AIR QUALITY ANALYSIS AND PREDICTION

IN TAMILNADU

Problem Statements:-

* Tamil Nadu faces significant challenges related to

air quality ,with various regions

experiencing high levels of pollution due to

industrial activities , vehicular emissions and

natural factors.

* Accurate analysis and timely prediction of air

quality are essential to mitigate the adverse

effects on public health and effect the

environment.

* This innovation document outlines a

comprehensive approach to address the

critical issues of air quality analysis and

prediction in Tamil Nadu.

* Poor air quality poses severe health and

environmental risks,making it imperative to

develop innovative solutions.

* This document presents a multifacted

strategy combining technology,data science,

and community engagement to tackle this

problem.

Innovative Solutions:-

 AirQuality Montoring Solutions

 Machine Learning Models

 Satelliate Data Integration

 Mobile Apps for Citizen Engagement

 Air Quality Index (AQI) Alerts

 Community-based Air Quality Monitoring

 Public Awareness Campaigns

 Early Warning Systems

Dataset Link:-

“https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014”

Dataset Details:-

The Air quality of India has been changing

gradually over a period of time.This is due to

increase in usage of automobiles and invasion

of industries.The usage of automobiles and the

presence of industrious in each state vary's with

respect to each district.So we must analyse how

much pollution does a particular state in India

is producing.To do that we need to analyse the

pollution level of some of the major district's

pollution level,so that the total contribution of

a states pollution can be determined andanalysed clearly .by doing so we can take action

accordingly in that specific area where there is

major air pollution is produced. This kind of

micro level analyse will also helps to predict

and visualize the air quality of all minor parts

of India.

Columns Details:-

Stn Code

State

City/Down/village

Location of Monitoring Station

Agency

Type of Location

SO2

NO2

RSPM/PM10

SPM

**Explanation about column details:**

* **Std Code:STD stands for Subscriber Trunk Dialling.To find STD code of any city from drop down select box.**
* **State:In this dataset we use the state TamilNadu.In That state We find the air quality analysis and prediction.**
* **City/Town/Village:In the dataset we use the city, town,villages like chennai,coimbatore,madurai,salem,thoothukudi.**
* **Location of Montoring Station: The location of montoring station in the dataset is Kathivakkam,Govt.High Secondary School,Thiruvottiyur,Madras Medical,etc...**
* **Agency:The agency we use in the datasets is TamilNadu,Thiruvottiyur Municipal Office, Chennai etc..**
* **Type of Location:In this datasets we use the location likes Industries Area,National Environmental Engineering Research Institute,Tamilnadu State Pollution Control Board etc..**
* **In the air quality analysis and prediction dataset the AQI has been estimated through a method used by US environmental protection agency for different criteria polluntants such as RSPM(Respirable Suspended Particulate Matter),SO2(Sulfur dioxide),NO2(Nitrogen dioxide) and SPM(suspended Particulate Matter).**

Libraries Used :

In this we use a libraries like Numpy,Pandas,Matplotlib,Seaborn and import the packages like warnings and fliters.

Way to install the libraries:

* Make sure Python and pip is preinstalled on your system.
* To check python version: “python --version”
* To check pip:”pip -V”
* Numpy:It is a python library,used to solve numerical problems.It stands for NumericalPython.It can be installed as follows “pip install numpy”.
* Matplotlib:Is a python library that helps to plot graphs.It used in data visualizations and graphical plotting.It can be installed as “pip install matplotlib”
* Pandas:Pandas is a python package that is used for data analysis and manipulation.Is a open source libaries that is built over numpy. It can be installed as “pip install pandas”
* Seaborn:Is a python data visualization library based on matplotlib.It can be installed as “pip install seaborn”.
* Scikit-Learn (sklearn): Provides tools for machine learning tasks like regression (for prediction), classification, and clustering. You can use regression models to predict air quality parameters.
* Statsmodels: For advanced statistical analysis and hypothesis testing, which can be valuable for understanding relationships and trends in air quality data.
* Mapping and Visualization:Folium,A library for creating interactive maps.Plotly, For creating interactive and customizable visualizations.

Train and test :

Now we train and test our dataset as follow:

* Data collection:Gather historical air quality data for various locations in Tamil Nadu. This data should include information on pollutants such as PM2.5, PM10, NO2, SO2, CO, O3, and meteorological data like temperature, humidity, wind speed, and direction.
* Data preprocessing:Clean the data by handling missing values, outliers, and duplicates.Perform data normalization or standardization to ensure that all features have similar scales.Create time-based features, such as time of day, day of the week, and season, which can be relevant for air quality prediction.
* Features Selection:Use feature selection techniques to identify the most relevant features for prediction.Correlation analysis and feature importance from machine learning models can help in feature selection.
* Model selection:Choose a suitable machine learning or deep learning model for air quality prediction. Common choices include linear regression, decision trees, random forests, support vector machines, or neural networks.
* Model trainig:Train the selected model on the training dataset using appropriate algorithms and hyperparameter tuning.Monitor the model's performance on the validation set to prevent overfitting.
* Model Evaluation:Evaluate the model's performance on the test set using appropriate evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or R-squared (R2) score.Use time series-specific evaluation techniques if applicable, such as time-series cross-validation or rolling origin validation.
* Model Deployment:Once you have a satisfactory model, deploy it to make real-time air quality predictions. This could involve creating a web-based dashboard or integrating it into existing monitoring systems.
* Ethical considerations:Be mindful of the ethical considerations surrounding data collection, privacy, and transparency in your

project.

For example:

import numpy as np

import pandas as pd

import matplotlib .pyplot as plt

import seaborn as sns

%matplotlib inline

plt.rcParams[‘figure.figsize’] = (10,7)

import warnings

warnings.filterwarnings(‘ignore’)

import os

Print(os.listdir(“../input))

data=pd.read\_csv(‘../input/india-air-quality-data/data.csv’,encoding =”ISO-8859-1”)

data.fillna(0, inplace = True )

data.head()

Output:

Rest of Explanation:

* Develop a system that sends real-time AQI alerts to residents' smartphones. These alerts can recommend actions such as avoiding outdoor activities, using air purifiers, or using public transport on days with poor air quality.
* Engage local communities in monitoring air quality. Provide them with low-cost air quality monitoring kits and training to collect data in their neighborhoods. This grassroots approach can supplement official monitoring efforts.
* Launch public awareness campaigns to educate residents about the health risks associated with poor air quality. Promote sustainable practices such as reducing vehicle emissions, planting more trees, and reducing waste burning.

Accuracy:

* Improving the accuracy by splitting the data on heavy varations
* From sklearn import the metrics ad mean square error.
* Mean squared error (MSE): MSE is a measure of the average squared difference between the predicted and actual values.
* Mean absolute error (MAE): MAE is a measure of the average absolute difference between the predicted and actual values.
* Median absolute error (MedAE): MedAE is a measure of the median absolute difference between the predicted and actual values.

Conclusion:

Implementing these innovative solutions will require collaboration between government bodies, technology companies, research institutions, and active participation from the community. By combining technological advancements with public engagement, Tamil Nadu can make significant strides in improving its air quality analysis and prediction efforts.