AIR QUALITY PREDICTION

INTRODUCTION

In an era of rising pollution, understanding and predicting air quality is crucial. Discover why accurate analysis and prediction are essential for public health, environmental sustainability, and effective policy-making.

Algorithm Selection:

Choose a suitable machine learning algorithm based on the nature of your project. Common choices include decision trees, support vector machines, neural networks, and more.

Linear Regression:

- Use for regression problems when you want to predict a continuous target variable.

Logistic Regression:

- Suitable for binary classification problems (e.g., spam detection).

```
Z = spam['EmailText']
y = spam["Label"]
z_train, z_test,y_train, y_test = train_test_split(z,y,)
```

Day level

To get AQI at day level, the AQI values are averaged over the hours of the day.

```
Df_station_hour = df
df_station_day = pd.read_csv(PATH_STATION_DAY)

df_station_day = df_station_day.merge(df.groupby(["StationId", "Date"])["AQI_calculated"].mean().reset_index(), on = ["StationId", "Date"])
df_station_day.AQI_calculated = round(df_station_day.AQI_calculated)
```

City level

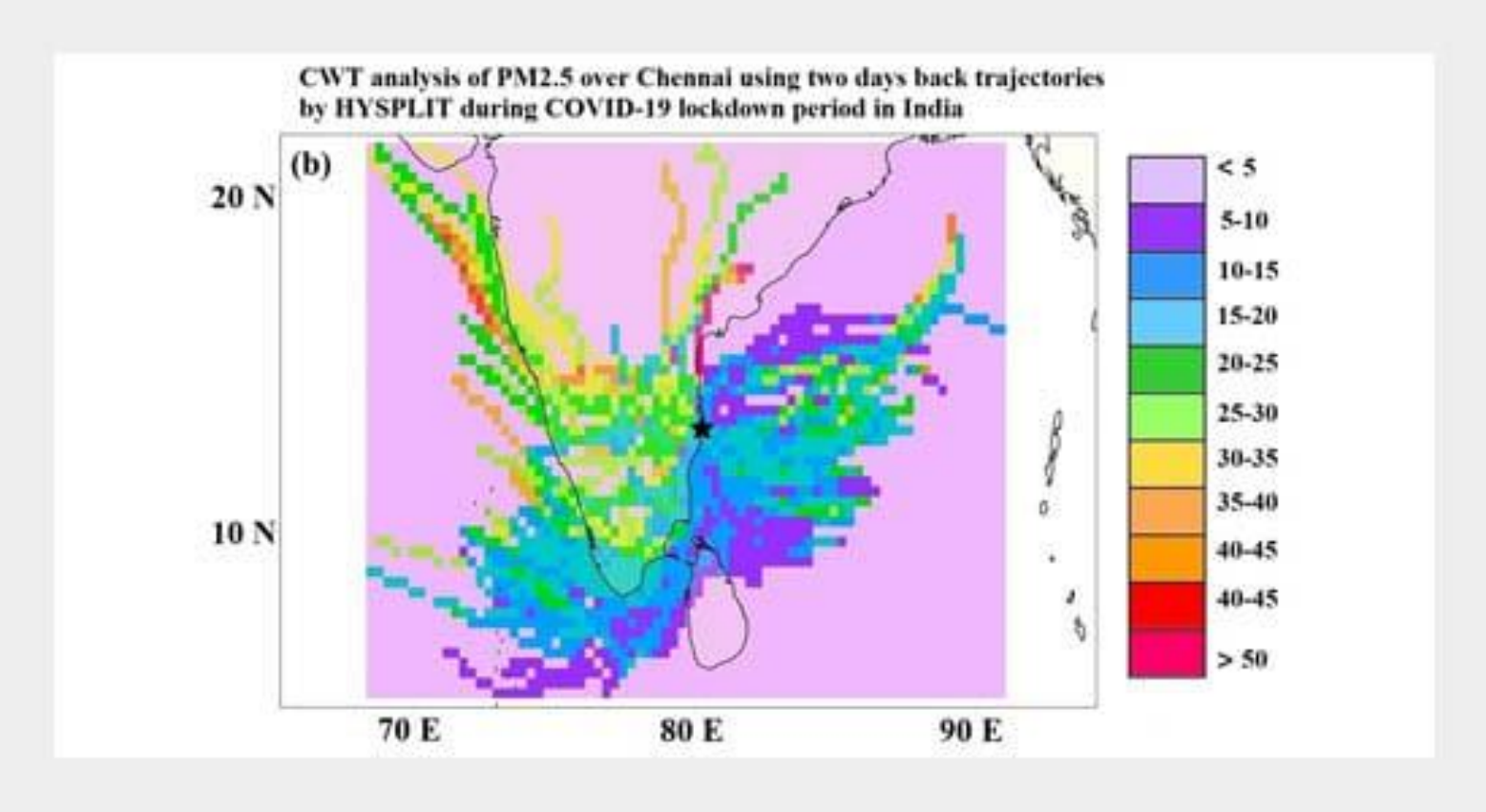
To get AQI at city level, the AQI values are averaged over stations of the city.

```
Df_city_hour = pd.read_csv(PATH_CITY_HOUR)
df_city_day = pd.read_csv(PATH_CITY_DAY)

df_city_hour[“Date”] = pd.to_datetime(df_city_hour.Datetime).dt.date.astype(str)

df_city_hour = df_city_hour.merge(df.groupby([“City”, “Datetime”])[“AQI_calculated”].mean().reset_index(), on = [“City”, “Datetime”])
df_city_hour.AQI_calculated = round(df_city_hour.AQI_calculated)

df_city_day = df_city_day.merge(df_city_hour.groupby([“City”, “Date”])[“AQI_calculated”].mean().reset_index(), on = [“City”, “Date”])
df_city_day.AQI_calculated = round(df_city_day.AQI_calculated)
```

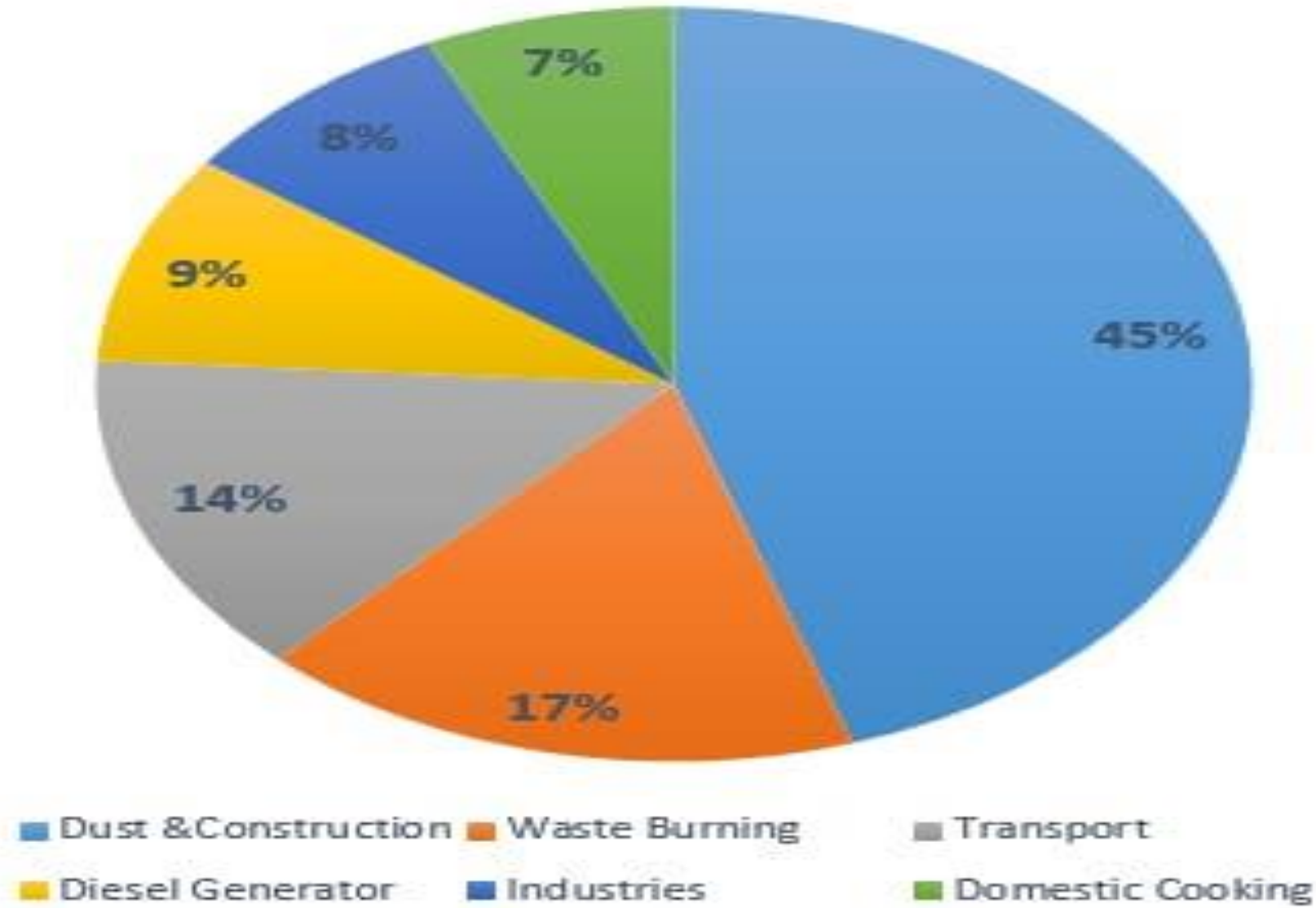
Data Visualization

datacount =sns.countplot(x ="location",data = tn);

datacount.set_xticklabels(datacount.get_xticklabels(), rotation=90);



Sources of Air Pollution



DATA COLLECTION

Data Gathering Methods:

Choose appropriate data gathering methods based on your data sources. This might involve writing scripts for web scraping, setting up data collection pipelines, configuring sensors, or manually entering data.

Data Format:

Determine the format in which you'll store the data. Common formats include CSV, JSON, databases, or structured files. Ensure that data is organized for easy processing.

Data Volume:

Decide how much data you need. The required amount of data depends on the complexity of your project and the chosen machine learning algorithm.

Data Cleaning:

After collecting data, it's common to perform data cleaning to handle missing values, outliers, and inconsistencies. This step is crucial for ensuring the quality of your dataset.

Data Labeling (if applicable):

If your project involves supervised learning, you may need to label the data by annotating it with the correct target values.

Documentation:

Maintain detailed documentation about your data gathering process. Include information about the data sources, collection dates, and any transformations applied.

CONCLUSION

- Continuous monitoring and prediction of air quality play a vital role in safeguarding public health and shaping effective policies. By adopting sustainable practices, raising awareness, and implementing stringent regulations, we can create a cleaner and healthier environment for future generations to thrive.