

ABSTRAK

Diabetes mellitus merupakan penyakit kronis dengan peningkatan kasus global, menjadikan deteksi dini sangat penting. Machine learning menawarkan pendekatan efektif untuk klasifikasi pasien diabetes. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Extreme Gradient Boosting (XGBoost) dalam mengklasifikasikan pasien diabetes menggunakan dataset Pima Indian Diabetes, khususnya untuk mengevaluasi pengaruh teknik normalisasi data. Penelitian ini membandingkan kinerja model XGBoost yang dilatih tanpa normalisasi, dengan normalisasi menggunakan MinMax Scaler, dan dengan normalisasi menggunakan RobustScaler. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC. Hasil menunjukkan bahwa model XGBoost tanpa normalisasi memberikan kinerja terbaik dengan akurasi 83,55%, recall 83,95%, F1-score 0,7816, dan AUC 0,8613. Penerapan MinMaxScaler dan RobustScaler tidak meningkatkan performa secara signifikan, bahkan dalam beberapa kasus menurunkannya hingga ke baseline (AUC=0.50), menandakan ketidakcocokan metode normalisasi tersebut dengan dataset atau model. Hal ini menunjukkan bahwa XGBoost secara intrinsik kuat terhadap skala fitur pada dataset ini. Penelitian ini menyimpulkan bahwa XGBoost merupakan algoritma yang andal untuk klasifikasi diabetes, bahkan tanpa proses normalisasi data.

Kata kunci: *Diabetes, XGBoost, Machine Learning, Normalisasi Data, Klasifikasi*

ABSTRACT

Diabetes mellitus is a chronic disease with a rising global incidence, making early detection very important. Machine learning offers an effective approach for classifying diabetes patients. This study aims to analyze and compare the performance of the Extreme Gradient Boosting (XGBoost) algorithm in classifying diabetes patients using the Pima Indian Diabetes dataset, particularly to evaluate the impact of data normalization techniques. This study compares the performance of XGBoost models trained without normalization, with normalization using MinMax Scaler, and with normalization using RobustScaler. Evaluation was conducted using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results show that the XGBoost model without normalization provides the best performance with an accuracy of 83.55%, recall of 83.95%, F1-score of 0.7816, and AUC of 0.8613. The application of MinMaxScaler and RobustScaler does not significantly improve performance, and in some cases even reduces it to the baseline (AUC=0.50), indicating a mismatch of these normalization methods with the dataset or model. This suggests that XGBoost is intrinsically robust to feature scaling in this dataset. This study concludes that XGBoost is a reliable algorithm for diabetes classification, even without data normalization processes.

Key word: *Diabetes, XGBoost, Machine Learning, Data Normalization, Classification*