

Anna University Regional Campus Coimbatore

Anna University: Chennai-600 025

DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING



IBM Naan Mudhalvan Phase 2 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:

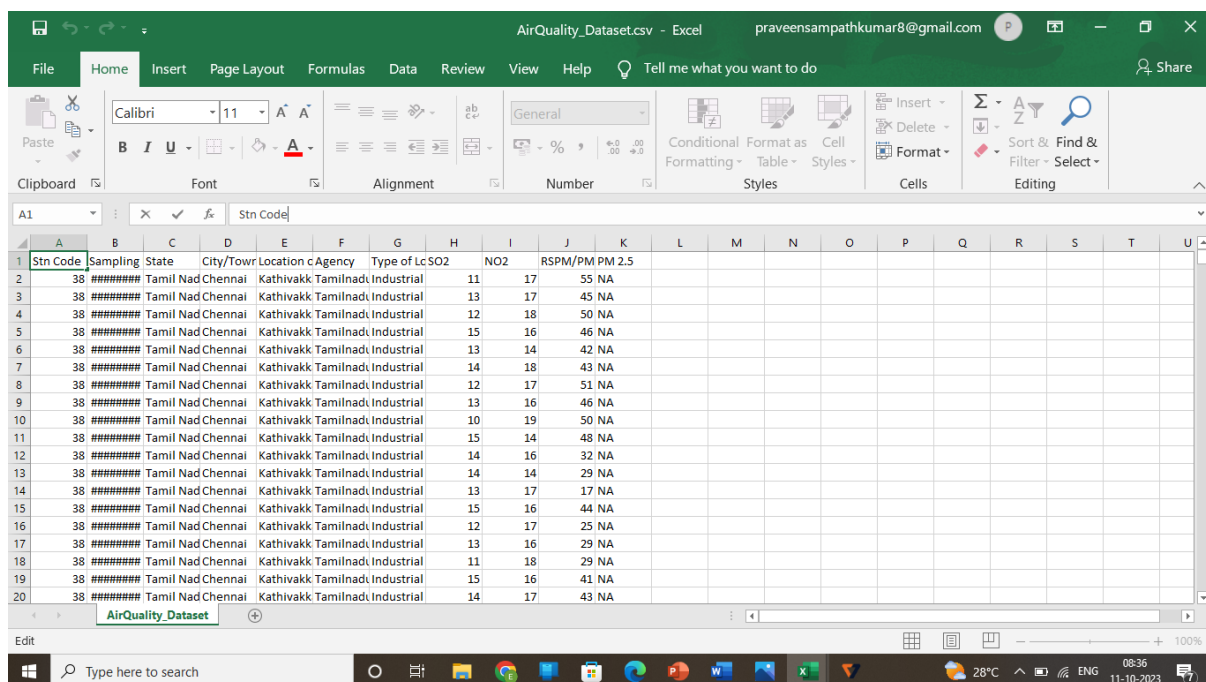
The Objective of the project is to analyze and visualize the air quality data from the various monitoring stations in Tamil Nadu. The given dataset contains the measurements of the various gases that release into the atmosphere. Some of the gases given in the dataset are Sulphur Dioxide(SO₂), Nitrogen Dioxide(NO₂) and Respirable particulate matter and these are measured in different cities, villages, towns. This Project aim is to gain the insight of air pollution trends, estimate the RSPM/PM₁₀ levels based on SO₂ and NO₂ levels

Description of dataset:

The link for the chosen dataset is mentioned 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. This contains the district wise dataset 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.



Stn Code	Sampling	State	City/Town	Location	Agency	Type of Loc	SO ₂	NO ₂	RSPM/PM 2.5
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11	17	55 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	13	17	45 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12	18	50 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15	16	46 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	13	14	42 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14	18	43 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12	17	51 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	13	16	46 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	10	19	50 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15	14	48 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14	16	32 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14	14	29 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	13	17	17 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15	16	44 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	12	17	25 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	13	16	29 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	11	18	29 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	15	16	41 NA
38	#####	Tamil Nad	Chennai	Kathivakk	Tamilnadu	Industrial	14	17	43 NA

Explanation:

The given dataset contains the different columns with their specific details. The different columns are “**stn code, sampling date, state, city/town, Location of monitoring stations, Agency, Type of location, SO2, NO2, RSPM/PM, SPM**”

To further proceeding of the project let us drop the unwanted columns that is unnecessary for analysis of the air quality in Tamil Nadu

The screenshot shows a Google Colab notebook titled "DATA_SCIENCE.ipynb". The file explorer on the left shows a folder named "sample_data" containing "AirQuality_Dataset.csv" and "India_Air_Quality_Data_For_TamilN...". The code cell contains the following Python code:

```
[1] import pandas as pd
[5] df=pd.read_csv("AirQuality_Dataset.csv")
[6] df.head()
```

The output of the code is a preview of the dataset:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO
0	38	01-02-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.

The bottom of the screenshot shows the Windows taskbar with the system clock at 08:03 on 11-10-2023.

The screenshot shows the same Google Colab notebook. The code cell now contains the following Python code:

```
"Stn Code", "Sampling Date", "Agency", "Location of Monitoring Station"], axis=1)
```

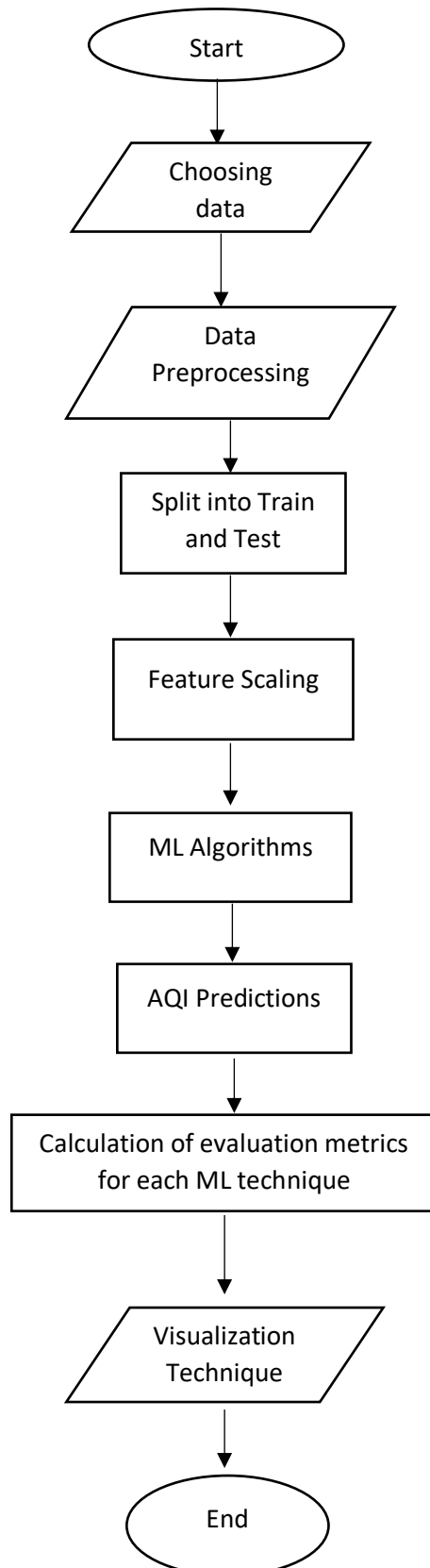
The output of the code is a preview of the dataset after dropping the specified columns:

	City/Town/Village/Area	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	Chennai	Industrial Area	11.0	17.0	55.0	NaN
1	Chennai	Industrial Area	13.0	17.0	45.0	NaN
2	Chennai	Industrial Area	12.0	18.0	50.0	NaN
3	Chennai	Industrial Area	15.0	16.0	46.0	NaN
4	Chennai	Industrial Area	13.0	14.0	42.0	NaN
...
2074	Residential, Rural and		45.0	48.0	400.0	NaN

The bottom of the screenshot shows the Windows taskbar with the system clock at 13:36 on 11-10-2023.

By this, we are going to use the specific columns for the preprocessing of the data to predict and analyze the air quality in various regions in Tamil Nadu.

Flow Chart for the proposed system



Conclusion:

In conclusion, this project focuses on analysing and predicting the air quality in Tamil Nadu. This will yield the valuable insights and outcomes. Through collection and analysis of historical air quality data, we could able to identify trends, seasonal variations, and the impact of various factors on air quality. Our predictive model, based on machine learning algorithms, demonstrated reasonable accuracy in forecasting air quality levels.