

ABSTRAK

Diabetes mellitus terus meningkat secara global, dengan 537 juta penderita pada 2021 dan diperkirakan mencapai 643 juta pada 2030. Dataset Pima Indians Diabetes (PIDDD) sering digunakan dalam studi prediksi risiko diabetes karena representasinya terhadap populasi berisiko tinggi, menjadikannya referensi penting dalam pengembangan model berbasis *machine learning*. Penerapan *machine learning* dalam sektor Kesehatan semakin berkembang, terutama dengan algoritma XGBoost untuk prediksi klinis. Namun, kompleksitas XGBoost menimbulkan tantangan dalam interpretabilitas, untuk itu digunakan metode SHAP untuk menjelaskan kontribusi variabel. Model XGBoost mencapai akurasi sebesar 83% pada data uji. Analisis SHAP mengidentifikasi variabel Glucose, BMI, dan Age sebagai prediktor paling signifikan untuk diabetes. Integrasi XGBoost dan SHAP menghasilkan model prediksi diabetes yang tidak hanya akurat tetapi juga dapat diinterpretasikan. Temuan ini menegaskan potensi *machine learning* dalam memfasilitasi deteksi dini dan meningkatkan pengambilan keputusan klinis, terutama dalam mengidentifikasi faktor risiko kunci bagi pasien yang berisiko diabetes.

Kata Kunci: Diabetes Mellitus, XGBoost, SHAP, Prediksi Risiko Diabetes