

---

**Peningkatan Algoritma C4.5 Menggunakan Ensemble Learning Untuk Mendeteksi Penyakit Ginjal**

**Agusviyanda<sup>1</sup>, Rita Novita<sup>2</sup>, Alfa Saleh<sup>3</sup>, Muhammad Jamaris<sup>4</sup>**

Informatika Kesehatan, Fakultas Kesehatan & Informatika, Institut Kesehatan Payung Negeri Pekanbaru<sup>1,2</sup>

Informatika, Fakultas Sains dan Teknologi, Universitas Samudra<sup>3</sup>  
Teknik Informatika, Universitas Sains dan Teknologi Indonesia<sup>4</sup>

Email: [agusviyanda27@payungnegeri.ac.id](mailto:agusviyanda27@payungnegeri.ac.id)<sup>1</sup>, [ritanovita@payungnegeri.ac.id](mailto:ritanovita@payungnegeri.ac.id)<sup>2</sup>,  
[alfasaleh@unsam.ac.id](mailto:alfasaleh@unsam.ac.id)<sup>3</sup>, [jamaris@usti.ac.id](mailto:jamaris@usti.ac.id)<sup>4</sup>

Corresponding Author : [agusviyanda27@payungnegeri.ac.id](mailto:agusviyanda27@payungnegeri.ac.id)

**ABSTRACT**

*Early detection of kidney disease is essential to prevent complications and improve patient outcomes. However, the diagnostic process faces challenges due to non-specific symptoms and the problem of imbalanced class distribution in patient data. This study aims to improve the performance of the C4.5 algorithm for kidney disease detection by integrating several advanced techniques. The methodology includes preprocessing using Label Encoder and Ordinal Encoder for categorical features, data balancing with the SMOTE-ENN method, and feature selection via LASSO. The C4.5 base model is then enhanced through ensemble learning with AdaBoost to boost predictive performance. Experimental results indicate that the AdaBoost+C4.5 combination significantly outperforms both the basic C4.5 and other ensemble methods previously studied. The proposed model achieves an accuracy of 99%, higher than the results of XGBoost and stacking ensemble approaches reported in recent literature. These findings demonstrate that the synergy of boosting, data balancing, and robust feature selection can effectively address the challenges of early kidney disease detection. In conclusion, the proposed machine learning framework shows strong potential as a clinical decision support system to aid timely and accurate diagnosis of kidney disease.*

**Keywords:** *Kidney disease, Machine learning, C4.5, Adaboost, SMOTE-ENN, LASSO.*

**I. Pendahuluan**

Penyakit ginjal masih menjadi salah satu penyebab utama morbiditas dan mortalitas di seluruh dunia (Cleto-Yamane et al., 2024). Deteksi dini penyakit ginjal sangat penting agar pasien dapat memperoleh penanganan yang tepat sehingga menurunkan risiko komplikasi dan kematian (Kumar et al.,

2023). Tantangan dalam diagnosis penyakit ginjal terletak pada gejala klinis yang sering tidak spesifik serta adanya kemiripan karakteristik dengan penyakit lain, sehingga keberadaan sistem pendukung keputusan berbasis teknologi sangat diperlukan untuk membantu dokter melakukan deteksi secara lebih akurat dan efisien (Singh et al., 2022).

Di era digital saat ini, machine learning telah terbukti memberikan kontribusi signifikan dalam bidang kesehatan, khususnya untuk menganalisis dan mengklasifikasikan data medis dalam skala besar (Sivaraman & Khanna, 2021). Berbagai algoritma machine learning telah diimplementasikan dalam deteksi penyakit ginjal, seperti Support Vector Machine (SVM), Random Forest, dan Decision Tree. Di antara metode tersebut, algoritma C4.5 decision tree banyak digunakan karena kemampuannya dalam menghasilkan model yang interpretable, menangani data numerik maupun kategorikal, serta mengatasi missing value dengan baik (Arif et al., 2023; Ghazal et al., 2023; Ibarra et al., 2022). Penelitian oleh (Mustafizur Rahman et al., 2024) menunjukkan bahwa C4.5 dapat mencapai akurasi 85% pada deteksi penyakit ginjal kronis. Namun, algoritma ini masih memiliki beberapa kelemahan, antara lain risiko overfitting, penurunan akurasi pada data tidak seimbang, dan ketidakstabilan pada data yang noisy.

Seiring berkembangnya penelitian, berbagai metode telah diusulkan untuk meningkatkan kinerja model klasifikasi medis, salah satunya dengan pendekatan ensemble learning seperti boosting. Boosting, khususnya AdaBoost, mampu meningkatkan akurasi model dasar dengan cara menggabungkan beberapa model lemah menjadi model yang lebih kuat dan stabil. Studi oleh (Damayanti & Herawati, 2024; Li et al., 2021) membuktikan bahwa kombinasi AdaBoost dengan decision tree secara signifikan meningkatkan akurasi serta mengurangi error pada klasifikasi data kesehatan yang kompleks dan tidak seimbang. Namun, hingga kini, kajian yang secara spesifik menguji penerapan AdaBoost

pada C4.5 untuk deteksi penyakit ginjal masih sangat terbatas, sehingga efektivitas metode ini dalam domain tersebut belum terungkap secara mendalam.

Selain pengembangan model, tahapan pra-pemrosesan data juga sangat krusial untuk memastikan kualitas input bagi algoritma machine learning. Pada penelitian ini, label diagnosis penyakit ginjal dikodekan menggunakan Label Encoder untuk mengubah label kelas menjadi format numerik yang dapat diproses oleh model (Ingkafi et al., 2023). Sementara itu, fitur-fitur kategorikal yang terdapat pada data pasien dikonversi menggunakan Ordinal Encoder agar setiap kategori dapat direpresentasikan dalam urutan numerik yang bermakna (Yustanti et al., 2023). Selanjutnya, untuk mengatasi masalah ketidakseimbangan data yang sering terjadi pada dataset medis, digunakan metode SMOTE-ENN (kombinasi SMOTE dan Edited Nearest Neighbours), yang efektif dalam menyeimbangkan distribusi kelas minoritas dengan menciptakan data sintetis serta membersihkan data outlier (Mubarak et al., 2023). Tahapan penting lainnya adalah seleksi fitur menggunakan LASSO (Least Absolute Shrinkage and Selection Operator) yang berfungsi untuk memilih fitur-fitur paling relevan dan mengurangi dimensi data sehingga proses pembelajaran menjadi lebih efisien serta menghindari overfitting (Liu et al., 2023).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menerapkan dan menganalisis peningkatan performa algoritma C4.5 melalui integrasi AdaBoost, dengan tahapan pra-pemrosesan menggunakan Label Encoder dan Ordinal Encoder, penyeimbangan data dengan SMOTE-ENN, serta seleksi fitur dengan LASSO, pada tugas deteksi penyakit ginjal.

Diharapkan, pendekatan ini dapat menghasilkan model yang lebih akurat, stabil, dan robust dibandingkan dengan metode C4.5 konvensional, sehingga dapat menjadi solusi efektif untuk mendukung proses diagnosis dini penyakit ginjal secara otomatis di lingkungan klinis.

## II. Landasan Teori

Sebelum membahas lebih jauh mengenai tahapan eksperimen dan implementasi pada penelitian ini, penting untuk memahami konsep-konsep dasar serta metode utama yang digunakan. Pada penelitian ini, beberapa pendekatan dan algoritma modern di bidang machine learning diterapkan secara terpadu guna meningkatkan akurasi dan keandalan deteksi penyakit ginjal berdasarkan data klinis. Penjelasan teoritis mengenai masing-masing metode sangat penting sebagai landasan ilmiah, agar pembaca dapat memahami alasan pemilihan metode serta kontribusi setiap teknik dalam keseluruhan sistem yang dibangun.

Adapun metode utama yang digunakan meliputi algoritma pohon keputusan C4.5, teknik boosting AdaBoost, metode penyeimbangan data SMOTE-ENN, serta teknik seleksi fitur LASSO. Uraian berikut akan membahas masing-masing metode secara ringkas namun komprehensif sebagai dasar pengembangan model pada penelitian ini.

### Algoritma C4.5

C4.5 merupakan salah satu algoritma pohon keputusan (decision tree) yang dikembangkan oleh J. Ross Quinlan sebagai pengembangan dari algoritma ID3. Algoritma ini bekerja dengan membagi dataset berdasarkan atribut yang memberikan nilai informasi tertinggi,

menggunakan ukuran Gain Ratio sebagai kriteria pemilihan atribut. Salah satu keunggulan utama C4.5 adalah kemampuannya menangani atribut numerik dan kategorikal, mengatasi missing value, serta menghasilkan pohon keputusan yang dapat dipangkas (pruning) untuk mencegah overfitting. Model pohon keputusan yang dihasilkan C4.5 sangat interpretatif sehingga memudahkan tenaga medis dalam memahami proses pengambilan keputusan oleh model. Dalam banyak studi, C4.5 telah terbukti efektif dalam berbagai tugas klasifikasi, termasuk di bidang kesehatan, karena kemudahannya interpretasi serta fleksibilitasnya dalam mengolah data yang kompleks (Pratama et al., 2022).

### AdaBoost (*Adaptive Learning*)

AdaBoost merupakan salah satu teknik ensemble learning berbasis boosting yang pertama kali diperkenalkan oleh Freund dan Schapire pada tahun 1997. Prinsip dasar AdaBoost adalah menggabungkan sejumlah model dasar (weak learners) secara iteratif untuk membentuk model akhir yang lebih kuat (strong learner). Setiap model dasar dilatih secara berurutan, di mana data yang sebelumnya salah diklasifikasikan akan diberikan bobot lebih besar pada iterasi berikutnya. Dengan cara ini, model akhir lebih fokus pada data-data sulit yang cenderung sering salah prediksi. Pada umumnya, model dasar yang digunakan adalah pohon keputusan sederhana (decision stump), namun dalam penelitian ini AdaBoost diintegrasikan dengan C4.5. Kelebihan utama AdaBoost adalah kemampuannya dalam meningkatkan akurasi model dan mengurangi risiko

overfitting, terutama pada dataset dengan kompleksitas tinggi dan distribusi kelas yang tidak seimbang (Piao et al., 2023; Sui & Ghosh, 2024).

**SMOTE-ENN (SMOTE + Edited Nearest Neighbours)**

SMOTE-ENN adalah teknik gabungan antara SMOTE (Synthetic Minority Over-sampling Technique) dan Edited Nearest Neighbours (ENN) yang dirancang untuk mengatasi permasalahan ketidakseimbangan kelas (class imbalance) dalam dataset. SMOTE bekerja dengan membuat data sintesis pada kelas minoritas dengan cara interpolasi antara sampel minoritas yang ada, sehingga distribusi data menjadi lebih seimbang. Setelah tahap over-sampling, ENN digunakan untuk membersihkan data dengan menghapus sampel yang kemungkinan merupakan outlier atau noise berdasarkan hasil voting dari tetangga terdekatnya. Dengan menggabungkan kedua teknik ini, SMOTE-ENN tidak hanya menambah jumlah data pada kelas minoritas, tetapi juga memastikan kualitas data hasil balancing lebih baik dan tidak mengandung noise yang berlebihan. Teknik ini sangat efektif untuk meningkatkan performa model klasifikasi pada kasus dengan distribusi kelas yang timpang, seperti pada data medis (Husain et al., 2025; Wang et al., 2023).

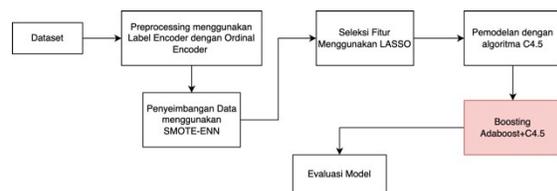
**LASSO (Least Absolute Shrinkage and Selection Operator)**

LASSO adalah salah satu metode regularisasi dan seleksi fitur yang banyak digunakan dalam regresi maupun klasifikasi. LASSO bekerja dengan menambahkan penalti berupa nilai absolut

koefisien (L1 regularization) ke dalam fungsi loss selama proses pelatihan model. Dengan adanya penalti ini, LASSO cenderung "menyusutkan" beberapa koefisien menjadi nol sehingga fitur yang kurang relevan akan dieliminasi secara otomatis dari model. Selain membantu mengurangi overfitting, LASSO juga meningkatkan interpretabilitas model karena hanya fitur-fitur paling signifikan yang dipertahankan. Dalam penelitian medis dan klasifikasi penyakit, penggunaan LASSO sangat berguna untuk memilih subset fitur terbaik dari data yang berdimensi tinggi dan kompleks (Cahaya et al., 2022; Pradyto & Raharja, 2023).

**III. Metode Penelitian**

Pada tahap ini dilakukan berbagai pengujian dengan mengikuti metode penelitian dibawah ini.



**Gambar 1.** Metodologi Penelitian

**Dataset**

Tahap pertama adalah pengumpulan dataset, yang berisi data rekam medis pasien terkait penyakit ginjal. Pada penelitian ini, dataset diambil dari platform Kaggle dan terdiri atas 1.089 record. Dataset tersebut memuat berbagai fitur numerik maupun kategorikal yang relevan dalam proses deteksi penyakit ginjal, seperti hasil pemeriksaan laboratorium, riwayat kesehatan, serta data demografi pasien. Keberagaman fitur dalam dataset ini memungkinkan model

machine learning untuk mengenali pola-pola penting yang berkontribusi pada proses diagnosis penyakit ginjal secara lebih komprehensif.

### **Preprocessing menggunakan Label Encoder dan Ordinal Encoder**

Pada tahap pra-pemrosesan data, seluruh fitur kategorikal yang akan digunakan sebagai label (target) dikonversi menggunakan Label Encoder, sedangkan fitur kategorikal lainnya dikonversi menggunakan Ordinal Encoder. Proses ini bertujuan agar semua data kategorikal dapat direpresentasikan dalam bentuk numerik sehingga dapat diproses oleh algoritma machine learning. Tahapan ini sangat penting untuk memastikan format data yang konsisten dan siap diproses pada tahap selanjutnya.

### **Penyeimbangan Data menggunakan SMOTE-ENN**

Ketidakseimbangan distribusi kelas sering terjadi pada dataset medis, di mana jumlah data pada kelas minoritas jauh lebih sedikit dibanding kelas mayoritas. Untuk mengatasi permasalahan ini, dilakukan penyeimbangan data menggunakan metode SMOTE-ENN (Synthetic Minority Over-sampling Technique–Edited Nearest Neighbours). SMOTE-ENN menggabungkan teknik oversampling (SMOTE) untuk menambah data pada kelas minoritas dan teknik cleaning (ENN) untuk menghapus data noise atau outlier, sehingga diperoleh dataset yang lebih seimbang dan bersih.

### **Seleksi Fitur Menggunakan LASSO**

Setelah data seimbang, dilakukan seleksi fitur menggunakan metode LASSO

(Least Absolute Shrinkage and Selection Operator). LASSO digunakan untuk memilih fitur-fitur yang paling berpengaruh terhadap target (label), dengan cara memberikan penalti terhadap koefisien regresi yang tidak signifikan hingga mendekati nol. Dengan seleksi fitur yang tepat, kompleksitas model dapat dikurangi, risiko overfitting diminimalisasi, dan akurasi model dapat meningkat.

### **Pemodelan dengan Algoritma C4.5**

Tahap selanjutnya adalah pemodelan data menggunakan algoritma C4.5, yaitu algoritma pohon keputusan yang mampu menangani data numerik dan kategorikal. Algoritma ini membentuk pohon keputusan berdasarkan entropi dan gain rasio, sehingga mampu memetakan pola klasifikasi penyakit ginjal secara jelas dan interpretatif.

### **Boosting dengan Adaboost+C4.5**

Untuk meningkatkan performa model, dilakukan proses boosting menggunakan AdaBoost pada algoritma C4.5. AdaBoost akan mengkombinasikan beberapa model C4.5 yang lemah (weak learner) menjadi satu model kuat (strong learner) dengan memberikan bobot lebih besar pada data yang sulit diklasifikasikan. Proses ini mampu meningkatkan akurasi dan kestabilan model dalam mendeteksi penyakit ginjal.

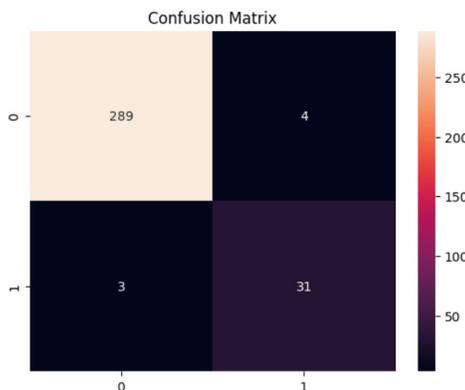
### **Evaluasi Model**

Tahap terakhir adalah evaluasi model, yaitu mengukur performa model yang telah dibangun dengan berbagai metrik seperti akurasi, presisi, recall, dan f1-score. Evaluasi ini dilakukan untuk

memastikan model hasil boosting benar-benar mengalami peningkatan performa dibandingkan model dasar (C4.5 tanpa boosting).

**IV. Hasil Dan Pembahasan**

Pada bagian ini, dipaparkan secara sistematis hasil penelitian yang telah diperoleh melalui serangkaian tahapan eksperimen, pemodelan, serta evaluasi algoritma yang diusulkan. Seluruh temuan dideskripsikan secara rinci, mulai dari hasil pra-pemrosesan data, penerapan teknik balancing dan seleksi fitur, hingga perbandingan performa model sebelum dan sesudah penerapan metode ensemble learning Adaboost pada algoritma C4.5. Selain itu, pembahasan mendalam juga dilakukan untuk menganalisis signifikansi setiap langkah terhadap peningkatan kinerja deteksi penyakit ginjal. Seluruh hasil penelitian ini kemudian dianalisis dan dikaitkan dengan penelitian terdahulu, sehingga dapat memberikan gambaran kontribusi serta keunggulan pendekatan yang diusulkan dalam studi ini. Berikut ini merupakan hasil confusion report dari algoritma C4.5.



**Gambar 2. Confusion Matrix Pengujian Algoritma C4.5**

Confusion matrix pada gambar 2 memberikan gambaran visual mengenai

performa prediksi algoritma C4.5 dalam mendeteksi penyakit ginjal. Dari hasil yang ditampilkan, dapat dilihat bahwa model mampu mengklasifikasikan mayoritas data dengan benar. Sebanyak 289 data yang sebenarnya negatif (tidak terkena penyakit ginjal) berhasil diprediksi dengan tepat, sementara hanya 4 data negatif yang salah diprediksi sebagai positif (false positive). Pada sisi lain, terdapat 31 data positif (terkena penyakit ginjal) yang berhasil diprediksi dengan benar oleh model, dan hanya 3 data positif yang salah diklasifikasikan sebagai negatif (false negative). Dengan jumlah prediksi benar yang sangat dominan baik pada kelas negatif maupun positif, confusion matrix ini menegaskan bahwa model C4.5 memiliki kemampuan deteksi yang tinggi dan tingkat kesalahan yang sangat rendah pada kedua kelas yang diuji. Hasil ini menunjukkan bahwa model sangat andal dan dapat digunakan untuk proses diagnosis awal penyakit ginjal secara otomatis. Kemudian gambar 3 merupakan classification report hasil pengujian algoritma C4.5.

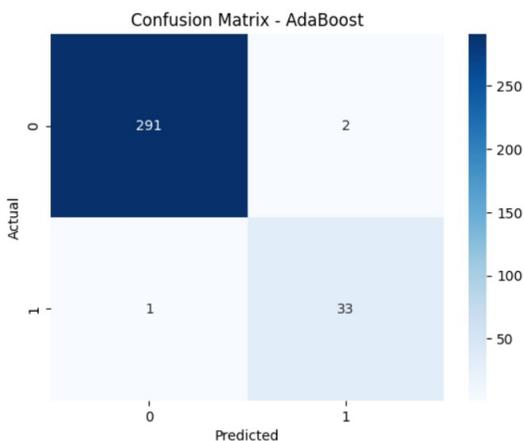
	precision	recall	f1-score	support
0	0.99	0.99	0.99	293
1	0.89	0.91	0.90	34
accuracy			0.98	327
macro avg	0.94	0.95	0.94	327
weighted avg	0.98	0.98	0.98	327
Accuracy : 0.9785932721712538				

**Gambar 3. Classification Report Pengujian Algoritma C4.5**

Classification report pada gambar 3 menampilkan beberapa metrik evaluasi utama untuk masing-masing kelas, yaitu precision, recall, f1-score, serta jumlah data pada masing-masing kelas (support). Untuk kelas 0 (misal: tidak terkena penyakit ginjal), precision dan recall mencapai 0.99, yang

menunjukkan bahwa model hampir selalu benar dalam memprediksi kelas ini dan sangat jarang salah. Untuk kelas 1 (misal: terkena penyakit ginjal), precision sebesar 0.89 dan recall 0.91, yang juga menunjukkan performa model cukup baik meskipun sedikit lebih rendah dibandingkan kelas 0. Secara keseluruhan, akurasi model mencapai 0.98 (atau 98%), sedangkan nilai rata-rata macro (macro avg) dan rata-rata tertimbang (weighted avg) pada precision, recall, dan f1-score berada pada kisaran 0.94 hingga 0.98. Nilai-nilai ini menunjukkan model C4.5 sangat andal dalam mengklasifikasikan kedua kelas, baik kelas mayoritas maupun minoritas.

Secara umum, baik dari classification report maupun confusion matrix, algoritma C4.5 terbukti sangat efektif untuk klasifikasi penyakit ginjal pada dataset yang digunakan. Model mampu meminimalisir kesalahan prediksi, baik pada kelas mayoritas maupun minoritas, sehingga dapat diandalkan untuk deteksi dini penyakit ginjal. Kemudian pengujian menggunakan Adaboost+C4.5 juga menunjukkan hasil yang cukup baik dan dapat dilihat pada gambar 4 dan 5.



**Gambar 4. Confusion Matrix Pengujian Algoritma Adaboost+C4.5**

Confusion matrix pada gambar 4 menunjukkan performa model C4.5 yang telah di-boosting dengan Adaboost dalam mendeteksi penyakit ginjal. Terlihat bahwa sebanyak 291 data negatif (tidak terkena penyakit ginjal) berhasil diprediksi dengan benar (true negative), hanya terdapat 2 kasus yang salah prediksi sebagai positif (false positive), sementara untuk kelas positif (terkena penyakit ginjal), 33 data diprediksi dengan benar (true positive) dan hanya 1 data yang salah prediksi sebagai negatif (false negative). Distribusi prediksi yang hampir seluruhnya berada pada diagonal utama mengindikasikan tingkat akurasi dan ketepatan model yang sangat tinggi dalam membedakan kedua kelas.

	precision	recall	f1-score	support
0	1.00	0.99	0.99	293
1	0.94	0.97	0.96	34
accuracy			0.99	327
macro avg	0.97	0.98	0.98	327
weighted avg	0.99	0.99	0.99	327

Accuracy : 0.9908256880733946

**Gambar 5. Classification Report Pengujian Algoritma Adaboost+C4.5**

Classification report pada gambar ini mendukung temuan pada confusion matrix dengan nilai precision, recall, dan f1-score yang sangat tinggi untuk kedua kelas. Untuk kelas 0 (negatif), precision dan recall mencapai 1,00 dan 0,99, sedangkan untuk kelas 1 (positif), precision dan recall masing-masing 0,94 dan 0,97. Nilai f1-score yang konsisten tinggi pada kedua kelas menandakan performa prediksi model yang seimbang. Akurasi total yang dicapai oleh model adalah 99%, menunjukkan bahwa kombinasi boosting Adaboost pada C4.5 mampu meningkatkan kinerja deteksi penyakit ginjal secara signifikan dibandingkan model C4.5 standar.

**Tabel 1. Perbandingan dengan Penelitian Terdahulu**

Peneliti	Algoritma	Objek	Akurasi
(Islam et al., 2023)	XGBoost Classifier	chronic kidney disease	98%
(Venkatesan et al., 2023)	XGBoost Classifier	chronic kidney disease	98%
(Van FC et al., 2025)	Stacking	Breast Cancer	97%
(Herianto et al., 2024)	Stacking	Graduati on Prediction	95%
Penelitian ini	Boosting (Adaboost + C4.5)	Deteksi Penyakit Ginjal	99%

Tabel 1 menampilkan perbandingan performa model dari beberapa penelitian terdahulu yang menggunakan algoritma machine learning untuk tugas klasifikasi di bidang kesehatan, khususnya pada deteksi penyakit kronis. Dalam tabel tersebut, Islam et al. (2023) dan Venkatesan et al. (2023) sama-sama menggunakan algoritma XGBoost Classifier untuk mendeteksi penyakit ginjal kronis dan berhasil mencapai tingkat akurasi sebesar 98%. Hasil ini menunjukkan bahwa XGBoost merupakan salah satu algoritma yang sangat kompetitif dalam menangani permasalahan klasifikasi medis dengan performa yang tinggi.

Selain itu, Van FC et al. (2025) dan Herianto et al. (2024) menerapkan pendekatan stacking ensemble pada dua kasus berbeda, yaitu deteksi kanker payudara dan prediksi kelulusan mahasiswa. Keduanya juga memperoleh tingkat akurasi yang sangat baik, masing-masing sebesar 97% untuk deteksi kanker payudara dan 95% untuk prediksi kelulusan. Hal ini memperkuat bukti bahwa strategi ensemble, baik stacking maupun

boosting, berperan penting dalam meningkatkan keakuratan model pada berbagai domain klasifikasi.

Pada penelitian ini, digunakan metode boosting dengan mengombinasikan Adaboost dan algoritma C4.5 untuk deteksi penyakit ginjal. Hasil eksperimen menunjukkan bahwa model yang dikembangkan berhasil mencapai akurasi tertinggi, yaitu 99%. Pencapaian ini lebih unggul dibandingkan penelitian-penelitian sebelumnya pada kasus serupa maupun kasus lain, baik yang menggunakan boosting (XGBoost) maupun stacking. Keberhasilan ini tidak lepas dari integrasi proses pra-pemrosesan, penyeimbangan data dengan SMOTE-ENN, serta seleksi fitur menggunakan LASSO, yang semuanya turut meningkatkan performa model secara signifikan. Dengan demikian, penelitian ini memberikan kontribusi nyata dalam pengembangan sistem deteksi penyakit ginjal berbasis machine learning, sekaligus menegaskan keunggulan pendekatan boosting Adaboost pada algoritma C4.5 di atas berbagai model pembanding yang ada dalam literatur.

## V. Kesimpulan Dan Saran

Penelitian ini berhasil membuktikan bahwa peningkatan algoritma C4.5 dengan integrasi teknik boosting Adaboost, balancing data dengan SMOTE-ENN, serta seleksi fitur menggunakan LASSO mampu menghasilkan model deteksi penyakit ginjal dengan performa sangat baik. Model gabungan Adaboost+C4.5 menunjukkan peningkatan akurasi, presisi, recall, dan f1-score dibandingkan model C4.5 murni, dengan akurasi mencapai 99%. Pencapaian ini juga melampaui hasil model-model pada penelitian terdahulu yang menggunakan XGBoost maupun stacking ensemble pada domain

serupa. Proses pra-pemrosesan yang sistematis, penyeimbangan data, dan seleksi fitur yang tepat turut berperan dalam memastikan model lebih robust dan minim overfitting. Temuan ini memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan berbasis machine learning untuk deteksi dini penyakit ginjal. Saran untuk penelitian selanjutnya adalah mengeksplorasi metode ensemble lain, pengujian pada dataset yang lebih besar dan beragam, serta integrasi data klinis multi-sumber untuk meningkatkan generalisasi dan akurasi model prediksi penyakit ginjal.

## VI. Daftar Pustaka

- Arif, M. S., Mukheimer, A. & Asif, D. (2023). Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model. *Big Data and Cognitive Computing*, 7(3). <https://doi.org/10.3390/bdcc7030144>
- Cahya, S. D., Sartono, B., Indahwati, I. & Purnaningrum, E. (2022). Performance of LAD-LASSO and WLAD-LASSO on High Dimensional Regression in Handling Data Containing Outliers. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, 6(4), 844. <https://doi.org/10.31764/jtam.v6i4.8968>
- Cleto-Yamane, T. L., Gomes, C. L. R., Koch-Nogueira, P. C. & Suassuna, J. H. R. (2024). Acute kidney injury requiring dialysis in children: a multicentric, emerging country perspective. *Pediatric Nephrology*, 39(7), 2253–2262. <https://doi.org/10.1007/s00467-024-06305-9>
- Damayanti, H. L. & Herawati, R. (2024). Enhancing Stroke Disease Prediction Performance Through a Fusion of Adaboost With C4.5 and K-Nearest Neighbor Algorithms. *Proxies*, 7(2), 102–114. <https://doi.org/10.24167/proxies.v7i2.12470>
- Ghazal, T. M., Ibrahim, A., Akram, A. S., Qaisar, Z. H., Munir, S. & Islam, S. (2023). Heart Disease Prediction Using Machine Learning. *2nd International Conference on Business Analytics for Technology and Security, ICBATS 2023*. <https://doi.org/10.1109/ICBATS57792.2023.10111368>
- Herianto, Kurniawan, B., Hartomi, Z. H., Irawan, Y. & Anam, M. K. (2024). Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction. *Journal of Applied Data Sciences*, 5(3), 1272–1285. <https://doi.org/10.47738/jads.v5i3.316>
- Husain, G., Nasef, D., Jose, R., Mayer, J., Bekbolatova, M., Devine, T. & Toma, M. (2025). SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models. *Algorithms*, 18(1), 1–16. <https://doi.org/10.3390/a18010037>
- Ibarra, R., León, J., Ávila, I. & Ponce, H. (2022). Cardiovascular Disease Detection Using Machine Learning. *Computacion y Sistemas*, 26(4), 1661–1668.

- <https://doi.org/10.13053/CyS-26-4-4422>
- Ingkafi, D. A., Aryana, G. A., Putra, A. K. & Kusumaningrum, R. (2023). Sentiment Analysis of The National Covid-19 Vaccination Program on Twitter Using The Bidirectional Encoder Representation From Transformer. *ICIC Express Letters*, 17(2), 201–208. <https://doi.org/10.24507/icicel.17.02.201>
- Islam, M. A., Majumder, M. Z. H. & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, 14(1), 1–12. <https://doi.org/10.1016/j.jpi.2023.100189>
- Kumar, K., Pradeepa, M., Mahdal, M., Verma, S., RajaRao, M. V. L. N. & Ramesh, J. V. N. (2023). A Deep Learning Approach for Kidney Disease Recognition and Prediction through Image Processing. *Applied Sciences (Switzerland)*, 13(6). <https://doi.org/10.3390/app13063621>
- Li, X., Chen, X. & Yuan, Z. (2021). Applicable model of liver disease detection based on the improved CART-AdaBoost algorithm. *2021 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2021*, 1177–1181. <https://doi.org/10.1109/ICAICA52286.2021.9498046>
- Liu, J., Ma, Y., Xie, W., Li, X., Wang, Y., Xu, Z., Bai, Y., Yin, P. & Wu, Q. (2023). Lasso-Based Machine Learning Algorithm for Predicting Postoperative Lung Complications in Elderly: A Single-Center Retrospective Study from China. *Clinical Interventions in Aging*, 18, 597–606. <https://doi.org/10.2147/CIA.S406735>
- Mubarak, M. M. R., Chrisnanto, Y. H. & Sabrina, P. N. (2023). Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce. *Enrichment: Journal of Multidisciplinary Research and Development*, 1(8), 463–477. <https://doi.org/10.55324/enrichment.v1i8.69>
- Mustafizur Rahman, M., Al-Amin, M. & Hossain, J. (2024). Machine learning models for chronic kidney disease diagnosis and prediction. *Biomedical Signal Processing and Control*, 87, 1–17. <https://doi.org/10.1016/j.bspc.2023.105368>
- Piao, C., Wang, N. & Yuan, C. (2023). Rebalance Weights AdaBoost-SVM Model for Imbalanced Data. *Computational Intelligence and Neuroscience*, 2023(1), 1–26. <https://doi.org/10.1155/2023/4860536>
- Pradyto, K. D. A. & Raharja, M. A. (2023). Implementasi Random Forest dengan LASSO Dalam Klasifikasi Penyakit yang Ditularkan Melalui Nyamuk. *JNATIA (Jurnal Nasional Teknologi Informasi Dan*

- 
- Aplikasinya*, 1(4), 1197–1202.  
[https://www.kaggle.com/datasets/richardbernat/vector-borne-disease-](https://www.kaggle.com/datasets/richardbernat/vector-borne-disease)  
Pratama, I. G. A. M., Astuti, L. G.,  
Widiartha, I. M., Putra, I. G. N. A.  
C., Pramatha, C. R. A. &  
Darmawan, I. D. M. B. A. (2022).  
Diagnosis Penyakit Ginjal Kronis  
dengan Algoritma C4.5, K-Means  
dan BPSO. *Jurnal-Elektronik-Ilmu-*  
*Komputer-Udayana*, 10(4), 371–  
381.  
<https://doi.org/10.24843/JLK.2022.v10.i04.p07>
- Singh, B., Arora, K. & Iyer, S. S. (2022).  
Chronic Kidney Disease Detection  
Using Machine Learning  
Regression Models. *ECS*  
*Transactions*, 107(1), 2191–2207.  
<https://doi.org/10.1149/10701.2191ecst>
- Sivaraman, K. & Khanna, V. (2021).  
Machine learning models for  
prediction of cardiovascular  
diseases. *Journal of Physics:*  
*Conference Series*, 2040(1).  
<https://doi.org/10.1088/1742-6596/2040/1/012051>
- Sui, Q. & Ghosh, S. K. (2024). Active  
Learning for Stacking and  
AdaBoost-Related Models. *Stats*,  
7(1), 110–137.  
<https://doi.org/10.3390/stats7010008>
- Van FC, L. L., Anam, M. K., Bukhori, S.,  
Mahamad, A. K., Saon, S. &  
Nyoto, R. L. V. (2025). The  
Development of Stacking  
Techniques in Machine Learning  
for Breast Cancer Detection.  
*Journal of Applied Data Sciences*,  
6(1), 71–85.  
<https://doi.org/10.47738/jads.v6i1.416>
- Venkatesan, V. K., Ramakrishna, M. T.,  
Izonin, I., Tkachenko, R. &  
Havryliuk, M. (2023). Efficient  
Data Preprocessing with Ensemble  
Machine Learning Technique for  
the Early Detection of Chronic  
Kidney Disease. *Applied Sciences*  
*(Switzerland)*, 13(5), 1–18.  
<https://doi.org/10.3390/app13052885>
- Wang, A. X., Chukova, S. S. & Nguyen,  
B. P. (2023). Synthetic minority  
oversampling using edited  
displacement-based k-nearest  
neighbors. *Applied Soft Computing*,  
148, 1–12.  
<https://doi.org/10.1016/j.asoc.2023.110895>
- Yustanti, W., Iriawan, N. & Irhamah.  
(2023). Categorical encoder based  
performance comparison in  
preprocessing imbalanced  
multiclass classification.  
*Indonesian Journal of Electrical*  
*Engineering and Computer*  
*Science*, 31(3), 1705–1715.  
<https://doi.org/10.11591/ijeecs.v31.i3.pp1705-1715>