Abstract

Air pollution poses a serious risk to our health, the environment, and the stability of our climate. It's primarily fueled by industrial growth, urban expansion, and emissions from vehicles. To effectively tackle this issue, we need accurate air quality forecasts that can raise public awareness and ensure compliance with regulations. This paper introduces Pollution Predictor, an innovative system that uses machine learning to predict key air quality metrics in real-time, such as the Air Quality Index (AQI) and concentrations of PM2.5 and PM10. By harnessing historical pollution data and employing advanced feature engineering along with deep learning algorithms, our approach captures the temporal and seasonal shifts in pollution patterns. Built with TensorFlow.js, this system is designed for efficient, scalable, and accessible web-based predictions without sacrificing accuracy. Unlike many existing methods that struggle with computational demands, limited real-time capabilities, and poor adaptability, Pollution Predictor strikes a balance that works well across various geographic areas. The platform also features interactive heatmaps to visualize forecasted pollution levels, enhancing understanding and decision-making for all stakeholders involved. Through rigorous evaluation using metrics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), we demonstrate the model's strength and reliability. It outshines traditional regression and hybrid deep learning methods in both accuracy and efficiency. This research adds to the growing field of AI-powered environmental monitoring tools, promoting sustainable urban development and smart city projects. Pollution Predictor lays the groundwork for future innovations in pollution forecasting by merging cutting-edge machine learning with a focus on user experience, effectively addressing key challenges in air quality management.

Conclusion

Data preprocessing is a crucial step in making sure that datasets are both reliable and useful, especially when it comes to environmental monitoring and machine learning. In this project, we set up a thorough preprocessing pipeline to tackle the data quality issues that often come with real-world pollution datasets. We kicked off the preprocessing by identifying and addressing missing values, using imputation techniques and standardizing the data to keep things consistent across various sources. To make the dataset more robust, we detected and removed outliers through statistical methods, which not only optimized the data volume but also lessened the impact of any unusual readings on model performance. We also tackled encoding issues—often caused by different data formats and input errors—to create a structured and uniform dataset. These extensive cleaning steps greatly enhanced the clarity and interpretability of our data visualizations, allowing for more accurate trend analysis and pattern recognition. The refined dataset also cut down on noise and errors, which directly boosted the accuracy and reliability of subsequent statistical analyses and machine learning model training. By focusing on thorough data preprocessing, this work underscores the vital importance of data quality in predictive analytics. A well-prepared dataset not only aids effective model learning but also supports scalable, real-time air quality forecasting. Ultimately, this approach lays a strong foundation for AI-driven environmental monitoring systems, paving the way for informed decision-making and sustainable urban development. The methodologies and best practices outlined here serve as valuable guidelines for future research aimed at enhancing pollution prediction and environmental data analysis.

Title: “Pollution Predictor: A Real-Time Air Quality Forecasting System Using Deep Learning and Jaya Optimization”