AIR QUALITY ANALYSIS IN TAMILNADU

PHASE 2 PROJECT :

MACHINE LEARNNG ALGORITHMS TO IMPROVE THE ACCURACY OF THE PREDICTIVE MODEL.

➢ **INTRODUCTION:**

Measuring the quality of air is important to prevent the environment from pollution and to live a healthy life. Air quality is a critical aspect of human health and environment. So the prediction must be accurate. To predict the air quality accurately some Machine Learning algorithms are used which will enhance the prediction.

➢ **BASICS OF AIR AND AIR QUALITY MEASUREMENT:**

❖ **WHAT IS AIR?**

Air is a combination of many gases and mixture particles. However there are many gases present in the air , there are two essential gases present in the air. They are oxygen and nitrogen.

❖ **FIVE MAJOR AIR POLLUTANTS?**

- GROUND LEVEL OZONE
- CARBON MONOXIDE
- SULFUR DI OXIDE
- NITROGEN DIOXIDE
- AIRBORNE PARTICLES AND AEROSOLES.

❖ **HOW AIR QUALITY IS MEASURED?**

- Air quality is measured with the Air Quality Index, or AQI. The AQI works like a thermometer that runs from 0 to 500 degrees. The AQI is a way of showing changes in the amount of pollution in the air.

- The AQI has six categories that communicate the level of health concern using specific colours. Code Green and Yellow means the air is generally safe for everyone. Code Orange is unhealthy for sensitive groups, including children, senior citizens, and people with heart and lung diseases. Code Red and Purple means the air is unhealthy for everyone, and Code Maroon is a health warning of emergency conditions.

Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0 to 50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health alert: everyone may experience more serious health effects.
Hazardous	301 to 500	Health warnings of emergency conditions. The entire population is more likely to be affected.

❖ **DRAWBACKS:**

- However the air quality index measures the quality of air, Sometime it do not give the accurate result due to some environment changes and lack of maintanence of the machine. So here we use the machine learning algorithms to measure the quality of air accurately.
- Traditional models like Linear Regression and Time Series Forecasting have been used in the past. However, they often fail to capture complex relationships in the data.

➢ **Machine Learning Algorithms And Approaches:**

❖ **Regression Model:**
- ✓ **Linear Regression:**

  - ✓ Linear Regression is Machine Learning Algorithm which is used to predict the accurate result between a dependent variable from one or more independent variable.
  - ✓ It is used in many fields like economics, medical, weather department, finance, psychology and much more.
  - ✓ It is a basic algorithm which is used as a baseline model.
- ✓ **Decision Trees:**

  - ✓ It is a Machine Learning Algorithm which is used for both classification and regression models.
  - ✓ It uses a tree along with the decision and leaf nodes, which will predict the results accurately based on the given datas.

- ✓ **Random Forests:**

  - ✓ Decision trees uses a single tree to predict a result for the given data, where the result will not be accurate.
  - ✓ To improve the accuracy Random Forest Algorithm are used.
  - ✓ In Random Forest Algorithm , it finalizes the result by combining many trees. It will predict the accurate result by filtering the results of many trees and give the final result.
- ✓ **Gradient Boosting Algorithm:**

  - ✓ Gradient Boosting Algorithm is a Machine Learning Algorithm which predicts and gives the accurate results by converting weak learners into stronger learners.

❖ **Neural Networks:**

- ✓ **Feed-Forward Neural Networks:**

  - ✓ In this method we get a results by passing an input through many nodes which will analyze and finalize the result.
  - ✓ The nodes do not form the cycle, where the input move only in one direction.

- ✓ **Recurrent Neural Networks:**

  - ✓ A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (nlp), speech recognition, and image captioning.Like feedforward recurrent neural networks utilize training data to learn.
  - ✓ It gains the information priorly about the data to give input and process the output.
  - ✓ It is useful for Time-Series Data

- ✓ **Long Short Term Memory Networks:**

  - It is useful for time series data.
  - It takes the data as a input from the output where it executed at last. With the help of the past output it takes as input and gives us the result.

- ✓ **Ensemble Learning:**

  - It is a method which combines the predictions from multiple models can yield the better results.

- ➢ **SOURCE DATA FOR PREDICTION:**

  - Lagged Values: Using previous air quality values.
  - Weather Data: Temperature, humidity, wind speed, etc.,
  - Temporal Data: Day of the week and seasons.
  - Spatial Data: Pollution and Traffic data.

- ➢ **Handling Missing Data:**

  - Complete Case (CC), Mean Substitution (MS), Last Observation Carry Forward (LOCF), Multiple Imputation (MI) are the methods used to handle the missing values in the data.
  - Another techniques like interpolation, imputation, or training a model to predict the missing values can be used.

- ➢ **Model Validation:**

  - Time series split – Useful for sequential data.
  - K-Fold Cross Validation – Provides a more generalized error estimate.

- ➢ **External Factors And Integration:**

  - Integration with IOT: Real Time data from IOT devices can be fed into the model for the real time predictions.
  - Feedback loop: Continuously training the data model can adapt to new changing pattern.

- ➢ **Conclusion:**

  Hence, the above algorithms and methods are a new way to predict the air quality more efficiently and accurately.

****