

ABSTRAK

Hipertensi adalah salah satu penyakit tidak menular yang sangat sering terjadi di berbagai negara dan bisa menyebabkan komplikasi penyakit lain dan juga penyakit jantung serta stroke. Situasi ini sering kali tidak menimbulkan gejala pada tahap awal, sehingga diperlukan metode yang dapat membantu dalam melakukan deteksi dan prediksi risiko secara lebih efektif. Perkembangan teknologi analisis data saat ini memungkinkan penggunaan metode machine learning untuk menemukan pola hubungan di antara variabel-variabel kesehatan dan memprediksi kemungkinan seseorang yang mengalami penyakit hipertensi ini. Penelitian kami ini bertujuan untuk menganalisis dan memprediksi risiko terkena penyakit hipertensi dengan menggunakan metode kombinasi yang menggabungkan algoritma Extreme Gradient Boosting (XGBoost) dan teknik penjelasan model berupa Shapley Additive Explanations (SHAP). Data yang digunakan pada penelitian ini merupakan data sekunder yang didapat dari platform Kaggle dengan jumlah 1008 data yang terdiri dari beberapa kelompok variabel seperti usia, indeks masa tubuh (BMI), status merokok, riwayat hipertensi, serta detak jantung. Model XGBoost ini digunakan untuk mengklasifikasikan risiko hipertensi, sedangkan metode SHAP digunakan untuk menjelaskan sejauh mana setiap variabel berkontribusi terhadap hasil prediksi yang dihasilkan oleh model tersebut. Penelitian ini juga menunjukkan bahwa pendekatan hybrid ini berhasil membuat model prediksi yang bagus dan juga membantu memahami dengan lebih jelas faktor-faktor apa saja yang memengaruhi risiko terjadinya hipertensi. Metode ini bisa menjadi salah satu cara yang baik dalam menganalisis data kesehatan dan membuat sistem prediksi penyakit yang menggunakan machine learning.

Kata kunci: *hipertensi, machine learning, XGBoost, SHAP, prediksi risiko*

ABSTRACT

Hypertension is a prevalent non-communicable disease globally, often leading to severe complications such as heart disease and stroke. This condition frequently remains asymptomatic in its early stages, necessitating more effective methods for detection and risk prediction. Advancements in data analysis technology now allow for the application of machine learning to identify complex patterns within health variables and predict the likelihood of hypertension. This study aims to analyze and predict hypertension risk using a hybrid approach that combines the Extreme Gradient Boosting (XGBoost) algorithm with Shapley Additive Explanations (SHAP) for model interpretability. The research utilizes secondary data sourced from Kaggle, comprising 1,008 records with variables including age, Body Mass Index (BMI), smoking status, medical history, and heart rate. While the XGBoost model is employed to classify hypertension risk, SHAP is utilized to explain the individual contribution of each variable to the model's predictions. The results demonstrate that this hybrid approach not only yields a high-performing predictive model but also provides a transparent understanding of the factors influencing hypertension risk. This methodology serves as a robust framework for analyzing healthcare data and developing machine learning-based disease prediction systems.

Keywords: *hypertension, machine learning, XGBoost, SHAP, risk prediction*