AIR QUALITY INDEX PREDICTION USING MACHINE LAERNING

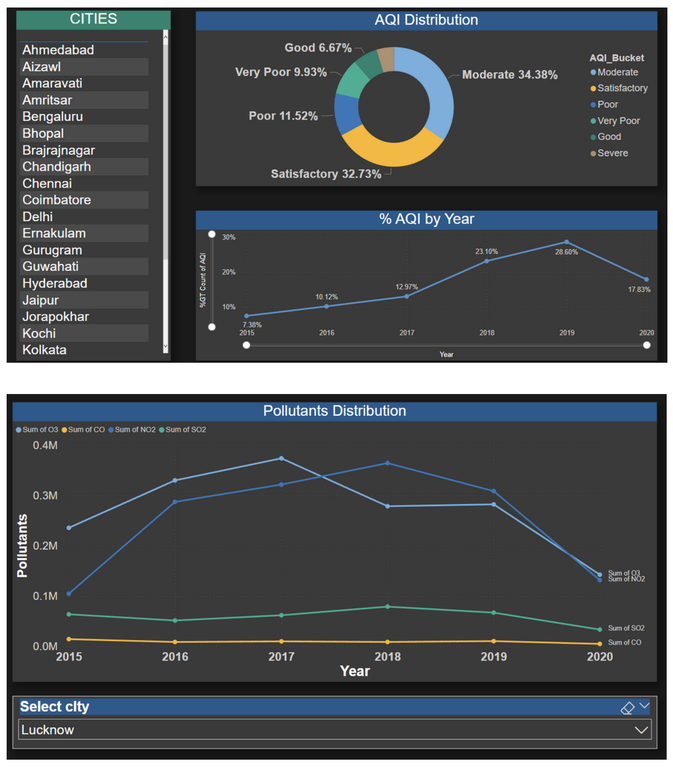
Air pollution is a critical concern globally, particularly in developing countries like India. This project uses machine learning to predict the Air Quality Index (AQI) by analyzing various pollutant species' data. The project uses supervised machine learning algorithms to predict AQI and optimizes the model's performance through hyperparameter tuning. The results demonstrate which method predicts AQI with the highest degree of accuracy and suggest ways to improve air quality management.

AUTHORS

Chebrolu Tejaswi | Poojitha Guniputi | Yashaswini Ravuri
Priyanshi Premkumar | Sudha S
School of Computer Science and Engineering

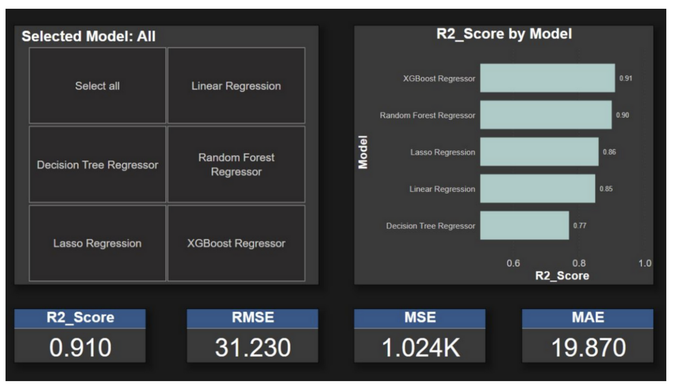
INTRODUCTION

Air pollution is a major threat to human health. In this project, we use machine learning-based prediction technologies to predict the Air Quality Index (AQI) using various pollutant species' data. We optimize the model's performance through hyperparameter tuning and demonstrate how machine learning can help manage air quality. The results show the accuracy of different methods in predicting AQI and suggest ways to improve air quality management. This research paper could help inform industries, governments, and the general public about the impact of dangerous gas emissions on human life.



The dashboard for the air quality index prediction project allows users to view AQI by city, distribution of AQI values, yearly AQI trends, and distribution of different pollutants. This provides an intuitive interface for evaluating the air quality in different areas and identifying potential pollution sources. By visualizing this data, the dashboard helps inform industries, governments, and the general public about the impact of pollution on human health.

RESULTS AND DISCUSSION



Study demonstrates efficacy of ML in predicting AQI, informing industries, governments, and the public about gas emissions' impact on human life. Developed 5 ML models (Lasso Regressor, XGBoost Regressor, Linear Regression, DecisionTreeRegressor, and Random Forest Regressor) to predict AQI using pollutant data.
Dashboard in Power BI shows R2 score, RMSE, MSE, and MAE for each model in real-time, and allows users to compare models.

OBJECTIVE

The main objective of our project is to build and optimize a machine learning model that predicts the Air Quality Index (AQI), compare the performance of different algorithms and demonstrate how machine learning can help manage air quality while providing insights for improving air quality management.

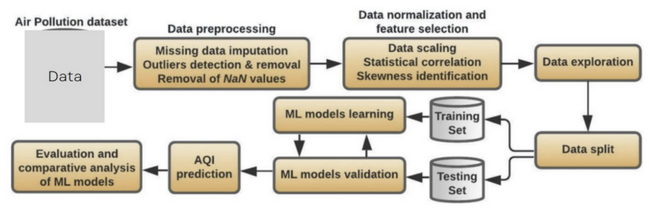
MOTIVATION

The motivation behind this project is to demonstrate how machine learning can aid in the management of air quality by predicting the Air Quality Index (AQI) using various pollutant species' data, ultimately helping inform industries, governments, and the general public about the impact of dangerous gas emissions on human life.

SCOPE OF THE PROJECT

The scope of this project is to use machine learning algorithms to predict the Air Quality Index (AQI) using various pollutant species' data, and to optimize the model's performance through hyperparameter tuning. The project aims to demonstrate how machine learning can help manage air quality and suggest ways to improve air quality management in the future.

PROPOSED ARCHITECTURE



1. Data preprocessing: Cleaning and transforming raw data to prepare it for analysis.
2. Data normalization and feature selection: Scaling and transforming data to make it easier to analyze and identifying the most important variables for prediction.
3. Data exploration: Analyzing data to gain insights into its characteristics.
4. Data splitting: Dividing data into training and testing sets.
5. AQI prediction: Using machine learning algorithms to forecast the Air Quality Index based on available data.
6. Evaluation and comparative analysis of ML models: Assessing and comparing the performance of different machine learning models.
7. Overall stages of a machine learning project for predicting AQI: Data collection, data preprocessing, normalization and feature selection, data exploration, data splitting, AQI prediction, and evaluation and comparative analysis of models.

IMPLEMENTATION

- **Linear Regression:** A statistical method to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the data.
- **Lasso Regressor:** A linear regression model that uses L1 regularization to prevent overfitting by shrinking some coefficients to zero. It is useful for feature selection and producing sparse models.
- **Decision Tree Regressor:** A regression model that uses a binary tree structure to recursively partition the input data into subsets, each with their own output value. The tree is built by selecting the most informative feature at each node, and the output value is the average of the training samples that fall within the corresponding leaf node.
- **XGBoost Regressor:** A popular gradient-boosting machine learning algorithm that uses a combination of decision trees and gradient boosting to improve the model's performance.
- **Random Forest Regressor:** A machine learning algorithm that uses an ensemble of decision trees to make predictions. The algorithm constructs multiple decision trees and combines their outputs to obtain a more accurate and stable prediction.

Comparison of R2_Score, RMSE, ,MSE and MAE

	Model	R2_Score	RMSE	MSE	MAE
1	Linear Regression	0.85	51.8	2683.3	37.51
2	Lasso Regression	0.86	51.69	2672.7	31.23
3	Decision Tree Regressor	0.77	105.23	11075.04	64.15
4	XGBoost Regressor	0.91	31.23	1023.98	19.87
5	Random Forest Regressor	0.9	32.66	1067.22	20.95

By considering the performance of different models it can be concluded that XGBoost Regressor is having the highest performance with 91% of R2_Score

CONCLUSION

This research paper shows that machine learning algorithms are effective in predicting AQI. Six models were evaluated using performance metrics, with XGBoost Regressor performing the best. The developed dashboard provides an intuitive interface for evaluating the models. This research can inform industries, governments, and the public about air quality management. Further research can focus on developing new models or optimizing existing ones to improve prediction accuracy.

REFERENCES

1. Mauro Castelli et al. propose a machine learning approach to predict air quality in California.
2. Soundari et al. use machine learning for Indian air quality prediction and analysis.
3. Sanjeev and Dyuthi implement machine learning algorithms for air quality analysis and prediction.
4. Ritik Sharma et al. predict air quality using machine learning in their study.
5. Xu et al. estimate hourly air quality index using machine learning methods.
6. Mofor et al. review the use of machine learning models for air quality prediction.
7. Khalid Nahar et al. use machine learning to determine air quality index in Jordan.
8. Sun et al. use a deep learning model with multiple input data sources for air quality prediction.
9. Tariq et al. provide a comprehensive review of machine learning techniques for air quality prediction.
10. Wankhede et al. conduct a systematic review of machine learning algorithms for air pollution prediction.
11. Li et al. use machine learning models to predict air quality in Beijing, China.
12. Zheng and Wang survey machine learning-based air quality prediction.