ABSTRACT

Recently, poor air quality has become a problem due to the many human activities that cause air pollution. Therefore, a system for predicting future pollutant concentrations is needed to increase public awareness and alertness. This research has created a prediction system using the LightGBM algorithm for $PM_{2.5}$ and CO_2 pollutant parameters with additional meteorological parameters as predictors. The large number of predictors used has the potential to cause overfitting, so this research proposes a parameter reduction method using PCA to increase prediction accuracy. Data for predictions were obtained from 3 measurement stations at Telkom University, Bandung, with a measurement period of approximately two months. The number of valid datasets is 918 for each of the five parameters at each measurement station, with data gaps filled using median values so that they can be used for predictions. The prediction results show that the best accuracy for $PM_{2.5}$ is at the Deli station, which uses PCA with a MAPE of 21.5%, and for CO_2, it is achieved at the Deli station without PCA with a MAPE of 4.8%. Based on its accuracy, PCA was less suitable if there were outliers in the dataset, but PCA was ideal for homogeneous datasets. Overall, the prediction results based on accuracy for $PM_{2.5}$ was included in the feasible category, and CO_2 was included in the accurate and very accurate category. To optimize prediction results, especially in the long term, it is necessary to retrain with a complete and up-to-date dataset to suit air conditions better.

Keywords: CO_2, LightGBM, PCA, $PM_{2.5}$