ABSTRACT

The advancement of Data Mining today plays a significant role in assisting decision-makers in selecting appropriate solutions to various problems. Data Mining involves analyzing data through the integration of statistical methods, mathematics, and machine learning techniques. In the context of Hybrid Data Mining, a key influencing factor is the generation of diverse classifiers to construct a generalized model. This study proposes the integration of two Data Mining techniques—clustering and classification—referred to as Hybrid Data Mining. By grouping data based on similarities and differences, this approach helps reduce data errors and enhances the accuracy of the classification models.

The Hybrid Data Mining method, which involves clustering data using the k-medoid algorithm followed by classification, demonstrates that three out of five models yield effective results. Using student learning data as test input, the hybrid approach shows varying impacts on model accuracy. For instance, the Decision Tree model initially achieves 90% accuracy, but after applying Hybrid Data Mining, the accuracy drops to 80%, indicating no performance improvement. In contrast, the Naive Bayes model improves from 86% to 92%, suggesting a 6% accuracy gain, making it well-suited for the hybrid method. On the other hand, the Random Forest model sees a decline in accuracy from 95% to 91% after hybridization, indicating it may not be compatible with this approach. Finally, the K-Nearest Neighbors (KNN) model shows improvement from 75% to 84%, signifying its suitability for use within the Hybrid Data Mining framework.

Keywords : Hybrid Data Mining, clustering, classification