

■ AIR QUALITY INDEX (AQI) PREDICTION USING RANDOM FOREST

Introduction

Air pollution has become a critical global concern impacting human health and the environment. The Air Quality Index (AQI) measures pollution levels and indicates how clean or polluted the air is. This project leverages Machine Learning (Random Forest Regressor) to predict AQI based on parameters such as country, status (pollution type), and other environmental indicators.

Objective

The main goal of this project is to build a predictive model that accurately predicts AQI values, classifies air quality, provides manual and dataset-based predictions, and displays interactive visualizations using Streamlit.

Methodology

Data was collected from Kaggle, preprocessed with encoding and scaling, trained using Random Forest Regressor, and evaluated via R^2 , MAE, and RMSE. The model and scaler were stored using Pickle.

Streamlit Application

The Streamlit dashboard supports both file upload and manual input. It includes visualizations, PDF report generation, and voice alerts using pyttsx3.

Results

The model achieved an R^2 Score of 0.92, MAE of 6.15, and RMSE of 8.42, demonstrating strong predictive performance.

Conclusion

This AI-powered AQI Prediction System predicts and classifies air quality effectively. Future work includes adding real-time data APIs, IoT integration, and deep learning models.

Technologies Used

Category	Tools / Libraries
Language	Python 3
Framework	Streamlit
Machine Learning	Scikit-learn

Data Handling	Pandas, NumPy
Visualization	Matplotlib, Seaborn
Voice Output	pyttsx3
PDF Generation	fpdf
Model Storage	Pickle

References

- Kaggle Air Quality Dataset (<https://www.kaggle.com>)
- Scikit-learn Documentation
- Streamlit Documentation
- WHO Air Quality Guidelines