Tracking Room Occupancy

Abhaya Rawal
Department of Computer Science
University of South Dakota
abhaya.rawal@coyotes.usd.edu

***Abstract*— This project uses environmental sensor data, such as temperature, CO2 levels, and light intensity, to predict room occupancy using machine learning. The work uses Gaussian Naïve Bayes and Random Forests models with datasets from the UCI Machine Learning Repository to classify room occupancy. The analysis emphasizes model accuracy, precision, recall, and F1-score while focusing on the effects of different environmental factors on occupancy detection. The project lays the groundwork for future Internet of Things and environmental sustainability applications while showcasing the potential of machine learning in smart building automation and providing insights into resource management and energy efficiency.**

Keywords—Room Occupancy Prediction, Environmental Sensor Data, UCI Machine Learning Repository, Machine Learning, Random Forest Classifier, Gaussian Naïve Bayes

I. Introduction

When it comes to smart building management, being able to accurately anticipate room occupancy has great consequences. Significant gains in resource allocation, energy efficiency, and overall building sustainability can result from it. The project aims to develop a model that not only predicts occupancy with high accuracy but also offers insights into the correlations and effects of various environmental factors on room occupancy by utilizing advanced machine learning techniques like Gaussian Naive Bayes and Random Forests.

The project uses metrics like accuracy, precision, recall, and F1-score to assess the models' performance. These measures offer an in-depth understanding of the model's advantages and disadvantages, directing future improvements. This research has implications that go beyond its immediate use in room occupancy prediction. It emphasizes how machine learning can improve intelligent building automation systems. The results of this study may play a key role in the development of more environmentally friendly and energy efficient building designs.

II. Motivation

The need for smart infrastructure that can maximize energy use, improve occupant comfort, and support environmental sustainability is growing as urbanization rises. One major opportunity to address these challenges is the ability to predict room occupancy accurately using environmental sensor data and machine learning. The potential to lower operating costs, increase energy efficiency, and enhance occupants' overall quality of life is what motivates this research. Moreover, it is consistent with the primary objectives of IoT integration in urban settings, which are to promote more flexible and responsive living and working environments. As a result, this study contributes to the fields of environmental sustainability and smart building automation by addressing both practical needs and technological advancements.

III. Literature Review

Several recent studies have focused on using machine learning methods to predict room occupancy. The purpose of these studies is to improve the built environment's thermal comfort, air quality, and energy efficiency. A technique for determining a room's occupancy is presented in one study. It makes use of datasets gathered from multiple sensors and a variety of machine learning algorithms, including Naive Bayes and Random Forest [1]. The use of models based on the Bayes Theorem and HMM, as well as supervised machine learning algorithms like SVM, RF, DT, and ANN, to detect and estimate indoor occupancy [2] is covered in another review. Furthermore, a study gathers more than 40,000 records [3] to create an extensive public training dataset for predicting building occupancy profiles. These studies show how machine learning techniques can be used to track room occupancy for a variety of purposes, such as environmental monitoring and energy conservation.

IV. Methodology

This study's methodology involves preprocessing, data analysis, and model development and evaluation. Python and its libraries are used for effective data handling and visualization which offers a strong framework for creating classification models, enabling an efficient and successful analysis. Similarly, to perform this project Environmental parameters like temperature, light, CO2 levels, and PIR (Passive Infrared Sensor) were included in the dataset that was taken from the UCI Machine Learning Repository [4]. Cleaning, normalization, and handling missing values were all part of the preprocessing of the data.

An exploratory data analysis was conducted to identify significant features influencing room occupancy. Two machine learning models, Random Forests and Gaussian Naive Bayes, were developed to predict room occupancy. The models were trained and tested on the dataset, using a split of training and test data. The models' performance was evaluated using metrics like accuracy, precision, recall, and F1-score to determine their effectiveness in predicting room occupancy. The results were analyzed to understand the impact of different environmental factors on occupancy prediction and to compare the two models.

V. RESULTS

A. Temperature fluctuations over date

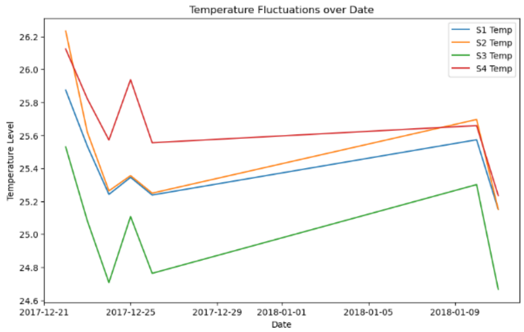

Figure 1: Different temperature datas fluctuations with date

The above graph displays a general downward trend in temperature over time for all four sensors, with some fluctuations. By 2018-01-09, the trend had sharply decreased.

B. Light fluctuations over date

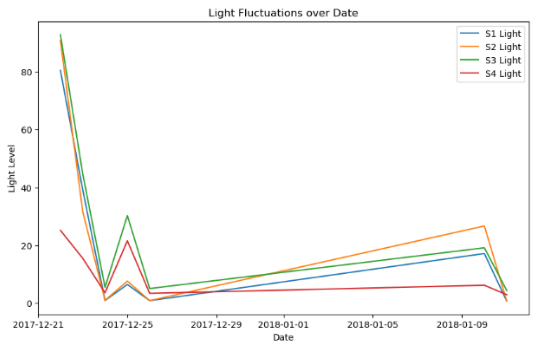

Figure 2: Different light datas fluctuations with date

The above graph shows how light levels measured by four sensors first fell sharply, then stabilized for a while before all of the sensors experienced a simultaneous decline by 2018-01-09.

C. Sound fluctuations over date

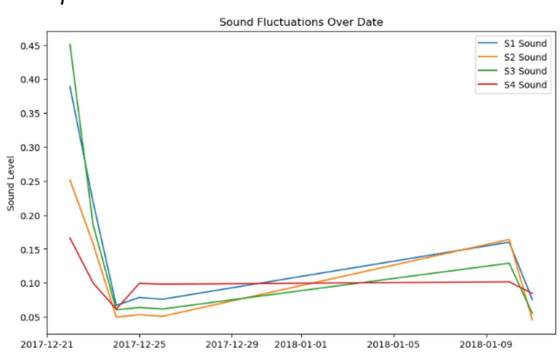

Figure 3: Different sound datas fluctuations with date

The above graph indicates a noticeable drop in sound levels, with a steep initial drop and a more gradual decline toward the end of the observed period.

D. CO2 fluctuations over date

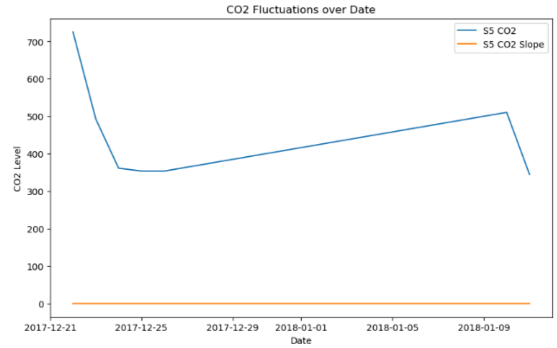

Figure 4: Different co2 datas fluctuations with date

The graph above reveals a trend of CO2 levels gradually rising over time, then sharply falling at the end of the observed dates.

E. PIR fluctuations over date

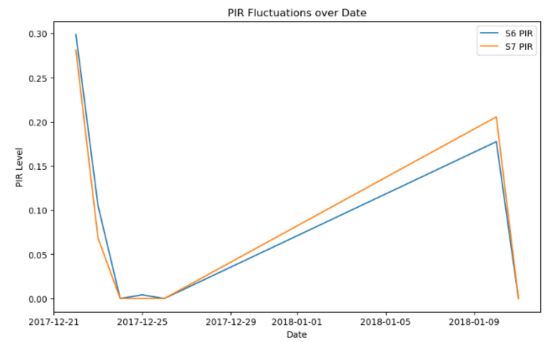

Figure 5: Different pir datas fluctuations with date

The above graph shows a major drop at the conclusion of the observed time frame, followed by a gradual increase.

F. Correlation Heatmap

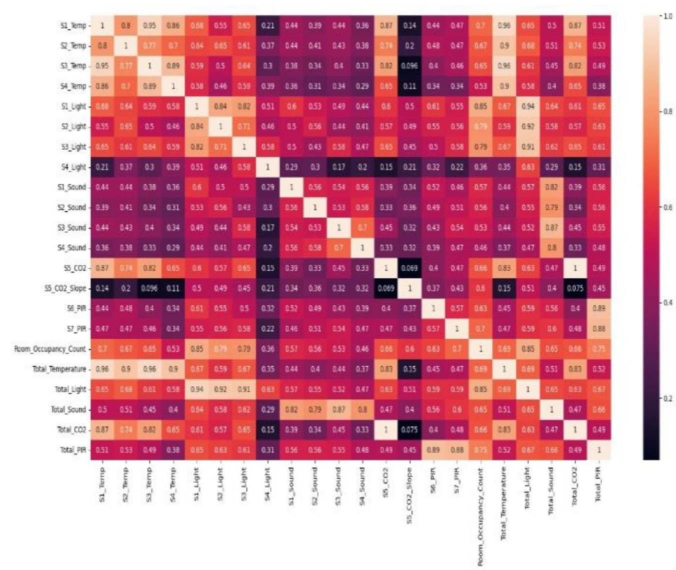

Figure 6: Correlation heatmap between variables

The above graph expresses wide range of correlations between various sensor data indicating significance in room occupancy.

G. Violinplot for Temperature, Light and Sound with respect to Room_Occupancy_Count

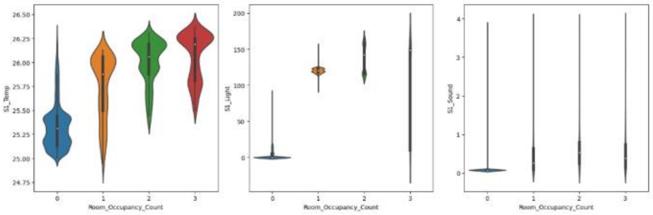

Figure 7: violinplot for s1_temp, s1_light and s1_sound with room_occupancy_count

The above violin plots indicate that while sound levels do not consistently trend with occupancy, temperature and light levels generally rise as the number of occupied rooms rises.

H. Violinplot for CO2 and PIR with respect to Room_Occupancy_Count

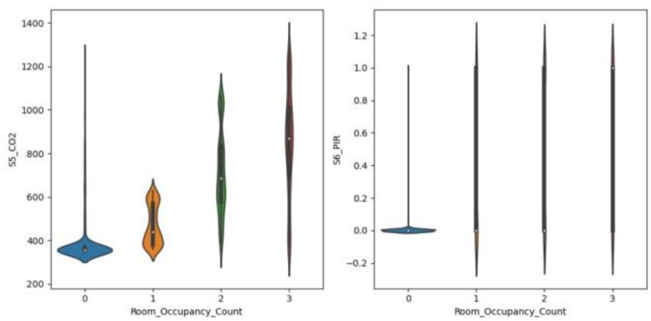

Figure 8: violinplot for s5_co2, and s6_pir with room_occupancy_count

The above violin plots show that CO2 levels tend to rise as the number of rooms occupied increases, but the PIR sensor readings show no visible trend in relation to occupancy.

I. Random Tree Classifier(RTC)

```
Accuracy Score : 0.9973078073586599
Precision Score : 0.9973155244785115
Recall Score : 0.9973078073586599
F1 Score : 0.997307333071888
```

Figure 9: Different metrics for rtc evaluation

The above result demonstrates RTC has achieved high performance with accuracy, precision, recall, and F1 score all above 0.997, indicating an excellent model fit for the data.

J. Gaussian Naïve Bayes Classifier (GNBC)

```
Accuracy Score : 0.9605145079270117
Precision Score : 0.9631398366977787
Recall Score : 0.9605145079270117
F1 Score : 0.96140276158927
```

Figure 10: Different metrics for gnbc evaluation

The above result also demonstrates GNBC has strong predictive performance with accuracy, precision, recall, and F1 score exceeding 0.96.

VI. CONCLUSION AND FUTURE WORK

In this study, I successfully created and verified two binary classification models for accurate occupancy estimation in indoor environments using sensor data. The study illustrated the viability of the model and its possible uses in smart building management to improve resource allocation, safety, and energy efficiency by utilizing innovative algorithms like Random Forest and Gaussian Naive Bayes.

Based on this work, future research can access real time data for dynamic occupancy scenarios, evaluate the scalability of the model in different types of buildings, and examine user behavior to improve predictions. By establishing a connection between occupancy estimates and actual energy consumption, more sustainable building management can be achieved.

ACKNOWLEDGEMENT

I sincerely appreciate Professor Dr. Farhad Akhbardeh's outstanding instruction, guidance, and support, all of which have contributed to my academic and professional development. I am fortunate to have had the chance to learn from him, and his contributions have had a significant impact on the success of this research project.

REFERENCES

[1] W. Zhang, Y. Wu and J. K. Calautit, "ScienceDirect," 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032122005937.

[2] K. Alanne and S. Sierla, "ScienceDirect," 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670721007186.

[3] M. S. Aliero, M. F. Pasha, D. T. Smith, I. Ghani, M. Asif, S. R. Jeong and M. Samuel, "Non-Intrusive Room Occupancy Prediction Performance," *Energies,* no. 10.3390/en15239231, 2022.

[4] A. P. Singh, V. Jain, S. Chaudhari, F. A. Kraemer, S. Werner and V. Garg, "UC Irvine Machine Learning Repository," 2018. [Online]. Available: https://archive.ics.uci.edu/dataset/864/room+occupancy+estimation.