

KLASIFIKASI PENYAKIT GINJAL KRONIS MENGUNAKAN *EXPLAINABLE BOOSTING MACHINE* DENGAN PENDEKATAN IMPUTASI *HYBRID KNN-* INTERPOLASI LINIER

Nama Mahasiswa : Abellia Putri Dwi Masita
NIM : 11221020
Dosen Pembimbing Utama : Bima Prihasto S.Si., M.Si., Ph.D.
Dosen Pembimbing Pendamping : Ramadhan Paninggalih S.Si., M.Si., M.Sc.

ABSTRAK

Penyakit Ginjal Kronis (*Chronic Kidney Disease/CKD*) memerlukan deteksi dini yang akurat untuk mendukung pengambilan keputusan medis. Penelitian ini mengevaluasi kinerja, interpretabilitas, dan ketahanan *Explainable Boosting Machine* (EBM) terhadap *Gaussian noise* dalam klasifikasi CKD. Penelitian menerapkan metode imputasi *hybrid* menggunakan *K-Nearest Neighbor* (KNN) dan interpolasi linier dengan variasi nilai $k = 3, 5, 7, 9, \text{ dan } 11$, serta menggunakan tiga skenario seleksi fitur, yaitu *full* dataset, top 10 fitur berdasarkan korelasi *Pearson*, dan fitur rekomendasi berdasarkan jurnal Mallamaci dan Tripepi (2024). Dataset dibagi menggunakan rasio 80:20 dan dilakukan *scaling* fitur numerik menggunakan *RobustScaler*. Penelitian membandingkan model EBM dengan *Logistic Regression*, *Random Forest*, dan *XGBoost* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta mengevaluasi pengaruh *Gaussian noise* sebesar 50% pada data *training* dan *testing*. Hasil penelitian menunjukkan bahwa model EBM mencapai performa terbaik pada skenario *full* dataset dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 1,00 pada seluruh variasi nilai k . Pada skenario top 10 korelasi *Pearson*, model EBM memperoleh *accuracy* antara 0,97 hingga 0,99 dan *F1-score* antara 0,98 hingga 0,99, sedangkan pada skenario fitur jurnal memperoleh *accuracy* sebesar 0,96 dan *F1-score* sebesar 0,97. Hasil tersebut menunjukkan bahwa EBM mampu menghasilkan performa yang setara dengan *Random Forest* dan lebih baik dibandingkan *Logistic Regression* pada sebagian besar skenario pengujian. Pengujian *Gaussian noise* menunjukkan bahwa penambahan *noise* pada data *testing* memberikan dampak terbesar terhadap performa model. Penurunan terbesar terjadi pada skenario fitur jurnal dengan *accuracy* sebesar 0,76 dan *F1-score* sebesar 0,83 pada $k = 3$. Meskipun demikian, model tetap mempertahankan nilai *recall* yang tinggi, yaitu antara 0,92 hingga 1,00 pada sebagian besar skenario pengujian. Interpretasi global model menunjukkan bahwa fitur serum *sc*, *hemo*, *rbcc*, *dm*, dan *sg* memberikan kontribusi terbesar terhadap klasifikasi CKD. Hasil penelitian menunjukkan bahwa model EBM tidak hanya menghasilkan performa klasifikasi yang tinggi, tetapi juga memberikan interpretasi yang transparan dan relevan secara klinis.

Kata kunci: CKD, EBM, imputasi data, interpretabilitas model, *Gaussian noise*



(Halaman ini sengaja dikosongkan)

INSTITUT TEKNOLOGI
KALIMANTAN