

PROJECT TITLE: AIR QUALITY ANALYSIS IN TN

PROBLEM STATEMENT:

Air pollution is a pressing issue in Tamil Nadu, affecting the health and well-being of its residents. The project aims to tackle this problem by analyzing and visualizing air quality data obtained from monitoring stations across the state. The primary objectives are:

Analyze Air Quality Trends: By examining historical data, we intend to identify patterns, trends, and variations in air quality over time. This will provide insights into the changing environmental conditions.

Identify Pollution Hotspots: Our goal is to pinpoint regions with consistently high pollution levels. This information is crucial for prioritizing interventions and directing resources where they are most needed.

Build a Predictive Model for RSPM/PM10 Levels: We plan to develop a predictive model that estimates RSPM/PM10 levels based on the concentrations of SO₂ and NO₂. This model will aid in forecasting air quality and assessing the impact of potential interventions.

Project Objectives:

Analyze air quality trends: Understand historical changes in air quality to identify long-term patterns and anomalies.

Identify pollution hotspots: Locate regions with consistently high pollution levels to prioritize intervention.

Build a predictive model for RSPM/PM10 levels: Develop a model using SO₂ and NO₂ data to estimate RSPM/PM10 levels.

Analysis Approach:

Data Collection: Gather air quality data from monitoring stations in Tamil Nadu.

Data Preprocessing: Clean and format the data, handle missing values, and ensure data quality.

Exploratory Data Analysis (EDA): Explore data to identify outliers, correlations, and potential patterns.

Data Visualization: Utilize appropriate visualization techniques to communicate findings effectively.

Predictive Modeling: Implement machine learning algorithms to create the RSPM/PM10 prediction model.

Evaluation: Assess model performance using appropriate metrics (e.g., RMSE, MAE).

Visualization Selection:

Line Charts: Use line charts to display trends in air quality over time.

Heatmaps: Create heatmaps to visualize pollution hotspots geographically.

Scatter Plots: Use scatter plots to explore relationships between variables, such as SO₂, NO₂, and RSPM/PM₁₀.

Bar Charts: Present comparisons of air quality levels in different regions.

Tools and Technologies:

Python: The primary programming language for data manipulation and analysis.

Pandas: Used for data manipulation, cleaning, and preprocessing.

Matplotlib and Seaborn: Employed for data visualization and chart creation.

Scikit-Learn: Utilized for implementing machine learning algorithms and model evaluation.

IBM Cognos: Integrated for advanced data analytics and business intelligence capabilities.