Anna University Regional Campus Coimbatore

Anna University: Chennai-600 025

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

IBM Naan Mudhalvan Phase1 Submission

Title: AIR QUALITY ANALYSIS AND PREDICTION IN TAMILNADU

Name : PRAVEEN S

Register Number: 710021106027

Department :B.E.ECE

Sem/year :V/III

Air Quality Analysis and Prediction in Tamil Nadu

Objective:

The objective of this project is to analyze and visualize air quality data from various monitoring stations in Tamil Nadu. The dataset contains measurement of Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), and Respirable Suspended Particulate Matter 10(RSPM/PM10) levels in different cities, towns, villages and areas. The project aims to gain insights into air pollution trends, identify areas with high pollution trends, identify areas with high pollution levels and cerate a predictive model to estimate RSPM/PM10 levels based on S02 and NO2 levels.

Abstract:

An index which is used to report air quality is called the air quality index (AQI). It measures the impact of air pollution on a person's health over a short period of time. The purpose of the AQI is to educate the public on the negative health effects of local air pollution. Air Pollution implies a great significant on environmental and health challenge and it demands on comprehensive and prediction efforts. This project,**" Air Quality Analysis and Prediction in Tamil Nadu,"** focuses on leveraging data science techniques to access and predict the air quality levels across various regions in Tamil Nadu, India

Data set description and Sample Data:

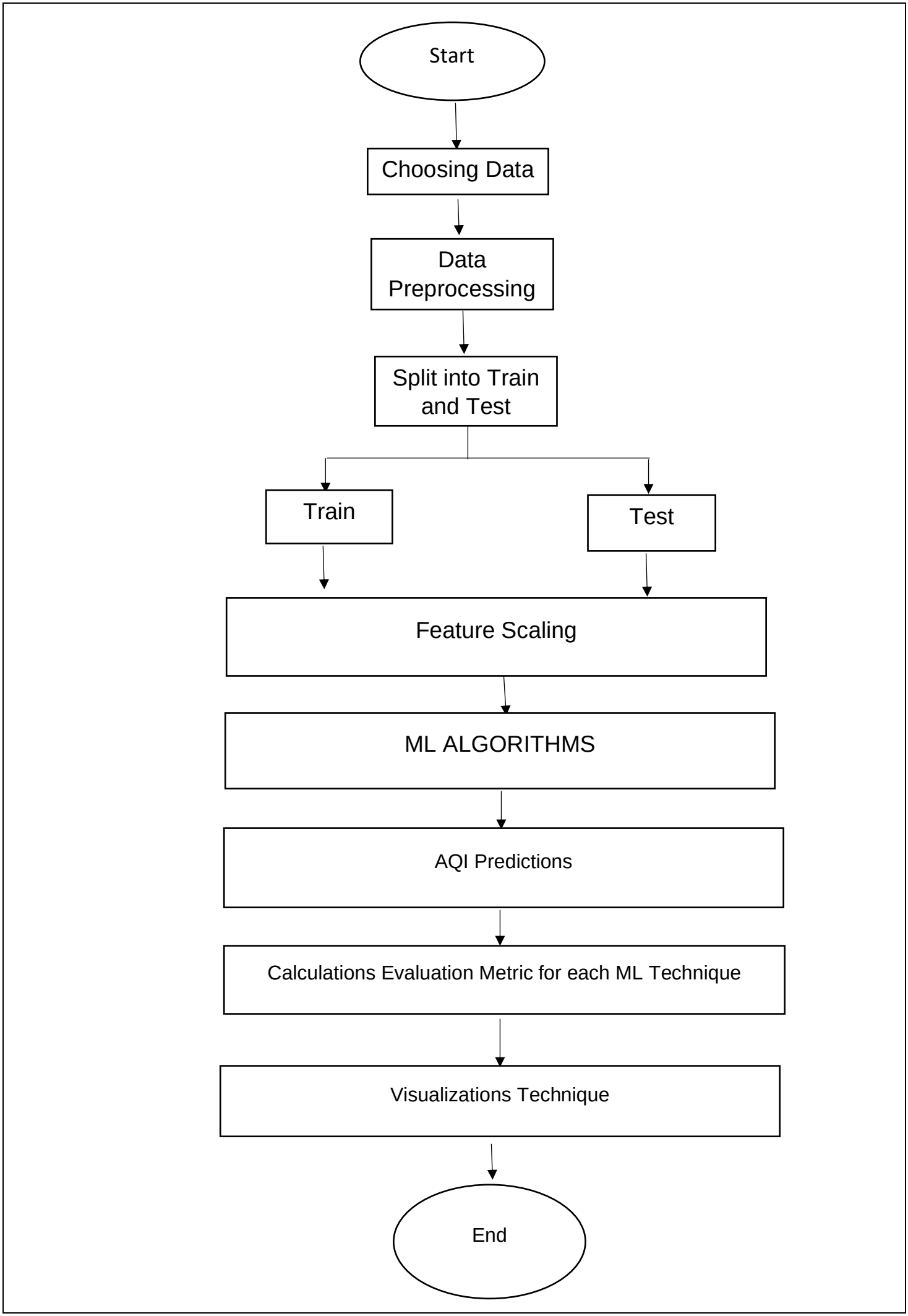
The link to the dataset for this chosen project is given below:

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

The above dataset contains the combined version of air quality of Tamil Nadu from 2014. This contains some district wise data for the prediction of air quality parameter in the state of Tamil Nadu. This data was released by the Ministry of Environment and Forests and Central Pollution Control Board of India under the National Data Sharing and Accessibility Policy (NDSAP).

Methodology:

Flow chart for the proposed system:


Start
Choosing Data
Data Preprocessing
Split into Train and Test
Train
Test
Feature Scaling
ML ALGORITHMS
AQI Predictions
Calculations Evaluation Metric for each ML Technique
Visualizations Technique
End

STEP1: Choosing a dataset

Choosing the proper dataset for implementing the project

STEP2: Data Pre-processing

In data pre-processing we have selected data for the analysis of air quality in the various district of Tamil Nadu. Each of the dataset was cleaned by remove null values of the chosen dataset. Microsoft Excel Software was used to remove unnecessary, irrelevant and erroneous data.

STEP3: Splitting of the dataset

The chosen datasets are split into training and test data. These are used to train the model and then test it against the original data. The values predicted by the machine learning algorithm and to predict accuracy of the data.

STEP4: Training the dataset

Empirical studies show the best results which are obtained if 80% of the data is used for training. Random sampling is used as a way to divide the data into train and test sections. It is widely accepted and is very popular.

STEP5: Testing the dataset

Empirical studies show the best results that are obtained if the remaining 20% of the data is used for testing. Random sampling is used as a way to divide the data into train and test sections. It is widely accepted and is very popular.

STEP6: Feature Scaling

The data should be normalized in order to make the dataset more flexible and more consistent. Standard Scalar from Scikit-Learn Library has been used to do so. It normalizes the features by deleting the mean and scaling the unit variance

STEP7: Applying various Machine learning techniques

After the normalization, we need to apply the various machine learning technique for analysing the data. Some of the machine learning

technique random forest regression, support vector regression which are used to analysis the air quality index.

STEP8: Applying ML technique-random forest regression

Random forest is a supervised machine learning algorithm that is used for classification and regression problems. It creates decision trees from several samples, using the majority vote for classification and the average in the case of regression. A random forest produces precise predictions that are easy to understand. Effective handling of large datasets is possible.

STEP 9: Calculation of evaluation metric for each ML techniques

The metrics used for the proposed work are R-SQUARE, MSE, RMSE, MAE, and the accuracy of various algorithm.

STEP 10: Determine the efficient Visualization techniques

Visualizations play a crucial role in conveying insights from air quality data analysis. Here are some visualization methods and techniques that can be employed in the **"Air Quality Analysis and Prediction in Tamil Nadu"**

- **Time Series Plots**- Plot historical trends of SO2, NO2 and RSPM/PM10 levels over time. Use **line charts** to illustrate daily, monthly or seasonal variations
- **Heatmaps**-Create heatmaps to visualize pollutant concentrations across different monitoring stations and geographical areas.
- **Scatter Plots**-Use scatter plots to explore correlations between air quality parameters.

CONCLUSION:

In conclusion, this project focuses on analyzing and predicting air quality in Tamil Nadu has yielded valuable insights and outcomes. Through the collection and analysis of historical air quality data, we are able to identify trends, seasonal variations, and the impact of various factors or air quality. Our predictive models, based on machine learning algorithms, demonstrated reasonable accuracy in forecasting air quality levels.