Naive Bayes for Smart Building Management: Predicting Workspace Occupancy

Mohammed Ennejjar[1*], Samira Chabaa[1,2], Mohammed Ali Jallal[1,3], and Abdelouhab Zeroual[1]

[1] I2SP Research Team, Physics Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakesh, Morocco.
[2] Industrial Engineering Department, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco.
[3] Univ. Grenoble Alpes, CEA, Liten, Campus Ines, 73375, Le Bourget-du-Lac, France.

Abstract. Occupancy detection plays a crucial role in building management, by improving living conditions and optimizing energy efficiency. So, our paper is a part of this perspective and is divided into two parts. Initially, we delve into the significance of detecting occupancy in buildings, emphasizing its positive impact on human well-being and productivity. Subsequently, the second section is dedicated on using the Naive Bayes Classifier (NBC) to predict occupancy in an office room using variables like temperature, humidity, humidity ratio, light, and CO_2 level. This approach demonstrates an impressive accuracy of 97.7%, underscoring the efficacy and the effectivness of this probabilistic classifier in managing building occupancy.

1 Introduction

Occupancy detection plays a crucial role in building management [1]. Indeed, a more precise occupancy prediction allows for more efficient energy management by intelligently adjusting the systems equipped in buildings such as Heating, Ventilation, and Air Conditioning systems (HVAC) and lighting systems to provide energy consumption management and improve the comfort level of occupants.

1.1 Reduction of energy consumption

Among the major advantages of predicting occupancy in buildings is its ability to reduce energy consumption (figure 1). Indeed, by adapting lighting and HVAC systems in real time to the real needs of occupants, it is possible to make substantial savings.

According to the literature [2], Lighting represents around 19% of the energy consumption of a tertiary building. Therefore, the occupancy prediction is crucial to reduce this consumption by turning off lights in empty spaces and adjusting their intensity according to the available natural light. In addition, concerning heating and air conditioning, this prediction also makes it possible to reduce energy consumption by adapting the temperature depending on the presence or absence of occupants. Some studies have shown that this approach can result in a decrease in energy consumption of 20 to 30% [3]. Furthermore, ventilation constitutes another significant energy consumption item in buildings. Also, by occupancy prediction, it is possible to modulate the air flow according to the presence of occupants, which results significant energy savings.

1.2 Improving occupant comfort and productivity

Occupancy prediction offers more than just reducing energy consumption; it also has numerous advantages for occupant comfort and productivity. For example, it helps to maintain optimal temperature and brightness in workspaces, thus promoting comfort, concentration, and, by extension, productivity. Likewise, it facilitates the reservation of workspaces and meeting rooms according to the needs of the occupants, thus avoiding situations of clutter and loss of time. Subsequently, this approach also helps to reduce stress and increase occupant concentration by providing them with an optimal working environment. Some studies have shown that companies investing in these aspects can benefit from increased productivity, reduced stress and sick leave, improved employee satisfaction, as well as greater attractiveness to talented employees [4,5].

1.3 Provide security in the building

Predicting occupancy helps monitor spaces proactively and prevent intrusions, accidents, and emergencies effectively. By studying occupancy patterns, we can spot unusual activities, enhance lighting in critical areas, and guide emergency responses during crises. Additionally, integrating this with current security systems boosts their effectiveness and cuts down on false alarms. Thus, forecasting occupancy contributes to making buildings safer for everyone.

*Corresponding author
E-mail address: m.ennejjar.ced@uca.ac.ma

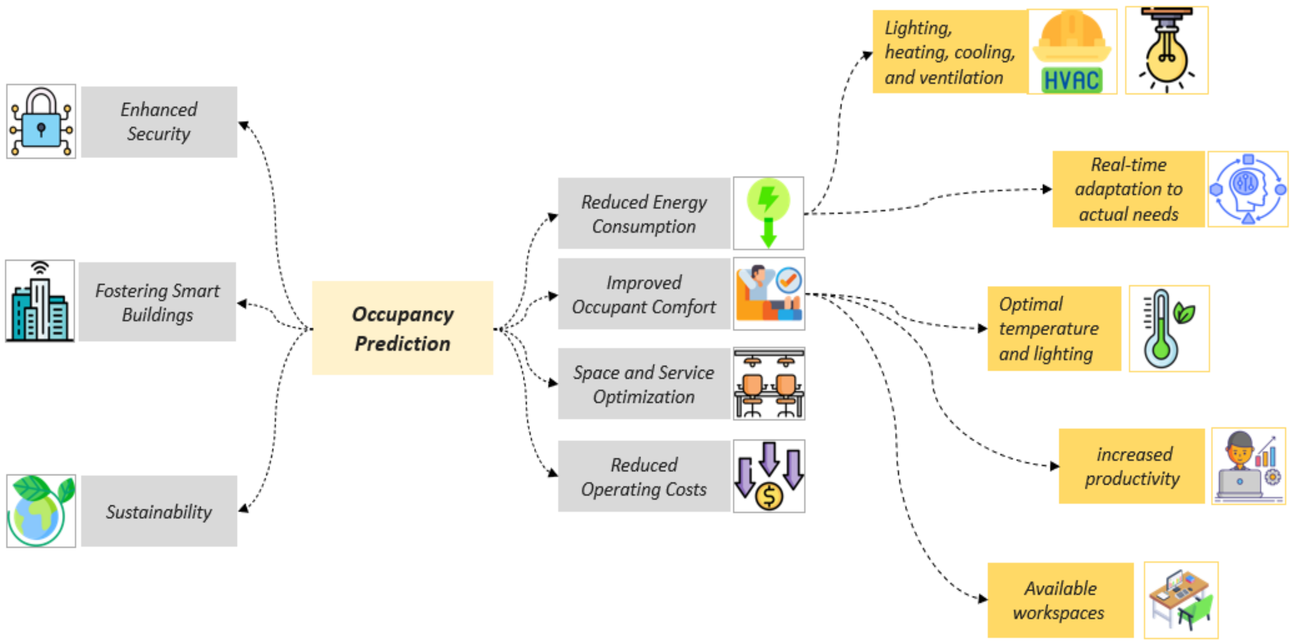

Figure 1: Benefits of Occupancy Prediction in Buildings.

The central objective of this paper is to show the importance of occupancy detection in the efficiency, comfort and productivity in building by treating as an application of the prediction of an office room occupancy using a probabilistic method based on simple reasoning which is Naive Bayes Classifier (NBC) based on Bayes' law.

The rest of this work is structured into three sections: the second section is dedicated to presenting the methodology and the materials. Subsequently, the obtained results are presented and discussed in the third section. Finally, the section four concludes our study with a summary.

2 Naive Bayes Classifier (NBC) for occupancy estimation

The Naive Bayes classifier is a widely used probabilistic model for classification tasks. It's based on Bayes' theorem [6] and makes a simplifying assumption of feature independence. In our study, we applied the Naive Bayes classifier to predict occupancy levels in an office environment using variables like Temperature, Humidity, Light, CO_2 level, Humidity Ratio, and Occupancy.

2.1 Dataset

For our study, we used a dataset that included measurements of temperature (x_1), humidity (x_2), humidity ratio (x_3), light (x_4), and CO_2 level (x_5). Specifically, the data was collected over a week, from February 02, 2015, to February 10, 2015, with measurements taken every minute. Furthermore, the office room dimensions were 5.85 meters by 3.50 meters by 3.53 meters [7]. Consequently, this comprehensive dataset provided a solid foundation for building our prediction model.

2.2 Model Formulation

The classifier predicts the probability of an instance x belonging to a certain class y using Bayes' theorem:
.

$$P(y|x) = \frac{P(x|y) \cdot P(y)}{P(x)} \tag{1}$$

Where:

- $P(y|x)$ is the probability of the class y given x.
- $P(x|y)$ is the probability of the instance x given the class y.
- $P(y)$ is the marginal probability of the class y.
- $P(x)$ is the marginal probability of instance x.

2.3 Parameter Estimation

We estimate the mean (μ) and standard deviation (σ) for each variable xi in each class y as follows:

$$\mu_{x_i|y} = \frac{1}{n_y}\sum_{j=1}^{n_y} x_{i,j} \tag{2}$$

$$\sigma_{x_i,y} = \sqrt{\frac{1}{n_y}\sum_{j=1}^{n_y}\left(x_{i,j} - \mu_{x_i|y}\right)^2} \tag{3}$$

Where: n_y is the number of instances in class y, and $x_{i,j}$ is the value of variable xi for the j^{th} instance of class y.

2.4 Probability Density Function

The probability density function for $P(x_i\ |y)$ follows a Gaussian distribution:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi}\sigma_{x_i|y}} exp\left(-\frac{\left(x_i - \mu_{x_i|y}\right)^2}{2\sigma^2_{x_i|y}}\right) \quad (4)$$

2.5 Prior Probabilities

The prior probabilities $P(y = 0)$ and $P(y = 1)$ are estimated based on the number of instances in each class using the following equations [8]:

$$P(y = 0) = P(0) = \frac{N_0}{N_y} \quad (5)$$

$$P(y = 1) = P(1) = \frac{N_1}{N_y} \quad (6)$$

Where N_y is the total number of instances, N_0 and N_1 are the number of instances in classes 0 and 1, respectively.

2.6 Prediction

Finally, we compute the probability of each class given the input x using the following equations:

$$P(y=0|x) = \frac{P(x_1|0).P(x_2|0).P(x_3|0).P(x_4|0).P(x_5|0).P(y=0)}{P(x|y=0)} \quad (7)$$

$$P(y=1|x) = \frac{P(x_1|1).P(x_2|1).P(x_3|1).P(x_4|1).P(x_5|1).P(y=1)}{P(x|y=0)} \quad (8)$$

$If\ P(y = 1|x) > P(y = 0|x)$, we predict $y = 1$; otherwise, $y = 0$.

This approach enables accurate estimation of occupancy levels in an office environment using the Naive Bayes classifier, particularly effective for handling multiple correlated features.

3 Evaluation metrics

We evaluate the performance of the binary classification model using several key metrics [9] such as:

Accuracy: it measures the overall proportion of correctly classified instances and it is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Where TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

Sensitivity (Recall): also known as recall that measures the model's ability to correctly identify all positive instances and it is calculated as:

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

Positive Predictive Value (PPV): also known as precision and it measures the proportion of predicted positive instances that are actually positive and it is calculated as:

$$PPV = \frac{TP}{TP + FP} \quad (11)$$

Specificity: Specificity measures the model's ability to correctly identify all negative instances. It is calculated as:

$$Specif\ icity = \frac{TN}{TN + FP} \quad (12)$$

Negative Predictive Value (NPV): it measures the proportion of predicted negative instances that are actually negative and it is calculated as:

$$NPV = \frac{TN}{TN + FN} \quad (13)$$

These metrics provide a comprehensive view of the performance of the binary classification model and help evaluate its ability to accurately predict occupancy levels in an office environment.

4 Results and discussion

The Gaussian Naive Bayes model was trained on the training data and evaluated using the testing data by the use of several performance metrics and the results are summarized in the table below:

Table 1: NBC performance in %.

Metrics	Dataset	
	Training data	Testing data
Accuracy	97.8	97.7
Sensitivity	97	94
Specifity	99.53	99
PPV	99.53	99
NPV	97.3	96

These metrics indicate that the model has an excellent ability to distinguish between occupied and unoccupied states, with an overall accuracy of 98%. The high sensitivity (99%) suggests that the model correctly detects almost all cases of occupancy, while the specificity of 97% shows that it also accurately identifies non-occupancy cases. Subsequently, the positive predictive value (95%) and negative predictive value (99%) confirm the robustness of the model in predicting occupancy and non-occupancy states.

The confusion matrix, presented in Fig.2 and Fig.3, illustrate the model's performance in terms of correct and incorrect classifications in the test phase and the training phase, respectively.

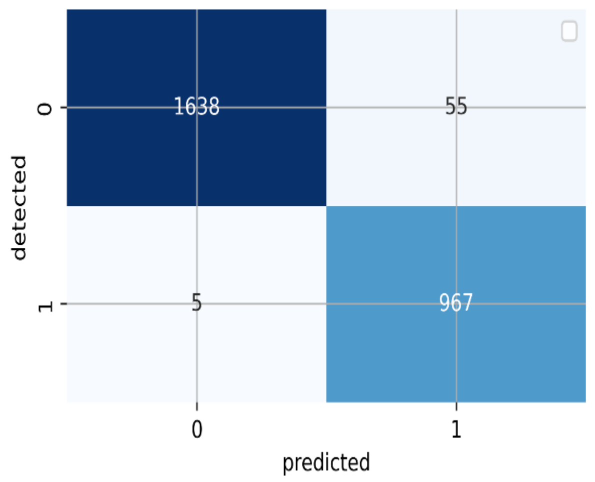

Figure 2: Confusion matrix (test data).

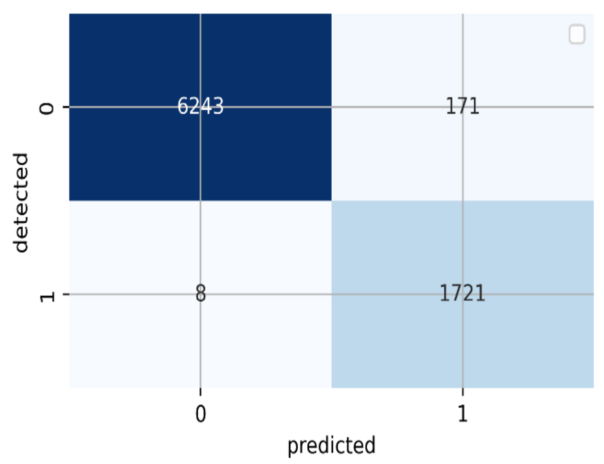

Figure 3 : Confusion matrix (training data).

The model correctly identified 1638 cases of non-occupancy (true negatives) and 967 cases of occupancy (true positives). It made 55 false predictions of non-occupancy (false positives) and 5 false predictions of occupancy (false negatives). These results show a strong ability of the model to correctly predict occupancy states with a low error rate.

The figure (4) compares the measured and predicted occupancy values for the testing data, which the red line represents the predicted values, while the black line represents the measured values. Thus, the model's predictions closely follow the actual measurements over different time periods, indicating a good fit of the model to the data.

The figure (5) shows the Receiver Operating Characteristic (ROC) curve to evaluate the performance of the binary classification model on occupancy. The curve shows a trade-off between sensitivity (true positive rate) and false positive rate for different model decision thresholds. With an AUC (Area Under the Curve) of 0.99,

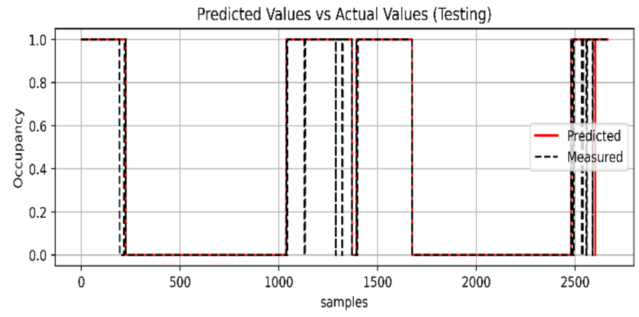

Figure 4: Comparison between predicted and measured occupancy.

the model demonstrates excellent discriminative ability, approaching the upper left corner of the graph, indicating a low false positive rate and a high true positive rate. These results reveal that the model performs extremely well at distinguishing between occupied and unoccupied classes, making accurate predictions while minimizing classification errors.

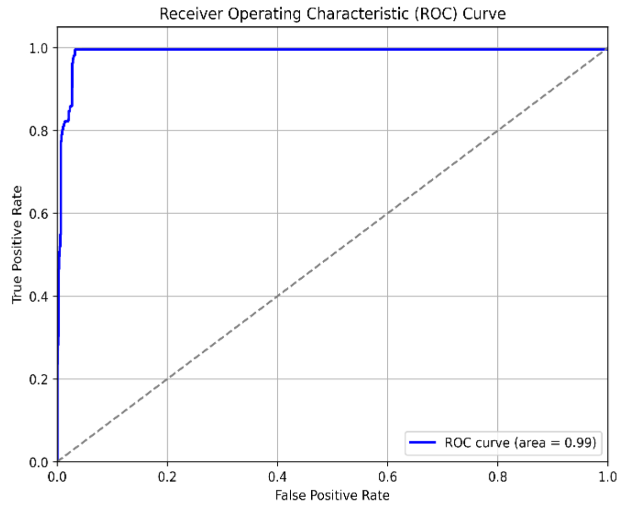

Figure 5 : Receiver Operating Characteristic Curve.

5 Comparative analysis of reported results in the literature.

From the literature, it is evident that our model achieved a remarkable accuracy of 97.7% during the testing phase. In the study referenced by [7], which utilized the same dataset, researchers applied four models to predict occupancy using inputs such as temperature, humidity, humidity ratio, light, and CO2 levels. Their results demonstrated that the Random Forest (RF) model achieved an accuracy of 95.53%, while the Linear Discriminant Analysis (LDA), the Gradient Boosting Machine (GBM) and Classification and Regression Tree (CART) models achieved accuracies of 97.9, 95.76% and 94.52%, respectively.

6 Conclusion

In this study, we highlight the crucial importance of predicting building occupancy for two essential aspects: reducing energy consumption and improving occupant well-being. Indeed, accurately predicting occupancy makes it possible to effectively adjust heating, ventilation and air conditioning systems, which helps minimize energy waste and reduce costs. In addition, by understanding occupancy patterns, it becomes possible to optimize the interior layout to guarantee the comfort of occupants and thus promote their productivity. So, to make this prediction, we used NBC (Naive Bayes Classifier) model which integrates various meteorological data. The obtained results are remarkable. Our model demonstrated exceptional accuracy, reaching a percentage of 97% for the test data. This performance confirms that even simple models, such as Bayesian networks, can provide very reliable results in predicting workspace occupancy.

References

1. L. Rueda, K. Agbossou, A. Cardenas, N. Henao, S. Kelouwani, A comprehensive review of approaches to building occupancy detection, Build. Environ. 180 (2020) 106966. https://doi.org/10.1016/j.buildenv.2020.106966
2. Attia, S., M. Hamdy, and S. Ezzeldin, 2017: Twenty-year tracking of lighting savings and power density in the residential sector. Energy Build., 154, 113–126. https://doi.org/10.1016/j.enbuild.2017.08.041
3. Y. Ding, X. Pan, W. Chen, Z. Tian, Z. Wang, and Q. He," Prediction Method for Office Building Energy Consumption Based on an Agent-Based Model Considering Occupant–Equipment Interaction Behavior," Energies, vol. 15, no. 22, p. 8689, 2022. https://www.mdpi.com/1996-1073/15/22/8689#
4. Goleman, W. C. Kim, R. A. Mauborgne, et C. M. Christensen," HBR's 10 Must Reads 2015: The Definitive Management Ideas of the Year from Harvard Business Review (with bonus McKinsey Award–Winning article 'The Focused Leader')," Harvard Business Review Press, 2015.
5. G. R. Newsham, J. A. Veitch, M. Q. Zhang, and A. D. Galasiu," Comparing better building design and operation to other corporate strategies for improving organizational productivity: a review and synthesis," Intelligent Buildings International, vol. 14, no. 1, pp. 3-22, 2022. https://doi.org/10.1080/17508975.2019.1588700
6. Koch, K. R., and Koch, K. R. (1990). Bayes' theorem. Bayesian Inference with Geodetic Applications, 4-8.
7. L. M. Candanedo and V. Feldheim," Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models," Energy and Buildings, vol. 112, pp. 28-39, 2016. https://doi.org/10.1016/j.enbuild.2015.11.071
8. Zhang, Z. (2016). Naive Bayes classification in R. Annals of translational medicine, 4(12).
9. Permatasari, T., Sandy, Y., Pratiwi, C., Damanik, K., and Silitonga, A. (2023, May). Naive Bayes Classifier (NBC) Application on the Nutritional Status of Adolescents in Medan. In Proceedings of the 4th Annual Conference of Engineering and Implementation on Vocational Education, ACEIVE 2022, 20 October 2022, Medan, North Sumatra, Indonesia.