

INTISARI

PEMODELAN PREDIKSI RISIKO PENYAKIT *STROKE* MENGUNAKAN METODE *RANDOM FOREST*

Oleh:

MUFIDAH NUR AMALIAH

mufidahnuramaliah@gmail.com

Stroke merupakan gangguan saraf akibat aliran darah ke otak yang terhambat, faktor risiko seperti hipertensi, usia, dan gaya hidup sangat berperan dalam meningkatkan peluang seseorang terkena *stroke*. Dalam penelitian ini data penyakit diambil dari repositori *Kaggle* dengan jumlah 4981 sampel data yang terdiri dari 10 kolom prediktor dan 1 kolom target. Penelitian ini bertujuan untuk memprediksi apakah seseorang beresiko terkena *stroke* menggunakan algoritma *Random Forest*, yang dikenal sebagai metode *Machine Learning* yang andal untuk data yang kompleks. Tahapan analisis menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM). *Tools* yang digunakan pada penelitian ini *google colaboratory*, *streamlit*, *rapid miner*. Model prediksi mencapai tingkat akurasi sebesar 95,33%. Namun, terdapat ketidakseimbangan kelas dengan 831 sampel untuk kelas 0 (tidak *stroke*) dan hanya 37 sampel untuk kelas 1 (*stroke*). Rata-rata performa model berdasarkan metrik *Precision*, *Recall*, dan *F1-Score* masing-masing adalah 47,66%, 50%, dan 48,80%.

Kata Kunci : Prediksi, *Stroke*, *Random Forest*, *Machine Learning*, CRISP-DM

ABSTRACT

MODELING STROKE DISEASE RISK PREDICTION USING RANDOM FOREST METHOD

By:

MUFIDAH NUR AMALIAH

mufidahnurmaliah@gmail.com

Stroke is a neurological disorder due to obstructed blood flow to the brain, risk factors such as hypertension, age, and lifestyle play a role in increasing a person's chances of having a stroke. In this study, disease data was taken from the Kaggle repository with a total of 4981 data samples consisting of 10 predictor columns and 1 target column. This study aims to predict whether a person is at risk of stroke using the Random Forest algorithm, which is known as a reliable Machine Learning method for complex data. The analysis stages use the Cross Industry Standard Process for Data Mining (CRISP-DM). The tools used in this research are google colab, streamlit, rapid miner. The prediction model achieved an accuracy rate of 95.33%. However, there is a class imbalance with 831 samples for class 0 (no stroke) and only 37 samples for class 1 (stroke). The average model performance based on Precision, Recall, and F1-Score metrics are 47.66%, 50%, and 48.80%, respectively.

Keywords: Prediction, Stroke, Random Forest, Machine Learning, CRISP-DM