Abstract

This study analyzes the performance of hybrid methods in improving accuracy on imbalanced data using Dengue Hemorrhagic Fever Case Data from 2017 to 2021 in Bandung City. The attributes used in this study consist of Total Population, Total Male, Elementary School Graduation, Junior High School Graduation, High School Graduation, College Graduation, Rainfall, Average Temperature, Humidity, Male Cases, Number of Cases, and Class. This research combines five Machine Learning methods, such as Decision Tree, Support Vector Machine, Artificial Neural Network, K-Nearest Neighbor, and Naïve Bayes. Hybrid Methods used in this research are Voting and Stacking methods. The oversampling methods used to handle imbalanced data in this study are Random Oversampling and Adasyn. The results show that Voting and Stacking without Random Oversampling and Adasyn get the same accuracy of 88,88%. While using Random Oversampling, voting gets an accuracy of 95,37% and stacking gets an accuracy of 96,29%. While using Adasyn, voting gets an accuracy of 94,44% and stacking gets an accuracy of 97,22%. Based on the results obtained, it can be concluded that the Random Oversampling and Adasyn Method can improve the performance of the Machine Learning hybrid method on imbalanced data. The contribution of this research is to provide information on the study and analysis of the implementation of the Random Oversampling and Adasyn methods in improving the performance of the Voting and Stacking methods in hybrid classification.

Keywords: Classification, Machine Learning, Hybrid Methods, Random Oversampling, Adasyn