

Long-Term Occupancy Monitoring and Data Analysis for large Building HVAC Energy Optimization^{*}

Stefan Knauth^[0000–0001–6873–472X], Mikail Aktürk, and Sebastian Blankenhorn

Faculty for Computer Sciences, Mathematics and Geomatics
HFT Hochschule für Technik Stuttgart – University of Applied Sciences
Stuttgart, Germany
`stefan.knauth@hft-stuttgart.de`
<https://www.hft-stuttgart.com/>

Abstract. This paper presents a sensor-based system for detecting room occupancy at a university to identify underutilized spaces and optimize energy consumption for heating and cooling. A novel approach is the categorization into 3 long term occupancy categories simplifying decision making for the operators. By distinguishing between frequently and infrequently used rooms, the system enables targeted control of HVAC systems, allowing for energy savings without affecting user comfort or academic scheduling. Data is collected from various types of sensors, including motion, temperature, humidity and CO₂ sensors, installed in lecture halls, labs, seminar rooms, and study areas. The collected data is processed and analyzed using machine learning techniques to detect occupancy patterns and usage anomalies. The system enables dynamic space management and informs facility managers, allowing unused rooms to be excluded from unnecessary climate control. Results show significant potential for energy savings without compromising user comfort. The approach demonstrates how smart building technologies can contribute to sustainable campus operations and efficient energy use. Employed ML algorithms reach an accuracy value of 91% and enable reliable categorization of occupancy states.

Keywords: Energy efficiency · Room occupancy detection · HVAC optimization.

1 Introduction

Efficient use of building space and energy resources is a growing priority for Universities seeking to reduce operational costs and meet sustainability goals. This paper presents a novel system designed for long-term occupancy analysis,

^{*} This work was kindly funded by the Carl Zeiss Foundation, Stuttgart, Germany, as part of the program “Klimaschutz am Campus: Gebäude nachhaltig nutzen” Project “NeuLand” with reference KSS-KLIWI-791-032.

tailored specifically to support non-expert users such as building superintendents and facility managers in making informed planning decisions.

Numerous studies have proposed sensor-based systems and Machine Learning detect room occupancy and optimize HVAC (heating, ventilation, and air conditioning) control in real time. For example, in [1] an ML approach is presented, however in a single living room deployment which is not comparable to the university situation. [4] employs a single PIR sensor again in a single room scenario. The paper focuses more on hardware performance issues. Paper [5] uses multiple sensors in a single room to estimate the number of persons with high accuracy, but with high deployment effort. A factory scenario in [3] is related to the university scenario and the authors have performed measurements for a week, but report issues with the IR camera technology. A building energy dashboard is proposed and investigated in [2], but no occupancy detection is performed. In general, most of the approaches in the literature focus on short-term occupancy detection and automated response systems, often targeting technical experts and facility engineers.

In contrast, the proposed system monitors room usage over weeks and months using distributed environmental sensors (E. g. motion detectors, temperature, humidity and CO₂ sensors). Besides the long term analysis and the targeted audience, another important feature is the machine learning based occupancy detection algorithm that remains robust even in the presence of incomplete or missing sensor data. It allows the system to deliver reliable insights despite common real-world limitations such as sensor faults or coverage gaps.

To ensure usability by non-specialist staff, the system includes an intuitive, web-based dashboard that visualizes long-term occupancy statistics, highlights underutilized rooms which are unnecessarily heated/air conditioned, and enables data-driven scheduling and space planning. Rather than automating control systems directly, our approach enables strategic decision-making aimed at long-term energy efficiency and optimized space utilization. This paper details the system architecture, machine learning methodology, and results from a deployment on a university campus.

2 Description of Data Analysis

2.1 Sensors

The system currently uses temperature, humidity, CO₂ level and movement (PIR sensor) data. The data is provided by microcontroller based battery/solar powered sensors of own manufacture. The larger sensor comprises an ESP32 microcontroller with WiFi and various other environmental sensors like temperature, humidity, illumination, air quality and more. The smaller one is Atmega328p based, comprises PIR, CO₂ and DHT22 temperature and humidity sensors, and runs on very low power, as long as the CO₂ sensor is activated only a few times per hour. Connection is established by LORA radio. The sensors are typically mounted at the ceiling such that they can not be easily accessed by unauthorized people (Figure 1).



Fig. 1. Sensor deployment. Left: Sensor prototype mounted at ceiling. Right: Classroom Situation.

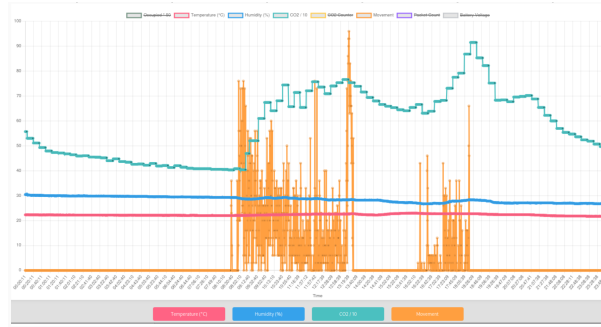


Fig. 2. Example data set, measured from 0:00 until 23:59 Red: Temperature, Blue: Humidity, Green: CO₂, Yellow: Movement.

Figure 2 represents a typical weekday dataset of a lecture room. Regular lectures took place between 8:00 and 13:00 and 16:00-18:00. Movement is already detected if students sit at their place and take notes or make else slight movement. People in the room let increase the CO₂ level continuously. Starting at about 10:00, ventilation took place by opening the windows. Due to loud outside noise, the windows were not opened all the time but only periodically, such that the CO₂ level stayed more or less constant. After the lecture, CO₂ decreased with a typical decay slope. Temperature increased with people being in the room, and the relative humidity decreased accordingly. On the right side, CO₂ stays high even when there is no movement. This is attributed to the limited coverage of the PIR sensor, which does comprise only parts of the hall, so that occupancy is detected by CO₂ only.

2.2 Data Labeling

For further processing, labels for annotation of the measurements were defined:

- **Label 0:** no people are present in a room.
- **Label 1:** the room is only sparsely occupied, e.g. the occupation is below 15% of the room capacity.
- **Label 2:** the room is occupied (Lecture, Training, Group Work etc.), e.g. the occupation is 15% or above.

Note that the value “15%” has been defined by observing actual situations at the authors university and other values might be more useful on other campus. With this definition, training- and validation data has been recorded and labeled, either by manual observation of the occupancy or by using the room schedules. Labeled data sets comprise also sets with missing CO₂ data and with missing PIR-movement data. The time granularity of the labeling is 15 minutes.

2.3 Evaluation of ML algorithms

Several occupancy detection ML algorithms have been investigated and most of them showed reasonable results (Figure 3):

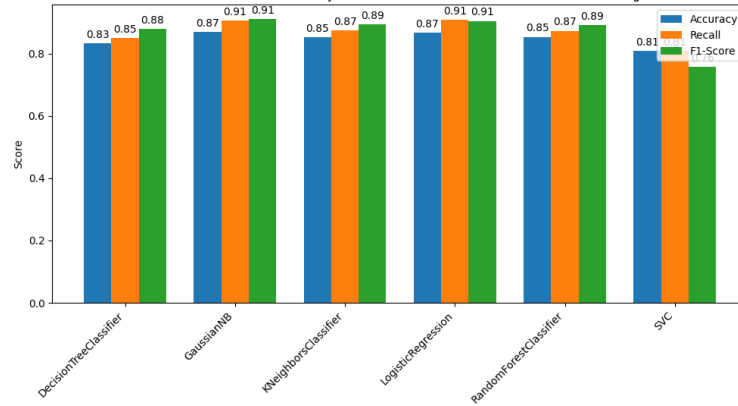


Fig. 3. Detailed evaluation of algorithms for model “difference”.

Particularly noteworthy are the GaussianNB and logistic regression models, which achieved an F1 score of around 91% and best values for the three evaluation metrics. Due to the excellent performance and simplicity of the model, logistic regression was chosen as it offers high efficiency and interpretability for binary classification tasks such as this one. Another advantage of logistic regression compared to Naive Bayes (GaussianNB) is that it makes fewer strong assumptions about the distribution of the data and is therefore more flexible with respect to different data sets

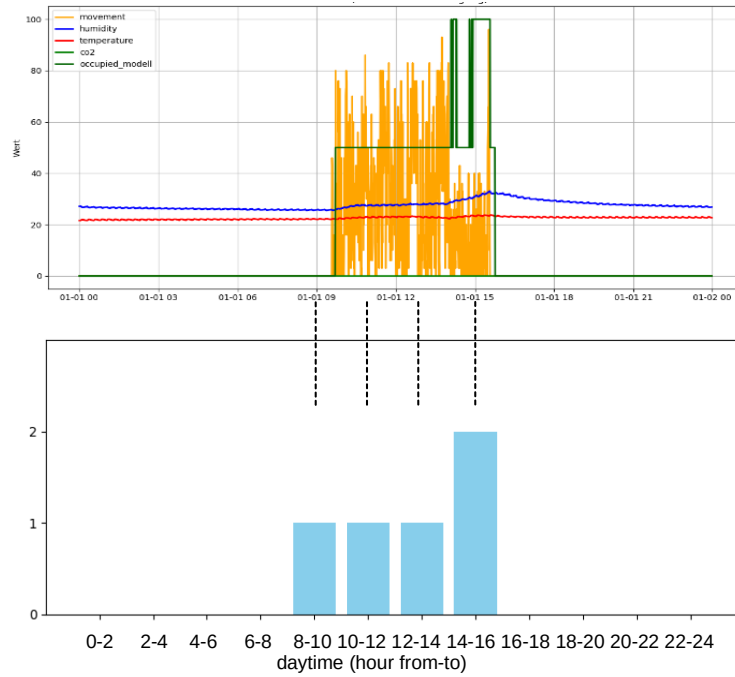


Fig. 4. Extraction of 2h occupation levels (0..2). Scenario: medium occupancy between 8..14 h (High movement, students coming and going, temp. and humidity staying constant). 14..16 h: High occupancy (Temp. and humidity rising but movement is lower since written exams have been negotiated).

2.4 Postprocessing

In order to further process the occupation data, a two-hour block occupation is extracted. It is categorized in 3 Levels (see also Figure 4). The two hour timespan reflects the typical occupation rhythm of classes and reduces the amount of data, which is helpful if long records of a whole semester are regarded. The levels are assigned according to the following rules:

- **Level 0:** No people are present in the room, or samples with label 1 were recorded for a maximum of 10 minutes. This allows short entries into the room, e.g. by cleaning staff, to be filtered out.
- **Level 1:** Samples with label 1 were recorded for more than 10 minutes. This is classified as “medium” occupancy.
- **Level 2:** Samples with label 2 occur for more than 30 minutes or at least 10 minutes of label 1 and 10 minutes of label 2 were recorded.

These Levels are used for generation of long-term statistics, which can be accessed by the Dashboard.

3 System and User Interface

3.1 System Overview

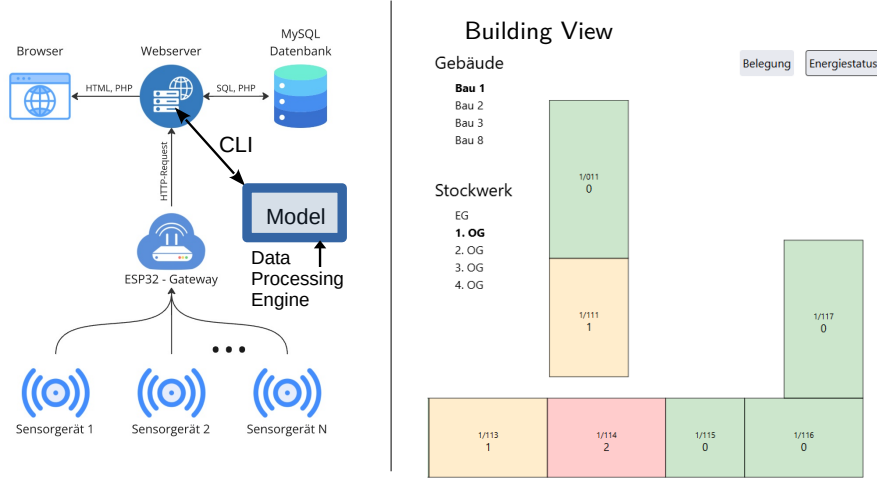


Fig. 5. Left: System overview. The data processing engine comprises the ML Model and algorithms. Right: User interface, selected view is “energy state”. Red: Unnecessary air conditioning / heating detected over the selected weeks.

The System comprises sensors, gateways, a server where the NodeJS / Angular application is running, a database, a processing engine for data analysis, and a web interface to allow connecting from a browser. The sensors transmit their data in a 30 sec. interval via LORA to the gateway, which forwards the message to the server. The server stores the data in the database and triggers the processing in order to get aggregated data with the corresponding 2 hour level information (Figure 5 left).

3.2 Dashboard

The server provides a user interface (Figure 5 right) which displays the long-term state of the rooms. Aggregated values of the 2 hour levels are compiled into colors. There are two different modalities: occupation view and energy waste view. The colors in energy waste view indicate:

- **green:** frequently used room - no energy wasted.
- **orange:** occupancy is below green threshold, possible indication of energy waste.
- **red:** Red: room is conditioned and usage is minor, action is encouraged.

The Time interval, e.g. begin- and end date of the interval to be analyzed, can be configured and will typically cover several weeks to allow a profound measure. Further views include current room occupation levels and displaying the recorded sensor data of a selected room.

4 Results and Conclusion

This paper presents a robust sensor-based system for long-term room occupancy monitoring in large educational buildings, demonstrating its potential to significantly improve HVAC energy efficiency. By combining environmental sensor data with machine learning algorithms, the system enables accurate long term occupancy classification even under conditions of partial data loss. The deployment on a university campus confirms that the approach is both technically feasible and practically useful. Through its user-friendly dashboard, the system empowers non-expert facility managers to make data-driven decisions regarding space utilization and climate control. Overall, the approach offers a scalable and sustainable solution for reducing energy consumption without compromising user comfort, and it represents a valuable contribution to the development of intelligent campus infrastructure

Acknowledgements This work was kindly funded by the Carl Zeiss Foundation, Stuttgart, Germany, as part of the program “Klimaschutz am Campus: Gebäude nachhaltig nutzen” Project “NeuLand” with reference KSS-KLIWI-791-032.

References

1. Aliero, M.S., Pasha, M.F., Smith, D., et al.: Non-intrusive room occupancy prediction performance analysis using different machine learning techniques. *Energies* **15**(23), 9231 (2022). <https://doi.org/10.3390/en15239231>
2. Berger, C., Calder, K., Cassway, S., Walton, C.: Developing a dashboard for high occupancy buildings. In: 2021 Systems and Information Engineering Design Symposium (SIEDS). pp. 1–6. Charlottesville, VA, USA (2021). <https://doi.org/10.1109/SIEDS52267.2021.9483791>
3. Kuutti, J., Saarikko, P., Sepponen, R.E.: Real time building zone occupancy detection and activity visualization utilizing a visitor counting sensor network. In: 2014 11th International Conference on Remote Engineering and Virtual Instrumentation (REV). pp. 219–224. Porto, Portugal (2014). <https://doi.org/10.1109/REV.2014.6784260>
4. Leech, C., Raykov, Y.P., Ozer, E., Merrett, G.V.: Real-time room occupancy estimation with bayesian machine learning using a single pir sensor and microcontroller. In: 2017 IEEE Sensors Applications Symposium (SAS). pp. 1–6. Glassboro, NJ, USA (2017). <https://doi.org/10.1109/SAS.2017.7894091>
5. Singh, A.P., Jain, V., Chaudhari, S., Kraemer, F.A., Werner, S., Garg, V.: Machine learning-based occupancy estimation using multivariate sensor nodes. In: 2018 IEEE Globecom Workshops (GC Wkshps). pp. 1–6. Abu Dhabi, United Arab Emirates (2018). <https://doi.org/10.1109/GLOCOMW.2018.8644432>