VIEH**G**ROUP

High Level Design (HLD)
Air Quality Index Prediction

Revision Number: 2.0

Last date of revision: 31 January 2022

Rajat Sharma

Akshay Lokhande

Aniket Nimaje

Document Version Control

Date Issued	Version	Description	Author
18/01/2022	1.0	Added Introduction and General description	Akshay Lokhande
21/01/2022	1.1	Added Design details and abstract	Rajat Sharma
22/01/2022	1.2	Created and organized whole document	Rajat Sharma
25/01/2022	1.3	Updated Design Details	Aniket Nimaje
27/01/2022	1.4	Checked and added Performance	Rajat Sharma
28/01/2022	1.5	Added Conclusion and References	Akshay Lokhande
31/01/2022	2.0	Updated diagrams in Design Details	Aniket Nimaje

Table of Content

Abstract

Pollution is major problem in 21st century. There is a lot of work going on to reduce the pollution. Air pollution is important issue in all over the world. In India it become worst in winter season in cities like Delhi and Gurugram. Air quality for a human being and all the living species is a vital part of life. Variation of air quality creates a high impact on human life; continuous monitoring and evaluation and investigation of AIR quality are critical. By using various factors air quality is calculated i.e., Air quality Index (AQI) mathematically. But it's also possible to calculate AQI by using machine learning technique.

Introduction

1. Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project

2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should be understandable to the administrators of the system.

3. Definition

The terms used in the projects are:

- AQI - Air Quality Index
- PM2.5 - Particulate Matter 2.5
- NO2 - Nitrogen Dioxide
- CO - Carbon Mono oxide
- SO2 - Sulphur Dioxide
- O3 - Ozone

General Description

1. Product Perspective

Air quality prediction webapp is machine learning based regression model which helps to determine air quality index of area. It can help to instruct the people about air quality so that they take precaution and can also identify the quality of the air present around them.

2. Problem Statement

To create the machine learning based solution to predict air quality index based on the parameters.

3. Problem Solution

Develop the web application to predict the air quality index and quality of air, which can help citizen to decide whether to go out or not and to alert the citizen in particular area if AQI is relatively high.

4. Further Improvement

The project can be extended by using smart IoT devices to make it portable and installing it near cities to track the AQI in real time and to alert people to weak mask if the pollution is high or quality of air is becoming worse.

5. Data Required

For training the model we need the data that consist of PM2.5, NO2, CO, SO2, O3 and AQI value of different cities with timestamp.

Data is completely depending upon our problem statement

6. Tools Used

- Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn are used to build the whole model.
- PyCharm and Visual Studio Code is used as IDE.
- For visualization of the plots, Matplotlib and Seaborn are used.
- Heroku is used for deployment of the model.
- Front end development is done using Stream lit.
- GitHub is used as version control system.

7. Constraints

The AQI prediction website should be user friendly. Different model to be created for different cities.

8. Assumptions

The main objective of the project is to ingest live data from IoT devices and AQI monitoring parameters for current air based on the training data used by using Machine Learning. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.
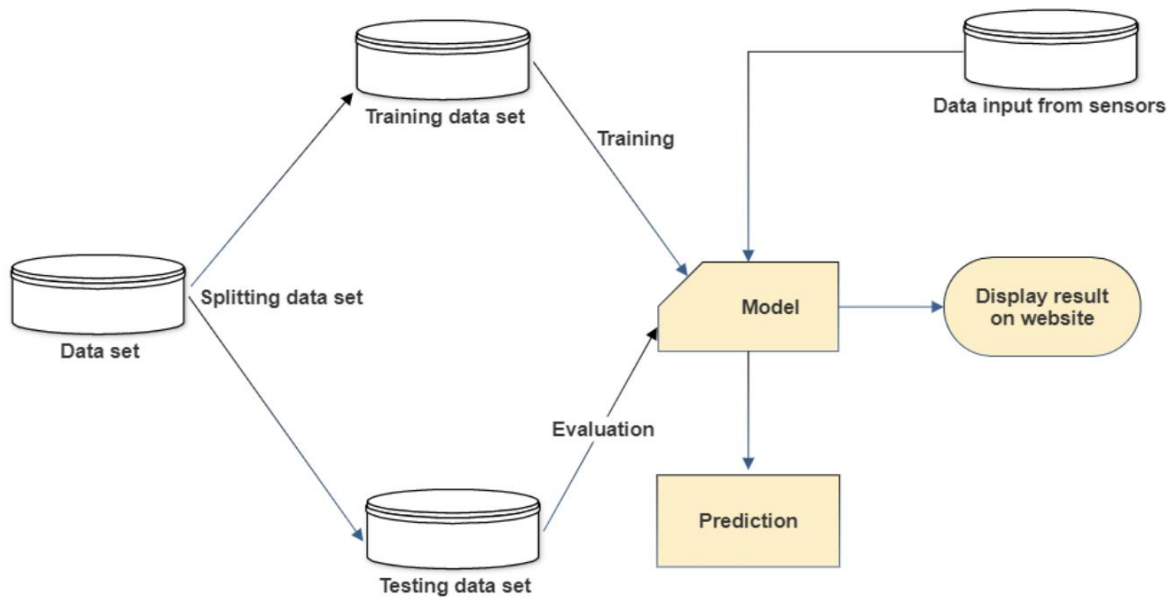
Design Details
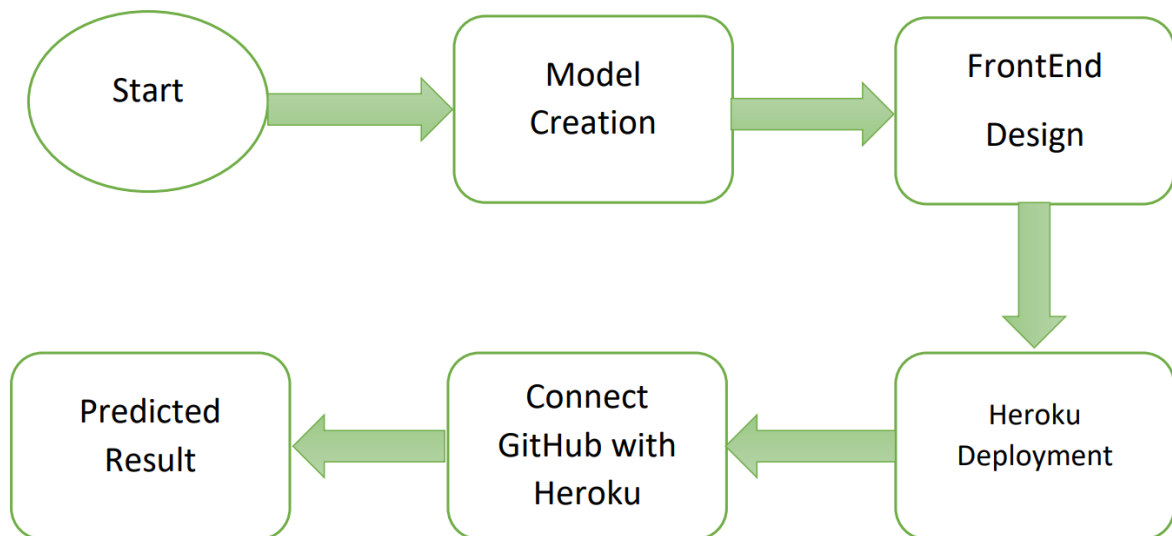
1. Process Workflow

For identifying the different types of anomalies, we will use a machine learning model. Below is the process flow diagram.
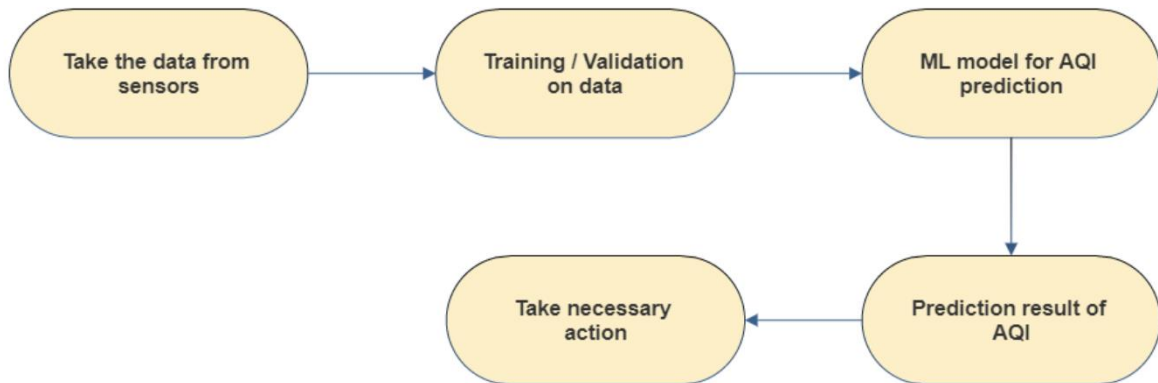
Model Training and Evaluation

Deployment Process

```
  ┌─────────┐      ┌──────────┐      ┌──────────┐
  │  Start  │ ───▶ │  Model   │ ───▶ │ FrontEnd │
  │         │      │ Creation │      │  Design  │
  └─────────┘      └──────────┘      └──────────┘
                                           │
                                           ▼
  ┌──────────┐     ┌──────────┐      ┌──────────┐
  │ Predicted│ ◀── │ Connect  │ ◀──  │  Heroku  │
  │  Result  │     │ GitHub   │      │Deployment│
  │          │     │ with     │      │          │
  │          │     │ Heroku   │      │          │
  └──────────┘     └──────────┘      └──────────┘
```

2. Error Handling

Initially we got error when connecting Heroku with GitHub which we solved. We then got error displaying the graphs and multiple text were not solving. All of the above errors have been resolved.

Performance

1. Reusability

AQI Prediction should be as accurate as possible, so that it will not mislead the user. Best possible model will be used to predict AQI. Since we have used Stream lit and documentation is available, our project follows reusability.

2. Application compatibility

Since we are using python and it is compatible with any platform, we follow Application compatibility

3. Resource utilization

At the initial stage, we were using high space to create the model. Once the model is created, our system only needs at least of 2GB RAM and 1 GB of storage to run the application smoothly. Whenever user tries to predict the AQI, system uses less than 10% of the processing power.

4. Deployment

The code is deployed in GitHub. The whole system is live and is hosted on Heroku.

Conclusion

This project proposes the machine learning model for AQI prediction. This model can be used for alerting the citizens when there is air quality high. It will be helpful to reduce disease and problems occur by air pollution. It also helps the government to take an action in that area against air pollution.

References

1. https://docs.streamlit.io/en/stable/
2. https://scikit-learn.org/stable/user_guide.html
3. https://numpy.org/doc/
4. https://seaborn.pydata.org/examples/regression_marginals.html
5. https://seaborn.pydata.org/examples/scatterplot_matrix.html
6. https://matplotlib.org/
7. https://pandas.pydata.org/docs/