AI Based Air Pollution Forecasting for PM2.5 Pollutants

Khushboo Mahawar
Department of AI and Data Science Engineering, School of Engineering and Technology Christ University, Bangalore, Karnataka, India

Dr. Poongodi C
Department of AI and Data Science Engineering, School of Engineering and Technology Christ University, Bangalore, Karnataka, India

Dr. Rekha V
Department of AI and Data Science Engineering, School of Engineering and Technology Christ University, Bangalore, Karnataka, India

Abstract:

Accurate air quality forecasting is critical to reducing health risks associated with poor air quality, and it also aids in urban planning. However, many locations do not have sufficient sensing infrastructure to generate reliable data and are reliant on fragmented datasets that are often noisy. [8] PM2.5 is an important pollutant in the air that we measure, and we use various forecasting methods to determine its level. Both time-series analysis and deep-learning techniques have been used to create models to forecast PM2.5 concentrations. We trained LSTM Networks,, ANN-LSTM hybrids and Prophets on daily PM2.5 data. The Bi-LSTM and LSTM Networks provided superior accuracy and performance when compared with all other models. This paper describes the creation of a unified, data-driven air-quality forecasting framework that can generate short-term forecasts based on historical air pollution records in areas where there is no air quality monitoring system in place. This framework utilises robust preprocessing, outlier detection, median/mode imputation, temporal feature extraction, a systematic exploratory analysis of the pollutant dynamics, and the development of seven different models of air quality forecasting using Linear Regression, Support Vector Regression, Random Forest, Gradient Boosting, XGBoost, Artificial Neural Networks, and LSTM time-series modelling. Experimental evaluation demonstrates that the nonlinear and deep learning models outperform the linear models by a significant margin, [10]with ANN achieving an approximately 93% accuracy rate and LSTM producing effective forecasts with the use of temporal dependencies. In contrast to sensor-only systems, the air quality forecasting framework described herein provides a cost-effective, scalable, and practical alternative to generating real-time alerts, decision-making support, and research insight from data-sparse environments, and is well-positioned for further integration with live application-programming interfaces, meteorological data, and hybrid technologies to continually improve predictive reliability.

Keywords:

Air pollution, machine learning, deep learning, time-series forecasting, LSTM, artificial neural networks, environmental monitoring, regression models, data preprocessing, predictive modeling.