Air Quality Prediction Using Machine Learning

RM Fernando[#], WMKS Ilmini, and DU Vidanagama

Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

[#]35-CS-18-0001@kdu.ac.lk

The main basis of human survival is Air. The Air Quality Index is the value that qualitatively describes the condition of air quality. The greater the Air Quality Index, the more threatening risk to human health and environment. In Sri Lanka, poor air quality is a huge concern, especially in cities like Colombo and Kandy. Accurate Air Quality prediction will minimize health issues that can occur due to air pollution. This research has attempted to identify the best-suited machine learning algorithm-based approach to predict accurate air quality based on PM2.5 concentration in Colombo. In order to identify the most influenced air pollution concentrations for the air quality prediction purpose, correlation analysis was conducted. In this research, PM2.5 was predicted in Colombo city using 4 related air pollution concentrations including SO_2 concentration, NO_2 concentration, PM2.5 concentration & PM10 concentration. In order to get higher prediction accuracy, the gathered dataset was pre-processed by prediction beforehand. The prediction model trained and tested using machine learning algorithms such as KNN, Multiple Linear Regression, Support Vector Machines, and Random Forest. Multiple Regression was identified as the most suited prediction model which was able to gain 94% higher accuracy.

Keywords: *air quality, concentration, correlations, machine learning, pollution*