ESW Project: Indoor and Outdoor Air Pollution

Team: 17
Archisha Panda
Dharmineni Yashaswinee
Paridhi Jain
Vyakhya Gupta

2 December 2023

1 Introduction

This report outlines the development of an 'Air Quality Monitor' designed to enhance awareness of air pollution. The system monitors key parameters such as PM 2.5, PM 10, temperature, and humidity both indoors and outdoors, with continuous data uploaded to ThingSpeak.

A notable feature is the caution system, transforming data into an Air Quality Index (AQI) with color-coded indicators for quick interpretation.The 'Air Quality Monitor' employs sensors for continuous monitoring of indoor and outdoor pollution levels. Daily trend in indoor pollution are analyzed, providing valuable insights into air quality variations within enclosed spaces. We have performed Diwali celebration pollution analysis, aiding in the assessment of the immediate impact of festive celebrations on air quality both indoors.

The amalgamation of real-time data collection, an intuitive caution system, and analysis offers a powerful tool for communities to assess both indoor and outdoor air quality. The insights derived from this system empower stakeholders to formulate informed decisions and strategies for mitigating the impact of air pollution on public health and the environment.

2 IOT Flow

1. **Sensor Data Acquisition:**
    - PM2.5, PM10, and AHT sensors measure air quality.
    - ESP32 preprocesses and calibrates data for accuracy.
2. **ThingSpeak Integration:**
    - ESP32 connects to ThingSpeak for data storage and analytics.

3. **User Interface (UI):**
    - UI displays real-time and historical air quality data.
4. **Alerts and Notifications:**
    - System alerts users if air quality exceeds thresholds.
5. **Analytics:**
    - Historical data enables trend analysis.

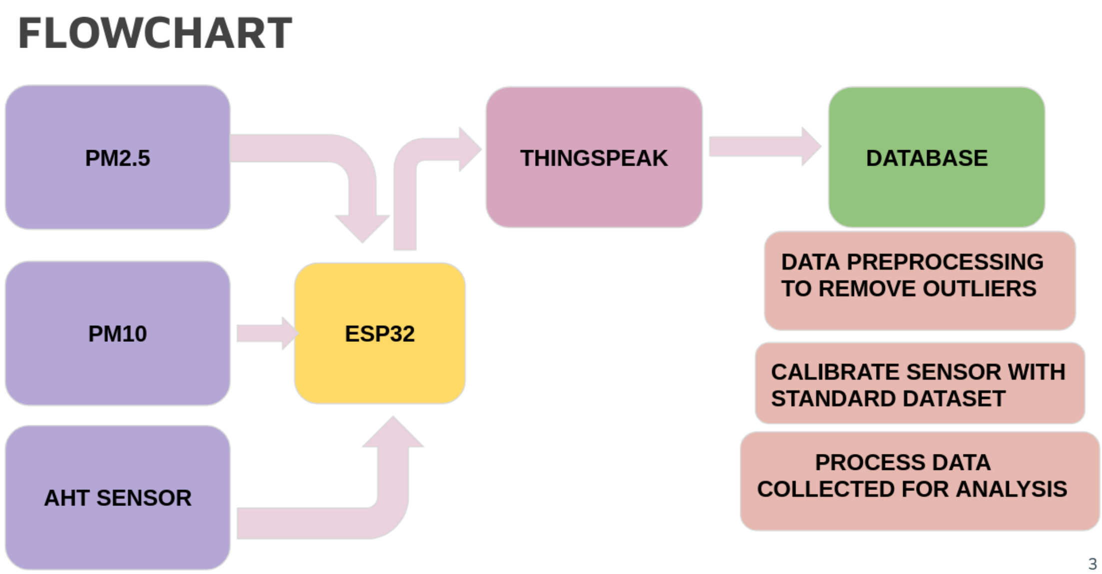

Figure 1: Flowchart

3 Circuit Diagram and Hardware

The hardware comprises PM2.5, PM10, and AHT sensors connected to an ESP32 microcontroller. The circuit involves sensor integration with ESP32 for data processing and calibration, ensuring accurate air quality measurements.

The constructed PCB were deployed outside New Boys Hostel (NBH) for outdoor circuit and in a hostel bedroom in Parijat Hostel for the indoor circuit. Data entries, including PM2.5, PM10, temperature, and humidity readings, were continuously collected and pushed to the ThingSpeak using HTTP for centralized storage and analysis.

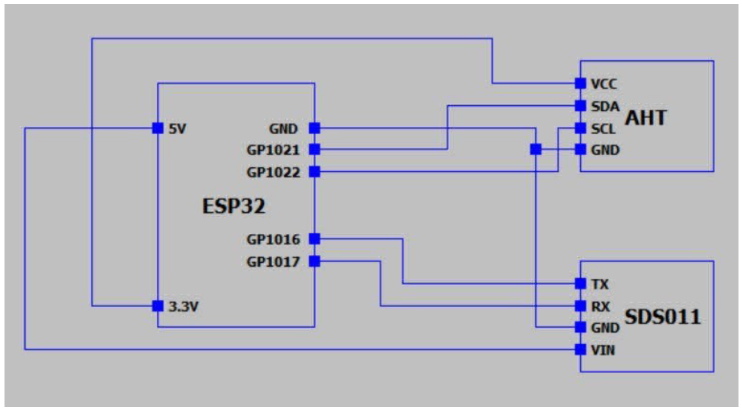

Figure 2: Circuit Diagram

Figure 3: Indoor Deployment

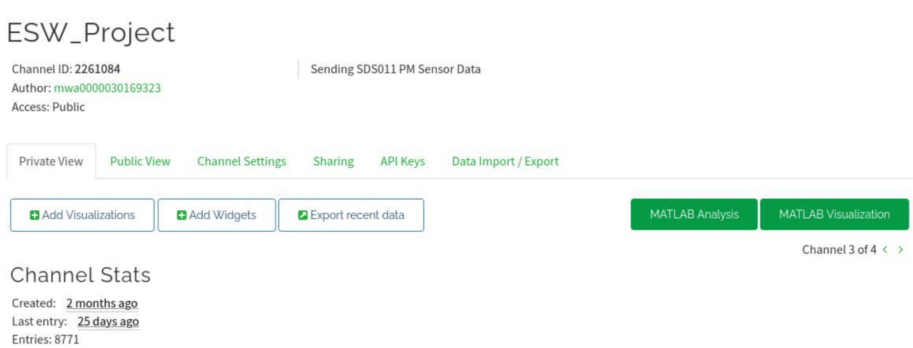

Figure 4: Thingspeak Channel

4 Data Visualization

The following plots illustrate the air quality data collected from ThingSpeak:

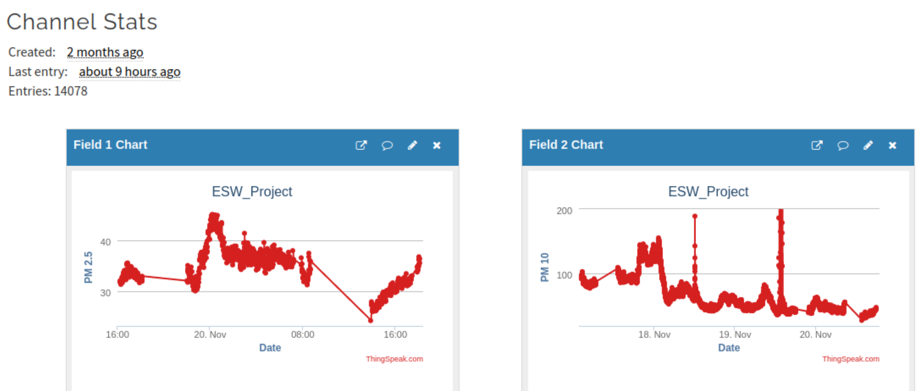

Figure 5: Indoor PM2.5 and PM10 Levels Over Time

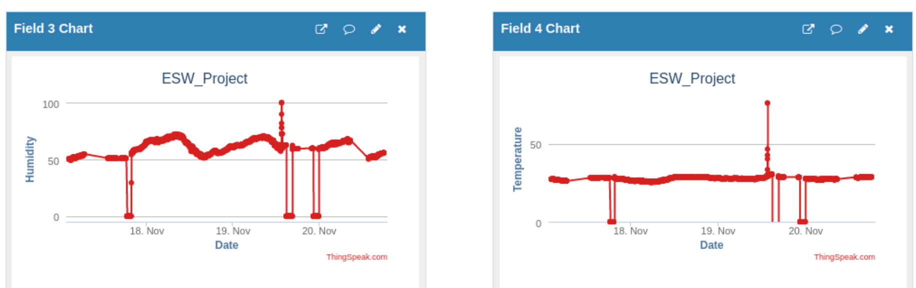

Figure 6: Indoor Temperature and Humidity Levels Over Time

Figure 7: Outdoor PM2.5 and PM10 Levels Over Time

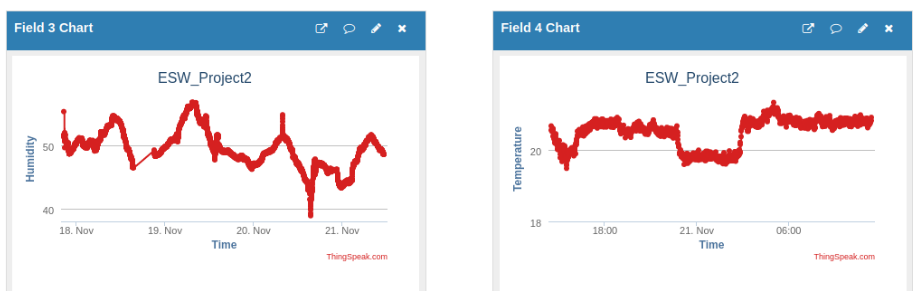

Figure 8: Outdoor Temperature and Humidity Levels Over Time

5 Data Preprocessing

1. **Duplicate Removal:** Entries with duplicate timestamps are removed to maintain data integrity.

2. **Outlier Removal:** Data points with relative humidity (RH) greater than or equal to 90 are filtered out to eliminate outliers.

3. **Interpolation:** Missing values in the dataset are interpolated to fill gaps and ensure a continuous time series.

4. **Irrelevant Data Removal:** Entries outside the interquartile range are excluded to focus on the most relevant data points.

5. **Retime:** The dataset is re-sampled to reduce volume and enhance efficiency in subsequent analyses.

6. **Moving Average:** A moving average is applied to maintain data consistency and smooth variations.

The preprocessed data is now refined and ready for in-depth analysis.


```
{
    "channel": {
        "id": 2261084,
        "name": "ESW_Project",
        "description": "Sending SDS011 PM Sensor Data",
        "latitude": "0.0",
        "longitude": "0.0",
        "field1": "PM 2.5",
        "field2": "PM 10",
        "field3": "Humidity",
        "field4": "Temperature",
        "created_at": "2023-09-05T09:27:59Z",
        "updated_at": "2023-09-26T09:57:41Z",
        "last_entry_id": 5458
    },
    "feeds": [
        {
            "created_at": "2023-10-03T04:43:07Z",
            "entry_id": 5359,
            "field1": "27.20",
            "field2": "60.00",
            "field3": "74.32",
            "field4": "27.28"
        },
        {
            "created_at": "2023-10-03T04:43:12Z",
            "entry_id": 5360,
            "field1": "28.20",
            "field2": "58.60",
            "field3": "74.87",
            "field4": "27.15"
        },
        {
            "created_at": "2023-10-03T04:43:17Z",
            "entry_id": 5361,
            "field1": "28.80",
            "field2": "57.80",
            "field3": "75.69",
            "field4": "27.07"
        },
        {
            "created_at": "2023-10-03T04:43:23Z",
            "entry_id": 5362,
            "field1": "28.70",
            "field2": "58.30",
            "field3": "76.05",
            "field4": "26.93"
        },
        {
            "created_at": "2023-10-03T04:43:27Z",
            "entry_id": 5363,
            "field1": "29.10",
            "field2": "61.00",
            "field3": "76.35",
            "field4": "26.85"
        },
```


Figure 9: Fetched Data

6 Data Calibration

The sensor data, collected from our database, underwent a preprocessing phase to eliminate outliers and ensure data consistency. This preprocessed data, along with a standard dataset, was employed for sensor calibration. Both datasets were aligned based on timestamps.

A linear relation between the local data (x) and the standard dataset (y) was established using a simple linear regression model: $y = mx + c$, where y

Date Time	Monitor ID	Location ID	pm_10	pm_25
13 Oct 2023 22:44	2	1	0.129	0.064
13 Oct 2023 22:45	2	1	0.127	0.064
13 Oct 2023 22:46	2	1	0.139	0.067
13 Oct 2023 22:47	2	1	0.13	0.063
13 Oct 2023 22:48	2	1	0.138	0.067
13 Oct 2023 22:49	2	1	0.13	0.065
13 Oct 2023 22:50	2	1	0.13	0.064
13 Oct 2023 22:51	2	1	0.13	0.065
13 Oct 2023 22:52	2	1	0.134	0.066
13 Oct 2023 22:53	2	1	0.134	0.067
13 Oct 2023 22:54	2	1	0.142	0.069
13 Oct 2023 22:55	2	1	0.132	0.062
13 Oct 2023 22:56	2	1	0.132	0.066
13 Oct 2023 22:57	2	1	0.128	0.064
13 Oct 2023 22:58	2	1	0.127	0.066
13 Oct 2023 22:59	2	1	0.135	0.066
13 Oct 2023 23:00	2	1	0.13	0.066
13 Oct 2023 23:01	2	1	0.136	0.067
13 Oct 2023 23:02	2	1	0.137	0.067
13 Oct 2023 23:03	2	1	0.127	0.064
13 Oct 2023 23:04	2	1	0.138	0.068
13 Oct 2023 23:05	2	1	0.135	0.067
13 Oct 2023 23:06	2	1	0.141	0.067
13 Oct 2023 23:07	2	1	0.136	0.067
13 Oct 2023 23:08	2	1	0.132	0.067
13 Oct 2023 23:09	2	1	0.14	0.067
13 Oct 2023 23:10	2	1	0.128	0.064
13 Oct 2023 23:11	2	1	0.128	0.067
13 Oct 2023 23:12	2	1	0.138	0.066

Figure 10: Standard Data

represents the standard dataset, x denotes the local data, m is the scaling factor, and c is the offset.

This calibration process ensures a robust and accurate mapping between the local sensor readings and the standardized dataset, enabling precise interpretation and application of the Aero PM sensor data.

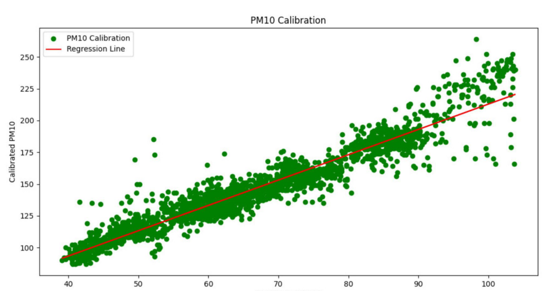

Figure 11: Regression Line PM10

- Scaling Factor (PM2.5): 1.0256618336813879
- Offset (PM2.5): 11.048211689089669
- Root Mean Square Error (RMSE) for PM2.5: 2.7046325971438927
- Scaling Factor (PM10): 1.9935275505072367

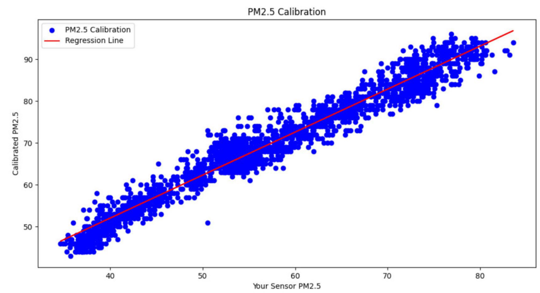

Figure 12: Regression Line PM2.5

- Offset (PM10): 13.568800635884315
- RMSE (PM10): 9.665478354202433

Our collected data points were plugged into the linear regression equation $y = mx + c$, derived from the calibration process, resulting in legitimate data for further analysis.

The coefficient of determination (R-squared) is a crucial metric for evaluating the calibration performance:

- R-squared (PM10): 0.9098807762004294
- R-squared (PM2.5): 0.9547599726165852

An R-squared value above 0.7 is generally considered a strong fit, indicating that the calibration explains a significant portion of the variance in the data. In our case, both R-squared values are well above 0.7, with the R-squared for PM2.5 reaching 0.9547599726165852, demonstrating that the SDS011 sensor is effectively calibrated to provide accurate readings.

This high R-squared value signifies that the calibration process has successfully captured and explained the relationship between the local sensor readings and the standard dataset, affirming the reliability of the sensor's measurements.

Coefficient of Variability Coefficient of variation is a relative measure of dispersion that is used to determine the variability of data.

$$CV = \frac{\sigma}{\mu}$$

where σ is the standard deviation and μ is the mean. This formula provides a measure of the relative variability in the data.

- Coefficient of Variation (CV) for PM2.5 data: 23.95%
- Coefficient of Variation (CV) for PM10 data: 24.78%
- Coefficient of Variation (CV) for Humidity data: 11.78%
- Coefficient of Variation (CV) for Temperature data: 12.67%

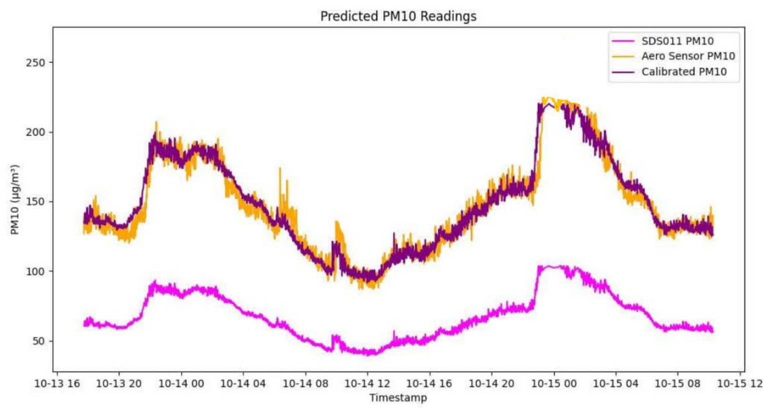

Figure 13: Regression Line PM10

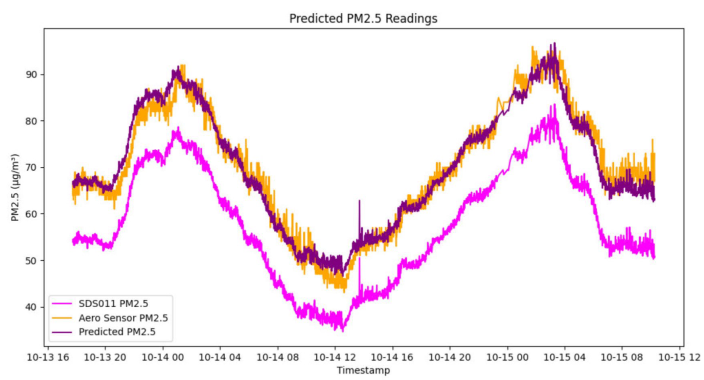

Figure 14: Regression Line PM2.5

7 Analysis

7.1 Correlation Between Indoor and Outdoor

The Pearson correlation coefficient (r) between two variables X and Y can be calculated using the following formula:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

where X_i and Y_i are individual data points, and $\bar{X}$ and $\bar{Y}$ are the means of the variables.

A heat map was generated to compare indoor and outdoor PM2.5 and PM10 levels. The indoor data was collected from a room with closed windows, providing insights into air quality within confined spaces.

The heat map visually represents the spatial distribution of particulate matter levels, allowing for a quick and intuitive comparison. Darker areas on the map indicate higher concentrations, enabling the identification of potential pollution sources and variations between indoor and outdoor environments.

This analysis contributes to a comprehensive understanding of air quality dynamics, particularly within enclosed spaces, and aids in formulating strategies to improve indoor air quality.

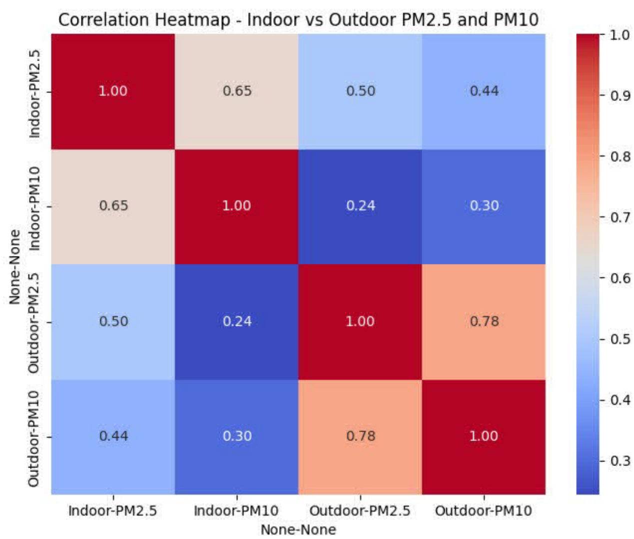

Figure 15: Heat Map

7.2 A Day in an Average Indian Household

7.2.1 Temperature and Humidity Patterns in an Indian Household

An analysis of the graph illustrates distinctive patterns of temperature and humidity fluctuations throughout a day in an average Indian household:

- **Morning Wake Up:** Characterized by a temperature drop and a rise in humidity.
- **Daytime and Cleaning:** Exhibits rising temperature and declining humidity, indicating increased activity.
- **Pooja and Cooking:** Results in a temperature hike with minimal humidity change, reflecting heat-producing activities.

Figure 16: Simulation of Puja using Incense

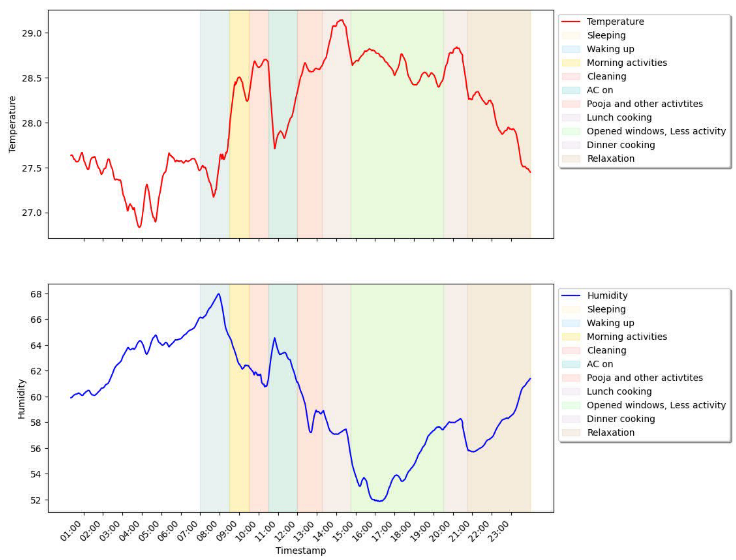

Figure 17: Temperature and Humidity over 24 hours

- **Opening Windows:** Leads to decreased humidity and a maintained temperature, indicative of ventilation effects.
- **Nighttime:** Shows a temperature decrease and a slight humidity rise as the household prepares for sleep.

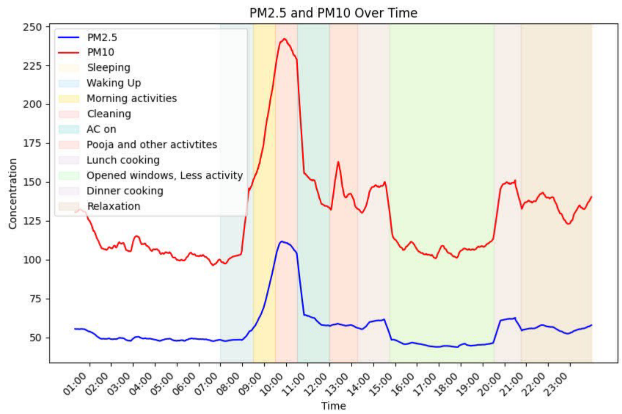

Figure 18: PM2.5 and PM10 over 24 hours

7.2.2 PM10 and PM2.5 Levels Throughout the Day

An examination of the PM10 and PM2.5 graph reveals distinct patterns in pollution levels during various activities within a household:

- **During Sleep:** Minimal pollution is observed.
- **Morning Activities and Cleaning:** Pollution levels gradually rise, with a notable spike during cleaning. This increase is likely due to the disturbance of settled dust, use of cleaning agents, and mechanical actions involved in the cleaning process.
- **AC Usage:** Maintains low PM10 and PM2.5 readings, suggesting effective air filtration.
- **Puja and Cooking:** Shows a slight increase in PM10, with a similar trend during dinner cooking, likely associated with cooking emissions. The burning of incense sticks contributes to an escalation in PM2.5 and PM10 levels, influencing air quality during their combustion.
- **During Less Activity:** Pollution levels decrease.
- **Relaxation at Night:** Moderate increase in pollution levels, indicating a shift in indoor activities.

These observed patterns provide insights into the daily activities and environmental conditions within an Indian household, contributing to a better understanding of the indoor pollution dynamics.

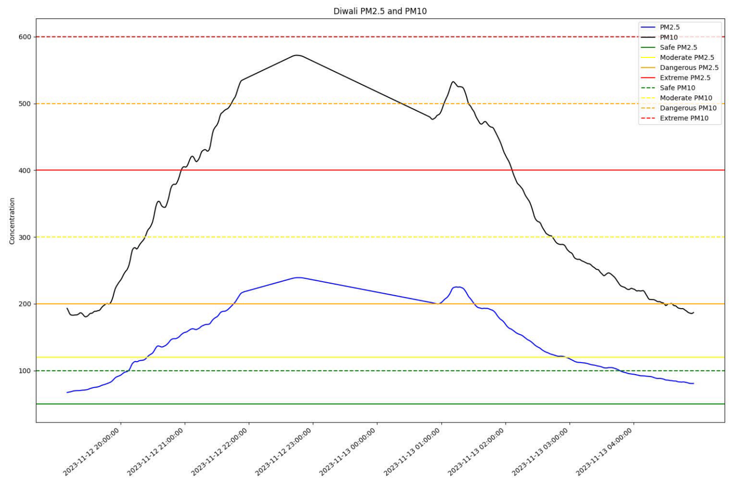

Figure 19: PM2.5 and PM10 on Diwali

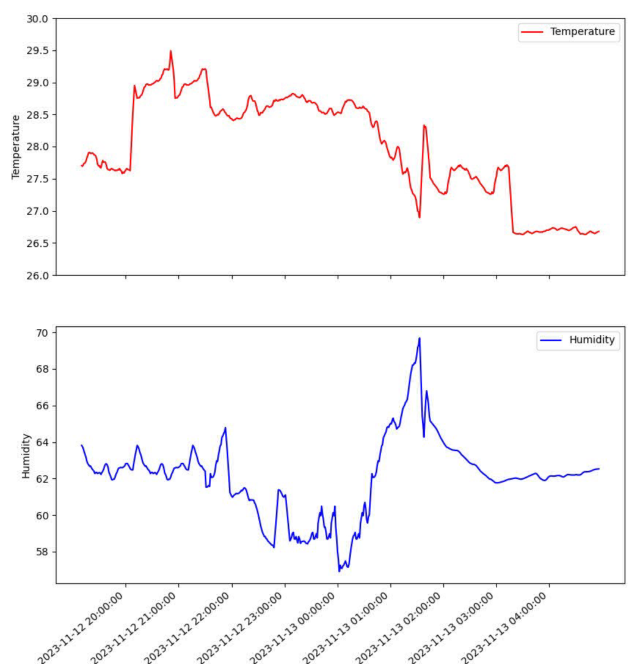

Figure 20: Temperature and Humidity on Diwali

7.3 Diwali

During Diwali, the traditional festival of lights involving the bursting of crackers, a noticeable surge in pollution levels was observed. The festive activities significantly contributed to higher concentrations of particulate matter, with PM2.5 registering at 231.57 and PM10 at 579.73.

This increase in pollution levels was accompanied by a temperature increase to 27.99°C. These changes in environmental conditions during Diwali highlight the impact of celebratory practices on air quality, emphasizing the need for environmentally conscious approaches during festive periods.

8 Dashboard

The dashboard designed for our project encompasses three main components:

8.1 Home

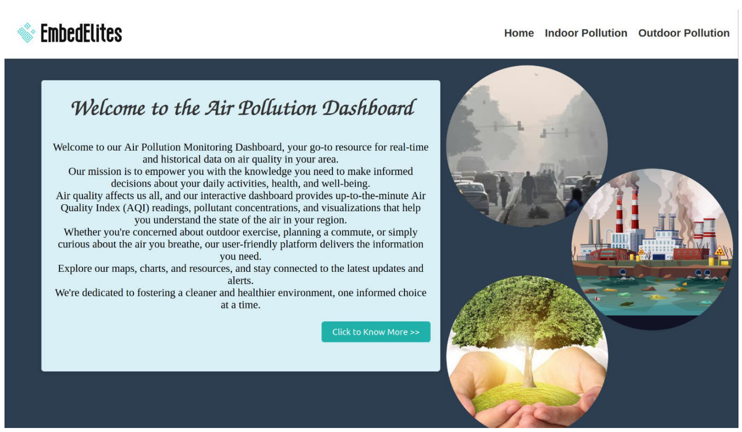

Figure 21: Dashboard Home

The home section serves as the central navigation point for accessing other components of the dashboard. Users can seamlessly navigate to different sections of the dashboard.

8.2 Realtime

Here, users can access up-to-the-minute data on PM2.5, PM10, temperature, and humidity. This real-time information is collected directly from sensors, published to ThingSpeak, and promptly displayed on the dashboard.

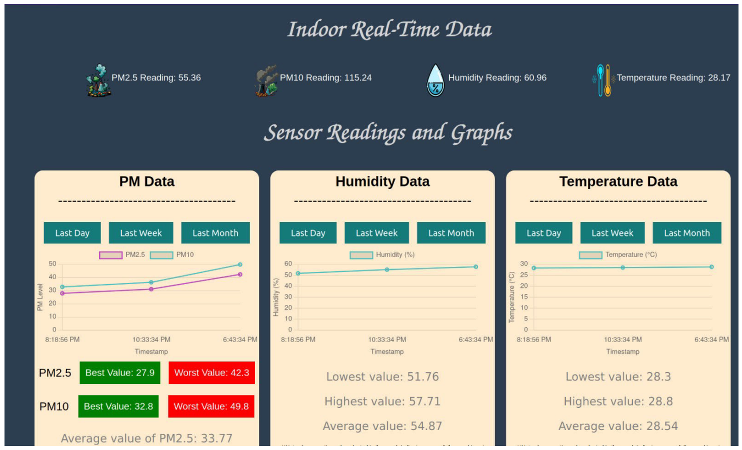

Figure 22: Dashboard RealTime

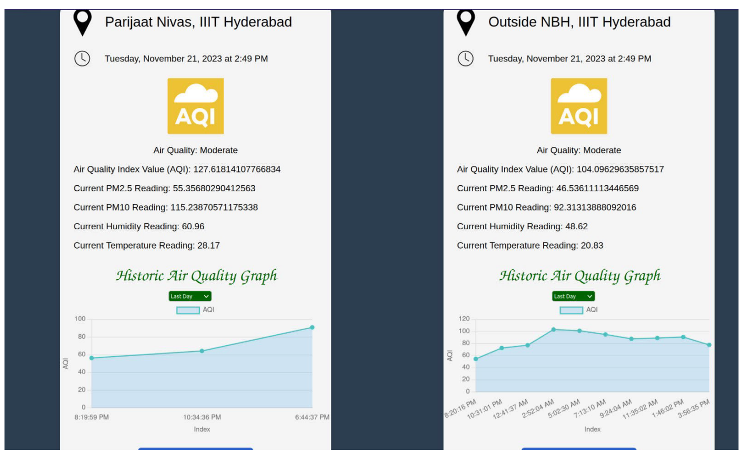

Figure 23: Dashboard Historical

8.3 Historical Analysis

This section presents data retrieved from ThingSpeak, showcasing trends over the last day, last week, and last month. This historical data offers valuable insights into air quality variations, aiding in a more comprehensive analysis.

In summary, the dashboard provides a user-friendly interface, allowing users to explore detailed project descriptions, view current sensor readings for both indoor and outdoor air quality, and analyze historical trends for informed decision-making.

8.4 Calculation of AQI

For PM2.5:

$$AQI_{PM2.5} = \frac{AQI_{high} - AQI_{low}}{Conc_{high} - Conc_{low}} \times (Conc_{PM2.5} - Conc_{low}) + AQI_{low}$$

For PM10:

$$AQI_{PM10} = \frac{AQI_{high} - AQI_{low}}{Conc_{high} - Conc_{low}} \times (Conc_{PM10} - Conc_{low}) + AQI_{low}$$

Where:

- AQI_{high} and AQI_{low} are the AQI breakpoints for PM2.5 and PM10, respectively.
- $Conc_{high}$ and $Conc_{low}$ are the corresponding concentration breakpoints for PM2.5 and PM10, respectively.
- $Conc_{PM2.5}$ and $Conc_{PM10}$ are the actual measured concentrations of PM2.5 and PM10, respectively.

Figure 24: AQI

9 Challenges Faced

Outdoor sensor deployment on Friday limited data points for initial analysis.

Frequent disconnection due to power shortages and network issues results in missing data.

10 Conclusion

Data collection done by us during Diwali highlighted the significant rise in pollution levels. This could be captured even by sensors deployed inside our room which is alarming. The PM2.5 and PM10 values shot to 231.57 and 579.73 respectively which is highly unsafe according to health standards.

For indoor pollution, we performed several activities like cooking, dusting, etc over the span of a day to simulate an average Indian household. The data collected showed variations in PM2.5 and PM10 and other parameters associated with these activities indicating extent to which even indoor pollution can impact our health.

Combining the results of indoor and outdoor data collection, we concluded that the correlation factor of a room not very well ventilated to outdoors is less than the correlation factor a well-ventilated room would achieve.

11 Appendix: Code Repositories

11.1 Sensor Implementation

Sensor Data Collection: https://github.com/ArchishaPanda/ESW-Air-Pollution/blob/main/ESQ_http.ino

11.2 Data Analysis

1. Data Preprocessing: https://github.com/ArchishaPanda/ESW-Air-Pollution/blob/main/calibration/pre.py
2. Calibration: https://github.com/ArchishaPanda/ESW-Air-Pollution/blob/main/calibration/calibr_new.py
3. Plots:https://github.com/ArchishaPanda/ESW-Air-Pollution/blob/main/Analysis/plot.py

11.3 UI

To run, download zip and open home.html on browser

UI: https://github.com/ArchishaPanda/ESW-Air-Pollution/tree/main/ui

12 References

1. https://www.sciencedirect.com/science/article/pii/S2210670719320141
2. https://circuitdigest.com/microcontroller-projects/air-quality-analyzer-using-arduino-
3. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health