

ABSTRAK

Diabetes melitus merupakan penyakit metabolik kronis dengan prevalensi yang terus meningkat di Indonesia, sehingga memerlukan metode deteksi dini yang praktis dan akurat. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, dan Extreme Gradient Boosting (XGBoost) dalam memprediksi risiko diabetes menggunakan data klinis statis yang disimulasikan dengan karakteristik noise dari perangkat spektroskopi near-infrared (NIR) non-invasif Glucosense (RMSE = 17,54 mg/dL). Dataset gabungan sebanyak 100.025 sampel diolah dengan teknik SMOTE untuk menangani ketidakseimbangan kelas dan dievaluasi menggunakan berbagai metrik performa klinis. Hasil eksperimen menunjukkan bahwa XGBoost mencapai kemampuan diskriminasi probabilistik terbaik dengan nilai ROC-AUC tertinggi (0,9763) dan F1-score tertinggi (0,7553), menjadikannya model paling seimbang untuk triase risiko umum. Sementara itu, SVM unggul pada nilai Recall (90,79%) yang cocok untuk skrining primer yang mengutamakan sensitivitas, dan Random Forest memberikan akurasi tinggi serta interpretabilitas fitur yang baik untuk Clinical Decision Support System. Analisis feature importance mengungkap bahwa kadar HbA1c, glukosa spektroskopi, dan usia merupakan kontributor utama dalam prediksi risiko. Secara keseluruhan, penelitian ini menegaskan potensi machine learning sebagai alat bantu skrining risiko diabetes yang andal berbasis data klinis non-invasif dengan pemilihan model yang disesuaikan pada kebutuhan spesifik layanan kesehatan.

Kata kunci : *Diabetes melitus; machine learning; prediksi risiko; spektroskopi NIR non-invasif; Random Forest; XGBoost*