ABSTRACT

Indonesia has a dominant tropical climate, resulting in limited temperature variations but diverse rainfall patterns. The variability of rainfall is closely associated with its impacts on various aspects of human life and business activities. Therefore, rainfall information is crucial in decision-making processes. However, conducting an analysis requires specific stages and methods. Hence, this research aims to find the best method between C4.5 and K-Nearest Neighbors, both of which are data mining algorithms, to classify rainfall data. Both algorithms are employed to build classification models based on relevant attributes. Subsequently, these models are tested and evaluated using various metrics such as Accuracy, Precision, Recall, and F1-Score. The research also applies Hyperparameter Tuning with the RandomizeSearchCV method to obtain the best parameters that can yield maximum accuracy. The research results indicate that both algorithms perform well in rainfall classification. Initially, based on the accuracy values obtained using the default parameters of both algorithms, C4.5 achieved a higher accuracy of 81.42%, while K-Nearest Neighbors scored only 78.10%. However, after utilizing the best parameters resulting from the application of Hyperparameter Tuning using RandomizedSearchCV, a significant change in accuracy occurred for K-Nearest Neighbors, reaching 83.37%, while C4.5 slightly increased to 82.56%.

Keywords: Rainfall, Indonesian climate, classification, C4.5 algorithm, K-Nearest Neighbor, Data mining.