

# High Level Design (HLD)

## Air Quality Index

## Document Version Control

Date Issued	Version	Description	Author
23/01/2023	1	HLD – V1.0	Diana Laveena DSouza

# Contents

Document Version Control .....	2
Abstract .....	4
1 Introduction .....	5
1.1 Why this High-Level Design Document? .....	5
1.2 Scope .....	5
1.3 Definitions .....	5
2 General Description .....	6
2.1 Productive Perspective .....	6
2.2 Problem Statement .....	6
2.3 Proposed Solution .....	6
2.4 Data Requirements .....	6
2.5 Tools Used .....	6
3 Design Details .....	8
3.1 Process Flow .....	8
3.1.1 Model Training and Evaluation .....	9
3.1.2 Deployment Process .....	10
3.2 Event Log .....	10
3.3 Error Handling .....	11
4 Performance .....	11
4.1 Reusability .....	11
4.2 Application Compatibility .....	11
4.3 Resource Utilization .....	11
4.4 Deployment .....	11
5 Conclusion.....	12

## Abstract

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. Based on that, we are supposed to take an action and control all the activities in every sense. To achieve this purpose, we have to build a model that will predict the quality of air for various reasons. I will be able to generate an alarm for the same to create a regulation and guidelines for industries and the general public. We can try to inform them about various dangerous gas emissions and their impact on human life.

# 1 Introduction

## 1.1 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 before coding and can be used as a reference manual for how the modules interact at a high level.

### The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface is implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
  - ◆ Security
  - ◆ Reliability
  - ◆ Maintainability
  - ◆ Portability
  - ◆ Reusability
  - ◆ Application compatibility
  - ◆ Resource utilization
  - ◆ Serviceability

## 1.2 Scope

The HLD documentation presents the system's structure, including the database architecture, application architecture, application flow and technology architecture. The HLD uses non-technical to mildly technical terms, which should be understandable to the system's administrators.

## 1.3 Definitions

<i>Term</i>	<i>Description</i>
<i>Database</i>	Collection of all the information monitored by this system
<i>IDE</i>	Integrated Development Environment

<i>GCP</i>	Google Cloud Platform
------------	-----------------------

## 2 General Description

### 2.1 Product Perspective

Air Quality Prediction is a machine learning model to predict the Air Quality Index.

### 2.2 Problem Statement

To create a regression model to predict the Air Quality Index of the cities in India.

- To detect the dangerous gas emissions that effect the human health.

### 2.3 Proposed Solution

- The solution proposed here is to build a regression model to detect the dangerous gas emissions that effect the human health.

### 2.4 Data Requirements

Data requirements depend entirely on our problem statement.

- We need a large amount of data.

We use the Cassandra database by inserting records into it and then exporting them to a CSV file for further use.

### 2.5 Tools used

Python programming language and frameworks such as NumPy, Pandas, and Scikit-learn are used to build the whole model.



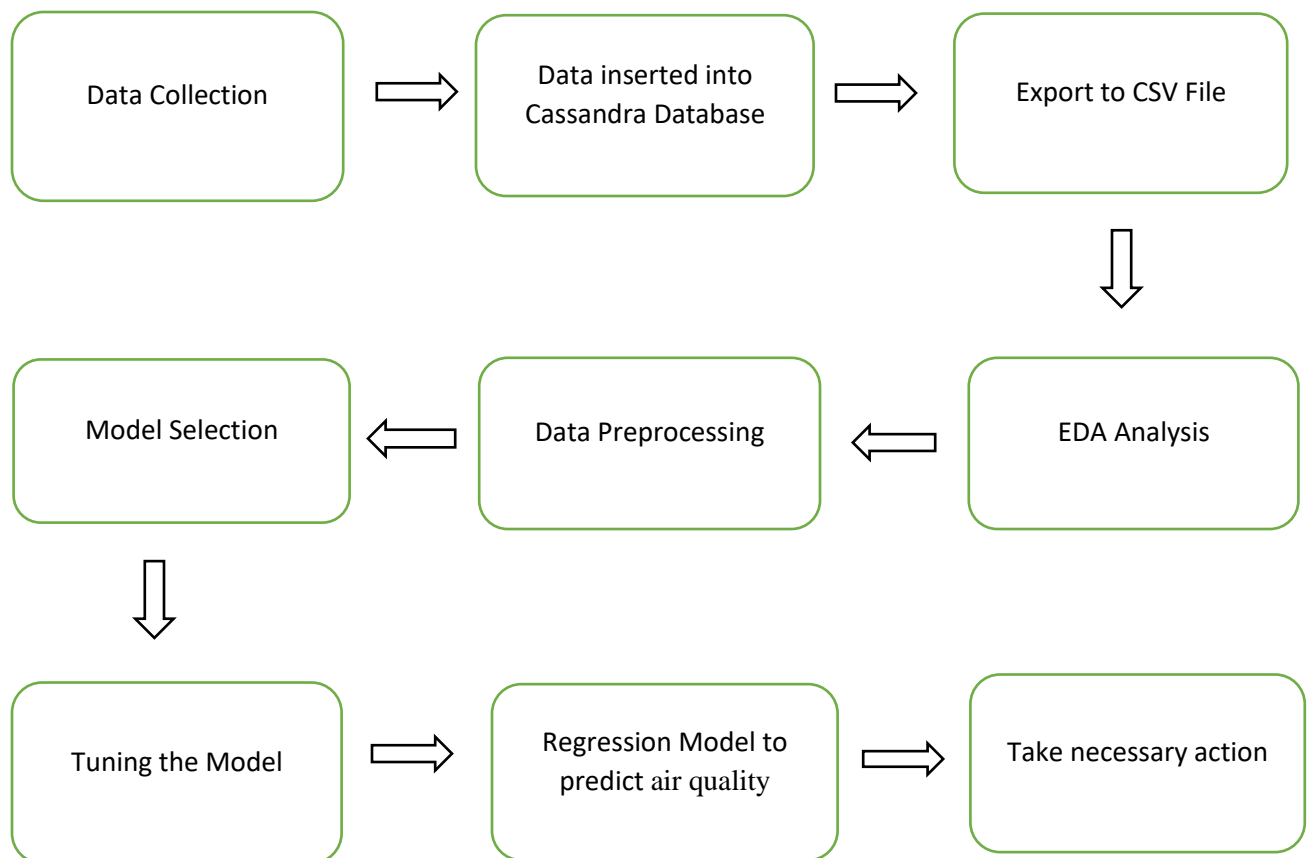
- PyCharm is used as IDE.
- Matplotlib, Seaborn are used for Data Visualization.
- Sci-kit learn, Xgboost are used for model building.
- GCP is used for the deployment of the model.
- Cassandra is used to retrieve, insert, delete and update the database.
- Front-end development is done using HTML.
- Python Flask is used for backend development.
- Imblearn is used for oversampling techniques.
- Github action to build CI/CD pipeline and report generated in DVC for visualization.

## 3 Design Details

### 3.1 Process Flow

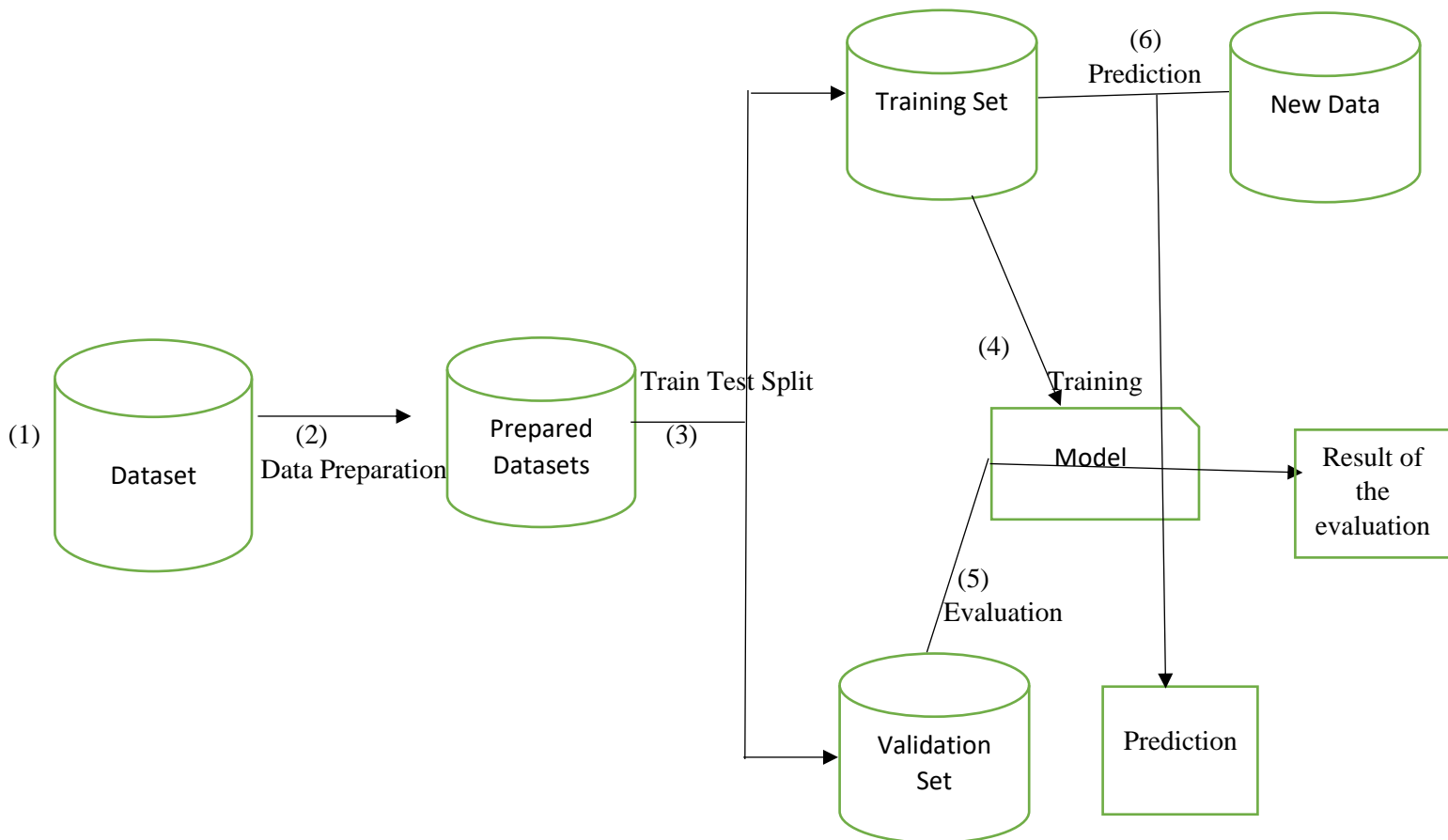
To predict the quality of the air, we will use the regression model. Below is the process flow diagram as shown below.

#### Proposed Methodology

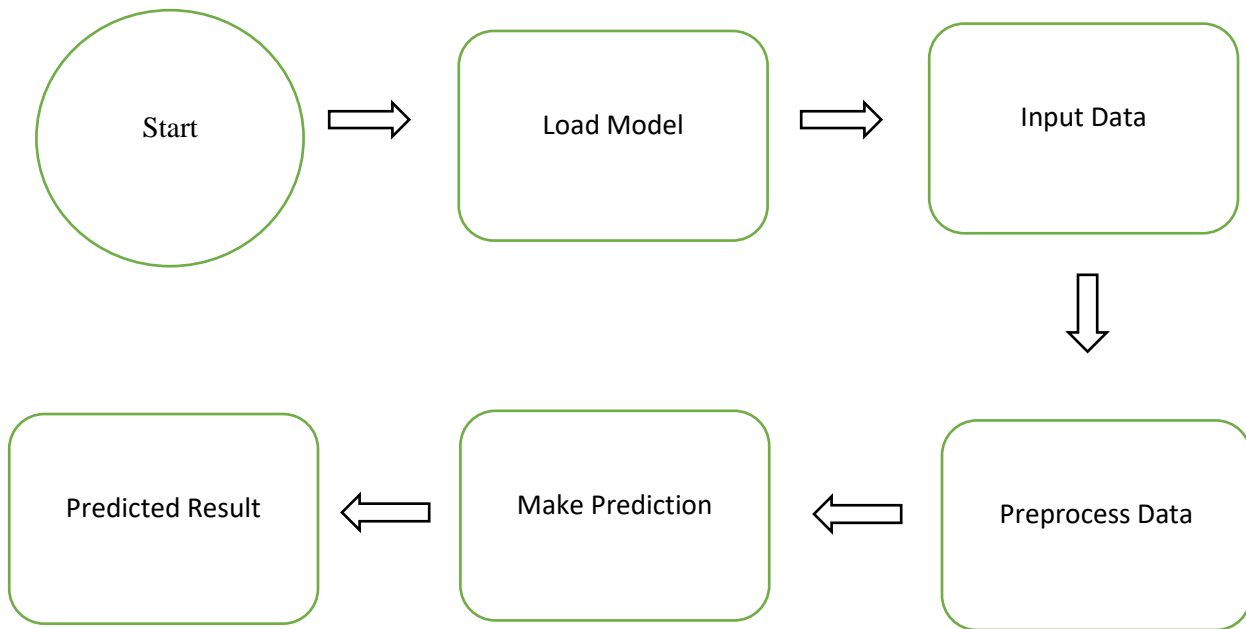




### 3.1.1 Model Training and Evaluation



### 3.1.2 Deployment Process



## 3.2 Event Log

The system should log every event, so the user will know what process is running internally.

#### Initial Step-By-Step Description:

1. The System identifies at what step logging is required.
2. The System should be able to log each system flow.
3. Developer can choose the logging method. You can choose database logging/File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues, so logging is mandatory too.

## **3.3 Error Handling**

Should errors be encountered, an explanation will be displayed as to what went wrong. An error will be defined as anything that falls outside the normal and intended usage.

# **4 Performance**

The Regression Model (Random Forest) is used to detect the quality of the air, which impact on human health.

## **4.1 Reusability**

The code written and the components used should have the ability to be reused with no problems.

## **4.2 Application Compatibility**

The different components for this project will use Python as an interface between them. Each component will have its task to perform, and it is the job of Python to ensure the proper transfer of information.

## **4.3 Resource Utilization**

When any task is performed, it will likely use all the preprocessing power available until that function is finished.

## **4.4 Deployment**



## 5 Conclusion

The Regression Model (Random Forest) is used to detect the quality of the air, which impact on human health.

## 6 References

1. <https://link.springer.com/article/10.1007/s13762-022-04241-5#:~:text=The%20ML%2Dbased%20AQI%20prediction,algorithms%20can%20deal%20with%20efficiently.>
2. <https://www.airnow.gov/aqi/aqi-calculator-concentration/>
3. <https://www.iqair.com/in-en/india/bihar/patna>
4. <https://www.kaggle.com/datasets/debanganthakuria/air-quality-index?select=train.csv>
5. <https://www.kaggle.com/datasets/sakethramanujam/cpcb-air-quality-index-data-2019>
6. <https://www.kaggle.com/datasets/yashpl11/air-quality-dataset>
7. <https://airquality.cpcb.gov.in/ccr/#/caaqm-dashboard-all/caaqm-landing/data>