

ABSTRAK

Stroke menjadi salah satu penyakit serius di seluruh dunia dengan dampak fatal yang menyebabkan kecacatan fisik serta mental. Deteksi dan prediksi dini stroke penting untuk mencegah dampak buruk pada kesehatan pasien. Dalam upaya meningkatkan kualitas prediksi dan deteksi stroke, penelitian ini berfokus pada analisis dan pemodelan prediktif menggunakan teknik Exploratory Data Analysis (EDA) dan algoritma klasifikasi, khususnya Support Vector Machine (SVM) dan Random Forest Classifier (RFC). Melalui EDA (*Exploratory Data Analysis*), kita bisa mendapatkan pemahaman yang lebih mendalam tentang faktor risiko dan indikator potensial yang terkait dengan stroke. Hasil yang didapat menunjukkan bahwa melalui analisis korelasi, ditemukan usia (umur) memiliki korelasi positif yang kuat dengan kejadian stroke, menunjukkan semakin tua usia seseorang, semakin tinggi risiko terkena stroke dalam *dataset* ini. Model RFC memiliki performa lebih baik dengan akurasi 91% dibandingkan SVM dengan akurasi 90% pada data uji setelah penyetelan hiperparameter. Kedua model efektif dalam mengidentifikasi risiko stroke. Dalam perbandingan ROC-AUC, keduanya menunjukkan kinerja yang sangat baik, SVM memiliki hasil 0.901 dan RFC memiliki hasil 0.910.

Kata Kunci: Stroke, Prediksi, EDA, Support Vector Machine, Random Forest Classifier

ABSTRACT

Stroke has become one of the serious diseases prevalent worldwide, leading to fatal consequences and causing physical and mental disabilities. Early detection and prediction of stroke are crucial in preventing adverse health impacts on patients. In an effort to enhance the quality of stroke prediction and detection, this research focuses on exploratory data analysis (EDA) and predictive modeling using classification algorithms, specifically Support Vector Machine (SVM) and Random Forest Classifier (RFC). Through exploratory data analysis (EDA), a deeper understanding of risk factors and potential indicators related to stroke can be obtained. The findings reveal a strong positive correlation between age and stroke incidence in this dataset, indicating that older age increases the risk of stroke. The RFC model demonstrated superior performance with an accuracy of 91% compared to SVM's accuracy of 90% on the test data after hyperparameter tuning. Both models effectively identify stroke risk. In the ROC-AUC comparison, both SVM and RFC models demonstrated excellent performance, with SVM yielding a result of 0.901 and RFC yielding a result of 0.910.

Keywords: Stroke, Prediction, Exploratory Data Analysis (EDA), Support Vector Machine, Random Forest Classifier.