

Project Title: Air Quality Assessment in Tamil Nadu

1. Problem Definition: The primary objective of this project is to conduct a comprehensive analysis of air quality data collected from monitoring stations in Tamil Nadu. Our specific goals are:

Objective 1: Air Quality Trends Analysis: We will examine how air quality parameters such as RSPM/PM10, SO₂ and NO₂ have evolved over time across different regions.

Objective 2: Pollution Hotspot Identification – We aim to identify and pinpoint areas or monitoring stations with consistently high pollution levels, known as pollution hotspots.

Objective 3: Predictive modelling – We will develop a predictive model. The model's purpose is to estimate RSPM/PM10 levels based on SO₂ and NO₂ levels, which will enable us to forecast air quality.

2. Design Thinking:

2.1 Analysis Objectives:

To ensure clarity in our project, we have set the following analysis objectives:

Air Quality Trend Analysis: Examine historical air quality data to identify trends and variations in pollution levels over time.

Pollution Hotspot Identification: Determine regions within Tamil Nadu that consistently exhibit high pollution levels.

Predictive modeling: Develop a predictive model that can estimate RSPM/PM10 levels based on SO₂ and NO₂ concentrations.

2.2 The Analysis Approach:

Data collection:

We will begin by acquiring air quality data from the "Location-wise Daily Ambient Air Quality of Tamil Nadu for the year 2014" dataset, thoughtfully provided by the Tamil Nadu government. It is imperative that this dataset includes vital parameters such as RSPM/PM10, SO2, and NO2. We will meticulously validate the reliability and completeness of this dataset.

Data preprocessing:

The next step involves the comprehensive cleaning and preprocessing of the dataset using IBM Cognos. Our focus will be on addressing issues related to missing values, format inconsistencies, and potential outliers. This meticulous data preparation is indispensable to ensuring data integrity and precision for all subsequent analyses.

Exploratory Data Analysis (EDA):

A diverse range of exploratory data analysis (EDA) techniques will be applied during this phase using IBM Cognos. Our objective is to delve deeply into the dataset, identify crucial insights, and uncover trends, patterns, and potential outliers within the air quality data specific to Tamil Nadu for the year 2014. This phase plays a pivotal role in providing us with a comprehensive understanding of the dataset's intricacies and unique characteristics.

Statistical analysis:

Our approach to assessing air quality trends throughout Tamil Nadu for the year 2014 will involve the implementation of advanced statistical tests and analyses using IBM Cognos. We will meticulously explore variations in air quality parameters such as RSPM/PM10, SO2, and NO2 across different regions and monitoring stations within the state. Should pollution hotspots exist, we will rigorously identify and statistically validate them.

Machine learning modeling:

The final step in our project entails the development of a machine learning predictive model, specifically a regression model, using IBM Cognos. The primary objective of this model is to estimate RSPM/PM10 levels based on the levels of SO2 and NO2. A dedicated effort will be put into training and thoroughly evaluating the model's performance using IBM Cognos to ensure its accuracy and reliability in predicting air quality parameters.

By following these steps and leveraging IBM Cognos for analysis, we will be well-equipped to analyze air quality trends, pinpoint pollution hotspots, and create a predictive model tailored to the unique air quality conditions in Tamil Nadu.

Team Members:

1. Hematharshini E – 2021115041

Email - hemae2512@gmail.com

2. Sabitha S – 2021115087

Email - sabitha.suresh15@gmail.com

3. Sandhya Shankar – 2021115090

Email - sandhya.shankar.2002@gmail.com

4. Sanmitha V.S – 2021115092

Email - sanmithasadhishkumar@gmail.com

5. Akash P – 2021115314

Email - akashpanneer2004@gmail.com