IDEATION PHASE

PROBLEM STATEMENT

|  |  |
| --- | --- |
| DATE | 23-09-2023 |
| TEAM ID | PROJECT\_224711\_TEAM2 |
| PROJECT TITLE | AIR QUALITY ANALYSIS AND PREDICTION IN TAMILNADU |

PROBLEM STATEMENT:

* The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu.Based on the current trends and the predictions from the TNPCB, the air quality in Tamil Nadu in 2022 is expected to be similar to or slightly worse than the air quality in 2021. The levels of SO2 and NO2 are expected to be particularly high in major cities such as Chennai, Coimbatore, and Madurai.
* The survival of mankind cannot be imagined without air. Consistent developments in almost all realms of modern human society affected the health of the air adversely. Daily industrial, transport, and domestic activities are stirring hazardous pollutants in our environment.

Problem statement template:

7WAYS TO CONTROL THE AIR POLLUTION:

* 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.

MISSING VALUES AND PERCENTAGE:

|  |  |  |
| --- | --- | --- |
| Dataset | Missing values | %of Total values |
| So2 | 3854 | 13.100000 |
| No2 | 3585 | 12.100000 |
| no | 3582 | 12.100000 |

ADVANTAGES OF AIR POLLUTION:

* Air pollution helps plants grow
* Air pollution slows climate change
* Air pollution keeps polluted cities cooler in the summer

OBJECTIVES OF AIR POLLUTION:

* To prevent adverse responses by all receptor categories exposed to the atmosphere
* To increase the risk of respiratory infections like,heart disease and lung cancer.
* It also increase the productivity creates employment and initiates innovation.

EXPLORATORY DATA ANALYSIS:

* This section of the present study deals with data exploration and analysis for finding various hidden patterns present in the dataset. Exploratory data analysis is the first step in data analytics which is performed before applying any ML model. Under this, the following important things are being analyzed
* Exploring statuses and trends of air pollutants over the past six years i.e. from 2015 to 2020;
* exploring the distribution of pollutants in the air along with top-six polluted cities with their average AQI values;
* estimating top four pollutants which are directly involved in increasing the AQI values.

FEATURE SELECTION:

* The *CPCB* dataset under study involves a specific parameter viz, AQI and government agencies use this parameter to alert people about the quality of the air and also practice forecasting it.
* According to the *National Ambient Air Quality Standards*, there are six AQI categories: good (0–50), satisfactory (51–100), moderate (101–200), poor (201–300), very poor (301–400), and severe (401–500)
* Scholars in the realm suggest that reducing input variables lowers the computational cost of modeling and enhances prediction performance. A correlation-based feature selection method has been exploited.
* The variables possessing the strongest correlation with the target variable are then filtered for further study. Since many ML algorithms are sensitive to outliers, any feature in the input dataset which does not follow the general trend of that data must be found.

CONCLUSION:

Prediction of air quality is a challenging task because of the dynamic environment, unpredictability, and variability in space and time of pollutants. The grave consequences of air pollution on humans, animals, plants, monuments, climate, and environment call for consistent air quality monitoring and analysis, especially in developing countries.