

IMPLEMENTASI ALGORITMA *RANDOM FOREST* DALAM PREDIKSI *STROKE* DENGAN PENGGUNAAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE*

Johanes Gerald

ABSTRAK

Stroke adalah salah satu masalah kesehatan global yang signifikan dan menjadi penyebab utama kecacatan serta kematian di seluruh dunia. Di Indonesia, penyakit *stroke* telah menjadi salah satu penyumbang penyakit paling mematikan. Menurut data pada profil kesehatan Indonesia tahun 2020, *stroke* menempati posisi ketiga dengan jumlah kasus sebanyak 1.789.261. Tujuan penelitian ini adalah untuk mengidentifikasi pasien yang berisiko tinggi terkena *stroke*. Algoritma yang digunakan adalah *Random Forest Classifier* dengan menggunakan *Synthetic Minority Oversampling Technique* untuk menyeimbangkan kelas data. Pada metode *Random Forest* menggunakan *Synthetic Minority Oversampling Technique* didapatkan hasil akurasi sebesar 95,61%, nilai presisi sebesar 93,66%, nilai recall sebesar 97,85%, dan nilai f1-score sebesar 95,71%. Sedangkan untuk model *Random Forest* didapatkan hasil akurasi sebesar 90,15%, nilai presisi sebesar 90,5%, nilai *recall* sebesar 90,15%, dan nilai *f1-score* sebesar 90,32%. Karena terjadi *imbalance class* maka algoritma *Random Forest* saja tidak tepat digunakan tanpa melakukan *resampling*. Sehingga *Random Forest – SMOTE* dapat digunakan sebagai salah satu algoritma untuk memprediksi *stroke*.

Kata kunci: Klasifikasi, *Stroke*, *Random Forest*, SMOTE.

**IMPLEMENTATION OF RANDOM FOREST ALGORITHM IN STROKE
PREDICTION USING SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE**

Johanes Gerald

ABSTRACT

Stroke is one of the significant global health issues, being a leading cause of disability and death worldwide. In Indonesia, stroke has emerged as one of the most fatal diseases. According to the 2020 Indonesia health profile data, stroke ranks third with a total of 1,789,261 reported cases. The aim of this research is to identify patients at high risk of stroke. The algorithm employed is the Random Forest Classifier utilizing the Synthetic Minority Oversampling Technique to balance the class data. In the Random Forest method using the Synthetic Minority Oversampling Technique, the results showed an accuracy of 95.61%, precision of 93.66%, recall of 97.85%, and an f1-score of 95.71%. Meanwhile, for the Random Forest model, the accuracy was 90.15%, precision was 90.5%, recall was 90.15%, and the f1-score was 90.32%. Due to class imbalance, using the Random Forest algorithm alone is not suitable without resampling. Therefore, Random Forest – SMOTE can be utilized as one of the algorithms to predict strokes.

Keywords: Classification, Stroke, Random Forest, SMOTE.