

# AQI Prediction Model



## Methodology:

The approach to predict the Air Quality Index (AQI) involves several critical steps:

### **Data Collection:**

Gathered historical AQI data from reliable environmental monitoring government sources, including weather conditions and pollution data.

### **Data Preprocessing:**

- **Data Cleaning:**

We have removed missing or erroneous present in the dataset.

- **Feature Engineering:**

Selected relevant features (e.g., NO<sub>2</sub> , CO levels) that impact AQI levels.

- **Feature Scaling:**

Standardized the features to ensure the model treats all variables equally.

## **Model Selection:**

Random Forest Regressor was chosen after evaluating various machine learning models, as it provided the best results for predicting AQI levels.

Random Forest is widely used in Machine Learning because it offers a high level of accuracy, is robust to overfitting, can handle both classification and regression tasks, is efficient with large datasets, and provides valuable insights into feature importance by automatically ranking the contribution of different features to the model's predictions

## **Model Training:**

The model would be trained using training data collected from CPCB(government pollution control website), with hyperparameters optimized using GridSearchCV to improve accuracy.

## **Technology to be Used:**

- Programming Language:

Python: Used for model development, data preprocessing, and evaluation.

- Machine Learning Algorithms:

Random Forest Regressor: A robust ensemble learning method for regression tasks, used due to its ability to handle complex datasets and capture non-linear relationships.

- Visualization:

Power BI: Data visualization tool.

## **Libraries and Tools:**

- scikit-learn: Used for implementing machine learning models and evaluation metrics.
- pandas: For data manipulation and preprocessing.
- NumPy: For numerical calculations and handling arrays.
- Matplotlib / Seaborn: For visualizing data and model performance.

- GridSearchCV: Used for hyperparameter tuning to optimize the model's parameters for better accuracy.

## Data Sources:

CPCB , [data.gov.in](http://data.gov.in)

Environmental datasets for AQI, weather conditions, and pollution data.



## Power BI Dashboard:



