Project : Air Quality Analysis and Prediction in Tamil Nadu

Phase 1: Problem Definition and Design Thinking

Problem Definition:

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The objective is to gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This project involves defining objectives, designing the analysis approach, selecting visualization techniques, and creating a predictive model using Python and relevant libraries.

Design Thinking:

Project Objectives:

1. Analyze Air Quality Trends: Explore historical air quality data to identify patterns and trends related to pollutants such as RSPM/PM10, SO2, and NO2.

2. Identify Pollution Hotspots: Locate areas in Tamil Nadu with consistently high levels of air pollution to prioritize interventions and regulations.

3. Build a Predictive Model: Develop a model using Python and relevant libraries to predict RSPM/PM10 levels based on SO2 and NO2 levels, aiding in proactive air quality management.

Analysis Approach:

1. Data Collection and Preparation:

- Download the air quality dataset from the provided link: [Dataset Link](https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014).

- Load the dataset into a Python environment and perform initial data cleaning, handling missing values, and organizing the data for further analysis.

2. Exploratory Data Analysis (EDA):

- Conduct exploratory analysis to understand the structure of the data, statistical summaries, and distributions of air quality parameters.

- Explore temporal trends and seasonality in air pollution levels.

3. Visualization Selection:

- Utilize visualization techniques such as line charts to demonstrate trends in air quality over time.

- Use heatmaps to visualize spatial variations in pollution levels across different regions in Tamil Nadu.

4. Predictive Modeling:

- Split the dataset into training and testing sets.

- Implement machine learning algorithms (e.g., regression models) to predict RSPM/PM10 levels based on SO2 and NO2 levels.

- Evaluate the model's performance using appropriate metrics like Mean Absolute Error, Mean Squared Error, and R-squared.

5. Documentation and Reporting:

- Document the entire process, including data preprocessing, analysis steps, visualization techniques used, and model development.

- Create a comprehensive report summarizing the findings, insights, and recommendations for effective air quality management in Tamil Nadu.

By following this approach, we aim to achieve the defined objectives and provide valuable insights into air quality trends and predictions to aid environmental decision-making and public health initiatives in Tamil Nadu.