Air quality analysis in tamilnadu using machine learning

UNDERSTANDING THE PROBLEM

In Tamil Nadu air pollution is widespread in urban areas where vehicles are the major contributors and in a few other areas with a high concentration of industries and thermal power plants.

Currently 65,052,004 people in Tamil Nadu are breathing toxic air that does not meet WHO's clean air guidelines. The district with the worst air pollution in Tamil Nadu is Chennai , where PM2. 5 is forecasted to be 42.4 µg/m3 .

WHY IS AIR POLLUTION SERIOUS PROBLEM IN ENVIRONMENT

Air pollution can damage crops and trees in a variety of ways. Ground-level ozone can lead to reductions in agricultural crop and commercial forest yields, reduced growth and survivability of tree seedlings, and increased plant susceptibility to disease, pests and other environmental stresses (such as harsh weather).

SIMPLE WAYS TO CONTROL AIR POLLUTION

1

Drive your car less. …

Keep your car in good repair. …

Turn off your engine. …

Don't burn your garbage. …

Limit backyards fire in the city. …

Plant and care for trees. …

Switch to electric or hand-powered lawn equipment. …

Use less energy.

COMMON SOLUTION FOR AIR POLLUTION

The most basic solution for air pollution is to move away from fossil fuels, replacing them with alternative energies like solar, wind and geothermal. Producing clean energy is crucial. But equally important is to reduce our consumption of energy by adopting responsible habits and using more efficient devices.

The environmental protection agency(EPA) tracks the commonly known criteria pollutants, example ground level ozone(O3),sulphur dioxide(SO), particulates matter(PM10) and (PM2.5),carbon monoxide(CO),carbondiaoxide(CO 2),nitrogen diaoxide(NO2).This substances are in compositions of a common index,called the air quality index(AQI),indicating how clean or polluted the air is currently are forecased to become in areas.

As the AQI increases,a higher percentage of population is exposed.

2

2

Recent researchers focus more on advanced statistical learning algorithms for air quality evaluations and air pollution prediction

The following machine learning (ML) algorithms are investigated, i.e., random forest, adaptive boosting (AdaBoost), support vector .

SOME OF AIR QUALITY MACHINE LEARNING PREDICTION METHODS

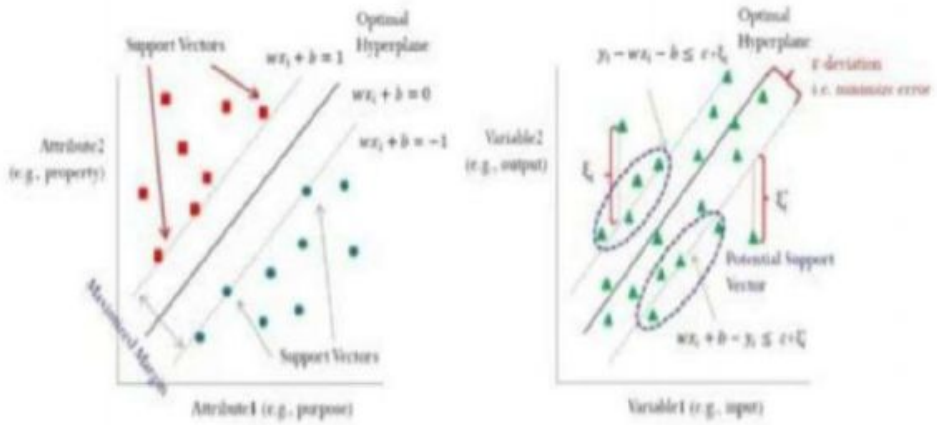
Machine learning involves computational methods which learn from complex data to build various models for prediction, classification, and evaluation. The study attempts to build forecasting models capable of efficient pattern recognition and self-learning. In this section, the underlying principle of five machine learning methods as the canonical procedure will be discussed respectively.

Support Vector Machine

Support vector machine, a supervised learning method for classification, regression, and outlier detection, constructs the hyperplane that acts as a boundary between distinct data points.Two distinctive versions of SVM are shown here. For classification problem , data points that lie at the edge of an area closest to the hyperplanes are considered as support vectors. The space between these two regions is the ma0rgin between the classes. Hyperplanes will determine the number of classes incurred in the dataset and the output of unseen data will be predicted according to which class holds the most similarity with the new data. An approximation of such hyperplane to a non-linear function is constructed at the maximal margin with linear regression.

3

3



Random Forest

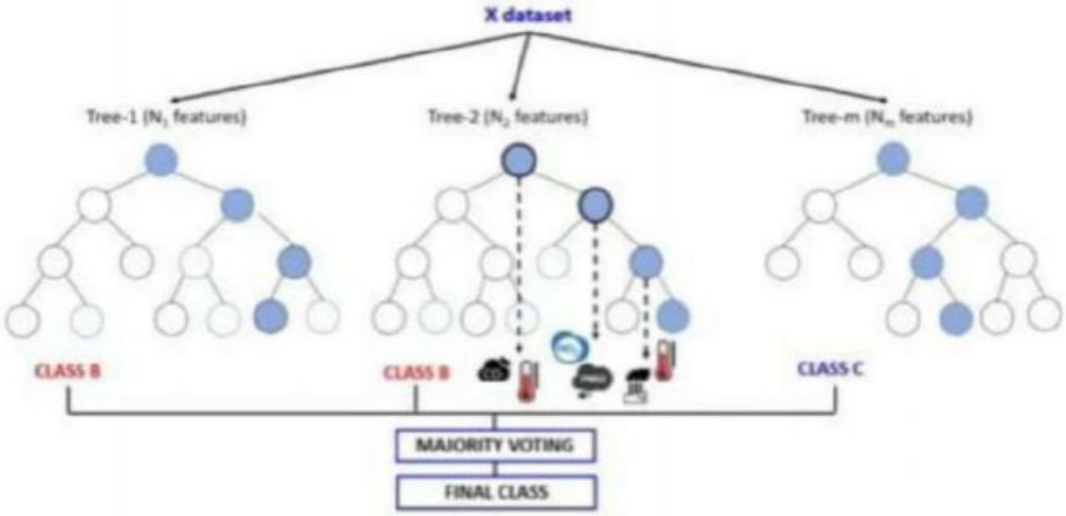
Another prominent machine learning method, random forest, a supervised learning ensemble algorithm, combines multiple decision trees to form a forest and the bagging concept, that latter adds the randomness into the model building.

The random selection of features is used to split the individual tree while the random selection of instances is used to create training data subset for each decision tree.

At each decision node in every tree, the variable from the random number of features is considered for the best split.

If the target attribute is categorical, random forests will choose the most frequent as its prediction. On the other hand, if it's numerical, the average of all predictions will be chosen.

4



Linear Regression

Linear regression is probably the method where most of the academicians started their first machine learning experience. Its main working principle lies behind the fitting of one or more independent variables with the dependent variable into a line in n dimensions. N usually denotes the number of variables within a dataset. This line is supposedly created as it would be minimizing the total errors when trying to fit all the instances into the line. Under machine learning, linear regression is equipped with the capability to learn continuously by optimizing the parameters in the model. Most commonly, optimization is carried out by a method

called gradient descent. It works by partially deriving the loss function and all parameters will be updated by subtracting the previous value with the derivative times a specified learning rate. The learning rate can be tuned by the simplest way, which is rule of thumb (trial and error), or a more

5

sophisticated rule, e.g., meta-heuristic. Another parameter that is left for tuning is the amount of generalization added to the model. Regularization is undergone as an effort to lessen the chance of overfitting and increase the robustness of the model. Two types of regularization used in linear regression are lasso and ridge regression. Lasso regularization will eliminate less important feature by letting the feature's coefficient to zero, and retain another more important one. Ridge regularization on the other hand will not try to eliminate a feature, but instead, tries to shrink the magnitude of coefficients to get a lower variance in the model.

DATA PREPROCESSING

The analysis of the readings begins with a crucial phase – data preprocessing. Various preprocessing operations precede the learning phase. At any particular time, one invalid variable will not affect the whole data group, and thus it will just be either marked blank or, where available.The missing values are treated by imputation to recover the corresponding values. Given the lack of spatial proximity of the readings to the original monitoring stations, the missing values are imputed for relative humidity, temperature, and rainfall, without using wind speed or wind direction. The next imputation process used the k-NN algorithm to substitute the rest of the invalid or missing data that did not qualify for the previous imputation process. Note that the percentage of missing values is lower than 1.3% in all three-station datasets.

Then, input and target data are normalized to eliminate potential biases; thus, variable significance won't be affected by their ranges or their units. All raw data values are normalized to the range of [0, 1]. Inputs with a higher scale than others will tend to dominate the measurement and are consequently

6

given greater priority. Normalization not only improves the model learning rate, but also supports k-NN algorithm performance because the imputation is decided by the distance measure.

MACHINE LEARNING CODING FOR AIR QUALITY ANALYSIS

Here's an example source code for an air quality analysis project using machine learning. This code demonstrates how to use the Random Forest algorithm to predict air quality based on various input features.

```
Import pandas as pd
From sklearn.ensemble import RandomForestRegressor
From sklearn.model_selection import train_test_split
From sklearn.metrics import mean_squared_error

# Load the dataset
Data = pd.read_csv('air_quality_dataset.csv')

# Split the dataset into input features (X) and target variable (y)
X = data.drop('air_quality', axis=1)
Y = data['air_quality']

# Split the data into training and test sets
```


7

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Initialize the Random Forest regressor

Rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)

Train the model

Rf_regressor.fit(X_train, y_train)

Make predictions on the test set

Y_pred = rf_regressor.predict(X_test)

Evaluate the model

Mse = mean_squared_error(y_test, y_pred)

Print('Mean Squared Error:', mse)

In this code, you need to replace 'air_quality_dataset.csv' with the path to your own dataset file. The dataset should contain columns for various input features (e.g., temperature, humidity, wind speed) and a target variable (air quality in this case).

Make sure you have the necessary libraries installed, such as pandas, scikit-learn, and numpy. You can install them using pip:

Pip install pandas scikit-learn numpy

8

This code uses the Random Forest algorithm (RandomForestRegressor) from scikit-learn. It splits the dataset into training and test sets using train_test_split, initializes the Random Forest regressor, trains the model using the training set, and makes predictions on the test set. Finally, it evaluates the model using mean squared error (MSE).

HOTSPOT AREA WHICH CAUSE AIR POLLUTION – COMMON EXAMPLE PLACES IN CHENNAI CONSIDERED BY GIVEN DATASET

38	18-12-14	Tamil Nadu	Chennai	Kathivakkam, Tamilnadu StaIndustrial Area
38	23-12-14	Tamil Nadu	Chennai	Kathivakkam, Tamilnadu StaIndustrial Area
38	30-12-14	Tamil Nadu	Chennai	Kathivakkam, Tamilnadu StaIndustrial Area
71	01-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	01-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	20-01-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	23-01-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	27-01-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	30-01-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	02-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	02-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	02-10-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	13-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	17-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	20-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	24-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	27-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	03-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	03-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	03-10-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	13-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	17-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	20-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	24-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	27-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	04-03-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	04-07-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	04-10-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	17-04-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	21-04-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	28-04-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	05-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	05-08-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	05-12-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	15-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	19-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	22-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	26-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	29-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	06-02-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	06-05-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	06-09-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	06-12-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	16-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	19-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area
71	26-06-14	Tamil Nadu	Chennai	Govt. High SchTamilnadu StaIndustrial Area

9

Stn Code	Sampling Date	State	City/Town/Vil	Location of M	Agency	Type of Locati
38	01-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	01-07-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	21-01-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	23-01-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	28-01-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	30-01-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	02-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	02-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	02-11-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	13-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	18-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	20-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	25-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	27-02-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	03-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	03-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	03-11-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	13-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	18-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	20-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	25-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	27-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	04-01-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	04-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	04-08-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	04-10-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	15-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	17-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	22-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	29-04-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	05-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	05-08-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	13-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	15-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	20-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	27-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	29-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	06-03-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	06-05-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	06-10-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	06-12-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	17-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	19-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	24-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area
38	26-06-14	Tamil Nadu	Chennai	Kathivakkam,	Tamilnadu Sta	Industrial Area

10

10

160	28-01-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	31-01-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	02-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	02-07-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	02-11-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	14-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	18-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	21-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	25-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	28-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	03-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	03-07-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	03-12-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	14-03-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	18-03-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	21-03-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	28-03-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	04-02-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	04-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	04-08-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	04-10-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	04-11-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	16-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	22-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	29-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	05-01-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	05-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	05-09-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	13-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	16-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	20-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	23-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	29-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	30-05-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	06-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	06-07-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	06-10-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	13-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	17-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	20-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	24-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	27-06-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	07-04-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	07-08-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	07-11-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R
160	15-07-14	Tamil Nadu	Chennai	NEERI, CSIR C	National Envir Residential, R

11

11

MODEL TRANING:

The data can be trained by splitting the data into training and testing sets.

INNOVATION:

Consider incorporating advanced techniques like deep learning, natural language processing, or reinforcement learning if they are suitable for specific problem.

IMPLEMENTATION:

Deploy the models into your sales and marketing processes to make real-time predictions and analysis.

USER INTERFACE:

Create a user-friendly interface for stakeholders to interact with the predictions and gain insights.

Interpretation; understand the insights gained from the models and use them to make informed business decisions.

12

12

STACKHOLDER COMMUNICATION:

Keep stackholder informed about performance of the model and any changes or improvement made.

This helps in maintaining alignment between the model and business objectives.

CONCLUSION

Air quality analysis assesses individual air pollutant level in the ambient air. Interpreting Air quality analysis may sound complicated, but when broken down it really is quite simple.

It is concluded that the all the four locations are getting polluted and may cause harmful ill effects to public, students in college and also the environment.

The exponential increase in vehicular usage and fossil fuels still makes this level worse day by day. Necessary steps must be taken in order to mitigate the particulate emissions from various sources, particularly from automobiles, which contribute the major source of particulates.

13