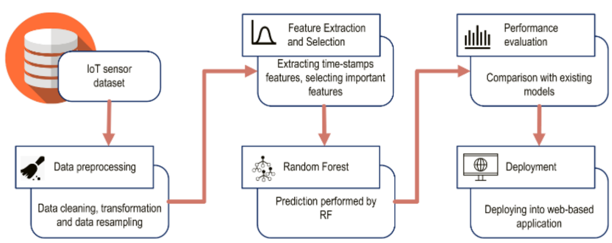

Fig. 1. Proposed machine learning model to detect room occupancy.

A. Dataset

While this research involved data collection and real-world application of the system, a public dataset was utilized to test and evaluate the proposed machine learning model. To construct a predictive model, a dataset suitable for indoor settings provided by a previous study was utilized [2]. This dataset was curated based on two distinct room conditions: one with occupants and another without occupants. For conditions with occupants, data reflecting room variables like temperature, humidity, light, humidity ratio, and CO_2 is stored in the database, and the associated sensor data is labeled as "1." Similarly, for uninhabited conditions or "label 0," sensor data representing this state is stored for subsequent analysis. A total of 8,143 sensor data points were gathered, featuring average values of 20.619 for temperature, 25.731 for humidity, 119.519 for light, 0.0038 for humidity ratio,

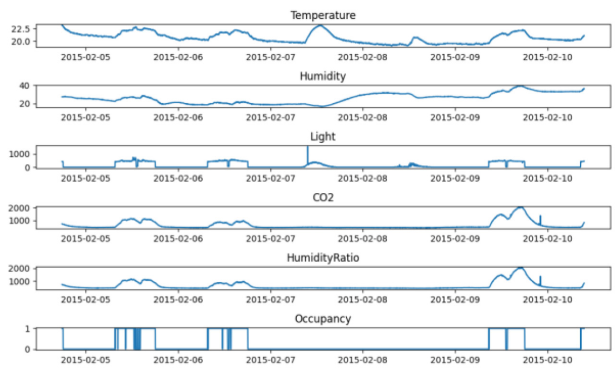

Fig. 2. Room occupancy dataset presented in time series charts.

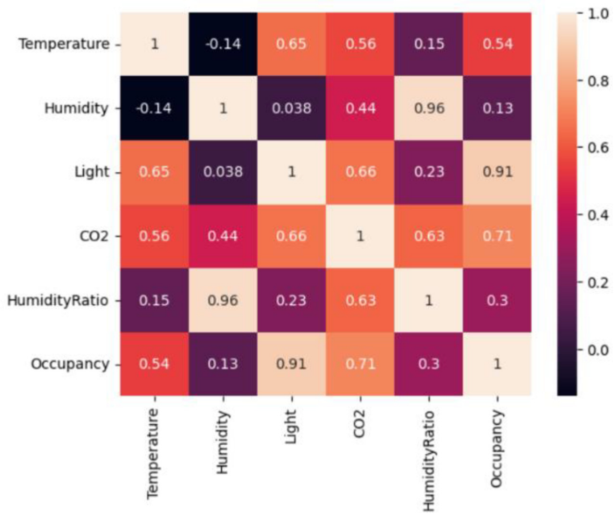

Fig. 3. Attribute correlation for room occupancy dataset.

and 606.546 for CO_2. Researchers have systematically compiled and stored data associated with the unoccupied room condition, as depicted in Fig. 2, for analysis.

The significance of features indicates the connection between attributes and the condition of the room. To explore this relationship, we employed Pearson correlation coefficient, ranging from -1 to +1. A negative or positive value signifies a negative or positive correlation, and a greater absolute value indicates a stronger correlation. Attributes with a high correlation to the output class can be employed as input features to enhance the accuracy of the prediction model. Fig. 3 illustrates that light, CO_2, and temperature exhibit a strong positive correlation with the class, while humidity demonstrates a weak correlation.

B. Timestamp Features

The gathered sensor data comprised temperature, humidity, light, humidity ratio, and CO_2, each recorded with corresponding date and time information at 1-minute intervals. Moreover, we introduced timestamp features in our study, specifically the “hour of the day” and “part of the day,” extracted from the date and time column. The delineation of "part of the day" encompassed categories like early morning ($4 < x \leq 8$), morning ($8 < x \leq 12$), afternoon ($12 < x \leq 16$), evening ($16 < x \leq 20$), night ($20 < x \leq 24$), and late night ($x \leq 4$). Through label encoding, we assigned numerical representations to these categories: "early morning" as 0, "morning" as 1, "afternoon" as 2, "evening" as 3, "night" as 4, and "late night" as 5. The inclusion of additional timestamp features, such as the hour of the day and part of the day, in our final dataset is illustrated in Fig. 4.

date	Temperature	Humidity	Light	CO2	HumidityRatio	Hour	Part	Occupancy
2015-02-04 17:51:00	23.18	27.2720	426.0	721.25	0.004793	17	3	1
2015-02-04 17:51:59	23.15	27.2675	429.5	714.00	0.004783	17	3	1
2015-02-04 17:53:00	23.15	27.2450	426.0	713.50	0.004779	17	3	1
2015-02-04 17:54:00	23.15	27.2000	426.0	708.25	0.004772	17	3	1
2015-02-04 17:55:00	23.10	27.2000	426.0	704.50	0.004757	17	3	1

Fig. 4. Additional timestamps feature in our dataset.

C. ANOVA Feature Selection

ANOVA feature selection, when used in conjunction with correlation analysis, offers a robust method for identifying the best features for machine learning models. ANOVA assesses the statistical significance of variance between different groups or classes within a dataset, enabling the identification of features that exhibit significant differentiation among the groups.

In this feature selection context, the F-value derived from ANOVA serves as a key metric [22]. The F-value reflects the ratio of variance between group means to variance within groups. A higher F-value suggests that the means of different groups are significantly different, indicating the importance of the associated feature in distinguishing between these groups. By considering F-values alongside correlation analysis, one can prioritize features that not only have significant variance across classes but also exhibit low correlations with each other. This comprehensive approach ensures the selection of non-redundant, highly informative features, enhancing the model's ability to capture relevant patterns and relationships in the data. Fig. 5 displays the significance of features determined by the F-value computed through ANOVA in our dataset. Ultimately, we opted for the five highest features—namely, light, CO_2, temperature, part of the day, and humidity ratio—for in-depth analysis.

D. Random Forest

The RF algorithm is a classification technique that combines decision trees [23]. This method involves

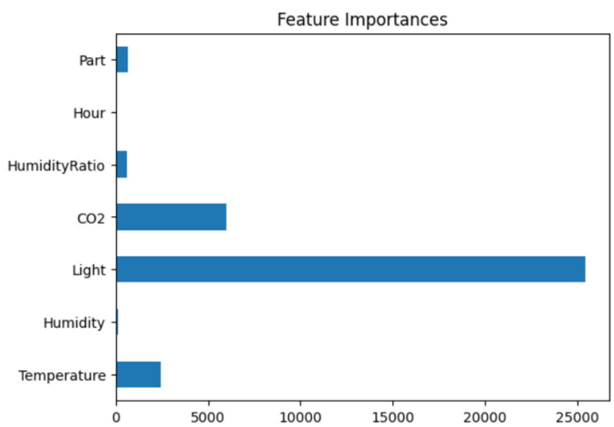

Fig. 5. The feature importance is presented by ANOVA for each feature.

constructing individual decision trees by randomly selecting a subset of attributes at each node for making splits. The process operates as follows: given a training dataset (D), the number of trees (T) in the model, a subspace dimension (d), and the available features (F), a bootstrapped sample (D_t) is derived from the original dataset (D). This sample includes certain records from the original dataset multiple times and excludes others. Subsequently, a subset (d) of attributes is randomly chosen from the bootstrapped sample (D_t) to be considered as candidates for splits at each node. The decision tree classifier is trained on this bootstrapped sample (D_t) along with the selected attributes (d), and it is grown to its maximum extent without pruning. This process is repeated for all trees within the forest. During the classification phase, each tree contributes its vote, and the most prevalent class is assigned as the predicted outcome.

RFs can address various challenges faced by decision trees, such as preventing overfitting and minimizing variance. The input attributes in this study include light, CO_2, temperature, humidity ratio, and part of the day. The random forest model consists of 100 trees (T), five attributes (F), and it utilizes the Gini index to make splits with the reduced number of attributes (d).

Machine learning models were employed for room occupancy classification, implemented in Python using XGBoost and Scikit-learn, with default parameters provided by Scikit-learn [24]. Model evaluation followed a hold-out method with a 70:30 ratio for training and testing.

Table 2 depicts the confusion matrix, a crucial instrument for evaluating a classifier's performance. Accurately classified data is denoted by true positive (TP) and true negative (TN), whereas inaccurately classified data is represented by false positive (FP) and false negative (FN). In this specific dataset, occupied rooms are designated the label 1, while empty rooms are assigned the label 0. Essential performance metrics, such as accuracy (%), precision (%), sensitivity or recall (%), and specificity (%), were calculated using the hold-out method. Table 3 showcases the performance metrics of the classification model.

Table 2. Classifier confusion matrix

		Predicted class	
		1	**0**
Actual class	1	TP	FN
	0	FP	TN

Table 3. Classifier model performance metrics

Metric	Formula
Accuracy	$\frac{(TP+TN)}{(TP+TN+FP+FN)}$
Precision	$\frac{TP}{(TP+FP)}$
Specificity	$\frac{TN}{(TN+FP)}$
Sensitivity/Recall	$\frac{TP}{(TP+FN)}$

IV. RESULT AND DISCUSSION

A. Performance of Machine Learning Models

In this study, we utilized supervised machine learning methods to identify room occupancy. This used input data from an IoT sensor device, including temperature, humidity, light, and CO_2. We introduced timestamp features such as “hour of the day” and “part of the day,” derived from the date and time column. Additionally, we applied a feature selection method based on ANOVA to choose essential features. Our study focused on assessing the performance of the proposed model in comparison to other machine learning models.

Various classification models, such as logistic regression (LR), k-nearest neighbor (KNN), decision tree (DT), adaptive boosting (AdaBoost), eXtreme gradient boosting (XGBoost), support vector machine (SVM), and multilayer

Table 4. Performance evaluation results

Method	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
Logistic regression	98.649	97.276	98.789	98.789	98.008
KNN	99.426	99.073	99.212	99.212	99.142
Decision tree	99.386	98.912	99.256	99.256	99.083
AdaBoost	99.222	98.538	99.152	99.152	98.841
XGBoost	99.549	99.222	99.431	99.431	99.326
SVM	98.853	97.431	99.273	99.273	98.315
MLP	99.386	98.716	99.469	99.469	99.087
Proposed model	99.713	99.467	99.676	99.676	99.571

perceptron (MLP), were employed. The results presented in Table 4 illustrate the diverse performances of these models in terms of accuracy, precision, recall, and F1-score. The findings indicated that our proposed model demonstrated superior performance compared to alternative models, achieving improvements of up to 99.713% in accuracy, 99.467% in precision, 99.676% in recall, 99.676% in specificity, and 99.571% in F1-score.

The performance of our machine learning model for room occupancy detection is outstanding, as all evaluation metrics exceed 97%. This includes accuracy, which indicates that the model correctly classifies room occupancy in over 98% of cases. Moreover, precision, a measure of the model's ability to avoid false positives, also exceeds 97%, ensuring minimal incorrect identifications of occupancy. The F1-score, which balances precision and recall, is also consistently above 98%, indicating the model's robustness and effectiveness in detecting room occupancy while maintaining a low rate of false alarms.

B. Impact of Feature Selection Method on Model Accuracy

Fig. 6 illustrates the impact of incorporating timestamp features and utilizing ANOVA-based feature selection on the accuracy of classification models. The results indicate a decline in accuracy for decision tree, AdaBoost, and XGBoost models. Conversely, the suggested RF model exhibits an enhanced accuracy of up to 0.164% when timestamp features are integrated and ANOVA-based feature selection is employed. This underscores that the removal of insignificant features led to an improvement in RF accuracy.

The RF model demonstrates a noteworthy improvement in accuracy, affirming the effectiveness of incorporating timestamp features and employing ANOVA-based feature selection. This enhancement implies that the selected features, influenced by the temporal dimension and ANOVA analysis, contribute positively to the RF model's ability to discern patterns and make accurate predictions. The marginal increase of 0.164% in RF accuracy may seem modest, but it highlights the sensitivity of the model to feature selection. Even small improvements can be significant, especially in applications where precision is crucial, such as occupancy detection systems. This finding emphasizes the nuanced interplay between feature relevance and model performance.

Furthermore, the success of the proposed RF model serves as a practical example of the potential benefits of feature optimization in machine learning applications. The ability to enhance accuracy by removing insignificant features not only improves the model's predictive capabilities but also contributes to a more streamlined and efficient system.

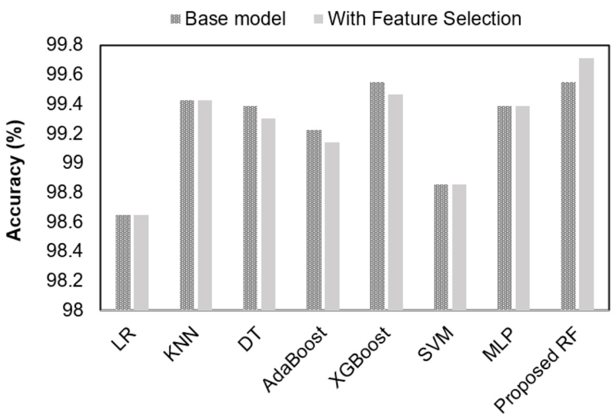

Fig. 6. Impact of timestamp and feature extraction method on model accuracy.

C. Practical Application

This study aims to develop a web-based system that employs a machine learning model to precisely detect room occupancy. By utilizing a machine learning-based predictive model, it is possible to accurately discern whether there is an occupant inside the room or not.

Occupancy detection employs various sensors, including temperature, humidity, light, and CO_2 sensors, through the implementation of IoT technology. In this study, the IoT system utilizes a wireless router as the IoT gateway and a mini-PC serving as both the IoT broker and server. Fig. 7 illustrates the IoT hardware components utilized in this research.

The web-based monitoring system was constructed using the PHP programming language and a MySQL database on the server side. Python was employed for the REST API and machine learning model. The predictive model was established using the Flask web framework and the Scikit-learn library on the server side. The interface depicted in Fig. 8 showcases the real-time display of sensor data, enabling managers to monitor the conditions within the room.

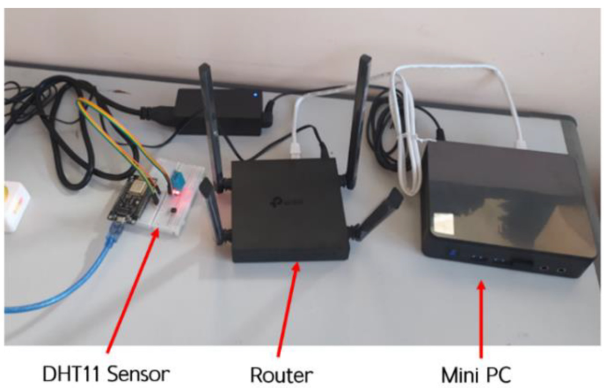

Fig. 7. IoT hardware for occupancy detection.

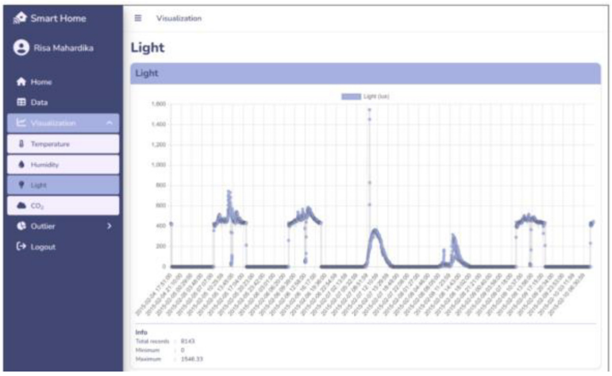

Fig. 8. Web-based occupancy detection monitoring system.

The development of a web-based system employing a machine learning model for precise room occupancy detection has numerous benefits that significantly impact various aspects of building management and resource optimization. One key benefit lies in enhanced energy efficiency. By accurately discerning whether a room is occupied or not, the system can intelligently control lighting, heating, and ventilation systems, ensuring they are activated only when necessary. This targeted approach to resource utilization not only contributes to substantial energy conservation but also aligns with sustainable and eco-friendly practices.

Additionally, the web-based monitoring system's real-time display of sensor data, as depicted in Fig. 8, empowers managers with valuable insights into the conditions within the room. This not only facilitates proactive decision-making but also enables a deeper understanding of room utilization patterns over time. Managers can leverage this information to make informed adjustments to space layouts, usage policies, or even building design, leading to more efficient and user-friendly environments.

Furthermore, the integration of a machine learning model into the occupancy detection system contributes to the overall system reliability. Traditional systems may be prone to false positives or negatives, leading to inaccuracies in occupancy detection. The predictive capabilities of machine learning models, when trained with diverse and representative datasets, can significantly improve the accuracy and reliability of occupancy detection. This, in turn, minimizes disruptions caused by incorrect occupancy assessments and enhances the overall effectiveness of building management strategies.

V. CONCLUSION

The proposed RF model, incorporating timestamps features and ANOVA feature selection, successfully identified room occupancy using a dataset obtained from IoT sensors measuring temperature, humidity, light, and CO_2 levels. Timestamp features, derived from the date and time column, including time of day and part of the day, were introduced. ANOVA analysis identified five crucial features —light, CO_2, temperature, humidity ratio, and part of the day—which were employed as inputs for the proposed RF model.

The results demonstrated that the proposed model surpassed alternative models (LR, KNN, DT, AdaBoost, XGBoost, SVM, and MLP) by impressive margins, achieving improvements of up to 99.713%, 99.467%, 99.676%, 99.676%, and 99.571% in accuracy, precision, recall, specificity, and F1-score, respectively. The trained model holds promise for integration into web-based systems for real-time house condition monitoring and room occupancy prediction.

Future work will involve evaluating IoT sensor performance under diverse conditions. Additionally, exploring extra IoT sensors as inputs for the classification model is warranted. Further research may include comparing the proposed model with other classification models and presenting findings on feature extraction and selection methods in the coming phases.

CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

ACKNOWLEDGMENTS

This work was funded by the Doctoral Competency Improvement Program Universitas Gadjah Mada Number 7743/UN1.P.II/Dit-Lit/PT.01.03/2023.

REFERENCES

1. A. N. Sayed, Y. Himeur, and F. Bensaali, “Deep and transfer learning for building occupancy detection: a review and comparative analysis,” *Engineering Applications of Artificial Intelligence*, vol. 115, article no. 105254, 2022. https://doi.org/10.1016/j.engappai.2022.105254
2. L. M. Candanedo and V. Feldheim, “Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models,” *Energy and Buildings*, vol. 112, pp. 28-39, 2016. https://doi.org/10.1016/j.enbuild.2015.11.071
3. Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, “Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives,” *Applied Energy*, vol. 287, article no. 116601, 2021. https://doi.org/10.1016/j.apenergy.2021.116601
4. D. Yoo, S. Chung, and J. Park, “Analysis and evaluation of channel-hopping-based MAC in industrial IoT environment,” *Journal of Computing Science and Engineering*, vol. 15, no. 4,

pp. 160-174, 2021. http://dx.doi.org/10.5626/JCSE.2021.15.4.160
5. N. S. Sani, I. I. S. Shamsuddin, S. Sahran, A. H. A. Rahman, and E. N. Muzaffar, "Redefining selection of features and classification algorithms for room occupancy detection," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 4-2, pp. 1486-1493, 2018. https://doi.org/10.18517/ijaseit.8.4-2.6826
6. Q. Chen, F. Jiang, X. Guo, J. Chen, K. Sha, and Y. Wang, "Combine temporal information in session-based recommendation with graph neural networks," *Expert Systems with Applications*, vol. 238, article no. 121969, 2024. https://doi.org/10.1016/j.eswa.2023.121969
7. J. Li, Y. Wang, and J. McAuley, "Time interval aware self-attention for sequential recommendation," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, Houston, TX, USA, 2020, pp. 322-330. https://doi.org/10.1145/3336191.3371786
8. S. I. Kampezidou, A. T. Ray, S. Duncan, M. G. Balchanos, and D. N. Mavris, "Real-time occupancy detection with physics-informed pattern-recognition machines based on limited CO_2 and temperature sensors," *Energy and Buildings*, vol. 242, article no. 110863, 2021. https://doi.org/10.1016/j.enbuild.2021.110863
9. T. H. Pedersen, K. U. Nielsen, and S. Petersen, "Method for room occupancy detection based on trajectory of indoor climate sensor data," *Building and Environment*, vol. 115, pp. 147-156, 2017. https://doi.org/10.1016/j.buildenv.2017.01.023
10. M. Parzinger, L. Hanfstaengl, F. Sigg, U. Spindler, U. Wellisch, and M. Wirnsberger, "Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection: using machine learning methods Random Forest and LASSO," *Building and Environment*, vol. 223, article no. 109313, 2022. https://doi.org/10.1016/j.buildenv.2022.109313
11. F. Demrozi, C. Turetta, F. Chiarani, P. H. Kindt, and G. Pravadelli, "Estimating indoor occupancy through low-cost BLE devices," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17053-17063, 2021. https://doi.org/10.1109/JSEN.2021.3080632
12. Y. Liu, T. Wang, Y. Jiang, and B. Chen, "Harvesting ambient RF for presence detection through deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 4, pp. 1571-1583, 2022. https://doi.org/10.1109/TNNLS.2020.3042908
13. C. Feng, A. Mehmani, and J. Zhang, "Deep learning-based real-time building occupancy detection using AMI data," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4490-4501, 2020. https://doi.org/10.1109/TSG.2020.2982351
14. K. Sun, P. Liu, T. Xing, Q. Zhao, and X. Wang, "A fusion framework for vision-based indoor occupancy estimation," *Building and Environment*, vol. 225, article no. 109631, 2022. https://doi.org/10.1016/j.buildenv.2022.109631
15. C. Sardianos, I. Varlamis, G. Dimitrakopoulos, D. Anagnostopoulos, A. Alsalemi, F. Bensaali, Y. Himeur, and A. Amira, "REHAB-C: recommendations for energy habits change," *Future Generation Computer Systems*, vol. 112, pp. 394-407, 2020. https://doi.org/10.1016/j.future.2020.05.041
16. Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations," in *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design*, Orlando, FL, USA, 2012.
17. Z. Liu, J. Zhang, and L. Geng, "An intelligent building occupancy detection system based on sparse auto-encoder," in *Proceedings of 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2017, pp. 17-22. https://doi.org/10.1109/WACVW.2017.10
18. H. Elkhoukhi, M. Bakhouya, M. Hanifi, and D. El Ouadghiri, "On the use of deep learning approaches for occupancy prediction in energy efficient buildings," in *Proceedings of 2019 7th International Renewable and Sustainable Energy Conference (IRSEC)*, Agadir, Morocco, 2019, pp. 1-6. https://doi.org/10.1109/IRSEC48032.2019.9078164
19. Z. Chen, C. Jiang, M. K. Masood, Y. C. Soh, M. Wu, and X. Li, "Deep learning for building occupancy estimation using environmental sensors," *Deep Learning: Algorithms and Applications*. Cham, Switzerland: Springer, 2020, pp. 335-357. https://doi.org/10.1007/978-3-030-31760-7_11
20. E. Hitimana, G. Bajpai, R. Musabe, L. Sibomana, and J. Kayalvizhi, "Implementation of IoT framework with data analysis using deep learning methods for occupancy prediction in a building," *Future Internet*, vol. 13, no. 3, article no. 67, 2021. https://doi.org/10.3390/fi13030067
21. K. P. Shirsat and G. P. Bhole, "Optimization-enabled deep stacked autoencoder for occupancy detection," *Social Network Analysis and Mining*, vol. 11, article no. 30, 2021. https://doi.org/10.1007/s13278-021-00730-6
22. Y. H. Wang, Y. F. Zhang, Y. Zhang, Z. F. Gu, Z. Y. Zhang, H. Lin, and K. J. Deng, "Identification of adaptor proteins using the ANOVA feature selection technique," *Methods*, vol. 208, pp. 42-47, 2022. https://doi.org/10.1016/j.ymeth.2022.10.008
23. L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324
24. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

Sahirul Alam https://orcid.org/0000-0002-3541-5922

Sahirul Alam received a Bachelor's degree in electrical engineering from Universitas Brawijaya, Malang, Indonesia, in 2013, a Master's degree in electrical engineering from Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2017, and a Doctoral degree in electrical engineering from Universitas Gadjah Mada, in 2022. He joined as a lecturer in the Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada in January 2022. His research interests include wireless networks, IoT, VANET, and cellular networks.

Risa Mahardika Sari https://orcid.org/0009-0003-4311-9183

Risa Mahardika Sari currently is a fresh graduate from the Department of Electrical Engineering and Informatics, Majoring in Software Engineering Technology from Vocational College, Universitas Gadjah Mada, Indonesia. Her research areas are artificial intelligence, machine learning, and web development.

Ganjar Alfian https://orcid.org/0000-0002-3273-1452

Ganjar Alfian currently works as an Assistant Professor at the Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Indonesia since 2022. Previously, he worked as an Assistant Professor at Dongguk University Seoul, Republic of Korea for five years. In July 2017, he was a short-term Visiting Researcher with the VSB-Technical University of Ostrava, Czech Republic. He received a Dr.Eng degree from the Department of Industrial and Systems Engineering, Dongguk University, Seoul, South Korea, in 2016. His research interests include artificial intelligence, RFID, IoT, machine learning, carsharing, and simulation. He is listed in the World's Top 2% Scientist 2022 by Stanford University.

Umar Farooq https://orcid.org/0000-0003-3432-7119

Umar Farooq is currently working as a lecturer of Business Analytics at the School of Strategy and Leadership, College of Business and Law, Coventry University, United Kingdom. Previously, he worked as an Assistant Professor for five years at Gulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan. He received his Master's and PhD in Industrial and Systems Engineering from Dongguk University, South Korea in 2017. Dr Umar has vast experience in the fields of supply chain management, operations management, and Business Analytics. His research interests are in RFID, IoT, Food Supply Chains, Machine Learning, and analytics.