

## ABSTRAK

Penyakit stroke adalah penyakit gangguan fungsi syaraf pada otak, terjadi secara mendadak, progresif, dan cepat sehingga banyaknya keterlambatan masyarakat dalam penanganan stroke. Gangguan tersebut disebabkan oleh gangguan peredaran darah di otak atau bisa juga disebabkan pecahnya pembuluh darah didalam otak. Untuk itu diperlukan cara untuk mengklasifikasikan kecenderungan pasien mengalami serangan stroke berdasar pada data pribadi pasien. Selain itu bagaimana cara *balancing data* jika data yang digunakan mengalami *imbalanced data* dan metode *balancing* manakah yang paling efektif diantara SMOTE, ADASYN, dan Random Oversampler.

Tujuan dari penelitian ini (menyelesaikan masalah) adalah melakukan klasifikasi pasien penyakit stroke menggunakan algoritma Random Forest. Oleh karena data pasien penyakit stroke yang didapat tidakimbang, maka akan dilakukan metode balancing data menggunakan teknik balancing SMOTE, ADASYN, dan Random Oversampler. Hasil akurasi yang didapat akan dibandingkan untuk menentukan akurasi yang terbaik.

Data yang digunakan adalah data stroke yang diambil dari kaggle.com yang akan diolah melalui beberapa tahap yaitu *selection, cleaning, transformation data, balancing data, data split testing training*, klasifikasi menggunakan *random forest*, dan hasil akurasi yang didapat.

Pada tahap pengujian dilakukan 12 pengujian menggunakan variasi jumlah pohon dan mendapatkan hasil akurasi tertinggi sebesar 85,67992% menggunakan metode balancing ADASYN dan variasi jumlah 2048 pohon menggunakan *KFold* 10.

Dengan adanya penelitian ini diharapkan dapat membantu untuk mengklasifikasikan kecenderungan pasien mengalami serangan stroke berdasar pada data pribadi pasien.

**Kata Kunci:** Stroke, *RandomForestClassifier* , Klasifikasi, *Balancing*, SMOTE, ADASYN, Random Oversampler

## ABSTRACT

Stroke is a disease of impaired nerve function in the brain, occurs suddenly, progressively, and quickly so that there are many delays in the handling of stroke. The disorder is caused by blood circulation disorders in the brain or can also be caused by rupture of blood vessels in the brain. For this reason, a way is needed to classify the patient's tendency to experience a stroke attack based on the patient's personal data. In addition, how to balance data if the data used is imbalanced data and which balancing method is the most effective among SMOTE, ADASYN, and Random Oversampler.

The purpose of this research (solving the problem) is to classify stroke patients using the Random Forest algorithm. Because the stroke patient data obtained is not balanced, the data balancing method will be carried out using the SMOTE, ADASYN, and Random Oversampler balancing techniques. The accuracy results obtained will be compared to determine the best accuracy.

The data used is stroke data taken from kaggle.com which will be processed through several stages, namely selection, cleaning, data transformation, data balancing, data split testing training, classification using random forest, and accuracy results obtained.

In the testing phase, 12 tests were carried out using variations in the number of trees and obtained the highest accuracy of 85.67992% using the ADASYN balancing method and a variation in the number of 2048 trees using KFold 10.

With this research, it is hoped that it can help to classify the tendency of patients to experience a stroke attack based on the patient's personal data.

***Keywords: Stroke, RandomForestClassifier, Classification, Balancing, SMOTE, ADASYN, Random Oversampler***