Performance of Machine Learning Algorithms for Predicting Air Pollution Parameters

Stan, Brianna (School: Lauder-Reut Educational Complex)

Air pollution is one of the world's most critical challenges today, generating numerous premature deaths and diseases and negatively impacting the environment. Therefore, forecasting air quality parameters is a solution to providing valuable information for local communities and authorities that helps them detect patterns, identify sources, and warn citizens against potential risks due to very high levels. This study analyzes and compares machine learning models like LSTM, a generally used recurrent neural network in this area, and ARIMA, SARIMAX, BVAR, and VAR, which are not common in the research literature about air prediction. The data used in the project contains measurements of different pollution parameters collected between 2001 and 2018 from 12 stations located in Madrid, Spain. A map of the station positions was generated using their coordinations and OpenStreetMap API. Predicting concentrations of NO2 and PM2.5 for the next 24 hours after the models were trained using the same existing time series data provided information about the algorithms' reliability, accuracy, and efficiency. The analysis of the prediction results of the models used MAPE, RMSE, MSE, and MAE as evaluation methods. The project's goal was to test the performance of several machine learning algorithms, develop a methodology for evaluating and comparing them and identify the most accurate one. Future work involves determining the variability of pollution predictions regarding climate data, including temperature, humidity, wind speed, and direction, building a local air pollution monitoring network using an original IoT device, and allowing residents to collect data and generate analyzed reports.

Awards Won:

Fourth Award of $500