AIR QUALITY ANALYSIS AND PREDICTION IN TAMIL NADU

TEAM MEMBERS

1.T Haley Pearl Caxmi - au913121106026

2.S Pritthiga - au913121106071

3.S Sandhya - au913121106089

4.D Haritha - au913121106302

5.R Vijaya Varshini - au913121106117

ABSTRACT:

The goal of this project is to examine and present air quality information gathered from monitoring stations situated in Tamil Nadu. The aim is to comprehend patterns in air pollution, pinpoint regions with elevated pollution levels, and construct a predictive model to estimate RSPM/PM10 levels by considering SO_2 and NO_2 levels. The project encompasses defining specific goals, devising an analytical strategy, choosing appropriate visualization methods, and building a predictive model utilizing Python and suitable libraries.

Design Thinking Steps:

1.Source data

Air quality data is gathered from the data set given. The data contains the gas levels in the atmosphere for various stations.

2.Data Preprocessing

This includes handling of missing values, outliers, and erroneous entries by imputing missing values, removing outliers, or using appropriate statistical methods to ensure data quality. Then, Normalizing or Standardizing the numerical features to bring them to a similar scale, making it easier for models to interpret and learn patterns from the data.

3.Feature Engineering

This process entails generating extra pertinent features to enhance the model's ability to predict outcomes. Within this scope, features related to time, like day of the week, month, or season, can be crafted to encompass patterns and trends in Air Quality data. Such features are instrumental in predicting future outcomes. It creates new features or transform existing ones based on domain knowledge to enhance the predictive power of the model.

4.Model selection

We use Atmospheric dispersion modelling. Atmospheric dispersion modelling is a technique used to simulate and predict the movement and spread of pollutants or substances in the atmosphere. It employs mathematical and computational models to simulate how these substances disperse and interact with the air, taking into account various environmental and meteorological factors.

Atmospheric dispersion modelling and regression for air quality analysis in Tamil Nadu involves utilizing various components. These include:

i. Data: Gathering information on weather conditions such as wind speed, temperature, and humidity to understand their impact on air quality dispersion.

ii. Geographical Data: Incorporating geographic details to account for the region's topography, which can influence how pollutants disperse.

iii. Chemical and Physical Properties: Understanding the properties of pollutants, including their chemical composition and physical characteristics, to model their behavior accurately.

iv. Modeling Algorithms: Employing mathematical and computational algorithms to simulate the dispersion of pollutants in the atmosphere based on the gathered data.

v. Analysis: Utilizing statistical regression techniques to establish relationships between air quality parameters and predictor variables, aiding in prediction and analysis.

5. Model Training:

The chosen time series forecasting model undergoes training using the prepared dataset. Throughout this training phase, the model acquires the ability to identify patterns and correlations within the data, enabling it to forecast future prediction accurately.

6. Evaluation

Evaluating air quality analysis and prediction using Root Mean Square (RMS) and Root Mean Square Error (RMSE) involves assessing the accuracy and precision of the forecasting models. Here’s how it’s typically done:

Root Mean Square (RMS):

Calculation: RMS is the square root of the average of the squared differences between the predicted and actual air quality values.

Formula:

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Value}_i)^2}$$

Interpretation: Lower RMS values indicate better alignment between predicted and actual values, showcasing the proximity of predictions to real-world observations.

Root Mean Square Error (RMSE):

Calculation: RMSE extends RMS by taking the square root of the mean of the squared errors (the differences between predicted and actual values).

Formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Predicted Value}_i - \text{Actual Value}_i)^2}$$

Interpretation: RMSE provides a measure of the average magnitude of the errors, conveying the model's overall accuracy. A lower RMSE signifies a more accurate model.

In air quality analysis and prediction:

i. Compute RMS and RMSE using the predicted air quality values and the corresponding actual measurements.

ii. Compare these metrics across different models or time periods to determine which model provides the most accurate predictions.

iii. A smaller RMS and RMSE indicate a better fit of the model to the observed data and, therefore, a more accurate prediction of air quality RMS and RMSE using the predicted air quality values and the corresponding actual measurements.

These metrics across different models or time periods to determine which model provides the most accurate predictions. A smaller RMS and RMSE indicate a better fit of the model to the observed data and, therefore, a more accurate prediction of air quality.

By leveraging RMS and RMSE, analysts can quantitatively assess and compare the performance of various forecasting models, aiding in the selection of the most effective model for air quality prediction