ABSTRACT

Stroke is a cerebrovascular disease (brain blood vessel disorder) caused by brain damage, characterized by the death of brain tissue. According to the World Health Organization (WHO), 70% of deaths are caused by stroke, making it the second leading cause of death globally. In Indonesia, stroke is also one of the highest causes of mortality. Based on this, the present study applies the K-Nearest Neighbor (KNN) and *Adaptive boosting* algorithms to predict stroke occurrence, utilizing the *SMOTE* technique to address data *imbalance*. However, this research does not only consider accuracy but also recall. The recall score measures how well the model detects all actual positive cases. In healthcare cases, it would be very dangerous if someone with a disease is classified as healthy; therefore, it is preferable for the model to have a high recall value. The study aims to identify the best accuracy performance between the KNN and *Adaptive boosting* algorithms and assess the impact of *SMOTE* on their accuracy. The algorithms predict stroke based on features including gender, age, hypertension, heart disease, marital status, work type, residence type, average glucose level, BMI, and smoking status. Using an 80% *training* and 20% *testing* data split without *SMOTE*, the K-Nearest Neighbor algorithm achieved the highest accuracy at 94%, while *Adaptive boosting* yielded 91%. After applying *SMOTE*, the best accuracy was achieved by *Adaptive boosting* at 89%, while K-Nearest Neighbor with *SMOTE* achieved 86%. Although the accuracy decreased after implementing *SMOTE*, there was an improvement in precision, recall, and F1-Score *value*s.

Keyword : *Adaptive boosting, K-Nearest Neighbor, Recall, SMOTE, Stroke.*