

ABSTRAK

Sholikhati, Manisha Elok, 2021, Klasifikasi Penyakit Stroke Menggunakan Metode SMOTE-XGBoost. Skripsi, Program Studi Statistika, Universitas Muhammadiyah Semarang, Pembimbing: I. Dr. Rochdi Wasono, M.Si, II. Prizka Rismawati Arum, S.Si, M.Stat

Stroke merupakan penyebab kematian nomor dua dan penyebab kecacatan nomor tiga di dunia. Deteksi dini penyakit stroke sangat penting. Jika pengobatan penyakit ini tertunda, sel-sel otak akan mati dalam beberapa menit. Klasifikasi merupakan model prediksi yang dapat memprediksi label kategori dari setiap pengamatan. Salah satu masalah klasifikasi adalah ketidakseimbangan dalam klasifikasi data. Tujuan dari penelitian ini yaitu membandingkan kinerja algoritma XGBoost dan SMOTE (Synthetic Minority Oversampling Technique)-XGBoost menggunakan software Python untuk mengklasifikasikan apakah pasien rentan terhadap stroke. Data yang digunakan dalam penelitian ini adalah data stroke dari situs Kaggle. Data penelitian dibagi menjadi data training dan testing (70%:30%). Kemudian dilakukan analisis menggunakan metode XGBoost dan SMOTE-XGBoost untuk mengklasifikasi pasien, apakah cenderung mengalami penyakit stroke atau tidak. Nilai AUC digunakan untuk mengukur dan membandingkan dua model klasifikasi penyakit stroke. Hasilnya menunjukkan bahwa nilai AUC dari kedua model hampir sama yakni nilai AUC model XGBoost sebesar 80,5% dan nilai AUC model SMOTE-XGBoost sebesar 81,5%. Oleh karena itu, dapat dikatakan bahwa penggunaan SMOTE kurang berhasil meningkatkan kinerja XGBoost dalam mengklasifikasikan status pasien stroke.

Kata kunci: stroke, SMOTE-XGBoost, XGBoost

ABSTRACT

Sholikhati, Manisha Elok, 2021, Classification of Stroke Using SMOTE-XGBoost Method. Undergraduate Thesis, Statistical Studies Program, Muhammadiyah University of Semarang, Advisor: I. Dr. Rochdi Wasono, M.Si, II. Prizka Rismawati Arum, S.Si, M.Stat

Stroke is the number two cause of death and the number three cause of disability in the world. Early detection of stroke is very important. If treatment of the disease is delayed, brain cells will die within minutes. Classification is a predictive model that can predict category labels from each observation. One of the problems of classification is the imbalance in data classification. The goal of the study was to compare the performance of the XGBoost and SMOTE-XGBoost algorithms using Python software to classify whether patients were susceptible to stroke. The data used in the study was stroke data from the Kaggle site. Research data is divided into training and testing data (70%:30%). Then the analysis was conducted using XGBoost and SMOTE (Synthetic Minority Oversampling Technique)-XGBoost methods to classify the patient, whether or not he was likely to have a stroke. AUC values used to measure and compare both models of stroke classification. Then analyzed using the XGBoost and SMOTE-XGBoost methods to classify patients, whether they tend to have a stroke or not. AUC values are used to measure and compare two models of classification of stroke. The results showed that the AUC values of both models were almost the same: the AUC value of the XGBoost model was 80.5% and the AUC value of the SMOTE-XGBoost model at 81.5%. Therefore, it can be said that the use of SMOTE is less successful in improving the performance of XGBoost in classifying the status of stroke patients.

Keywords: stroke, SMOTE-XGBoost, XGBoost