

## ABSTRAK

Penyakit kronis seperti diabetes mellitus, kanker, stroke, gagal ginjal kronis, hipertensi, dan penyakit jantung merupakan masalah kesehatan utama yang semakin meningkat di Indonesia, terutama pada kelompok usia lanjut. Upaya preventif menjadi langkah penting untuk menekan angka penderita di masa depan, salah satunya melalui pemanfaatan data hasil *general checkup* yang dianalisis untuk mendeteksi potensi risiko penyakit. Dalam praktiknya saat ini, dokter menggunakan wawancara medis, pemeriksaan fisik, dan pemeriksaan penunjang yang diproses melalui *clinical reasoning* untuk menghasilkan sebuah diagnosis penyakit. Namun, dengan semakin kompleks dan besarnya volume data medis, proses ini memerlukan dukungan dari model *machine learning* untuk meningkatkan akurasi dan efisiensi pengambilan keputusan klinis.

Namun demikian, sebagian besar penelitian terdahulu memanfaatkan *machine learning* hanya untuk memprediksi satu jenis penyakit, tanpa mempertimbangkan integrasi pengetahuan klinis secara langsung dalam proses *preprocessing* maupun pemodelan prediktif. Selain itu, tantangan besar lainnya adalah permasalahan kualitas data medis, seperti adanya *missing values*, outliers, dan data yang *imbalanced*. Oleh karena itu, disertasi ini mengusulkan pendekatan baru dengan mengintegrasikan *domain knowledge* dalam proses *preprocessing* dan pengembangan model ensemble learning berbasis *tree-based classifiers* yang dimodifikasi.

Penelitian ini menggunakan dataset GCU dari Yakes Telkom periode 2019–2021, yang terdiri dari data pemeriksaan untuk enam penyakit kronis. Tahapan penelitian mencakup: 1) eksplorasi data dan penanganan permasalahan data imbalanced berbasis Weighted SMOTE yang dirancang berdasarkan *domain knowledge*, 2) pengembangan model ensemble learning dengan memasukkan bobot fitur yang diperoleh dari *domain knowledge* ke dalam algoritma Random Forest, XGBoost, dan AdaBoost, dan 3) evaluasi kinerja model menggunakan metrik F1-Score, Balanced Accuracy Score, dan ROC-AUC.

Hasil penelitian menunjukkan bahwa pendekatan yang diusulkan mampu meningkatkan performa prediksi secara signifikan dibandingkan model tanpa pembobotan. Model dengan integrasi *domain knowledge* dalam Weighted SMOTE dan ensemble learning menghasilkan rata-rata F1-Score lebih dari 85%, yang lebih tinggi dibandingkan baseline model tanpa *domain knowledge* dengan peningkatan performansi sebesar 15%. Selain itu, model ini mampu memprediksi lebih dari satu jenis penyakit kronis, serta memberikan *insight* bagi dokter dalam menyusun tindakan preventif yang lebih tepat sasaran. Kontribusi utama dari disertasi ini adalah pengembangan metode *preprocessing* dan model prediksi berbasis *machine learning* yang terintegrasi dengan *domain knowledge* untuk meningkatkan akurasi dan keandalan prediksi penyakit kronis. Pendekatan ini diharapkan dapat memberikan manfaat nyata bagi pasien, tenaga medis, serta pihak asuransi dalam rangka mendukung pengambilan keputusan klinis yang lebih baik dan efisien.

**Kata Kunci:** Penyakit Kronis, Domain Knowledge, Machine Learning, Imbalanced Data, Ensemble Learning