Air Quality Analysis Using Machine Learning

Problem Statement

Air pollution is a significant environmental and health issue, contributing to respiratory diseases, climate change, and reduced quality of life. Traditional monitoring systems are expensive and provide limited coverage. This project aims to leverage machine learning techniques to analyze air quality data, predict pollution levels, and identify key contributing factors. By utilizing historical and real-time environmental data, the model enhances forecasting accuracy, supporting environmental policies and public health initiatives.

Aim

To develop a machine learning-based system for analyzing air quality, predicting pollution levels, and identifying key contributing factors using historical and real-time environmental data, enabling better decision-making for public health and environmental management.

Learning Objectives

- Understand air pollution, its causes, and its impact on health and the environment.
- Collect, preprocess, and analyze air quality datasets effectively.
- Apply exploratory data analysis (EDA) to uncover trends and correlations.
- Implement machine learning models for air pollution prediction.
- Evaluate model performance using appropriate metrics.
- Identify key factors influencing air quality.
- Develop an interactive dashboard for real-time air quality predictions.
- Deploy the model using Streamlit for accessibility.
- Explore the role of AI in environmental policy-making.
- Gain hands-on experience with Python, machine learning libraries, and data visualization tools.

About the Project

This project focuses on developing a machine learning-based air quality analysis system. The

system utilizes historical and real-time air quality data, including pollutant concentrations (PM2.5, PM10, NO2, SO2, CO, O3) and meteorological parameters (temperature, humidity, wind speed, etc.). Various ML models, such as regression and deep learning algorithms, are used to predict air quality index (AQI) values and classify air pollution levels. The findings are visualized through an interactive dashboard, helping policymakers and the public make informed decisions about air quality and pollution control strategies.

Datasets Used

Description:

The dataset contains air quality measurements from various locations, including pollutant concentrations, meteorological data, and temporal information. The data is sourced from government agencies, OpenAQ, Kaggle, and other reliable repositories.

Target:

Air Quality Index (AQI): A standardized indicator representing overall air pollution levels based on pollutant concentrations.

Tools and Techniques

1. Python

Libraries:

- Pandas, NumPy - Data processing and manipulation
- Matplotlib, Seaborn, Plotly - Data visualization
- Scikit-learn, TensorFlow, Keras - Machine learning model development
- Streamlit - Deployment of an interactive dashboard

2. Machine Learning Models

- Linear Regression - For predicting AQI values
- Random Forest, Decision Trees - For feature importance analysis
- Neural Networks (Deep Learning) - For advanced prediction models
- Support Vector Machines (SVM) - For classification of air quality categories

3. Deployment Tools

Streamlit:

Used for building an interactive web application to visualize air quality predictions and insights.

Findings and Insights

1. Data Preparation

- Collected air quality data from multiple sources.
- Preprocessed data by handling missing values and normalizing features.
- Conducted exploratory data analysis (EDA) to identify trends and correlations.

2. Model Training and Evaluation

- Trained multiple machine learning models and evaluated them using metrics like RMSE, MAE, and R^2.
- Identified the most effective model for accurate AQI prediction.

3. Predictive Insights

- Determined key contributors to air pollution (e.g., PM2.5 and NO2 being major pollutants in urban areas).
- Predicted air quality trends based on historical and real-time data.

4. Interactive Interface

- Developed a Streamlit-based dashboard for real-time air quality monitoring.
- Users can input location data and get instant air quality predictions.

Conclusion

This project demonstrates the effectiveness of machine learning in air quality analysis and prediction. By leveraging ML algorithms and real-time data, the model provides valuable insights into air pollution trends, enabling proactive measures for pollution control. The interactive dashboard enhances accessibility, allowing policymakers and the public to make data-driven decisions for environmental management. This project highlights the potential of AI in improving air quality forecasting and supporting sustainable urban development.