Air Quality Analysis in Tamil Nadu

Date	26-10-2024
Team ID	719
Project Name	Air Quality Analysis in Tamil Nadu

Table of Contents:

1. Introduction:

Air pollution is a growing concern worldwide, and its adverse effects on human health and the environment are well-documented. In the Indian state of Tamil Nadu, rapid urbanization and industrialization have led to an increase in air pollution levels, raising serious public health and environmental issues. To address this problem, a comprehensive project has been initiated to analyze and visualize air quality data from monitoring stations located throughout Tamil Nadu. The project's primary goal is to gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate Respirable Suspended Particulate Matter (RSPM) and Particulate Matter (PM10) levels based on the concentrations of Sulfur Dioxide (SO2) and Nitrogen Dioxide (NO2).

2. Problem Statement:

The project involves establishing air quality sensors in cities, identifying pollution sources, developing an accessible app, and optimizing transportation routes to enhance air quality, public health, and environmental sustainability.

3. Project Objectives:

- **Data Collection:** Gather air quality data from monitoring stations across Tamil Nadu. This data will include parameters such as RSPM, PM10, SO2, and NO2 levels, along with geographical information.
- **Data Analysis**: Perform exploratory data analysis (EDA) to understand the distribution of air pollutants, detect outliers, and identify trends and patterns.
- **Visualization:** Utilize data visualization techniques to represent air quality data geospatially and temporally. This will help identify pollution hotspots and understand pollution trends over time.
- Identification of High-Pollution Areas: Determine areas with consistently high pollution levels and investigate the factors contributing to this pollution.
- **Predictive Model:** Develop a predictive model, likely using machine learning techniques, to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This model will be valuable for forecasting air quality and identifying areas that require immediate attention.

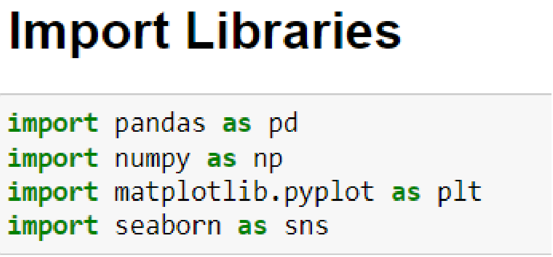
4. Steps Involved in Model Evaluation:

4.1. Data Collection:

- First, ensure you have access to air quality data from monitoring stations in Tamil Nadu. Obtain this data from reliable sources, such as government agencies or environmental organizations. Ensure that the dataset contains the relevant information, including RSPM/PM10, SO2, NO2 levels, and station locations.

4.2 Import Libraries:

- Start by importing the required libraries. In this case, you'll use Pandas for data manipulation.



Import Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

4.3. Load the Dataset:

- This step involves loading your air quality dataset into your Python environment. The dataset should be in a format that Pandas can easily handle, such as a CSV file.

Loading Dataset

```
data = pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')
```

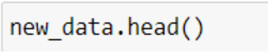
- The read_csv () function is used to load a CSV (Comma-Separated Values) file into a Pandas DataFrame. You specify the file path within the parentheses.
- The result of this operation is a DataFrame, which is a tabular data structure that's similar to a spreadsheet. It allows you to work with your data in a structured and flexible way.

4.4. Explore the Dataset:

Before diving into data preprocessing, it's important to understand your dataset. You can use various Pandas functions to explore it:

data.head ():

- This function displays the first few rows of your dataset, giving you a glimpse of its structure.



```
new_data.head()
```

data.describe ():

- It provides basic statistical information about your data, including measures like mean, standard deviation, and quartiles for numerical columns.

```
new_data.describe()
```

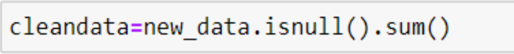
data.columns:

- This helps you see the names of all the columns in your dataset.

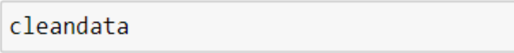
```
new_data.columns
```

data.isnull ().sum ():

- This checks for missing values in each column, showing you how many missing values exist in each.



```
cleandata=new_data.isnull().sum()
```



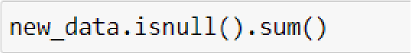
```
cleandata
```

4.5. Data Pre-processing:

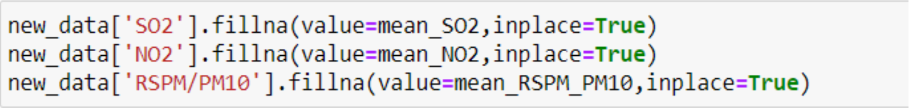
- Data preprocessing is crucial for ensuring the quality and usability of your data:

Handle Missing Values:

- Check for missing values in your dataset and decide on an appropriate strategy to handle them. You can fill missing values using methods like forward-fill, backward-fill, mean, median, or simply remove rows with missing values.



```
new_data.isnull().sum()
```



```
new_data['SO2'].fillna(value=mean_SO2,inplace=True)
new_data['NO2'].fillna(value=mean_NO2,inplace=True)
new_data['RSPM/PM10'].fillna(value=mean_RSPM_PM10,inplace=True)
```

Data Transformation:

- If your dataset contains date or time columns, convert them to the datetime data type for time-based analysis.

“# Example: Convert a date column to datetime
data['Date'] = pd.to_datetime (data['Date'])”

Data Cleaning:

- Inspect your data for inconsistencies, outliers, or irregularities. Ensure that the data is clean and standardized. This may include dealing with irregular units, correcting typos, or removing duplicates.

4.6. Predictive Model training:

- Choose Support Vector Machine (SVM) for regression and classification tasks, handling complex data relationships.
- Split the data into training and testing sets.
- Train the model using preprocessed dataset and target variables.
- Train the model on the training data and evaluate its performance on the test data using relevant metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

5. Conclusion:

Air quality data from Tamil Nadu stations shows a rise in pollution, posing a significant environmental and public health challenge. High pollution areas, characterized by heavy industrial activity, increased vehicular emissions, and population density, are disproportionately affected. Addressing these areas requires targeted interventions, stricter regulations, and public awareness campaigns. A predictive model has been developed to estimate RSPM/PM10 levels, enabling environmental authorities to make informed decisions and improve air quality.