Project Title: Air Quality Prediction for Urban Areas

Problem Statement:

You have been assigned the task to develop a machine learning model that can predict air quality in urban areas. The project involves various stages, including web scraping, exploratory data analysis (EDA), training a machine learning model, and saving the model for future use.

1. Web Scraping:

- Identify a reliable source of air quality data for urban areas. This could be a public API, government database, or any other reputable source.

- Implement a web scraping script to extract historical air quality data for different cities or regions of interest. The data should include attributes such as pollutant concentrations (e.g., PM2.5, PM10, CO, NO2, etc.), meteorological conditions (e.g., temperature, humidity, wind speed), and timestamp.

2. Exploratory Data Analysis (EDA):

- Perform data cleaning and preprocessing to handle missing values, outliers, and inconsistent data.

- Conduct a comprehensive exploratory data analysis to gain insights into the dataset.

- Visualize the data using appropriate graphs and plots to understand the relationships between different features and air quality.

3. Feature Engineering:

- Select relevant features that are likely to influence air quality.

- Perform feature engineering techniques such as encoding categorical variables, handling numerical features, and creating new derived features if necessary.

- Consider incorporating additional data sources, such as satellite imagery or geographical information, to enhance the predictive power of the model.

4. Machine Learning Model Training:

- Split the dataset into training and testing sets.

- Choose an appropriate machine learning algorithm (e.g., regression) for predicting air quality.

- Train the model using the training data and evaluate its performance on the testing data.

- Fine-tune the model by adjusting hyperparameters to improve its performance.

5. Model Evaluation and Deployment:

- Evaluate the trained model's performance using suitable evaluation metrics (e.g., mean squared error, R-squared).

- Save the trained model to a file for future use.

- Develop a simple web interface or command-line application where users can input relevant meteorological conditions, and the model predicts the corresponding air quality levels.

Additional Tasks (Optional):

- Implement feature scaling or normalization techniques if required.

- Compare the performance of different machine learning algorithms.