

Course Project Proposal

AirVise

October 11, 2024

1 Project Title

AirVise: Air Quality Prediction

2 Team Member

- Faisal Ilyas (2022-CS-63)

3 Objective

The goal of this project is to accurately predicts air quality index (AQI) levels based on meteorological data, enabling policymakers and the public to make informed decisions about health and environmental policies.

4 Background and Motivation

Air pollution has become a significant global concern, affecting public health and the environment. Accurate prediction of air quality can help in understanding pollution trends and formulating effective measures to combat air quality deterioration. By leveraging historical air quality and meteorological data, this project aims to build a predictive model that will provide real-time AQI forecasts, assisting individuals and organizations in making better health-related decisions.

5 Proposed Methodology

The project will follow these steps:

5.1 Data Collection

I will gather historical air quality data from reliable sources, such as government databases or public datasets on Kaggle, including features like temperature, humidity, and wind speed.

5.2 Data Analysis

Using Pandas, I will perform exploratory data analysis (EDA) to understand the dataset and uncover relationships between different variables. This will involve visualizing data distributions and correlations to gain insights into air quality patterns.

5.3 Data Preprocessing

In this phase, I will:

- Handle missing values using `SimpleImputer` from Scikit-learn.
- Normalize and standardize the data with `StandardScaler`.
- Split the dataset into training and testing sets to evaluate model performance.

5.4 Model Selection

I will experiment with various regression algorithms to predict AQI levels, including:

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor
- Support Vector Regressor

I will assess the performance of each model and select the best one based on evaluation metrics.

5.5 Model Evaluation

I will evaluate the model's performance using:

- Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to gauge prediction accuracy.
- Cross-validation techniques to ensure the model's robustness and reliability.

6 Tools and Technologies

I will utilize:

- Python programming language
- Libraries: Pandas, NumPy, Matplotlib, Scikit-learn, and Joblib for saving and loading the model.

7 Expected Outcomes

I aim to develop a robust air quality prediction model with a prediction accuracy of at least 90%. The model will provide real-time AQI predictions based on user inputs, enabling proactive health and safety measures.

8 Challenges

Some potential challenges may include:

- Managing missing or inconsistent data in the dataset.
- Ensuring the model generalizes well to new, unseen data.

9 References

- Data Source: <https://www.kaggle.com/discussions/accomplishments/510447>

10 Conclusion

By the end of this project, I aspire to create a functional air quality prediction model that will assist individuals and organizations in monitoring air quality and making informed decisions regarding health and environmental policies.