ABSTRACT

Comparison of KNN, SVM, and Random Forest Model Performance in Air Quality Classification in Jakarta

By

Ernawati Prahesta Kurnia Sari

21110011

The decline in air quality in Jakarta has serious impacts on public health, requiring effective analytical methods to classify air pollution levels. This study aims to evaluate and compare the performance of three classification algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF), in classifying air quality in Jakarta. The evaluation was carried out using performance metrics such as accuracy, precision, recall, and F1-score. The data used were obtained from air quality measurements in Jakarta based on the Air Pollution Standard Index (ISPU). Based on the comparison results, the Random Forest model achieved the best performance with an accuracy of 92%, precision of 66%, recall of 64%, and F1-score of 65%, while the KNN model achieved an accuracy of 89%, precision of 64%, recall of 58%, and F1-score of 60%, and the SVM model achieved an accuracy of 88% with precision, recall, and F1-score values of 65%, 54%, and 58%, respectively. However, the dataset used was imbalanced, leading to reduced performance on minority classes. To address this issue, the SMOTE method was applied to the Random Forest model. After applying SMOTE, the accuracy decreased to 83%, but precision, recall, and F1-score increased to 87%, 83%, and 82%, respectively, indicating improved performance balance across classes despite the reduction in overall accuracy. The results of this study are expected to provide recommendations for classification models that can be implemented in machine learning-based air quality monitoring systems.

Keywords: air quality, classification, k-nearest neighbors, machine learning, random forest, support vector machine