Project Title: Predicting Air Quality for Sustainable Urban Living

1. Project Idea:

- Problem Statement: Poor air quality is a major environmental and health concern, especially in urban areas. Predicting air quality can help policymakers and citizens take preventive measures.
- Objective: Develop a predictive model that forecasts air pollution levels (e.g., PM2.5, PM10, CO, NO2) using historical air quality data and meteorological conditions.

2. Relevance to Sustainable Development Goals (SDGs):

- SDG 3 (Good Health and Well-being): Reducing air pollution can decrease respiratory diseases and improve public health.
- SDG 11 (Sustainable Cities and Communities): Predicting air quality helps create cleaner and healthier urban environments.
- SDG 13 (Climate Action): Monitoring pollution contributes to mitigating environmental impacts and climate change.

3. Literature Examples:

1. Wang et al. (2020) - "Air Quality Prediction Using Machine Learning Algorithms"
    - The study used machine learning models (Random Forest, LSTM, and XGBoost) to predict air pollution levels in China.
    - It demonstrated that deep learning models, particularly LSTMs, improved prediction accuracy compared to traditional statistical methods.
2. Zhang et al. (2021) - "Deep Learning for Air Pollution Forecasting"
    - The paper introduced CNN-LSTM hybrid models for forecasting PM2.5 concentrations.
    - The study found that incorporating meteorological data improved prediction performance.

3. Description of Data:

- Dataset Source:
  - U.S. EPA Air Quality Data platform
- Data Format: CSV files with daily/hourly air pollution readings.
- Preprocessing Steps we will conduct:
  - Handling missing values and outliers.
  - Normalizing and scaling pollution levels.
  - Feature engineering (adding weather conditions, time-based trends).