Applied DataScience :

Project Name: Air Quality Analysis and Prediction in Tamilnadu

Project Description: To Develop a Machine learning algorithms like random forest and Using Pandas and Numpy Libraries to predict and calculate the air quality in Tamilnadu.

Phase 3: Development Part 1

Description :

Begin building the Air quality prediction model by loading and preprocessing the dataset.

Load the Air Quality dataset and preprocess the data for analysis.

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

Working Procedure :

To load and preprocess the Air Quality dataset in 2014 from Kaggle, we can use the following steps:

Step 1:

Install the necessary Python libraries

Step 2 :

Load the dataset

# Load the dataset from the Kaggle website

Step 3 :

Explore the dataset

# Print the first 5 rows of the dataset

# Print the basic information about the dataset

Step 4 :

Preprocess the data

Handle missing values: There are no missing values in the dataset.

Convert categorical features to numerical features:

# Define a function to convert categorical features to numerical features

# Encode the Genre feature

# Encode the Language feature

Step 5 :

Scale the numerical features

# Define a function to scale numerical features

Step 6 :

Split the dataset into training and test sets.

Conclusion:

We have now loaded and preprocessed the Air Quality Analysis dataset for analysis. The next step is to build a machine learning model to predict Air quality

Program for an above steps :

In[1] : import pandas as pd

# Load the dataset from the Kaggle website

In [2] : air\_quality= pd.read\_csv(<https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014>)

Out[2] :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stn Code | Sampling Date | State | City/Town/Village/Area | Location of Monitoring Station | Agency | Type of Location | SO2 | NO2 | RSPM/PM10 | PM 2.5 |
| 38 | 1/2/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 11 | 17 | 55 | NA |
| 38 | 1/7/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 17 | 45 | NA |
| 38 | 21-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 12 | 18 | 50 | NA |
| 38 | 23-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 15 | 16 | 46 | NA |
| 38 | 28-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 14 | 42 | NA |
| 38 | 30-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 14 | 18 | 43 | NA |

... ... ... ... ... ... ...

In [3] : air\_qualitu.head()

Out [3] :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stn Code | Sampling Date | State | City/Town/Village/Area | Location of Monitoring Station | Agency | Type of Location | SO2 | NO2 | RSPM/PM10 | PM 2.5 |
| 38 | 1/2/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 11 | 17 | 55 | NA |
| 38 | 1/7/2014 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 17 | 45 | NA |
| 38 | 21-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 12 | 18 | 50 | NA |
| 38 | 23-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 15 | 16 | 46 | NA |
| 38 | 28-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 13 | 14 | 42 | NA |
| 38 | 30-01-14 | Tamil Nadu | Chennai | Kathivakkam, Municipal Kalyana Mandapam, Chennai | Tamilnadu State Pollution Control Board | Industrial Area | 14 | 18 | 43 | NA |

<class 'pandas.core.frame.DataFrame'

# Print the basic information about the dataset

In [4] : air\_quality.info()

Out [4] :

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 stncode 105 non-null object

1 date 105 non-null int64

2 State 105 non-null object

3 Village 105 non-null object

4 SO2 105 non-null int64

5 NO2e 105 non-null float64

dtypes: float64(1), int64(2), object(3)

memory usage: 5.0+ KB

# Check for missing values

In [5] : netflix\_originals.isnull().sum()

Out [5] :

--- ------ -------------- -----

0 stncode 0

1 date 0

2 State 0

3 Village 0

4 SO2 0

5 NO2. 0

This means that the training set contains SO2samples and the test set contains NO2 samples.