Paper ID: IRCUWU2024-238 *Oral*

Development of a machine learning model for air quality forecasting: leveraging long-term meteorological data analysis to predict air quality index in Colombo District

R.M.S.I. Rathnayaka✉, I.U. Hewapathirana

Software Engineering Teaching Unit, University of Kelaniya, Kelaniya, Sri Lanka

✉*isuruni.rathnayaka0508@gmail.com, +94704478064*

Air quality is a critical aspect of environmental health, directly impacting individuals and the broader ecosystem. Therefore, real-time monitoring and understanding the factors influencing air quality are crucial. The most typical reasons for air pollution are vehicle emissions, organic waste burning, and petroleum refining. However, other factors have arisen as causes of air pollution. Although meteorological factors are natural phenomena, they have been changing detrimentally due to human actions. Extreme meteorological events may significantly influence air quality. In Sri Lanka, a region with its own set of environmental challenges, understanding the dynamics of air quality is important. Over the past decade, Sri Lanka has witnessed notable shifts in weather patterns, with potential implications for human well-being. Available data indicates that Colombo often experiences high levels of air pollution. Recognizing these factors, this research introduced a model for real-time forecasting of the Air Quality Index (AQI) based on meteorological factors, emphasizing the Colombo district. The research focused on the period from 2020 to 2023, using a dataset that includes daily meteorological factors, wind speed, temperature, atmospheric pressure, rainfall, and relative humidity, alongside daily AQI values for the Colombo district. A temporal analysis identifies long-term trends and patterns in air quality. The study leveraged five machine learning algorithms: Linear Regression, Random Forest Regression, Gradient Boosting Regression, Support Vector Regression, and Long Short-Term Memory Network to develop models for predicting air quality based on meteorological factors. It also evaluated the performance of these machine learning models using metrics such as Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and R-Squared to determine each model's reliability in predicting the AQI. In conclusion, this research aims to discuss the role of weather variables in shaping air quality in the Colombo district. The outcomes contribute to understanding air quality in Sri Lanka and the broader global discourse on utilizing advanced technologies for environmental monitoring and forecasting. With insights into the predominant weather factors influencing air quality, decision-makers can formulate policies to improve the region's air quality based on seasonal weather pattern changes.

Keywords: Air quality index; meteorological factors; machine learning models; Colombo district; environmental monitoring

Underlined is the presenting author.