Enhancing Study Space Utilization at UCL: Leveraging IoT Data and Machine Learning

Abstract: University College London (UCL) has more than 51,000 students along with a challenge to find available study space. UCL has improved its Internet of Things (IoT) infrastructure to better analyze space occupancy in order to solve this. Current systems offer real-time information, but they are only available in crowded places. Therefore, this study presents a data analysis system that recommends the best study areas by using Machine Learning (ML) and historical data. The main library, student center, and science library at UCL are the top three study places. The goal of this study is to transform raw occupancy data from these venues into useful insights. The development of an independent system written in Python was facilitated by research into dataset optimization, ML regressions, popularity patterns, and feature analysis. This technique uses historical trends to forecast occupancy rates for the future.

Keywords: Data, Internet of Things, Machine Learning, Python, Space

1. Introduction

Academic institutions face a major difficulty when it comes to study space crowding. These institutions are responsible for managing resources to ensure operational efficiency and improve the quality of the student experience. The scarcity and high expense of study areas make administering major universities like University College London (UCL), which has over 51,000 students, more difficult. Since even little efficiencies can result in significant cost savings, efficient space utilization becomes a crucial goal. These savings have the ability to affect graduation rates, student access, research funding, and student services in addition to supporting academic and research activities. UCL has resorted to technology solutions, including the Internet of Things (IoT) and people-counting systems, to address these difficulties. However, the IoT is a network of physical objects that can gather and share data because they are outfitted with sensors and internet access. Therefore, in order to maximize the usage on campus areas, UCL’s IoT network incorporates overhead cameras and under-desk sensors with motion detection. IoT technology is becoming a useful tool for organizing study areas since it provides ever-more-effective ways to track occupancy and foot movement. Whereas, people-counting technology, an expanding topic in academia, counts people in a space using automation, real-time monitoring, and data analysis. By giving organizations, the ability to allocate space based on data, this technology can greatly enhance space management. For example, by deploying an Artificial Intelligence (AI) [1]–[13]-based optimization algorithm that assigned classes based on expected attendance rather than enrollment numbers, the University of New South Wales (UNSW) recorded a 10% reduction in room expenditures. In addition to lowering expenses, people-counting technology can improve how classroom and study spaces are assigned, help with investment choices, and enhance the campus environment in general.

Therefore, the principal objective of this study is to create a digital platform that facilitates students in locating and scheduling their study sessions on campus. This solution makes use of the current IoT infrastructure as well as historical occupancy data to forecast and suggest study spaces. The main output is the development of a data analysis system that can forecast study spaces using Machine Learning (ML) [14]–[16] and extrapolate occupancy trends. However, the study is essentially based on a thorough methodology that starts with the definition of design
goals and success criteria and ends with the development of an infrastructure concept that makes occupancy data easier to interpret. ML models are then trained on historical data, together with thorough feature analysis and occupancy studies, to create a prediction method. By taking these actions, we hopes to offer a solution that not only solves the pressing issue of where to study, but also makes use of technology to improve the general efficacy and efficiency of space utilization on campus. In order to improve the student experience and help the university achieve its operational objectives, the study aims to provide a comprehensive, data-driven approach to managing campus infrastructure through the integration of IoT and people-counting technologies.

The paper is as follows in next section we will see the related works. In Section 3, the materials and methods are presented. In Section 4, the implementation of the system is discussed. In Section 5, the experimental analysis is conducted. In Section 6, the discussion is presented based on the findings of the study and we conclude the paper in Section 7 with some conclusions and future works.

2. Related Works

The modern era's economic expansion and scientific and technological breakthroughs have highlighted the value of careful design and planning in study spaces. Despite this emphasis, creating and organizing study spaces often encounters formidable obstacles according to [17]—include the failure to achieve optimal resource utilization, the irrational planning and design of study spaces facilities, the absence of scientific structuring in layouts, and the lack of interconnectedness among diverse study spaces functions. [18] study of university students’ travel habits in the Netherlands raises a number of important issues, one of which is the impact of daily commutes’ greenhouse gas emissions to air pollution. This challenge highlights a bigger one [19]—how community commercial service areas are arranged spatially can have a significant impact on sustainability outcomes. This strategy improves the efficiency of using public resources while also assisting in the reduction of environmental pollution. It lessens pressure to expand construction into green spaces and improves community areas' overall utilization. These kinds of expansions usually result in a reduction in biodiversity and green vegetation coverage. Communities can easily incorporate sustainable study spaces infrastructure, encouraging resource-saving and energy-efficient activities, by carefully planning the layout of commercial service facilities according to [20]. Therefore, repurposing study spaces presents a special chance to cultivate a sustainable development-supporting community. Energy-efficient design, the use of renewable energy sources, and the landscape improvements can all help achieve this and create a more sustainable study space environment. Institutions are taking more and more actions to make their study spaces more sustainable according to [21]. These steps include publishing sustainable development reports, conducting regular evaluations, and implementing environmental management systems that adhere to guidelines like ISO 14001 and the Eco-Management and Audit Scheme (EMAS).

The above studies address several aspects of sustainable study spaces development that has made low energy usage in community commercial service areas a key issue. However, in order to improve energy efficiency in these study spaces, scholars are researching sustainable techniques. They place special emphasis on the integration of renewable energy sources, architectural design, and the important role that occupant behavior plays in promoting energy efficiency. This emphasizes the need for strict evaluation metrics and monitoring systems in order to determine the efficacy of sustainability initiatives. The lack in thorough theoretical and practical research on the design of study spaces persists—despite the increasing corpus of study in this area. Progress has been achieved in this field by a few prominent studies such as [21]. For instance, [22] evaluated the accessibility and integration of study spaces within colleges in 2021. Additionally, this study looked at the movement patterns of the students, the distribution of heat along the pathways they took, the satisfaction ratings of the participants, and the conceptual designs. Using space syntactic analysis, [23] focused on creation of a consistent and reliable assessment index for these spaces, the identification of sustainable development strategies for the design of study spaces that support energy-saving objectives, and the identification of the particular attitudes and preferences that community users have towards incorporating sustainable development practices in study spaces. Similarly, [24] intends to investigate how ML might be used to provide metrics that evaluate study spaces and to suggest improvements. Their research analyzes survey data and evaluates quantitative data using the communities of Shandong Jianzhu University as a case study. The goal is to provide a thorough understanding of how sustainable practices can be successfully incorporated into the design of study spaces.
3. Materials and Methods

Based on utilization and functionality, UCL uses four different systems to monitor and manage space consumption. These systems are separated into primary and secondary categories. Under-desk sensors and flow counters are a few of the principal systems for people counting—electronic turnstiles and a card-based attendance tracking system are among the secondary systems for security and attendance. The flow counters and under-desk sensors are made by FMSystems—a business that specializes in workplace management systems. Using Passive Infrared (PIR) technology—the under-desk sensors—sense motion at each desk to determine occupancy. Above doorways, flow-counters use infrared cameras to count individuals coming in and going out, giving a constant stream of occupancy data. According to [25], the cost of upkeep for these systems is around £204,000. In order to improve campus facility management and planning, UCL additionally makes use of the online platform provided by FMSystems to conduct comprehensive analyzes of time and space utilization. The UCL administration has not made considerable use of these systems for thorough occupancy analysis—despite their capacity to generate occupancy data. This suggests a gap in the utilization of current resources for space optimization. According to [25], there is still much space for improvement in UCL’s occupancy analysis methodology. In order to properly utilize occupancy data for study space optimization—the institution must overcome a number of obstacles, which calls for a more integrated strategy that combines physical and digital solutions. Adoption of alternative technologies, like those provided by Terabee—a smart sensor firm offering people-counting cameras based on time-of-flight light-detection-and-ranging (LIDAR) technology, is one way that UCL’s people-counting capabilities could be improved. These sensors are a good substitute for UCL’s current infrared-based systems since they are easy to install and can transmit data wirelessly. With JSON messages, they can send data to any web server, enabling more personalized IoT applications. UCL’s student app, "UCL Go!"—is currently available for download. However, it is accessible and displays real-time occupancy using under-desk sensor data, but it is not able to provide historical trends or study space recommendations based on prediction. The necessity for a more advanced system that can provide thorough occupancy insights is highlighted by this constraint. On the other hand, some colleges have included Occupuspace’s Wi-Fi tracking technology, which uses its Waitz app to display occupancy data in real-time as well as historical data. Like "UCL Go!"—the Waitz app does not, however, provide predictive advice for future occupancy—exposing a prevalent flaw in the majority of occupancy tracking apps available today. Therefore, UCL needs to think of a comprehensive plan that incorporates digital and physical technology in order to properly handle the overcrowding of study spaces. In addition to improving the people-counting data’s reliability and precision, this technique should make sure that students can readily access and utilize the information. Examples of such technologies include Terabee’s LIDAR for precise people counts and advanced data analysis platforms with predictive modelling capabilities. Feedback systems would also be beneficial to such a system, enabling ongoing improvement depending on user experience and occupancy trends. UCL can establish a more effective, flexible, and student-friendly study space management ecosystem by coordinating these technological and operational improvements with the requirements and preferences of the student body. UCL can significantly improve the sustainability and efficiency of its campus areas by establishing more advanced tools for data analysis and presentation, as well as by improving the precision with which occupancy data is collected. This advancement in space management promises to improve student experience while also supporting institutional aims of resource optimization and sustainability.

4. Implementation

Clear design objectives and success criteria were established at the outset of the study—which paved the way for an organized design approach as shown in Fig. 1. To begin with, an infrastructure concept was established, laying the foundation for the analysis of occupancy data. In order to understand how space is used over time, it is necessary to establish historical trends and patterns, which was accomplished in this phase by analyzing raw data. Then, the emphasis turned to creating a prediction algorithm. Regression models were trained utilizing the

---

2 https://fmsystems.com/
3 https://www.terabees.com/
4 https://occupuspace.io/
5 https://waitz.io/
historical data at this step, along with a thorough feature analysis and a comparison of the respective occupancy rates. These trials were essential in enhancing the model's ability to predict and ensuring that the system could provide trustworthy recommendations regarding study space availability. The last stage of the study was creating a mobile application to display the occupancy forecasts. The goal of this study was to design an application that was easy to use and simple, giving priority to user interface and user experience considerations. In order to help students make better decisions about study area selection, the app was made to effectively communicate the analysis results and make occupancy information simply accessible and comprehensible. The study followed a methodical approach during these phases, making sure that every part was created in accordance with the predetermined objectives and success criteria, culminating in a full occupancy management solution.

![Flowchart of the overall design process](image)

**Fig. 1. Flowchart of the overall design process**

### 4.1 Further Improvements

A complex infrastructure that integrates digital and physical systems is essential for addressing UCL data analysis needs. In order to properly handle and analyze the data, this infrastructure needs to be able to support extensive data collection and have a large amount of computational power. The Gallagher SpeedStiles system is the primary component of physical infrastructure that UCL uses to collect data. This system was chosen due to its current integration inside the university, internal security control, and efficacy in gathering large amounts of data. Other people-counting technologies at UCL, like flow-counters and under-desk sensors, are either area-specific or have difficulty extracting data. Nonetheless, the SpeedStiles system offers a constant flow of information on people's movements, covering important entry and exit locations all throughout the campus. While it does gather personally identifiable details—however—to protect people's privacy, all personal identifiers—such as student ID numbers—are anonymized during the data processing process. Therefore, the study suggests using cloud computing services, especially Google Firebase or Amazon Web Services (AWS), to support the functionalities of mobile applications and data analysis on the digital front as shown in Fig. 2. These cloud platforms were picked because of their flexibility and adaptability, which are essential for changing with the needs of data analysis and storage. These services offer an affordable solution for data storage and processing power because of their "pay-as-you-go" pricing model, which guarantees that the infrastructure can be scaled as needed. This digital

---

6 https://www.sine.co/integrations/gallagher/
infrastructure is necessary to host the mobile application and analyze the massive amounts of data produced by the physical sensors. It also allows real-time occupancy data analysis, which is crucial for giving students fast and accurate study space recommendations. Without the mobile app fully developed and launched, it is difficult to estimate the exact maintenance expenses for this infrastructure. On the other hand, early projections indicate that the expenses would be reasonable. The systems expected maintenance expenses in 2023 were predicted to be between £200 and £400, translating into an anticipated yearly cost of approximately £5,000. This estimate takes into consideration the fact that the data processing is expedited and that cloud computing resources are used effectively, both of which may significantly reduce overheads associated with data storage and computational needs. Physical and digital systems are well-integrated in the projected infrastructure at UCL that would assist data processing. While the digital infrastructure, which makes use of cloud computing services like Google Firebase or AWS, offers a scalable, reasonably priced platform for data processing and application hosting, the physical infrastructure, which is centered around the Gallagher SpeedStiles, guarantees the reliable acquisition of raw data.

![Diagram of infrastructure architecture](image)

Fig. 2. An infrastructure architecture that makes use of both new cloud computing services and current IoT systems

5. Experimental Analysis

Since, data requests at UCL are time-consuming, the study's primary focus was on the three main study areas at the university. Python code was built to extrapolate historic occupancy using SpeedStiles data, which is prone to random and systematic mistakes. These mistakes are the result of things like malfunctions, planned maintenance, and UCL security actions. Depending on the severity of the inaccuracies, data points were manually removed to reconcile the differences. An instance of random errors necessitated human correction due to an atypically high net influx of 200–500 individuals as shown in Fig. 3. In order to address these irregularities, the occupancy numbers had to be adjusted, as seen in the Fig. 4. Indicating the significance of time and day as predictive variables—Fig. 5—showed distinct occupancy trends, with longer peak hours and shorter troughs, as well as maintained popularity during weekends. By computing average occupancy in 10-minute increments, a trade-off between computational efficiency and accuracy was made in order to handle the irregular data collection intervals as shown in Fig. 6. When examining how space is used, it was determined that relative crowding of spaces was more important than total occupancy. To see the occupancy distribution and preferences across the study areas, comparative scatterplots were made as a result—as shown in Fig. 7. These graphs showed that students showed a clear preference during lower to moderate occupancy levels, but as occupancy rose, this tendency began to more closely resemble other regions. This choice was probably affected by the regular operating hours, which were more fixed than other places' schedules. In order to analyze the occupancy data distribution, Gaussian tests were performed—the Anderson-Darling (AD) test was used due to its effectiveness in handling vast sizes of data. A modelling strategy that could manage such data features was required, as the test results verified non-Gaussian occupancy trends. Therefore, using Support Vector Regression (SVR), Random Forest (RF), and linear regression methodologies—ML methods for occupancy prediction were investigated as shown in Fig. 8. When it came to precisely predicting real occupancy trends and preserving reduced error rates, the RF model performed better than the others. Its resilience against outliers and capacity to handle non-Gaussian and non-linear data were cited as the reasons for RF's outstanding performance. The amount of data needed for future forecasts to be correct was...
determined by looking at training set size. Every extra week of training data increased the model's accuracy, especially once reading week data was added as shown in Fig. 9. The improvement was seen in the model's capacity to more precisely forecast occupancy for the current and next terms, demonstrating the efficacy of the RF approach in identifying the intricate patterns of occupancy data.

Fig. 3. Example of random error correction

Fig. 4. Updated example of a random error

Fig. 5. The science library, student center, and main library's daily trends
Fig. 6. The main library, student center, and scientific library's comparative scatterplots

Fig. 7. Occupancy level histograms during business hours

Fig. 8. ML predictions
6. Discussion

By making use of the current Gallagher SpeedStiles, the UCL infrastructure design accomplishes its primary objectives and avoids incurring significant new infrastructure costs. The predicted £5,000 cost of digital infrastructure maintenance is significantly less than the current £204,000 yearly maintenance expense—therefore—this strategy offers a major financial advantage. Nevertheless, there are a number of flaws and restrictions with the Gallagher SpeedStiles technology. The accuracy of data analysis is impacted by random and systematic errors mostly because it is not optimized for accurate people counting. Moreover, its use is limited to locations that need ID card exits—which reduces the range of study rooms it can keep an eye on. Even with these problems, the SpeedStiles are still more advantageous than other systems at the moment—even though under-desk sensors have more promise in the future. To extract historical trends and forecast future occupancy, the autonomous data analysis system uses a set of procedural Python scripts. The scripts work automatically—specific parameters, such as time filters, are the only thing that need to be entered. However, certain random errors require manual repair. A range of statistics and scatterplots are used to demonstrate the system's success as its forecast accuracy and consistency are evaluated based on predetermined criteria. With an 85% consistency rate over six investigations, the model achieves more accuracy in specific areas and time periods. For example, by the third week, the model achieves the 85% consistency requirement during the main library and science library assessments, and it eventually improves to about 95%. But for the student centre, this need isn't fulfilled until much later in the semester—after reading week data has been taken into account. The model's sensitivity to data volume and variability is demonstrated by this delay, where performance peaked following reading week and then declined as a result of smaller test sizes. The student centre is where the model performs the worst, improving more slowly and never reaching 90% consistency. The strong demand and steady availability of the student centre are characteristics that make occupancy forecast difficult, which is why there is this inconsistency. Furthermore, the data included in the model's current study comes exclusively from the 2022–2023 academic year—the first complete in-person year following the COVID-19 pandemic. Having access to data from later years will probably improve the model's ability to forecast outcomes, which implies that the original performance reviews might have been too harsh. Moreover, the requirement of a complete term for improving the prediction model draws attention to still another drawback. In comparison to the entire dataset, forecasts for certain weeks are significantly less accurate when specific variables like holiday periods or reading weeks are not taken into account. For instance, the model exhibits difficulties during particular weeks in Term 1, highlighting the difficulties in forecasting occupancy during irregular times. Although Gallagher SpeedStiles' current infrastructure provides an affordable basis for UCL's occupancy research, its shortcomings highlight the need for more advanced counting techniques and improved predictive models. Whilst the current system is financially sustainable and functioning, it has to be improved in terms of accuracy and granularity in order to better meet the changing needs of the university.

7. Conclusion and Future Works

Using just three weeks of training data, the UCL data analysis system successfully forecasted study space occupancies and met success criteria. User-flow and concept illustrations were used to illustrate the concept, while
the infrastructure design effectively integrated current UCL systems at a minimal additional cost of £5,000 per year. Nonetheless, difficulties persist, especially with the SC forecasts, which require seven weeks of instruction because of their great popularity and constant use. The ideas are still in their early stages and need to be developed significantly before they can be used in real-world scenarios. The model for occupancy analysis provided by this study has bigger implications since it might potentially save the cost of commercial sensors for any organization using electronic turnstile systems. Therefore, it is recommended that future studies in occupancy analysis take into account variables such as test periods and student enrollment, and evaluate annual performance over longer durations. In addition to ML, investigating ARIMA models may provide insight into methodological simplicity and efficiency. Under-desk sensor testing may also increase study space monitoring and enhance data accuracy. To enhance the system and gain a deeper comprehension of user requirements, practical applications should continue with small-scale user trials and prototype development. When planning new UCL expansions or architectural projects, predictive skills may also help with energy, space, and safety planning optimization.

References


