Pertanika J. Sci. & Technol. 34 (1): 79 - 94 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Occupancy-based Association Analysis of Indoor Environment Quality and Energy Consumption in a Smart Office at Tropical Region

Soon Chong Johnson Lim[1], Boon Tuan Tee[2]*, Rajesh Sunkari[3], Peng Wah Siew[3], and Ming Foong Lee[4]

[1]*Faculty of Mechanical Engineering,Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia*
[2]*Faculty of Mechanical Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia*
[3]*Seikou Systec Sdn. Bhd., 58200 Kuala Lumpur, Malaysia*
[4]*Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia*

ABSTRACT

As the building sector transitions towards sustainability, there has been a growing emphasis and research on the interplay and balance between occupants' well-being and energy consumption. This paper investigated the relationship between indoor environmental quality (IEQ) parameters, energy consumption, and occupancy in a smart office environment equipped with sensor devices in a tropical region using data mining techniques, specifically clustering and association rule mining (ARM). The aim was to detect opportunities for energy savings and IEQ improvements. Our methodology, based on an extensive collection of sensor-based data, relates energy consumption and IEQ parameters to human occupancy and translating these associations into rules. Key findings from the mined association rules included identifying benchmark patterns based on occupancy and detecting anomalies. Anomalous rules highlighted potential inefficiencies, such as high lighting or medium power consumption during periods of very low or no human presence, pointing towards opportunities for energy savings. Rules also revealed situations with high CO_2 concentration and warm temperatures associated with medium or high occupancy, suggesting opportunities for IEQ improvement through ventilation optimisation. This study demonstrates the capability of the ARM algorithm to uncover nuanced relationships among occupancy, power consumption, and indoor environmental conditions and provides useful indications

ARTICLE INFO

Article history:
Received: 17 December 2024
Accepted: 03 October 2025
Published: 20 January 2025

DOI: https://doi.org/10.47836/pjst.34.1.04

E-mail addresses:
johnson@uitm.edu.my (Soon Chong Johnson Lim)
tee@utem.edu.my (Boon Tuan Tee)
rajesh.sunkari@seikousystec.com (Rajesh Sunkari)
siew.pw@gmail.com (Peng Wah Siew)
mflee@uthm.edu.my (Ming Foong Lee)
* Corresponding author

e-ISSN: 2231-8526

towards potential energy savings and improvements in IEQ. It highlights the potential of sensor-collected, data-driven strategies for building operational efficiency and sustainability.

Keywords: Data mining, energy consumption, energy savings, indoor environment quality, knx, occupancy, sensor, smart office

INTRODUCTION

In the pursuit of global sustainability, the building industry stands as a significant contributor to energy consumption and greenhouse gas emissions. In Malaysia, the hot and humid tropical climate consumes approximately 48% of the total electrical energy at commercial and office buildings (Syed Yahya et al., 2015). As a result, building energy consumption has been a primary area of focus in building research to ensure the sustainability of energy. Achieving energy efficiency in buildings while ensuring occupant comfort and a healthy indoor environment is a significant challenge facing researchers and practitioners (Anand et al., 2022; Halhoul Merabet et al., 2021; Mofidi & Akbari, 2020; Ngarambe et al., 2020; Syed Yahya et al., 2015).

While occupant engagement in building energy efficiency programmes presents a cost-effective solution (Pisello & Asdrubali, 2014), a lack of clear instructions for occupants to change and enhance their energy-use behaviour is the primary obstacle to its effectiveness. A plausible approach to tackling this problem is to research data-driven methodologies that can directly analyse the energy-consuming activities of occupants and transform them into organized knowledge and practical recommendations. Previously, there are numerous studies that have highlighted the importance of incorporating occupancy-based controls and monitoring systems to optimise energy use based on real-time occupancy patterns (Anand et al., 2022; Mofidi & Akbari, 2020; Syed Yahya et al., 2015). In this regard, some studies have attempted to propose various methods, including data mining (DM) techniques (Ashouri et al., 2018; Yu et al., 2011), and occupants behaviour-based reasoning (Ahmadi-Karvigh et al., 2018), to establish a correlation between their activities and energy usage.

Despite the importance of balancing act between comfort and energy in buildings, particularly in tropical climates, and while data-driven approaches and occupancy-based methods show promise, a fundamental gap remains in literature. To the best of our knowledge, despite its importance, the interrelationship between these aspects (i.e., occupancy behaviour, IEQ, and energy consumption) remains understudied, especially in the tropical climate. Addressing this crucial gap is essential because previous research attempting to connect these areas has significant limitations that prevent complete understanding. Specifically, a significant drawback of many methodologies is that dynamic changes in occupancy patterns are not considered when estimating energy savings.

Furthermore, previous research has predominantly concentrated on energy consumption, overlooking the interrelationship between energy usage, thermal comfort, and indoor environmental quality (IEQ), especially with regards to occupancy behaviour (Tien et al., 2022). While review studies on occupant comfort exist (Halhoul Merabet et al., 2021; Mofidi & Akbari, 2020; Šujanová et al., 2019), empirical studies involving extensive and long-term sensor data collection in real-world settings, such as those in Oluwatayo and Pirisola (2021), are generally lacking. Most relevant studies limit their attention to examining human comfort or energy use concerning particular elements of IEQ parameters, and there is a notable lack of comprehensive research that connects all these factors (occupancy, IEQ, and energy) to provide a complete understanding of their complex interaction (Mofidi & Akbari, 2020; Verma et al., 2023). This highlights the need for data-driven research to uncover cause-and-effect relationships between occupancy behaviour, IEQ, and energy consumption (Ngarambe et al., 2020).

As the building sector is currently undergoing a transformation towards more intelligence with extensive sensor field deployment, this paper presents a study of occupancy-based association analysis of IEQ and energy consumption. We conducted the study in a smart office space that has deployed KNX-based sensor technology. The purpose of this study is to examine the correlation between human occupancy, IEQ, and energy consumption, with the aim of identifying potential energy savings and improving IEQ for the well-being of occupants. We reckoned that identifying specific patterns, associations, and anomalies within the complex interplay of the aforementioned factors requires suitable data mining techniques like Association Rule Mining (ARM), which can reveal relationships between disparate parameters. Crucially, the novelty of this paper lies in the unique application of ARM, validated by continuous, granular, real-world sensor data, specifically within the understudied context of a hot and humid tropical office building, providing empirical evidence absent in much of the existing literature.

LITERATURE REVIEW

Indoor environmental quality (IEQ) plays a significant role in the well-being and productivity of building occupants (Lan et al., 2014; Mofidi & Akbari, 2020; Tharim et al., 2017). IEQ includes several elements such as thermal comfort, visual comfort, and indoor air quality (IAQ). Past studies have emphasised the significance of these aspects in establishing a comfortable and healthy indoor environment (Ascione et al., 2021; Lim et al., 2017; Mofidi & Akbari, 2020; Tharim et al., 2017). A comfortable thermal environment, achieved through factors such as temperature, humidity, and air velocity, is essential for occupant satisfaction and can significantly influence productivity levels (Mofidi & Akbari, 2020). In addition, visual comfort, influenced by factors such as daylighting, glare control and artificial lighting, is equally important for good visual performance and satisfaction

(Lim et al., 2017). Maintaining an optimal IAQ through adequate ventilation rates can also reduce indoor air pollutants and contribute to occupant health and productivity (Anand et al., 2022; Ascione et al., 2021; Tharim et al., 2017). Multiple studies have highlighted the strong connection between IEQ, occupant well-being and productivity (Halhoul Merabet et al., 2021; Mofidi & Akbari, 2020; Ngarambe et al., 2020; Tharim et al., 2017).

However, the challenge lies in achieving energy efficiency while simultaneously maintaining occupant comfort and IEQ. Balancing energy savings with occupant needs has been a continuous concern in building design and operation (Syed Yahya et al., 2015; Verma et al., 2023). Many studies have focused on optimising individual IEQ parameters, such as thermal comfort, in relation to energy consumption. For example, one way of achieving this is through passive design strategies to minimise energy consumption while maintaining thermal comfort (Tien et al., 2022; Verma et al., 2023). However, there is a growing recognition of the need for a more holistic approach that considers the interactions between all IEQ parameters and their combined impact on energy use (Dong et al., 2023; Mehmood et al., 2019; Mofidi & Akbari, 2020; Verma et al., 2023). This requires integrating advanced technologies, such as computational intelligence, optimisation methods, and behaviour modelling techniques, into building design and operation (Mofidi & Akbari, 2020). The goal is to create intelligent buildings that are adaptive to dynamic occupant needs and environmental conditions while minimising energy consumption.

While the implementation of energy-efficient building systems, including heating, ventilation and air-conditioning (HVAC) systems and lighting controls, is of paramount importance, it is imperative that these systems are designed and operated with a primary focus on the comfort and health of the occupants. Recent studies have emphasised the importance of incorporating occupancy-based controls and monitoring systems to optimise energy use based on real-time occupancy patterns (Anand et al., 2022; Mofidi & Akbari, 2020; Syed Yahya et al., 2015). These typically involve strategies such as adjusting ventilation rates based on occupancy levels (Anand et al., 2022), utilising daylighting to reduce reliance on artificial lighting (Lim et al., 2017), and implementing personalised ventilation approaches to enhance both thermal comfort and IAQ (Anand et al., 2022). The integration of data-driven approach and techniques, such as computational intelligence, optimisation methods, and behaviour modelling techniques, is crucial towards the realisation of intelligent buildings that effectively balance energy efficiency with occupant well-being (Mofidi & Akbari, 2020; Verma et al., 2023).

While AI and ML show promises for enhancing building performance in energy efficiency, thermal comfort, and IAQ prediction, a critical gap persists in their holistic application integrating these factors with dynamic occupancy behaviour. Current research often isolates energy efficiency or thermal comfort (Halhoul Merabet et al., 2021; Hong et al., 2020; Mehmood et al., 2019; Ngarambe et al., 2020), overlooking the interdependence of energy use, IEQ, and occupancy patterns (Tien et al., 2022). Furthermore, most ML/

DL studies remain simulation-based, lacking real-world validation via post-occupancy evaluation (Tien et al., 2022). This gap is compounded by insufficient datasets capturing diverse building types and occupancy-IEQ dynamics (Halhoul Merabet et al., 2021; Metwally et al., 2022) and a scarcity of research integrating AI-based comfort models into building controls to quantify real-world energy-occupancy-IEQ causality (Ngarambe et al., 2020). Addressing these limitations is vital for advancing intelligent buildings that holistically optimise energy and occupant well-being.

METHODOLOGY

The study was performed in a three-story smart office building. This building was specifically chosen for this study because it is a smart office equipped with a comprehensive KNX-based building automation system (BAS) and an extensive network of sensor devices (Tee et al., 2023). This system was implemented as the core technology for energy management, with various sensors (including power, motion, lighting, and presence sensors) installed to monitor and control building elements. This sensor infrastructure enables the continuous collection of detailed, real-world data on human occupancy, indoor environmental quality parameters (temperature, CO_2), and energy consumption, which is essential for the data-driven analysis using clustering and association rule mining employed in this research.

The study was performed in the ground-floor office area of the building which covers an area of 81.32 m^2 (around 875 sq.-ft.) with cubicle partition seating that houses 15 engineers or technical staff. The area is served by three units of 1 HP (745 W) split air-conditioning units. The working hours of the office area are 09:00–18:00 hours, from Monday to Saturday. Sensor-wise, there is a Modbus power sensor module made by EVC installed to record detailed energy use for air conditioning, plug-points, and lighting. In addition, a KNX-based True Presence Multisensor made by Steinel (hereinafter presence sensor) is also installed at the ceiling of the office space, as shown in Figure 1. The sensor is capable of accurate recording of 360° human presences up to 15 meters (seated and walking) and also indoor parameters that are limited to air temperature, relative humidity (RH), carbon dioxide (CO_2) concentration, and volatile organic compound (VOC) concentration. To ensure data reliability, the sensors were calibrated prior to the data collection period using a standardised, fully calibrated IAQ meter (TSI 7545 IAQ-Calc Indoor Air Quality Metre).

For the purpose of the experiment, all the data from KNX devices and Modbus devices is recorded with a measurement frequency of 5 minutes. The detailed data logging process started on 1st March 2024, and all the sensor data is logged through a network gateway (by Netx Automation) towards an Internet server, where all the data is hosted. The logged parameters are power consumption of air-conditioning units, plug points, and lighting; occupancy; and IEQ parameters (CO_2 and VOC concentration, air temperature, and humidity).

Figure 1. Location of air-conditioning unit and presence sensor

In this study, the analysis of IEQ parameters and energy consumption is based on occupancy data. While ongoing data collection captures patterns over a longer duration, this preliminary study focusses on the analysis of one month of data, specifically 5-minute interval data collected in March 2024, with slightly over 8,700 observations per parameter. This month was selected to illustrate the application of our methodology and demonstrate the types of insights that can be derived from this granular sensor data. For data pre-processing, data aggregation is performed for each parameter to produce hourly data on a daily basis during only working hours (Monday to Saturday). The hourly presence probability is calculated to perform a timing split (or period) based on general office working timing. This will facilitate the study of the daily occupancy pattern and its subsequent correlation with changes in IEQ parameters and energy consumption. For the segmentation of timing into designated periods, averaged presence probability is used to detect key changes in occupancy patterns. Upon determining the time period, all corresponding data in the period are grouped together for period-based analysis.

In each period, clustering of each parameter data is performed as a data preparation step for association rule mining. A clustering algorithm, *k*-means clustering (Han et al., 2023), a widely used and efficient method suitable for partitioning data into distinct categories, is applied to all the parameters of presence probability, energy consumption, and IEQ of each period. For the optimal number of clusters (k), the standard elbow method (Han et al., 2023) is adopted to find the optimal k value, which categorises each hourly parameter into designated categories. Next, association rule mining (ARM) is performed to find the association among the different categories of different parameters. The ARM algorithm used in this study is frequent pattern growth (FP-Growth) (Han & Pei, 2000), a divisive-based frequent item discovery algorithm based on FP-Tree generation that is efficient for discovering frequent item sets and generating interpretable association rules. FP-Growth's main merits lie in its efficiency and scalability for association rule mining, primarily due to its ability to avoid candidate generation and to reduce database scans, which are significant drawbacks of algorithms like *Apriori*. The purpose of rule discovery is to discover benchmarking rules (i.e., frequently occurring patterns) and abnormal rules that may lead to energy wastage (e.g., low occupancy coupled with high power consumption) or to subpar IEQ (e.g., high occupancy patterns with poor air quality) or both.

To ensure the relevance of the discovered rules, a multi-stage filtering process was applied to the raw ARM output. Rules were initially filtered to include the presence cluster (PR) in the antecedent and exclude it from the consequent, aligning with the study's occupancy focus. Redundancy filtering, including pruning of subsumed or directly implied rules, was performed. Subsequently, standard statistical metrics were used for filtering: rules were selected based on a minimum Support of 0.1 (indicating frequent patterns), a minimum Confidence of 0.7 (indicating strong predictive power), and a Lift value of 1.0 or greater (indicating a positive association). Finally, a domain-specific filtering step identified the most "interesting" abnormal rules from the statistically significant set that directly suggested potential energy wastage or IEQ issues based on domain knowledge. For the purpose of presenting such a relationship, all parameter data shall be plotted in conjunction with human presence in a descriptive plot. Then, the generated rules from ARM shall be examined carefully to discover some of the interesting rules that lead to anomalies.

RESULTS AND DISCUSSION

The purpose of analytics in this study is to examine the pattern of power consumption and IEQ parameter changes based on occupancy. For the purpose of illustration, the results from the selected period of 2024-03-11 to 2024-03-17 are presented to showcase our methodology. This week was specifically chosen as it is representative of the typical daily and weekly patterns observed throughout the month analysed of March 2024 and clearly demonstrates the types of benchmark and abnormal rules discoverable by our data-driven approach. The details of each step of analysis are presented below.

Human Presence Analysis and Activity Segmentation

Human presence data is basically represented as a binary 0 or 1 value, where 0 means no human presence and 1 represents otherwise. For presence analysis, since occupancy sensing is conducted every 5 minutes, each presence data is aggregated into hourly data

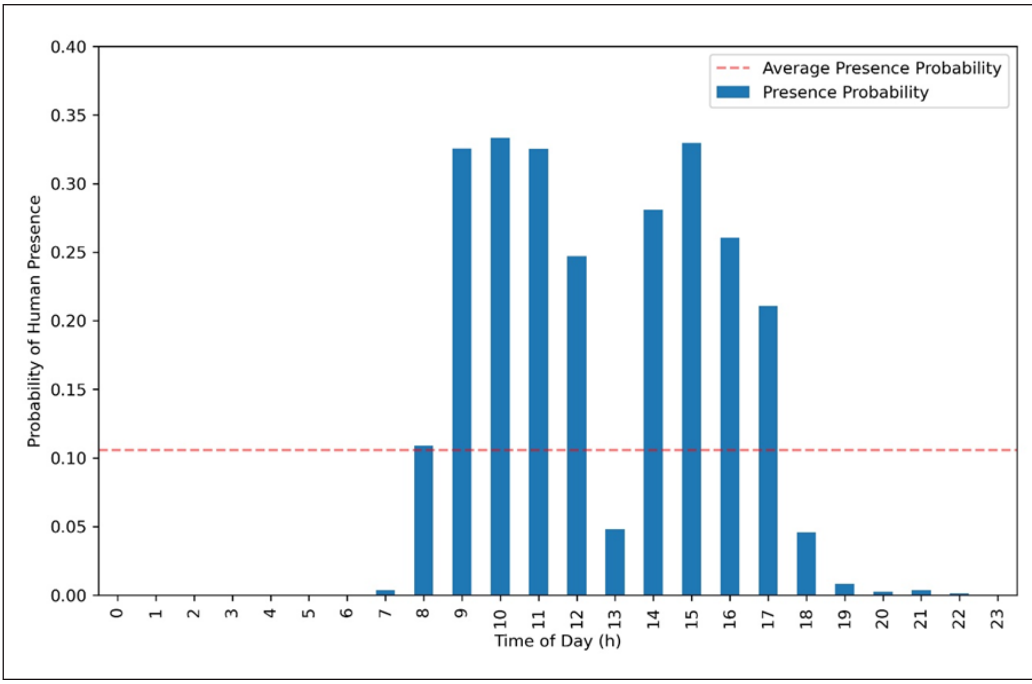

Figure 2. Probability of human occupancy by hour of day (working day)

to determine the presence probability, which is a mean value for assessing how frequently a space is occupied. In this study, such hourly data are calculated and aggregated daily during working days for the whole month of March 2024. The results of such a presence probability calculation are summarised in Figure 2. From the figure, it can be observed that mainly the office area has occupants starting from 0800–1200 hours, with significantly lower occupancy at 1300 hours (lunch break), higher occupancy from 1400–1800 hours, and low occupancy from 1800 onwards, which is in line with the working hours. In order to perform occupancy period segmentation, the average presence probability is plotted in the presence plots and used as a threshold to determine the reasonable number of change points in a day. As a result, the activity period can be segmented into four time periods: (a) Period 1 (P1): 0000–0700 hours; (b) Period 2 (P2): 0800–1300 hours; (c) Period 3 (P3): 1400–1800 hours; and (d) Period 4 (P4): 1900–2300 hours.

IEQ and Power Consumption Analysis

Indoor air quality and thermal comfort are crucial parts of indoor environmental quality, and the physical measurements are usually associated with the carbon dioxide concentration, temperature, air velocity, and RH. Building standards and guidelines throughout the world set those parameters as the reference for indoor building design. In this study, the reference standards are based on ASHRAE Standard 55 and Standard 62, along with Malaysia's MS1525 for comparison. The recommended value for each parameter is listed in Table 1. Figure 3 shows the overall data plot that has been recorded, including the IEQ parameter and power consumption with human presence on the office floor (ground floor) during the selected period. It is noticeable that the temperature range is higher than the recommended values of ASHRAE Standard 55 and MS1525. This observation is different for RH as the recorded data is within the acceptable range according to MS1525 but slightly higher than the recommended range in ASHRAE Standard 55.

Another observation for IEQ is through volatile organic compounds (VOCs) and CO_2 concentration levels. VOCs are organic chemicals that easily vapourise when they are at

Table 1
Summary of recommended parameter values according to relevant standards

Indoor parameter	Standard	Recommended value
Temperature	ASHRAE-55:2017 MS1525:2019	22.2-26.7°C 24-26°C
Relative Humidity (RH)	ASHRAE-55:2017 MS1525:2019	30-60% 50-70%
CO_2 Concentration	ASHRAE-62.1:2019	Not exceeding 700 ppm
VOC Concentration	ASHRAE-62.1:2019	Not exceeding 10 ppm (or 10,000 ppb)

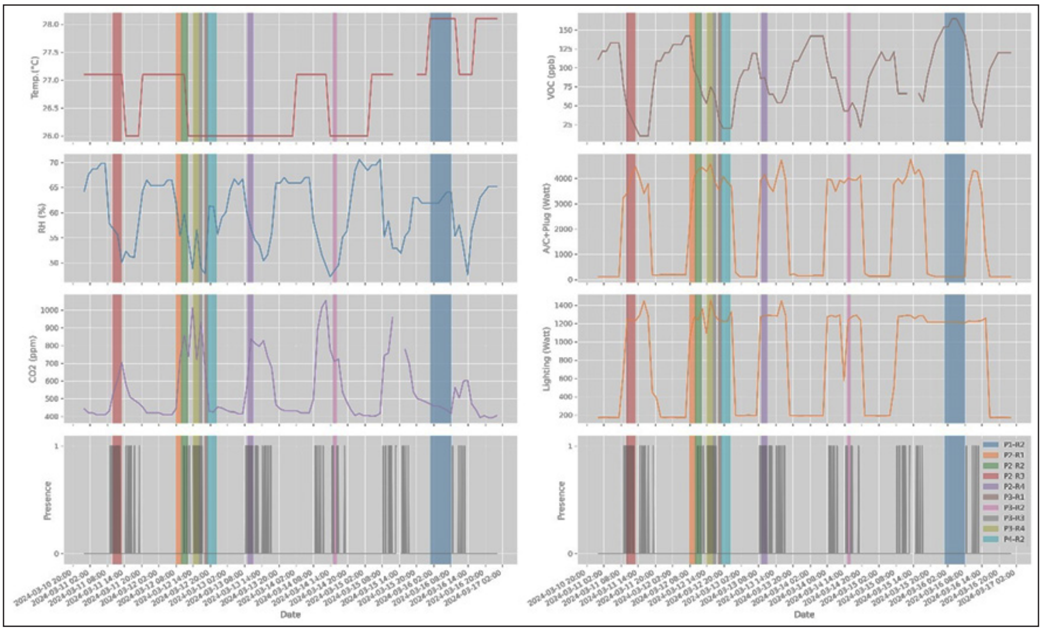

Figure 3. IEQ and power consumption parameter plots with human occupancy for selected period with discovered association rules

room temperature. Based on the plot, it can be concluded that the VOC levels were still far below the recommended ASHRAE 62.1. On the other hand, CO_2 concentrations can be observed to be very high during high occupancy, exceeding the recommended values under the ASHRAE 62.1 Standard. Power consumption wise, it is in general aligned with human presence, where power consumption for air conditioning, plug-points, and lighting is significantly higher during occupancy. However, there are also situations that show otherwise, where some of these time periods are highlighted in accordance with rules discovered through ARM. The results of this analysis will be discussed in the next section.

Cluster Analysis

Table 2 shows a summary of clustering results, with parameter labels and the corresponding units indicated. In summary, the occupancy, power, and lighting parameters were clustered using the k-means algorithm, and the elbow method suggested the appropriate *k* for each parameter.

Table 2
Summary of clustering results

Parameter (Label) (Unit)	Cluster	Centroid values			
		Mean	Std. dev.	Min.	Max.
Presence (PR) (Prob.)	VLOW	1.3×10^{-16}	-	-	-
	LOW	0.12	-	-	-
	MID	0.27	-	-	-
	HIGH	0.49	-	-	-
Power (POWER) (Watt)	VLOW	154.4	15.7	138.6	187.7
	LOW	1866.0	1386.0	268.6	3661.9
	MID	3952.0	401.2	3289.9	4443.3
	HIGH	5993.4	798.9	4571.5	6925.9
Light (LIGHT) (Watt)	LOW	188.2	9.8	178.2	206.8
	MID	717.3	341.6	292.8	1188.0
	HIGH	1260.5	37.1	1205.8	1311.9
Temp. (TEMP) (°C)	COLD	24	-	-	-
	GOOD	24 – 26	-	-	-
	WARM	26	-	-	-
R.H. (RH) (%)	DRY	50	-	-	-
	GOOD	50 – 70	-	-	-
	WET	70	-	-	-
CO_2 (CO_2) (ppm)	LOW	< 500	-	-	-
	GOOD	500 – 700	-	-	-
	HIGH	≥ 700	-	-	-

Occupancy data is clustered based on the presence probability value as the feature, which results in four categories from very low (VLOW) to high (HIGH) activities. The centroid of each cluster is as indicated. For both power (air conditioning and plug-points, as POWER) and lighting power consumption (LIGHT), clustering is based on four features: mean, standard deviation, minimum, and maximum values. These features are useful to capture temporal variation while reducing the dimensionality of clustering. The hourly mean temperature (TEMP) and RH data are clustered based on the recommended values by the MS1525 standard, where the GOOD cluster indicates comfortable levels. Likewise, CO_2 concentration mean data are clustered based on the ASHRAE-62.1 standard recommended values. VOC data are excluded from clustering, since overall they are far below the standard recommended values.

Mined Association Rules

Table 3 summarises selected association rules that are discovered using ARM and based on presence probability (PR) under different time periods. The ARM process identifies both benchmark rules, representing frequently occurring and expected patterns of building operation relative to occupancy, and abnormal rules, highlighting deviations or anomalies. While benchmark rules provide necessary context, our discussion prioritises these abnormal rules as they directly indicate potential energy wastage or subpar IEQ conditions and offer the most actionable insights for building optimisation. To aid explanation, some of the corresponding plots are highlighted in Figure 3.

For instance, P1-R1 is presented as a benchmark rule with high support and confidence values. It shows that when PR is very low, all the other parameters (explained using cluster label hereinafter), such as LIGHT, POWER, and CO_2, are also either low or very low, and TEMP is warm. This frequently observed pattern serves as a baseline for normal unoccupied periods. An example period for this to happen is from 0100–0600 hours of 2024–03-11 to 2024–03-15 without PR. In contrast, P1-R2 indicates an abnormal rule discovered where LIGHT is HIGH when PR is VLOW. As highlighted in Figure 3, this happens on 2024-03-16, 0100–0800 hours, where upon inspection there are no human presence captured by presence sensor. This presents an opportunity for power consumption savings.

For Period 2 (P2), P2-R1 shows very low PR with high LIGHT (e.g., 2024-03-12, 0800–1000 hours), suggesting a clear opportunity for power savings. The elevated lighting power consumption during this period, even without occupants, points to potential inefficiencies in the lighting control system. Such a situation can be possibly mitigated with an installation of occupancy-based lighting controls to ensure lights are off when space is unoccupied. P2-R2 indicates low PR with medium POWER and high LIGHT (e.g., 2024-03-12, 1000–1200 hours). The intermittent presence during this period suggested power saving opportunities. P2-R3 associates medium PR with warm TEMP and rising

Table 3
A summary of selected association rules discovered based on occupancy

Period	Rule	Antecedents	Consequents	Support	Conf.	Lift
P1	R1	PR-VLOW	LIGHT-LOW, POWER-VLOW, TEMP-WARM, CO2-LOW	0.9412	0.9412	1.0000
	R2	PR-VLOW, LIGHT-HIGH	POWER-VLOW	0.4118	1.0000	1.0625
	R3	PR-VLOW, POWER-LOW	LIGHT-MID, CO2-LOW, TEMP-WARM	0.1176	1.0000	8.5000
P2	R1	PR-VLOW, LIGHT-HIGH	TEMP-WARM, RH-GOOD	0.3571	1.0000	1.0000
	R2	PR-LOW, POWER-MID, LIGHT-HIGH	TEMP-GOOD, RH-GOOD	0.2667	1.8667	3.5000
	R3	PR-MID, POWER-MID	LIGHT-HIGH, CO2-GOOD, TEMP-WARM, RH-GOOD	0.2000	1.5000	3.5000
	R4	PR-HIGH	POWER-MID, LIGHT-HIGH, TEMP-WARM, RH-GOOD	0.1333	1.0000	2.8000
P3	R1	PR-VLOW	POWER-MID, LIGHT-HIGH, TEMP-WARM, RH-GOOD	0.4545	0.7143	1.9643
	R2	PR-LOW, POWER-MID, LIGHT-HIGH	TEMP-WARM, RH-GOOD	0.2727	1.0000	1.3750
	R3	PR-MID	POWER-MID, LIGHT-HIGH, TEMP-WARM, CO2-HIGH	0.2727	1.0000	2.7500
	R4	PR-HIGH	POWER-MID, LIGHT-HIGH, CO2-HIGH	0.1667	1.0000	2.7500
P4	R1	PR-VLOW	POWER-VLOW, LIGHT-LOW, CO2-LOW, TEMP-WARM	0.9167	0.9167	1.0000
	R2	PR-VLOW, POWER-MID	LIGHT-HIGH, CO2-GOOD, RH-GOOD	0.3846	1.6667	3.6111

CO_2 concentration (e.g., 2024-03-11, 1000–1300 hours), suggesting IEQ improvements. P2-R4 discovered high PR with warm TEMP and rising CO_2 concentration (e.g., 2024-03-13, 0900–1100 hours). Both P2-R3 and P2-R4 suggest improvement opportunities in the IEQ aspects, especially for reducing the high CO_2 concentration, with adequate ventilation for a more conducive work environment.

Similarly, Period 3 (P3) also discovered rules that suggest multiple improvement opportunities. For instance, P3-R1 found that LIGHT is high after office hours (e.g., 2024-03-12, 1800–1900 hours). P3-R2 also discovered high LIGHT with medium POWER consumption despite low PR during the afternoon (e.g., 2024-03-14, 1500–1600 hours). Both cases of P3-R1 and P3-R2 suggest power saving opportunities. Next, P3-R3 indicates unsatisfactory IEQ conditions with high CO_2 and warm TEMP under medium PR (e.g., 2024-03-12, 1600–1700 hours). P3-R4 also suggest unsatisfactory IEQ conditions with high CO_2 associated with high occupancy (e.g., 2024-03-12, 1400–1600 hours). This indicates that ventilation optimisation strategies may need to be considered to improve air quality.

Lastly, for Period (P4) which is the nighttime after office hours, P4-R1 is a benchmark rule where very low PR is associated with very low POWER and LIGHT during night time (e.g., 2024-03-11, 2000–2400 hours, not shown in Figure 3), which is expected. However, P4-R2 indicates high LIGHT with medium POWER is associated with very low PR (e.g., 2024-03-12, 1900–2200 hours). This rule is a clear signal for power-saving opportunities, where energy-saving measures for lighting, air conditioning and plug-points need to be considered.

Overall, the identified association rules, shown in the shaded regions of Figure 3, provide insights into certain combinations of IEQ indicators and power consumption patterns associated with occupancy at various times throughout the day. They demonstrate the capability of the ARM algorithm to identify nuanced relationships among occupancy, power consumption and indoor environmental conditions in the office. These insights are essential for understanding the impact of occupancy behaviour on indoor environmental quality and energy consumption patterns, which may guide strategies for optimising building operations to enhance occupant comfort and minimise energy usage.

CONCLUSION

This study demonstrates a data-driven approach for uncovering actionable insights to optimise energy consumption and indoor environmental quality (IEQ) in real-world smart buildings. Leveraging extensive sensor data and data mining techniques, specifically clustering and association rule mining (ARM), within a tropical smart office setting, we successfully moved beyond general correlations to identify specific, critical anomalies. These anomalies, which include instances of high lighting or medium power consumption during periods of minimal occupancy and opportunities for IEQ improvement like high CO_2 levels during medium or high occupancy, represent concrete targets for efficiency enhancements. The study demonstrates that analysing granular sensor data with ARM can reveal hidden inefficiencies and IEQ issues that may be less apparent or dynamic through intermittent traditional energy audit methods, thus complementing existing approaches and providing building operators with a powerful tool for more efficient and sustainable operation.

The significance of this study lies in providing an empirical demonstration of how data from smart building sensor systems can be effectively analysed using data mining to yield actionable insights. This approach offers unique strengths that complement traditional energy audit methods, particularly in its ability to provide the granularity and continuous monitoring capabilities required to detect dynamic relationships and specific operational anomalies identified by our data-driven analysis. By translating complex data into understandable association rules, we provide a powerful tool for building operators to identify targets for energy saving measures and IEQ enhancements, contributing to more efficient and sustainable building operation. Crucially, the identified abnormal rules (e.g., high LIGHT and VLOW PR or high CO_2 with high PR) serve as direct, actionable triggers that can be mapped onto a BAS, such as the KNX-based system deployed in this study. For example, a BAS could be automatically programmemed to switch off lights or adjust ventilation rates immediately upon detection of these rule-based anomalies, thereby closing the loop between data discovery and automated efficiency response.

We acknowledge, however, that the scope of this work, being a preliminary study, is constrained by the analysis of only one month of data. While this duration was sufficient to demonstrate the methodology, the generalisability and seasonal stability of the discovered patterns require further long-term investigation. As this preliminary study only focussed on a specific area, our next actions involve extending the analysis to multiple zones, developing and testing advanced predictive modelling algorithms (such as deep neural networks) to accurately predict IEQ degradation (e.g., rising CO_2 level) and potential energy waste based on dynamic occupancy patterns. A practical integration of our data-driven prediction into the building automation system control strategies, such as modulating HVAC or ventilation rates to preemptively mitigate unsatisfactory IEQ conditions before occupant comfort is compromised, is also planned. Such a predictive control integration represents the crucial next step in advancing smart building operational efficiency. This research addresses a crucial gap in providing real-world evidence and a practical methodology for achieving energy savings and IEQ improvements based on dynamic occupancy behaviour in smart buildings, particularly in tropical climates.

ACKNOWLEDGEMENT

This research was supported by Universiti Teknikal Malaysia Melaka, Universiti Tun Hussein Onn Malaysia (UTHM) and Seikou Systec Sdn. Bhd. through industrial research grant (vot M065).

REFERENCES

Ahmadi-Karvigh, S., Ghahramani, A., Becerik-Gerber, B., & Soibelman, L. (2018). Real-time activity recognition for energy efficiency in buildings. *Applied Energy*, *211*, 146-160. https://doi.org/10.1016/j.apenergy.2017.11.055

Anand, P., Cheong, D., & Sekhar, C. (2022). A review of occupancy-based building energy and IEQ controls and its future post-COVID. *Science of The Total Environment*, *804*, 150249. https://doi.org/10.1016/j.scitotenv.2021.150249

Ascione, F., De Masi, R. F., Mastellone, M., & Vanoli, G. P. (2021). The design of safe classrooms of educational buildings for facing contagions and transmission of diseases: A novel approach combining audits, calibrated energy models, building performance (BPS) and computational fluid dynamic (CFD) simulations. *Energy and Buildings*, *230*, 110533. https://doi.org/10.1016/j.enbuild.2020.110533

Ashouri, M., Haghighat, F., Fung, B. C. M., Lazrak, A., & Yoshino, H. (2018). Development of building energy saving advisory: A data mining approach. *Energy and Buildings*, *172*, 139-151. https://doi.org/10.1016/j.enbuild.2018.04.052

Dong, J., Schwartz, Y., Mavrogianni, A., Korolija, I., & Mumovic, D. (2023). A review of approaches and applications in building stock energy and indoor environment modelling. *Building Services Engineering Research and Technology*, *44*(3), 333-354. https://doi.org/10.1177/01436244231163084

Halhoul Merabet, G., Essaaidi, M., Ben Haddou, M., Qolomany, B., Qadir, J., Anan, M., Al-Fuqaha, A., Abid, M. R., & Benhaddou, D. (2021). Intelligent building control systems for thermal comfort and energy-efficiency: A systematic review of artificial intelligence-assisted techniques. *Renewable and Sustainable Energy Reviews*, *144*, 110969. https://doi.org/10.1016/j.rser.2021.110969

Han, J., & Pei, J. (2000). Mining frequent patterns by pattern-growth: Methodology and implications. *ACM SIGKDD Explorations Newsletter*, *2*(2), 14-20. https://doi.org/10.1145/380995.381002

Han, J., Pei, J., & Tong, H. (2023). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Hong, T., Wang, Z., Luo, X., & Zhang, W. (2020). State-of-the-art on research and applications of machine learning in the building life cycle. *Energy and Buildings*, *212*, 109831. https://doi.org/10.1016/j.enbuild.2020.109831

Lan, L., Wargocki, P., & Lian, Z. (2014). Thermal effects on human performance in office environment measured by integrating task speed and accuracy. *Applied Ergonomics*, *45*(3), 490-495. https://doi.org/10.1016/j.apergo.2013.06.010

Lim, G.-H., Hirning, M. B., Keumala, N., & Ghafar, N. A. (2017). Daylight performance and users' visual appraisal for green building offices in Malaysia. *Energy and Buildings*, *141*, 175-185. https://doi.org/10.1016/j.enbuild.2017.02.028

Mehmood, M. U., Chun, D., Zeeshan, Han, H., Jeon, G., & Chen, K. (2019). A review of the applications of artificial intelligence and big data to buildings for energy-efficiency and a comfortable indoor living environment. *Energy and Buildings*, *202*, 109383. https://doi.org/10.1016/j.enbuild.2019.109383

Metwally, E. A., Farid, A. A., & Ismail, M. R. (2022). Development of an IoT assessment method: An interdisciplinary framework for energy efficient buildings. *Energy and Buildings*, *254*, 111545. https://doi.org/10.1016/j.enbuild.2021.111545

Mofidi, F., & Akbari, H. (2020). Intelligent buildings: An overview. *Energy and Buildings*, *223*, 110192. https://doi.org/10.1016/j.enbuild.2020.110192

Ngarambe, J., Yun, G. Y., & Santamouris, M. (2020). The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls. *Energy and Buildings*, *211*, 109807. https://doi.org/10.1016/j.enbuild.2020.109807

Oluwatayo, A. A., & Pirisola, O. H. (2021). Impact of passive energy strategies on workers productivity in tropical climate. *IOP Conference Series: Materials Science and Engineering, 1107,* 012210. https://doi.org/10.1088/1757-899X/1107/1/012210

Pisello, A. L., & Asdrubali, F. (2014). Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies. *Applied Energy*, *133*, 224-235. https://doi.org/10.1016/j.apenergy.2014.07.049

Šujanová, P., Rychtáriková, M., Sotto Mayor, T., & Hyder, A. (2019). A healthy, energy-efficient and comfortable indoor environment, a review. *Energies*, *12*(8), 1414. https://doi.org/10.3390/en12081414

Syed Yahya, S. N. N., Ariffin, A. R. M., & Ismail, M. A. (2015). Building energy index and students' perceived performance in public university buildings. In A. Sayigh (Ed.), *Renewable energy in the service of mankind Vol I: Selected topics from the World Renewable Energy Congress WREC 2014* (pp. 541-550). Springer International Publishing. https://doi.org/10.1007/978-3-319-17777-9_48

Tee, B. T., Lim, S. C. J., Siew, P. W., & Lee, M. F. (2023). Application of sensor technology for energy consumption analysis: A case study in a smart office building. In *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 0913-0917). IEEE.https://doi.org/10.1109/IEEM58616.2023.10406624

Tharim, A. H. A., Samad, M. H. A., & Ismail, M. (2017). Relationship between indoor environmental quality (IEQ), occupant's satisfaction and productivity in GBI rated office building using SEM-PLS. *Pertanika Journal of Social Science and Humanities*, *25*(S), 319-329.

Tien, P. W., Wei, S., Darkwa, J., Wood, C., & Calautit, J. K. (2022). Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality – A review. *Energy and AI*, *10*, 100198. https://doi.org/10.1016/j.egyai.2022.100198

Verma, A., Gupta, V., Nihar, K., Jana, A., Jain, R. K., & Deb, C. (2023). Tropical climates and the interplay between IEQ and energy consumption in buildings: A review. *Building and Environment*, *242*, 110551. https://doi.org/10.1016/j.buildenv.2023.110551

Yu, Z., Haghighat, F., Fung, B. C. M., Morofsky, E., & Yoshino, H. (2011). A methodology for identifying and improving occupant behaviour in residential buildings. *Energy*, *36*(11), 6596-6608. https://doi.org/10.1016/j.energy.2011.09.002