**A close up of a logo

Description automatically generated**

**Braude College of Engineering  
Department of Software Engineering**

Capstone Project – Phase A

25-2-D-11

**An Integrated Internet of Things Monitoring System**

**Robomo 2.0**

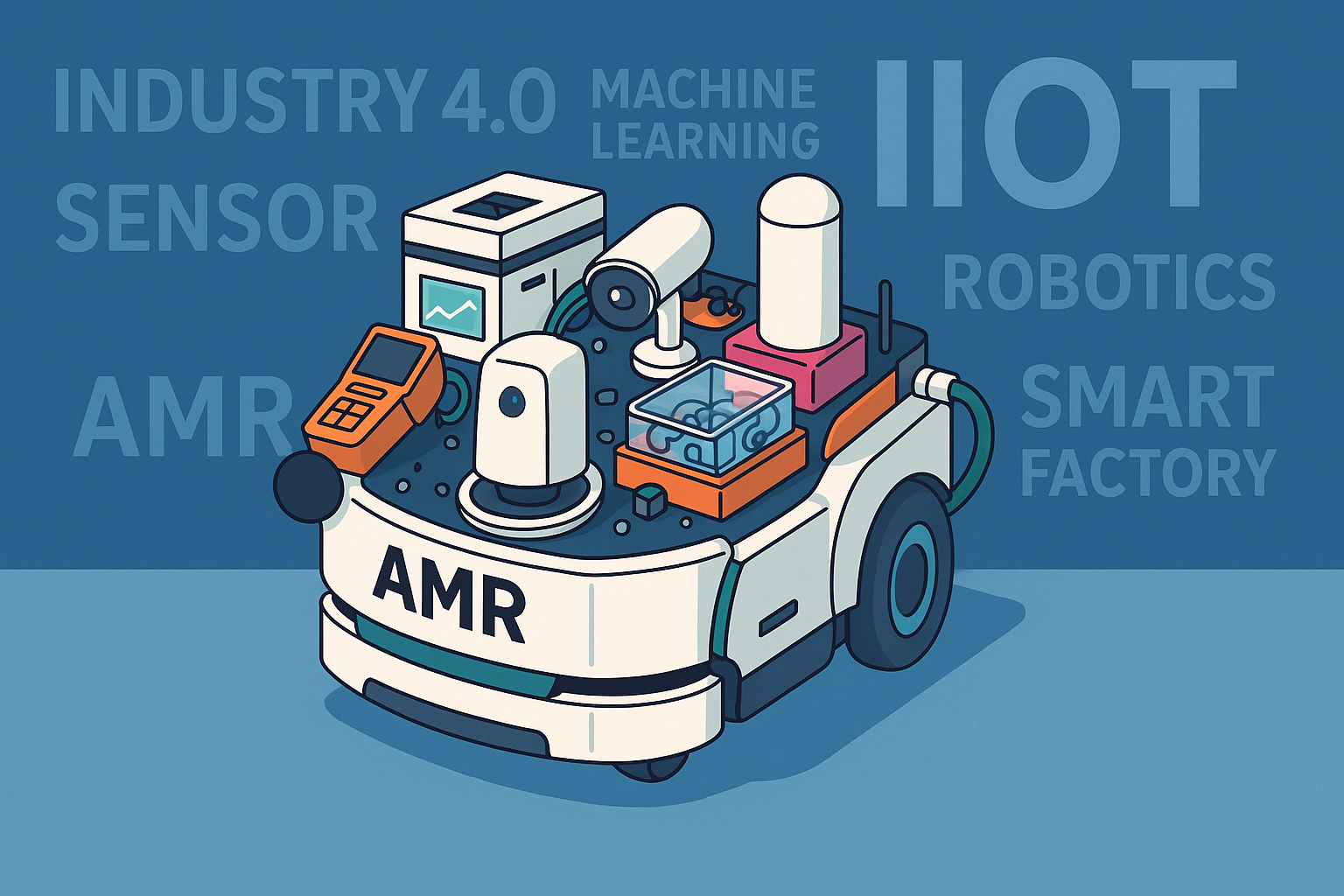
[**GIT**](https://github.com/Mohammed19J/Robomo_2.0)

Collaboration with TAMK University

BY:  
Mohammed Jaber  
Amal Kandeel

Advisor:  
Dr. Naomi Unkelos Shpigel

Date:  
May 2025



**Table of Contents**

[**Abstract** 3](#_Toc204608830)

[**1. Introduction** 4](#_Toc204608831)

[**2. Literature Review** 4](#_Toc204608832)

[**2.1 Smart Sensing for Indoor Monitoring** 4](#_Toc204608833)

[**2.2 Real-Time Occupancy Detection and Prediction** 5](#_Toc204608834)

[**2.3 Machine Learning Algorithms for Occupancy Estimation** 5](#_Toc204608835)

[**2.4 Real Time Monitoring in IoT** 5](#_Toc204608836)

[**2.5 IIoT** 6](#_Toc204608837)

[**2.6 Agile Development of Smart Sensor Systems** 6](#_Toc204608838)

[**3. Research** 6](#_Toc204608839)

[**3.1 Scheduled Meeting** 6](#_Toc204608840)

[**3.1.1 Meetings Summary** 7](#_Toc204608841)

[**3.1.1.1 Meeting Number 1 - 17.6.24** 7](#_Toc204608842)

[**3.1.1.2 Meeting Number 2 - 27.6.24** 7](#_Toc204608843)

[**4. Engineering Process** 8](#_Toc204608844)

[**4.1 Development Process** 8](#_Toc204608845)

[**4.2 User-Centered Design Approach** 8](#_Toc204608846)

[**4.2.1 Regular Customer Meetings** 8](#_Toc204608847)

[**4.2.2 Prototype-Driven** 8](#_Toc204608848)

[**4.2.3 Machine Learning Model Selection and Validation** 9](#_Toc204608849)

[**4.2.4 Conclusion** 10](#_Toc204608850)

[**4.3 Workflow** 10](#_Toc204608851)

[**4.3 Architecture Diagram** 11](#_Toc204608852)

[**4.4 Technologies Review** 12](#_Toc204608853)

[**4.4.1 Websocket** 12](#_Toc204608854)

[**4.4.2 Client-Side Technologies** 12](#_Toc204608855)

[**4.4.2.1 React** 12](#_Toc204608856)

[**4.4.2.2 Javascript** 12](#_Toc204608857)

[**4.4.2.3 Tailwind Css** 12](#_Toc204608858)

[**4.4.3 Server Technologies** 13](#_Toc204608859)

[**4.4.3.1 Node.js** 13](#_Toc204608860)

[**4.4.3.2 Fiware** 13](#_Toc204608861)

[**4.4.3.3 Firebase** 13](#_Toc204608862)

[**4.4.3.4 MQTT** 13](#_Toc204608863)

[**5 Work Artifacts** 13](#_Toc204608864)

[**5.1. Requirement** 13](#_Toc204608865)

[**5.1.1 Functional Requirements** 13](#_Toc204608866)

[**5.1.2 Non-Functional Requirements** 15](#_Toc204608867)

[**5.2. Use Case** 16](#_Toc204608868)

[**5.3. Activity Diagram** 18](#_Toc204608869)

[**5.4. Website Screens** 20](#_Toc204608870)

[**6. Expected Achievements** 22](#_Toc204608871)

[**6.1. Challenges** 23](#_Toc204608872)

[**6.2. Success Criteria** 23](#_Toc204608873)

[**6.3. Evaluation** 24](#_Toc204608874)

[**7. Testing Plan** 25](#_Toc204608875)

[**7.1. Scope** 25](#_Toc204608876)

[**7.2. Objectives** 25](#_Toc204608877)

[**7.4. Test Cases** 25](#_Toc204608878)

[**8. AI Tools and Prompts** 27](#_Toc204608879)

[**8.1. AI Usage and Results** 27](#_Toc204608880)

[**8.2. AI Prompts and Responses** 27](#_Toc204608881)

[**9. References** 29](#_Toc204608882)

### **Abstract**

In recent years, smart sensor systems have been increasingly deployed in educational environments to enable real time environmental monitoring and space utilization analysis. In collaboration with TAMK University, our project builds upon an existing IIoT (Industrial Internet of Things) framework and extends it with intelligent capabilities. Specifically, we developed a system that uses environmental sensors measuring CO₂, temperature, humidity, and VOC (Volatile Organic Compounds) to estimate room occupancy using machine learning.

We inherited a working web application from a previous student team and expanded it by integrating data driven features, including a machine learning-based agent capable of analyzing air quality data to detect occupancy, predict health index values, and even trigger alerts for fire/smoke or HVAC adjustments. To achieve this, we trained and tested models like Random Forest, Support Vector Regression (SVR), and LSTM (Long Short-Term Memory) on both public and collected datasets.

The result is a real-time monitoring system that presents predictions through an intuitive dashboard and preserves user privacy, making it suitable for sensitive environments such as exam rooms. This project reflects the strength of interdisciplinary collaboration between Braude College and TAMK, showcasing the potential of AI-powered IIoT solutions for smart academic infrastructures.

### **1. Introduction**

This project is part of a collaboration between Ort Braude College and TAMK University in Finland, aimed at advancing smart campus systems through sensor data and intelligent analytics. Our objective was to build upon an existing IIoT system and introduce a machine learning agent capable of enhancing the value of collected data. Specifically, we focused on predicting whether a room is occupied based on air quality measurements, rather than relying on intrusive methods such as cameras.

To do this, we used real time data from Robomo units mobile robots equipped with environmental sensors and centralized it using WebSocket and MQTT protocols. Our system tracks CO₂ levels, temperature, humidity, PM sensors, and VOCs, and uses these features to infer room occupancy. This is particularly relevant for spaces like exam rooms where privacy must be maintained, and traditional surveillance is not ideal.

Throughout development, we followed Agile principles, held regular meetings with our supervisor from TAMK (Mr. Kari Naakka), and iteratively improved the application. We tested different machine learning algorithms (including Random Forest, SVR, and LSTM) on both public and collected datasets. Our goal wasn’t to detect the exact number of people, but rather to classify whether the room is occupied or empty, enabling smarter HVAC (Heating, Ventilation, and Air Conditioning) control and potential energy savings.

### **2. Literature Review**

#### **2.1 Smart Sensing for Indoor Monitoring**

Indoor air quality (IAQ) sensing plays a central role in our system. Sensors that measure CO₂, temperature, humidity, volatile organic compounds (VOCs), and particulate matter (PM) are mounted on a mobile robot (Robomo) to detect changes in the environment that are often associated with human activity. For instance, CO₂ levels tend to increase when people are present due to exhalation, while VOCs and temperature shifts can indicate pollution or discomfort.

Unlike traditional fixed sensors, Robomo can autonomously navigate through different areas, offering localized and flexible coverage. This mobility addresses common blind spots found in static installations and provides more detailed and representative environmental data. These sensor readings form the basis for our smart features like occupancy detection, health index calculation, and fire/smoke detection.

#### **2.2 Real-Time Occupancy Detection and Prediction**

Estimating room occupancy in real time based on environmental data provides a privacy-respecting alternative to visual surveillance. As people enter or leave a room, the robot senses variations in CO₂, VOCs, and temperature. By training machine learning models on these patterns, the system can predict whether the room is occupied or vacant — even without cameras.

This functionality enables downstream decisions like automatically adjusting HVAC settings or logging usage patterns. Our system uses time-synchronized data streamed through WebSockets and MQTT for low-latency processing, allowing predictions to appear on the dashboard in real time. We validated our approach using public datasets and are actively collecting live data for model improvement.

#### **2.3 Machine Learning Algorithms for Occupancy Estimation**

To detect and predict occupancy, we experimented with several machine learning models:

* **Random Forests** – Known for robustness and ability to handle non-linear feature relationships. Good for interpretability and small datasets.
* **Support Vector Regression (SVR)** – Strong at predicting continuous outputs like people count when the data has clear margins.
* **LSTM (Long Short-Term Memory networks)** – Ideal for learning long-range temporal dependencies in time-series sensor data.
* **Polynomial Regression and Linear Models** – Sometimes used in small spaces with few variables and clear thresholds.

Preprocessing steps such as denoising, smoothing (e.g., Kalman filters), or computing CO₂ change rates (∆CO₂/∆time) enhance performance. Most studies report good accuracy and low MAE (Mean Absolute Error) even with limited training data, especially when sensor placement is consistent.

#### **2.4 Real Time Monitoring in IoT**

Robomo operates as a real time IoT system. It streams sensor data continuously to a centralized server using MQTT and WebSockets. This ensures users receive up to date information on air quality and occupancy, which is essential for immediate decision making. The robot’s mobility combined with constant sensing creates a rich, live dataset that can trigger automatic recommendations (e.g., increase ventilation) or alerts (e.g., smoke detection).

Compared to traditional IoT setups, our mobile sensing unit brings a spatial dimension, detecting localized problems that might go unnoticed with wall-mounted sensors. The real-time nature of this monitoring increases responsiveness and allows for proactive actions via an integrated AI agent.

#### **2.5 IIoT**

Our system is an example of an IIoT application a network of interconnected, intelligent sensors and devices designed to monitor and act upon physical environments. Robomo is equipped with modular sensors and connected to backend systems that log, analyze, and respond to environmental data. This allows real-time feedback loops that adjust HVAC systems and alert users.

Our use of open protocols like MQTT and platforms like Firebase ensures interoperability with future infrastructure. The system is designed to be scalable more robots or rooms can be added without redesigning the whole architecture. This modularity makes it a candidate for integration into smart buildings or campus-wide systems.

#### **2.6 Agile Development of Smart Sensor Systems**

We followed Agile principles during development. Starting with weekly planning meetings (especially with TAMK advisor Mr. Kari Naakka), we defined features incrementally first verifying sensor connections, then occupancy detection, then expanding to fire detection and health indexing.

This iterative approach allowed us to react to testing results, dataset limitations, and stakeholder feedback. Early experiments showed headcount prediction was inaccurate, leading us to focus on binary occupancy. Later, we expanded into HVAC automation and fire detection based on feedback. Each sprint delivered a new module or enhancement, ensuring fast adaptation to findings and goals.

### **3. Research**

#### **3.1 Scheduled Meeting**

Throughout this project, we held scheduled meetings with Senior Lecturer Mr. Kari Naakka from TAMK University, alongside our advisor, Dr. Naomi Unkelos Shpigel. These meetings were crucial for guiding our work and refining our project goals. In each session, we presented our latest progress — starting from taking over the system and setting it up, to sharing our research on machine learning approaches for occupancy detection using air quality sensors.

We used these meetings not only to report on what we had accomplished but also to discuss challenges we faced, such as the difficulty of accurately estimating the number of people in a room based on sensor data alone. These discussions led to valuable advice from Mr. Naakka, including the idea of focusing on presence detection and using the data to control HVAC systems, detect smoke and fire, and contribute to energy-saving efforts. Each meeting ended with clear action points and next steps, helping us maintain steady progress and align our work with both TAMK and Braude College expectations.

#### **3.1.1 Meetings Summary**

#### **3.1.1.1 Meeting Number 1 - 17.6.24**

During our first meeting, which was held via Microsoft Teams, the participants included Mohammed, Amal, Mr. Kari Naakka from TAMK University, and our advisor, Dr. Naomi Unkelos Shpigel. We discussed the direction of the project and evaluated the existing web application developed by a previous team. From there, we began shaping our own contribution. One of the key ideas we introduced was applying a machine learning algorithm to the sensor data we had collected. Mr. Naakka was impressed by this direction, and we had an in-depth discussion about which type of sensor would be most suitable for training the model. We ultimately chose the air quality sensor, as we already had a large dataset available and could continue collecting real-time data for training and prediction.

In addition, we briefly shared another project we had developed in Unity using augmented reality. Mr. Naakka was enthusiastic about the concept but advised us, for the time being, to focus solely on the machine learning component and to expand the range of use cases supported in the web application related to the robotic systems.



#### **3.1.1.2 Meeting Number 2 - 27.6.24**

During our second and final meeting with Mr. Kari Naakka from TAMK University, we presented our research findings and experimental results. Initially, our focus was on precise occupancy detection—estimating the exact number of occupants in a room using available environmental sensors. However, our research concluded that accurate occupant counting required additional methods beyond our existing sensor capabilities, such as cameras or Bluetooth connectivity.

Consequently, we shifted our focus toward occupancy level detection (empty, low, medium, high), air quality indexing, and smoke and fire detection, utilizing environmental sensor data. Our extensive experimentation involved several machine learning models, including KNN, Random Forest, XGBoost, LightGBM, SVM, and Logistic Regression, evaluated on credible datasets from Kaggle and Mendeley Data. Random Forest emerged as the most effective model across all three use cases due to its accuracy, interpretability, and scalability.

Mr. Naakka expressed strong approval and support for these alternative approaches, affirming the project's direction. We discussed further implementation strategies and the integration of the selected ML model into our system architecture. This final meeting effectively defined and validated the next steps of our project's development.

### **4. Engineering Process**

#### **4.1 Development Process**

Our engineering process primarily focused on developing a robust machine learning (ML) agentic system, leveraging real-time sensor data from the TAMK API. This agent system was designed and initially tested using Google Colab, preparing it for future integration into the project's frontend and backend infrastructure in the upcoming semester. Our ML models address three critical use cases: occupancy prediction, fire and smoke detection, and computation of the Air Quality Index (AQI) on a standardized scale from 0 to 100, following established AQI calculation guidelines. Additionally, data utilized in the development and evaluation of these models was archived to support future enhancements and retraining.

Key challenges during this phase included ensuring reliable preprocessing of sensor data, addressing noise and synchronization issues, and maintaining efficient inference capabilities. We structured a preprocessing pipeline optimized for ML model performance and developed a model-selection strategy that dynamically chooses the most appropriate ML algorithm for specific analyses.

#### **4.2 User-Centered Design Approach**

Our project adopted a User-Centered Design (UCD) methodology, ensuring alignment with customer expectations and usability requirements. This approach significantly influenced our development decisions, prioritizing functionality, interpretability, and responsiveness in our system design.

#### **4.2.1 Regular Customer Meetings**

Regular meetings with Kari Naakka, the TAMK project manager and primary stakeholder, were pivotal in identifying and refining the project's objectives and use cases. Our initial meeting involved ideation and capturing Kari's vision and desired functionalities. Subsequent research and experimentation informed a detailed implementation plan, which we validated and confirmed during our second meeting. Feedback from these interactions directly influenced our ML model objectives and approach.

#### **4.2.2 Prototype-Driven**

Our prototype-driven approach facilitated iterative validation of ML model capabilities and user expectations:

* Initial prototypes validated basic model functionality and accuracy.
* Intermediate prototypes included refined predictions and anomaly detection capabilities.
* Final prototype iterations emphasized model scalability and interpretability, addressing stakeholder preferences for clarity and ease of use.

#### **4.2.3 Machine Learning Model Selection and Validation**

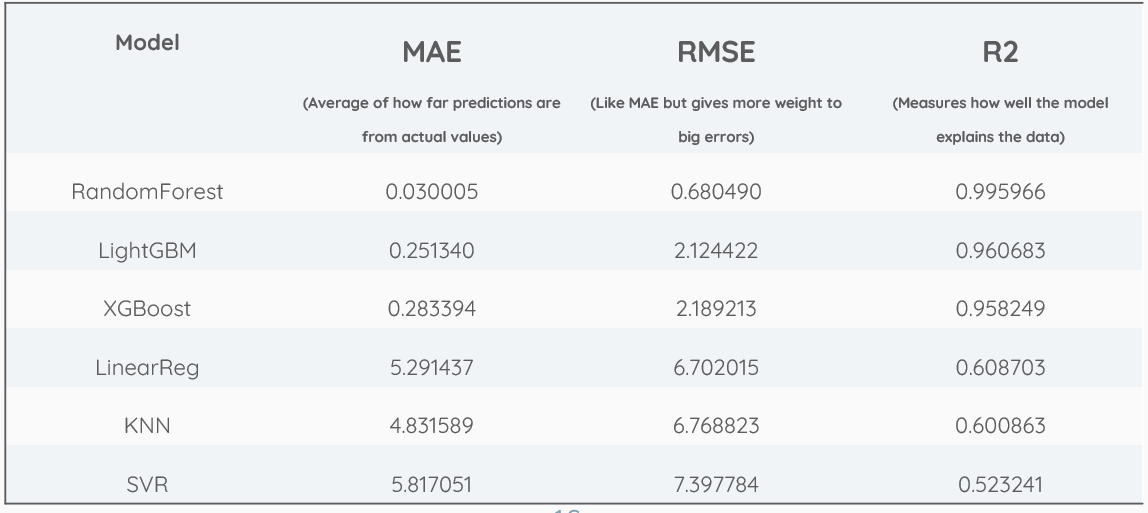
Extensive experimentation was conducted to determine optimal machine learning algorithms for each use case (occupancy prediction, fire/smoke detection, and AQI calculation). We evaluated at least six ML algorithms using credible datasets from Kaggle and Mendeley Data, carefully matching our project needs. Each algorithm was assessed using standard evaluation metrics, leading us to select Random Forest as the most consistently high-performing model across all scenarios due to its accuracy, scalability, and clear interpretability.

* **Room occupancy estimation**: multiple classification models were evaluated. As shown in the table below, the Random Forest and K-Nearest Neighbors (KNN) models demonstrated the highest performance, with Random Forest being selected for its balance of accuracy and interpretability.

A screenshot of a computer screen

AI-generated content may be incorrect.

* **Air Quality Classification**: For calculating the Air Quality Index (AQI), a regression task, several models were tested. The Random Forest model significantly outperformed others, achieving the lowest Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) and the highest R-squared (R2) value, indicating a superior fit to the data.

****

* **Smoke Detection**: In the critical task of smoke detection, various models were assessed for their classification accuracy. While LightGBM and XGBoost showed marginally higher metrics, the Random Forest model also achieved exceptional performance (over 99.9% accuracy) and was chosen for consistency across all use cases.

**A screenshot of a graph

AI-generated content may be incorrect.**

#### **4.2.4 Conclusion**

Our iterative, user-centered development and thorough validation processes resulted in the selection and preparation of robust ML models tailored to our project's requirements. The clear emphasis on usability, interpretability, and future scalability provides a strong foundation for the successful integration and deployment of these models in the subsequent development phases.

#### **4.3 Workflow**

At the start of the project, we held an introductory meeting with our fellow students, Or and Dima, from whom we inherited the initial system architecture and foundational infrastructure. During this meeting, essential information such as credentials and system access was exchanged, and the backend structure was transitioned to our management.

Following this, we initiated discussions with our project manager at TAMK University, Mr. Kari Naakka, to identify potential enhancements for the project. Kari initially suggested exploring precise occupancy detection—estimating the exact number of people present—using available environmental sensors. We conducted comprehensive research into this idea and discovered that accurate occupant counting was not feasible solely with our available sensors, as it would require additional methods like cameras or Bluetooth connectivity checks.

However, our research identified alternative valuable approaches, such as occupancy level detection (empty/low/medium/high), air quality indexing, and smoke and fire detection. We extensively tested these new concepts using publicly available datasets from platforms like Kaggle and Mendeley Data, applying various machine learning models including KNN, Random Forest, XGBoost, LightGBM, SVM, and Logistic Regression. These models were evaluated using metrics such as accuracy, precision, recall, F1-score, MAE, RMSE, and R².

Our results demonstrated that Random Forest consistently delivered outstanding performance across all three use cases, particularly noted for its accuracy, scalability, and interpretability. We presented these findings and suggested use cases to Kari during our second and final meeting, where he expressed strong approval and confirmed his interest in proceeding with these applications.

Following this validation, we focused on designing the system's architecture and workflow, clearly defining functional and non-functional requirements, and preparing detailed documentation for our capstone presentation. This foundational work positions us effectively for integrating the ML-driven agentic system into the existing project infrastructure during the upcoming development phase.

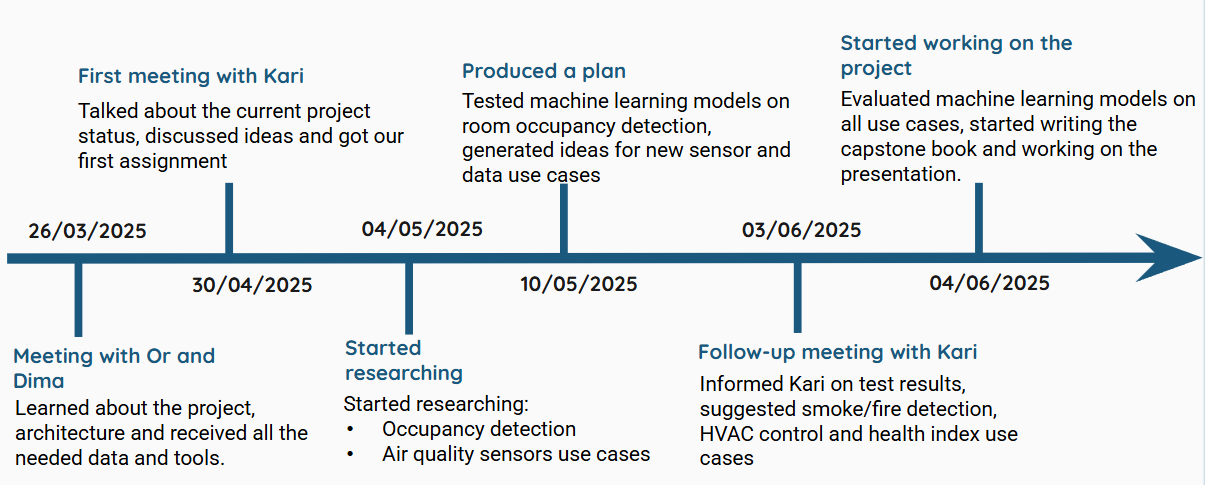


figure 1. Workflow diagram

#### **4.3 Architecture Diagram**

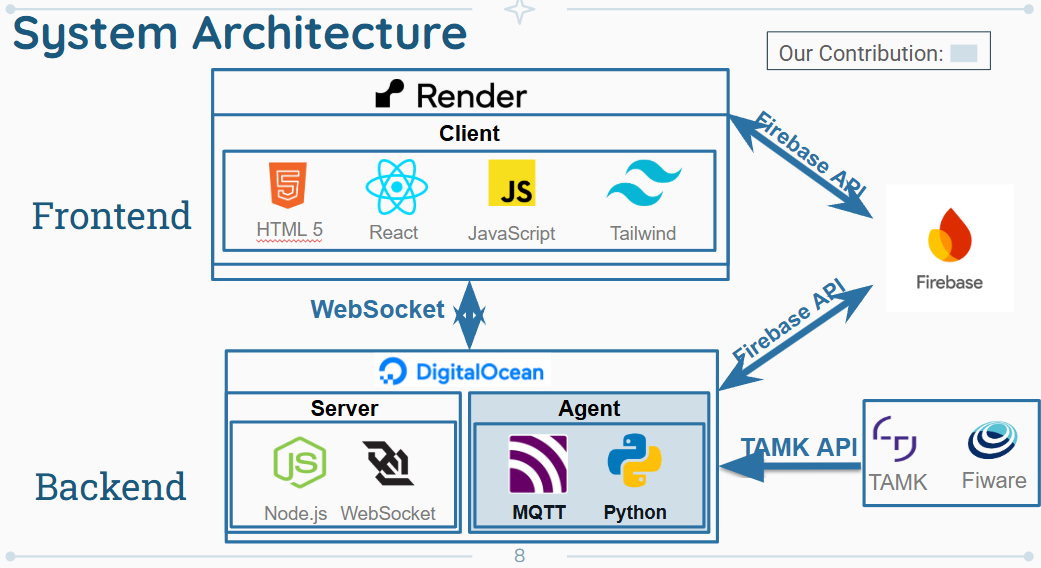
****

figure 2. Architecture diagram

The diagram above illustrates the system architecture of our project, highlighting the integration of frontend and backend components to deliver a scalable, real-time monitoring solution. The frontend, deployed on Render, is built using HTML5, React, JavaScript, and Tailwind CSS, providing users with an interactive interface to access real-time data and system insights. It communicates with the backend via WebSocket for continuous, low-latency data exchange. The backend, hosted on DigitalOcean, consists of a Node.js server that manages WebSocket communication and processes incoming data, alongside a Python-based agent that handles MQTT messages and external API calls. The agent interacts with the TAMK API and Fiware to retrieve sensor data and context information, while Firebase is used not only for real-time data storage and notifications but also to log agent actions for future system improvements and analysis. This architecture supports seamless integration between user interfaces, sensor data collection, cloud storage, and external services, enabling a robust and privacy-conscious smart monitoring system.

#### **4.4 Technologies Review**

Our project integrates several modern technologies to facilitate real-time data collection, processing, visualization, and machine learning model integration. These technologies provide a responsive and scalable system that supports continuous data flow and user interaction.

#### **4.4.1 Websocket**

WebSocket is employed as a real-time, bidirectional communication protocol, essential for seamless data exchange between the server and the client dashboard. Unlike traditional HTTP requests, WebSocket maintains an open connection, allowing instantaneous transmission of sensor data and ML model predictions, ensuring users receive timely updates reflecting the current environment.

#### **4.4.2 Client-Side Technologies**

#### **4.4.2.1 React**

React forms the foundation of our frontend interface, leveraging a component-based architecture to facilitate efficient management of complex user interfaces. React components dynamically update in response to new data, enabling real-time visualization of sensor data, predictions, and analytics without necessitating full page reloads.

#### **4.4.2.2 Javascript**

JavaScript is central to our frontend interactivity, enabling dynamic behavior and real-time responsiveness within React components. It facilitates data processing, visual updates, and user interaction handling, supporting a fluid and interactive user experience.

#### **4.4.2.3 Tailwind Css**

Tailwind CSS is utilized for precise and flexible styling, relying on utility-first CSS classes rather than predefined UI components. This approach allows us to build custom, responsive, and maintainable designs aligned specifically with our project's requirements, maintaining a cohesive and modern user interface.

#### **4.4.3 Server Technologies**

#### **4.4.3.1 Node.js**

Node.js serves as the backend's core technology, providing an efficient, lightweight, event-driven environment tailored for real-time data processing. Its non-blocking architecture effectively handles continuous data streams from sensors, supporting seamless WebSocket communication and interaction with machine learning algorithms. Additionally, Node.js powers our backend API endpoints and overall data handling processes.

#### **4.4.3.2 Fiware**

FIWARE supports the management and structuring of contextual information from sensor data. Its standardized APIs facilitate scalability and integration into smart environments or smart city infrastructures, enhancing our system's modularity and adaptability for future expansions.

#### **4.4.3.3 Firebase**

Firebase Realtime Database stores real-time sensor data, machine learning predictions, and logs of system activities. The automatic synchronization of Firebase across clients ensures consistent, up-to-date information for users, further supporting secure user authentication and role-based access control.

#### **4.4.3.4 MQTT**

MQTT (Message Queuing Telemetry Transport) is a lightweight publish-subscribe messaging protocol ideally suited for IoT applications, particularly in environments with variable network conditions. In our project, MQTT facilitates efficient and reliable data transfer between the sensor-equipped robotic units and the server, supporting real-time streaming of environmental data essential for backend processing and machine learning predictions.

### **5 Work Artifacts**

#### **5.1. Requirement**

#### **5.1.1 Functional Requirements**

|  |  |
| --- | --- |
| **No.** | **Requirement** |
| 1 | The system shall detect room occupancy using CO₂, temperature, humidity, and VOC sensors. |
| 2 | The system shall display real-time sensor data and Health Index on the web dashboard. |
| 3 | The system shall compute and display a Health Index score that summarizes air quality conditions. |
| 4 | The system shall trigger alerts for abnormal air quality conditions (e.g., high CO₂ levels, smoke detection via VOC/temperature spikes). |
| 5 | The system shall recommend actions (e.g., increase ventilation) when air quality deteriorates or a room is unoccupied. |
| 6 | The system shall log all sensor data, predictions, alerts, and recommendations for review and analysis. |
| 7 | The system shall provide a live map showing robot location and sensor status. |
| 8 | The system shall allow users to generate graphs for selected attributes and view historical data. |
| 9 | The system shall send notifications when manual intervention is required (e.g., when HVAC cannot be automatically controlled). |
| 10 | The system shall support fire/smoke detection logic as part of safety features. |
| 11 | The system shall allow dynamic selection and comparison of multiple devices and sensors on the dashboard. |
| 12 | The system will allow the administrator to approve new sign-ups. |
| 13 | The system will allow users to select devices from a map. |
| 14 | The system will allow users to choose which devices are shown on the map. |
| 15 | The system will allow users to choose attributes for each device. |
| 16 | The system will allow users to copy selected data. |
| 17 | The system will allow users to mark an attribute as a favorite. |
| 18 | The system will allow users to hide a device's attributes. |
| 19 | The system will allow users to expand the information window. |
| 20 | The system will allow users to minimize the information window. |
| 21 | The system will allow users to add graphs of an attribute to the compare window. |
| 22 | The system will allow users to remove graphs from the compare window. |
| 23 | The system will allow users to view the device in 360 degrees. |
| 24 | The system will allow users to close the selected device. |
| 25 | The system will allow access to the site through mobile devices. |
| 26 | The system will support dark mode functionality. |
| 27 | The system will allow users to log in. |
| 28 | The system will allow users to sign up. |

table 1. Functional Requirements

#### **5.1.2 Non-Functional Requirements**

|  |  |  |
| --- | --- | --- |
| **No.** | **Requirement** | **type** |
| 1 | Dashboard updates should occur within 1 second of receiving new sensor data. | Performance |
| 2 | Code modularity should support updates with minimal downtime (<5 minutes per month). | Maintainability |
| 3 | The site should be mobile friendly (WCAG 2.2 compliant). | Adaptability |
| 4 | The system should always provide up-to-date data. | Maintainability |
| 5 | The site should load within 3 seconds. | Efficiency, Performance |
| 6 | The site will save cookies. | Usability |
| 7 | The site will work properly even if none of the devices are available. | Fault tolerance |
| 8 | The system should support up to 5000 concurrent users without performance degradation. | Scalability |
| 9 | Only authorized users should be able to access data with MFA. | Security |
| 10 | The system should comply with data protection regulations (e.g., GDPR). | Compliance |
| 11 | New modules or features can be integrated with no more than 10% modification of existing code. | Extensibility |
| 12 | The system should provide interactive features with response time < 1 second for queries. | Interactivity |
| 13 | At least 80% of system components should be reusable in similar projects. | Modularity |

table 2. Non-Functional Requirements

#### **5.2. Use Case**

The system’s functionality is captured by two primary actors depicted in the Use Case diagrams:

**User**

* Sign in to and sign out from the web dashboard, managing authentication and session control.
* Sort and group devices, and toggle the visibility of individual devices on the dashboard map.
* Adjust the data refresh interval to regulate the frequency of real-time updates.
* Activate or deactivate the autonomous ML agent to start or pause automated analysis.
* Select devices and pin specific sensor attributes for detailed monitoring and comparison.
* Generate comparison graphs of chosen attributes across user-defined time windows.
* View real-time outputs such as occupancy levels, Air Quality Index readings, and system alerts.

A diagram of a company

AI-generated content may be incorrect.

=

Figure A. UseCase diagram – User

**Agent**

* Fetch live sensor data from the TAMK API, using WebSocket and MQTT protocols.
* Analyze incoming environmental data and route it to the appropriate ML workflows.
* Predict room occupancy levels (empty, low, medium, high) based on sensor measurements.
* Compute the Health Index (AQI) and, when applicable, extend agent actions to control HVAC systems.
* Detect fire or smoke conditions using particulate matter (PM) thresholds and issue emergency alerts.
* Log all model predictions, alerts, and control actions in Firebase for auditing and future retraining.
* Trigger external system responses, including sending notifications and executing HVAC adjustments.

A group of blue ovals with black text

AI-generated content may be incorrect.

Figure B. UseCase diagram – Agent

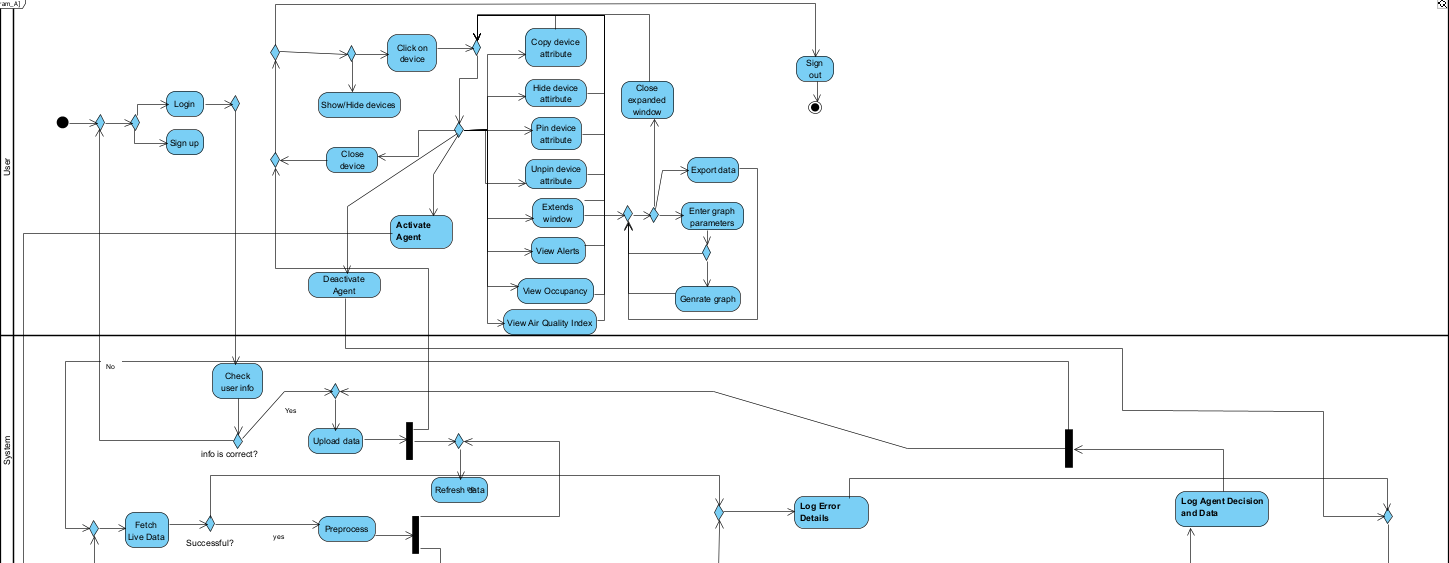
These use case diagrams clearly delineate the interactions between end users and the autonomous agent, capturing both manual dashboard operations and automated decision-making workflows.

#### **5.3. Activity Diagram**

To further elucidate the dynamic behavior of the system, two swimlane activity diagrams model the distinct workflows of the end user and the autonomous agent.

**User Activity Flow:**

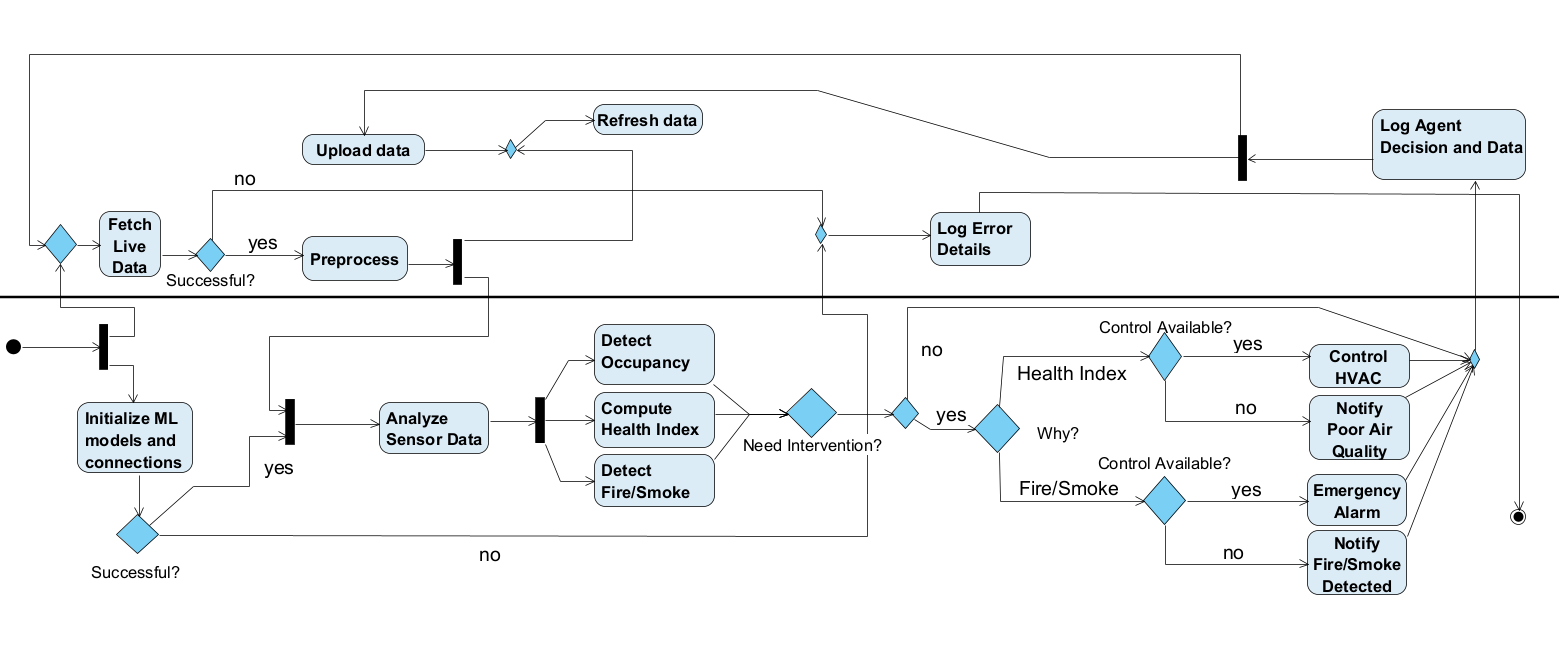
1. **Authentication**: The user signs in (or signs up) to access the dashboard and may sign out to end the session.
2. **Device Management**: The user toggles device visibility on the map, selects devices, and closes expanded device views.
3. **Attribute Interaction**: Within an expanded device window, the user can copy, pin, hide, or unpin individual sensor attributes.
4. **Agent Control**: The user activates or deactivates the ML agent to begin or pause automated analysis.
5. **Data Visualization**: The user exports attribute data, inputs parameters (time range and chart type), and generates custom graphs.
6. **Real‑Time Monitoring**: Throughout the session, the user can view occupancy levels, Air Quality Index readings, and receive system alerts as provided by the agent.

****

Part A. Activity diagram - User and System

**Agent Activity Flow:**

1. **Initialization**: The agent initializes machine learning models and establishes required connections; if initialization fails, error details are logged and the process terminates.
2. **Data Acquisition**: The agent fetches live sensor data via MQTT or WebSocket; failed retrieval triggers error logging and an attempt to refresh data.
3. **Preprocessing**: Successful data fetches undergo cleaning and formatting to prepare for ML analysis.
4. **ML Inference**: Preprocessed data is fed into three parallel ML workflows:
   * **Occupancy Detection** (empty/low/medium/high)
   * **Health Index Computation** (AQI scale 0–100)
   * **Fire/Smoke Detection** (emergency threshold checks)
5. **Intervention Decision**: A decision gateway evaluates whether intervention is required:
   * **Air Quality Intervention**: If AQI exceeds predefined thresholds, the agent either issues HVAC control commands (if available) or notifies users of poor air quality.
   * **Fire/Smoke Intervention**: If fire/smoke is detected, the agent triggers an emergency alarm (if available) or sends an alert to users.
6. **Logging**: All agent decisions, actions, and input data are recorded in Firebase for auditing, performance analysis, and future model retraining.
7. **Loop Continuation**: The workflow returns to the data acquisition step, ensuring continuous, real‑time monitoring and responsiveness.



**System**

**Agent**

Part B. Activity diagram - Agent and System

These activity diagrams confirm the system’s robustness and clarity by explicitly modeling both manual user interactions and automated agent-driven processes, highlighting error handling, decision logic, and continuous operation loops.

#### **5.4. Website Screens**



figure 3. Login screen

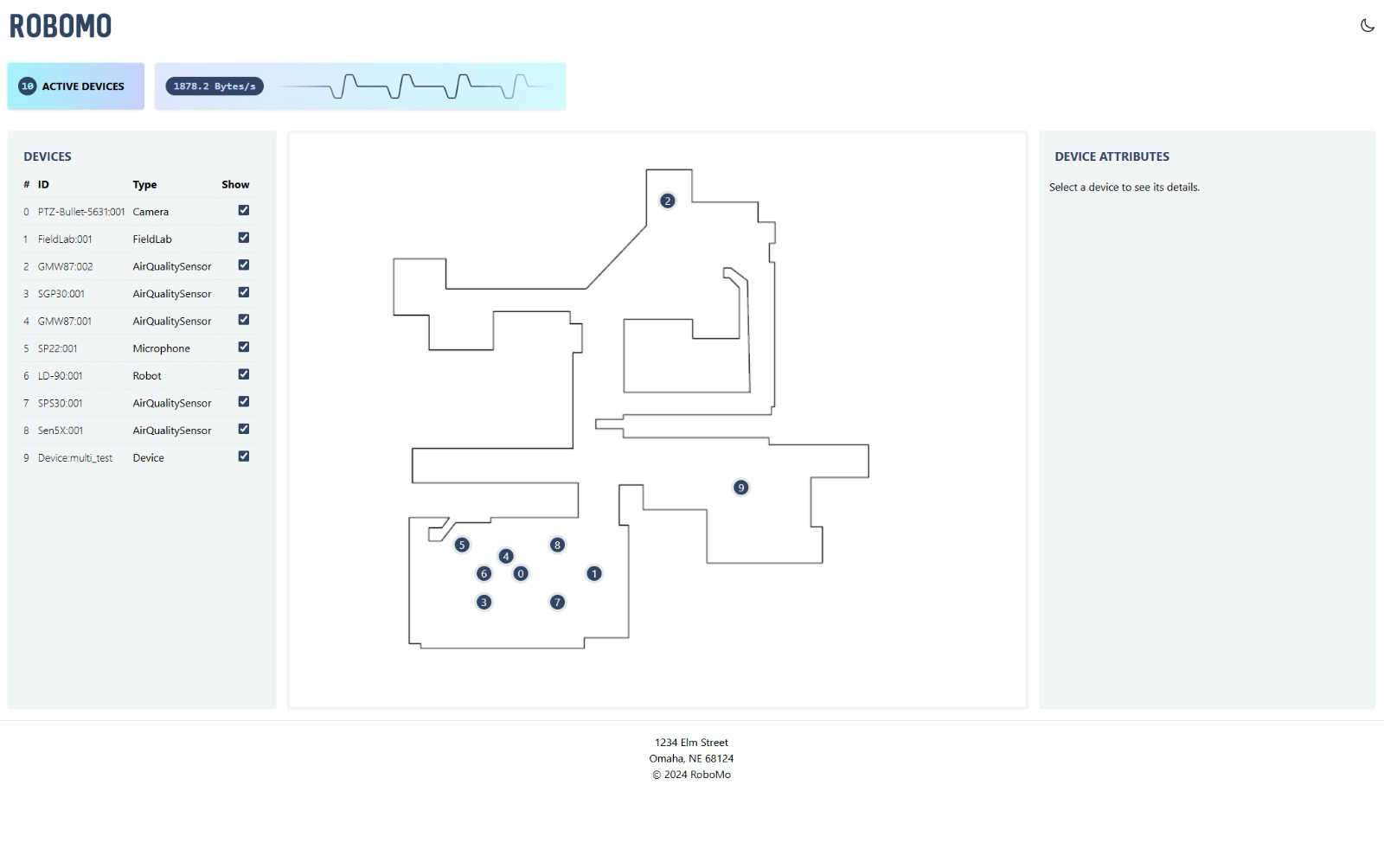


figure 4. Home screen



figure 5. Selecting device screen



figure 6. Expanded window



figure 7. Generating graph from expanded screen

### **6. Expected Achievements**

* Develop a Smart Robot Prototype: Create a functional robot integrated with a variety of sensors for comprehensive environmental monitoring.
* Implement Machine Learning Models: Utilize machine learning to analyze sensor data for occupancy detection, health index calculation, and smoke/fire detection.
* Ensure Mobile Monitoring Capability: Enable the robot to autonomously navigate and collect data from different locations, providing more extensive coverage than stationary sensors.
* Design for Modularity and Scalability: Create a system that allows for the easy addition of new sensors and updated machine learning algorithms in the future.
* User-Friendly Interface: The system will feature an intuitive dashboard for users to easily monitor live sensor data and the calculated Health Index.

#### **6.1. Challenges**

* **Dataset Limitations:**
  + The available datasets did not perfectly match the specific lab environment.
  + There was no pre-labeled data for head-counting using the project's specific set of sensors.
  + Finding datasets that met all the specific requirements for the project proved to be difficult.
* **TAMK Misalignment:**
  + Coordination with the TAMK team was challenging due to their break coinciding with the project's semester.
  + There was a lack of precise details regarding the TAMK lab's dimensions and the locations of HVAC units.
* **Robot Availability:** 
  + The robot was not always operational when needed for data collection.

#### **6.2. Success Criteria**

* **Optimized Air Quality:** Maintain CO₂ levels below 1000 ppm and achieve a "good" rating for VOCs (under 500 μg/m³) for 90% of the time.
* **Enhanced Health Index:** Ensure the health index score remains above 80 for at least 90% of the day.
* **Accurate Occupancy and Automation:** Achieve over 99% accuracy in occupancy detection, leading to an estimated 10-30% in energy savings through HVAC automation.
* **Increased Productivity and Coverage:**
* Lead to a 5-10% increase in productivity and a 20-40% reduction in complaints regarding air quality.
* The robot's mobile monitoring should identify 20-40% more problem areas than fixed sensors.
* **System Performance:**
* The dashboard will update with sensor data within one second.
* The system will maintain at least 95% uptime during operational hours.

#### **6.3. Evaluation**

The project's success will be evaluated through a two-pronged approach: **technical performance validation** and **user-centric assessment**.

**Technical Performance Validation**

This phase will measure the system against the quantitative success criteria defined in the presentation.

* **Model Accuracy:** The performance of the machine learning models for occupancy, air quality classification , and smoke detection will be re-validated using real-world data collected by the robot in the TAMK lab environment.
* **Data Latency:** We will measure the time from a sensor detecting a change to the corresponding update appearing on the dashboard, ensuring it meets the sub-1-second goal.
* **System Uptime:** An automated monitoring service will be used to track the availability of the web dashboard and backend services to ensure the target of 95% uptime is met.

**User-Centric Assessment**

This phase focuses on the usability and effectiveness of the solution for its intended users (e.g., lab managers, facility operators).

* **Task-Based Usability Testing:** A small group of representative users will be given specific scenarios to complete using the dashboard. Example tasks include:
  + "Identify the current Health Index in the main lab and determine if any pollutants are at high levels."
  + "A smoke alert has been triggered. Locate the source of the alert on the map."
  + "Check if the system thinks the lab was occupied yesterday afternoon."
* **Qualitative Feedback:** Following the usability tests, users will be interviewed or asked to complete a short questionnaire to gather qualitative feedback on the interface's design, clarity, and overall usefulness in managing the indoor environment.

### **7. Testing Plan**

#### **7.1. Scope**

The testing scope covers the entire system architecture:

* **Frontend**: The React-based user dashboard, including all UI components, data visualizations, and WebSocket communications.
* **Backend**: The Node.js server, including API endpoints, data processing logic, and the MQTT message broker integration.
* **ML** **Agent**: The Python agent responsible for fetching live data, running predictions for occupancy, health index, and fire/smoke, and logging actions.
* **End**-**to**-**End**: The complete data flow from a sensor reading to a prediction and visualization on the dashboard.

#### **7.2. Objectives**

* To verify that all functional requirements, such as occupancy detection, alert generation, and data logging, operate as designed.
* To ensure the system meets its non-functional requirements for performance, scalability, and usability.
* To identify, document, and resolve bugs across the full technology stack.
* To validate the system's stability during continuous, long-term operation.

**7.3. Testing Approach**

* **Unit Testing**: Individual functions and components will be tested in isolation. We will use Jest for the React frontend components and PyTest for the Python ML agent's functions (e.g., the AQI calculation logic).
* **Integration Testing**: We will test the interfaces between system components, such as verifying that the backend correctly processes data received from the MQTT broker and forwards it to the ML agent.
* **End-to-End (E2E) Testing**: We will simulate a complete user scenario, from having the robot collect sensor data to verifying that the correct alerts and dashboard updates occur in the browser.
* **Manual Testing**: We will manually execute the test cases for the user interface to ensure the workflows are intuitive and the user experience is seamless.

#### **7.4. Test Cases**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Test ID | Precondition | Expected Result | Description |
| 1 | **FetchSensorData** | The agent is active and connected to the network. | Live sensor data is successfully received and displayed on the dashboard. | The agent initiates a request to fetch live data from the integrated sensors. |
| 2 | **PredictOccupancy** | The system has successfully received sensor data (CO₂, temp, humidity, etc.). | The system accurately predicts whether the room is occupied or empty. | The system analyzes the sensor data using the trained machine learning model to predict occupancy. |
| 3 | **ComputeHealthIndex** | The system has successfully received and analyzed sensor data. | A health index score is calculated and displayed on the dashboard. | The system aggregates various environmental parameters into a single health index score. |
| 4 | **DetectFire/Smoke** | The system has successfully received sensor data, including Particulate Matter (PM) readings. | The system correctly identifies the presence of smoke or fire. | The system analyzes sensor data to detect anomalies indicative of fire or smoke. |
| 5 | **SendHealthIndexAlert** | The computed Health Index falls below a predefined threshold. | An alert is sent to the user, and the system may trigger HVAC control. | The system sends a notification about poor air quality and may automatically adjust the HVAC system. |
| 6 | **SendFire/SmokeAlert** | The system detects a fire or smoke event. | An emergency alarm is set off, and a high-priority alert is sent. | The system triggers an immediate emergency alarm and notifies the user of the fire/smoke detection. |
| 7 | **LogActions** | The agent has performed an action (e.g., data fetching, sending an alert). | The action and relevant data are logged for historical analysis. | The system logs all agent decisions and data for future review. |
| 8 | **UserInteraction-ViewData** | A user is logged into the dashboard. | The user can view the live air quality index, occupancy status, and alerts. | A user navigates the interface to view different data visualizations. |
| 9 | **UserInteraction-ControlDevice** | A user is logged in and viewing device attributes. | The user can click on a device to see its detailed attributes. | A user selects a specific device on the map or list to view its data. |
| 10 | **UserInteraction-ActivateAgent** | A user is logged in with appropriate permissions. | The agent is successfully activated or deactivated. | A user toggles the agent's operational status through the UI. |

table 3. Test cases

### **8. AI Tools and Prompts**

During our work on part a, we used AI tools such as: ChatGPT, Claude and Gemini. We used these different AI tools in a variety of areas such as checking syntax and spelling, conducting literature introduction research as well as in the practical part to check on various technologies and methods.

#### **8.1. AI Usage and Results**

* **Use**: We used the AI tools to look for datasets for our 3 use cases (Occupancy Detection, Fire/Smoke Detection and Air Quality Index).

**Result**: They provided us with many datasets from different websites like Kaggle, Mendeley Data, UCI Machine Learning Repository, Hugging Face.

* **Use**: We used AI tools to help us build the machine learning models testing code on google colab, and tested the datasets we choose there on different ML models and evaluated each model to check which is the best for our use cases.

**Result**: We solved bugs easily, tested the models on the datasets using only our sensors and got our results back in a table format with each evaluation value metric, thus we were able to find out which model was the best for us.

* **Use**: We used AI tools to learn which methods are used today to control HVAC systems and to learn more about the sensors we have and their capabilities.

**Result**: The AI tools explained two ways (MQTT, FIWARE), they taught us about each sensor we have and what is it useful for.

#### **8.2. AI Prompts and Responses**

**ChatGPT**

We used ChatGPT extensively for multiple stages of the project, including research, system design, and code generation.

**Prompts Used:**

* "Find datasets with CO₂, temperature, humidity, VOC, or PM values for indoor occupancy detection."
* "How can HVAC systems be controlled in a smart building using IoT technologies?"
* "Write a Python script that loads a CSV file, runs multiple machine learning models on selected features, evaluates their performance, and displays the results in a table. The code should run in Google Colab."

**Results:**

* ChatGPT provided a list of relevant datasets from **Kaggle**, **Mendeley Data**, and the **UCI Machine Learning Repository**. After reviewing them, we selected three datasets that best matched our use cases: **Occupancy Detection**, **Fire/Smoke Detection**, and **Air Quality Index Prediction**.
* In response to the HVAC control prompt, ChatGPT introduced us to **MQTT** and **FIWARE**:
  + **MQTT (Message Queuing Telemetry Transport):** A lightweight protocol ideal for sending sensor data from devices to servers with low overhead and high efficiency.
  + **FIWARE:** A platform offering standardized APIs to manage context data in smart systems, making it easier to integrate real-time sensor input with applications like HVAC control.
* The Python script generated by ChatGPT in Colab helped us test various machine learning models on our datasets using only the features our sensors provided. The script evaluated each model using metrics like accuracy, MAE, RMSE, and R², and summarized the results in a comparison table. This allowed us to select the best model (Random Forest) for each use case.

**Gemini**

Gemini was mainly used to supplement our dataset search.

**Prompts Used:**

* "Suggest datasets with environmental sensor data (CO₂, VOC, humidity) suitable for indoor occupancy estimation."

**Results:**

* Gemini returned a broad list of dataset sources across Kaggle, Mendeley Data, and Hugging Face. Some datasets were not relevant, but it helped us filter out and choose those that aligned best with our specific sensor types.

**Claude**

Claude was used specifically for debugging and optimizing the Python code generated in ChatGPT.

**Prompts Used:**

* "Fix this error in my Colab notebook when running Random Forest on my CSV data."
* "Why is my AQI calculation giving NaN values in some rows?"
* "How can I optimize this code block to loop through models and print results more cleanly?"

**Results:**

* Claude helped us quickly resolve syntax and logic issues in the Colab notebook, including handling NaN values in the dataset, fixing model training errors, and improving how results were formatted and printed.
* It also suggested efficient practices for code modularity, such as wrapping evaluation steps into functions and using loops to handle multiple models, improving maintainability for future testing.

### **9. References**

1. Ahmed, K. P. (2022). “IoT Indoor Air Quality Dataset.”. [Kaggle](https://www.kaggle.com/datasets/khajaahmed1/iot-indoor-air-quality?utm_source=chatgpt.com)
2. Candanedo, L. M., & Feldheim, V. (2016). “Accurate occupancy detection of a office room from light, temperature, humidity and CO₂ measurements using statistical learning models.” Energy & Buildings, 112, 28‑39. [ResearchGate](https://www.researchgate.net/publication/285627413_Accurate_occupancy_detection_of_an_office_room_from_light_temperature_humidity_and_CO2_measurements_using_statistical_learning_models?utm_source=chatgpt.com)
3. Contractor, D. (2021). “Smoke Detection Dataset.”. [Kaggle](https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset?utm_source=chatgpt.com)
4. Demertzi, V., Demertzis, S., & Demertzis, K. (2023). An Overview of Privacy Dimensions on the Industrial Internet of Things (IIoT). Algorithms, 16(8), 378.
5. FIWARE Foundation, What is FIWARE?. [Online]. Available: <https://www.fiware.org/about-us/>
6. [FIWARE, the standard that the IoT needs – TM Forum](https://www.tmforum.org/press-and-news/fiware-standard-iot-needs/)
7. Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE access, 8, 108952 108971.
8. Google Colab. ML models implementation. [Colab](https://colab.research.google.com/drive/1qoP6RXCS9k5gaaQEZeUasZUP2cZ-Uvrv?usp=sharing)
9. Google, Firebase Realtime Database. [Online]. Available:<https://firebase.google.com/docs/database>
10. Grieco, L. A., Rizzo, A., Colucci, S., Sicari, S., Piro, G., Di Paola, D., & Boggia, G. (2014). IoT-aided robotics applications: Technological implications, target domains and open issues. Computer Communications, 54, 32-47.
11. [IEEE SA - IEEE 828-2012](https://standards.ieee.org/ieee/828/5367/)
12. IETF, The WebSocket Protocol, RFC 6455, 2011. [Online]. Available:

<https://tools.ietf.org/html/rfc6455>

1. Kamilaris, A., & Botteghi, N. (2020). The penetration of Internet of Things in robotics: Towards a web of robotic things. Journal of ambient intelligence and smart environments, 12(6), 491-512.
2. Khan, A., Aziz, S., Bashir, M., & Khan, M. U. (2020, March). IoT and wireless sensor network based autonomous farming robot. In 2020 international conference on emerging trends in smart technologies (ICETST) (pp. 1-5). IEEE.
3. Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A., & Qureshi, B. (2020). An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors, 20(21), 6076.
4. Li, P., & Liu, X. (2019, July). Common sensors in industrial robots: A review. In Journal of Physics: Conference Series(Vol. 1267, No. 1, p. 012036). IOP Publishing.
5. Meta. (2023). “React documentation.” <https://react.dev> [React](https://react.dev/?utm_source=chatgpt.com)
6. Mozilla MDN, JavaScript Guide. [Online]. Available: <https://developer.mozilla.org/en-US/docs/Web/JavaScript>
7. Nayyar, A., Puri, V., Nguyen, N. G., & Le, D. N. (2018). Smart surveillance robot for real-time monitoring and control system in environment and industrial applications. In Information Systems Design and Intelligent Applications: Proceedings of Fourth International Conference INDIA 2017 (pp. 229-243). Springer Singapore.
8. Node.js Foundation, Node.js Documentation. [Online]. Available: <https://nodejs.org/en/docs>
9. OASIS, MQTT Version 3.1.1 Protocol Specification, 2014. [Online]. Available: <https://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html>
10. OASIS. (2019). MQTT Version 5.0. Standard specification. [OASIS Open](https://www.oasis-open.org/standard/mqtt-v5-0-os/?utm_source=chatgpt.com)
11. Ray, P. P. (2016). Internet of robotic things: concept, technologies, and challenges. IEEE access, 4, 9489-9500.
12. [React](https://react.dev/)
13. Singh, M., Fuenmayor, E., Hinchy, E. P., Qiao, Y., Murray, N., & Devine, D. (2021). Digital twin: Origin to future. Applied System Innovation, 4(2), 36.
14. [Tailwind CSS: What It Is, Why Use It & Examples](https://blog.hubspot.com/website/what-is-tailwind-css)
15. Tailwind Labs, Tailwind CSS Documentation. [Online]. Available: <https://tailwindcss.com/docs>
16. Vela, A., Alvarado‑Uribe, J., & Ceballos, H. G. (2021). “Fitness‑Gym and Living‑room Occupancy Estimation Data” (Version 3) [Mendeley Data]. [Mendeley Data](https://data.mendeley.com/datasets/kjgrct2yn3/3?utm_source=chatgpt.com)
17. Vermesan, O., Bröring, A., Tragos, E., Serrano, M., Bacciu, D., Chessa, S., ... & Bahr, R. (2022). Internet of robotic things–converging sensing/actuating, hyperconnectivity, artificial intelligence and IoT platforms. In Cognitive hyperconnected digital transformation (pp. 97-155). River Publishers.
18. [Web Content Accessibility Guidelines (WCAG) 2.2](https://www.w3.org/TR/WCAG22/)
19. [What is FIWARE?](https://www.fiware.org/2011/05/17/what-is-fiware/)
20. [What is web socket and how it is different from the HTTP? - GeeksforGeeks](https://www.geeksforgeeks.org/what-is-web-socket-and-how-it-is-different-from-the-http/)
21. [What Is JavaScript: A Beginner's Guide to the Basics of JS](https://www.hostinger.com/tutorials/what-is-javascript)
22. [What Is Node.js? A Complete Guide for Developers](https://www.simplilearn.com/tutorials/nodejs-tutorial/what-is-nodejs)
23. Zhang, W., Wu, Y., & Calautit, J. K. (2022). “A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment.” Renewable & Sustainable Energy Reviews, 167, 112704. [ResearchGate](https://www.researchgate.net/publication/361650719_A_review_on_occupancy_prediction_through_machine_learning_for_enhancing_energy_efficiency_air_quality_and_thermal_comfort_in_the_built_environment)