Student Name:J Saleth Harison

Register Number: 620123106096

Institution: AVS engineering college

Department: Electronics and communication engineering

Date of Submission: 30.04.2025

1.Problem Statement

Air pollution poses significant health and environmental risks worldwide. Traditional methods of monitoring and predicting air quality are often reactive and lack real-time precision. There is a growing need for a proactive system that can accurately predict air quality levels using advanced machine learning techniques to inform public health decisions and environmental policy.

2.Objectives of the Project

To develop a machine learning model that accurately predicts air quality levels

- *(e.g., AQI).*
  To analyze historical and real-time environmental data for pollutant trends.
-
- *To identify key factors contributing to poor air quality using feature importance techniques*
  .
- *To provide visual insights and forecasts through dashboards and reports.*
  To support government and environmental agencies in timely decision-making.
-

- Geographic scope: Focus on urban cities with high pollution levels (e.g., Delhi, Beijing, Los Angeles).
- Pollutants covered: PM2.5, PM10, NO2, SO2, CO, and O3.
- Timeframe: Hourly and daily air quality predictions.
- Use cases: Public health alerts, smart city planning, environmental research, and awareness campaigns.

4.Data Sources

Public datasets:

- Openal
- UCI Machine Learning Repository – Air Quality Data Set
  EPA AirData
- World Air Quality Index Project

Real-time APIs:

- OpenWeatherMap API
- BreezoMeter or AQICN APIs (for real-time AQI)

5.High-Level Methodology

- Data Collection & Preprocessing :
  Gather historical and real-time air quality data.
  Handle missing values, outliers, and normalize the dataset.
- *Exploratory Data Analysis (EDA) :*
  Identify trends, correlations, and key pollutants.
-

Feature Engineering :

- Derive new variables (e.g., weather features, temporal patterns).
- Model Development :
  Use regression (for AQI prediction) or classification (AQI categor y).
- Algorithms: Random Forest, XGBoost, LSTM (for time series), or CNN-LSTM hybrids.
  Model Evaluation :
- Metrics: RMSE, MAE, R2 Score, Accuracy (for classification).
  Cross-validation and hyperparameter tuning.

Visualization and Deployment :

- Build dashboards (e.g., using Dash, Tableau, or Power BI).
  Deploy model via Flask/Django API or cloud platforms (e.g., AWS, GCP).

6.Tools and Technologies

- Languages: Python, R

- Libraries: scikit-learn, pandas, NumPy, matplotlib, seaborn, XGBoost, TensorFlow/Keras
- Frameworks: Flask/Django for deployment
- Cloud: AWS S3, Lambda, EC2 or Google Cloud AI
- Visualization: Power BI, Tableau, or Plotly Dash
- Version Control: GitHub
- APIs: OpenWeatherMap, AQICN, BreezoMeter

7.Team Members and Roles

Saleth Harison J - Project Lead & Data Scientist

Oversees the project, performs data analysis,builds and evaluate ML models.

Thirupathi E - Data Engineer

Collects and preprocess data, integrates APIs, manages storage and data pipelines.

Mourish Kanna V - Full Stack Developer

Develops and deploys the application (backend + frontend), integrates ML models.

Sakthivel D - Visualization & Q A Specialist

Creates dashboards, reports and visual insights; tests model accuracy and app performance