



Klasifikasi Penyakit Jantung Menggunakan *Extreme Gradient Boosting (XGBoost)*

Dimas Bagus Darmawan^{1*}, Ibnu Rafli Azahwa², Rendy Wijaya Saputra³, Robby Septiadi⁴,
Perani Rosyani⁵

^{1,2,3,4,5}Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl.Raya Puspipetek No. 46, Kel.Buaran,
Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}dims2602@gmail.com, ²azahwadimas@gmail.com, ³aldi.pwijayaa@gmail.com,
⁴robbyseptiadi4@gmail.com, ⁵dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak - Jantung adalah salah satu organ dalam tubuh manusia yang berfungsi sebagai pemompa darah dan oksigen ke seluruh tubuh. Menurut data World Health Organization (WHO), sekitar 17,9 juta orang meninggal setiap tahun akibat penyakit kardiovaskular yang menyerang jantung (Rika Widianita, 2023). Penelitian ini melakukan klasifikasi penyakit jantung menggunakan algoritma berbasis machine learning, yaitu Extreme Gradient Boosting atau XGBoost. Algoritma XGBoost dipilih karena kemampuannya yang menjanjikan dalam melakukan klasifikasi (Kurnia et al., 2023). XGBoost telah digunakan oleh banyak peneliti untuk mendapatkan hasil yang optimal dalam berbagai kasus machine learning. Dalam penelitian ini, model machine learning yang dikembangkan menggunakan XGBoost dibandingkan dengan model lain yang telah diterapkan sebelumnya, seperti Stacking, Random Forest, dan Majority Voting. Hasil pengujian menunjukkan bahwa XGBoost mampu mencapai performa yang baik dalam seluruh metrik evaluasi, termasuk akurasi (Murdiansyah, 2024).

Kata Kunci : Penyakit Jantung, *Machine Learning*, *XGBoost*

Abstract - The heart is one of the organs in the human body that functions as a pump for blood and oxygen throughout the body. According to data from the World Health Organization (WHO), around 17.9 million people die each year from cardiovascular diseases that attack the heart (Rika Widianita, 2023). This study classifies heart disease using a machine learning-based algorithm, namely Extreme Gradient Boosting or XGBoost. The XGBoost algorithm was chosen because of its promising ability to perform classification (Kurnia et al., 2023). XGBoost has been used by many researchers to obtain optimal results in various machine learning cases. In this study, the machine learning model developed using XGBoost was compared with other previously applied models, such as Stacking, Random Forest, and Majority Voting. The test results showed that XGBoost was able to achieve good performance in all evaluation metrics, including accuracy (Murdiansyah, 2024).

Keywords: Heart Disease, *Machine Learning*, *Xgboost*

1. PENDAHULUAN

Penyakit kardiovaskular adalah penyebab utama kematian di dunia. Pada tahun 2016, jumlah kematian akibat penyakit ini mencapai 17,9 juta orang (WHO, 2017). Di Amerika Serikat, penyakit kardiovaskular menyumbang 836.846 kematian per tahun, setara dengan satu dari tiga kematian, atau sekitar 2.300 kematian setiap hari, yang berarti satu kematian setiap 38 detik (Fadlillah, Sucipto, dan Amestiasih, 2021). Di Indonesia, penyakit jantung juga menjadi penyebab utama kematian. Penyakit jantung adalah masalah kesehatan global yang serius, dan jumlah kasusnya diperkirakan akan meningkat seiring dengan pola hidup dan tingkat stres masyarakat modern yang cenderung buruk (Rika Widianita, 2023).

Kardiovaskular merujuk pada sistem yang melibatkan jantung dan pembuluh darah, dan penyakit kardiovaskular mencakup kondisi yang menyerang sistem ini, seperti penyakit jantung dan stroke. Faktor risiko penyakit kardiovaskular antara lain usia, kebiasaan merokok, pola makan, konsumsi alkohol, kelebihan berat badan, aktivitas fisik rendah, dan tekanan darah tinggi (Handika Permana, Rakhmat Umbara, dan Kasyidi, 2024).

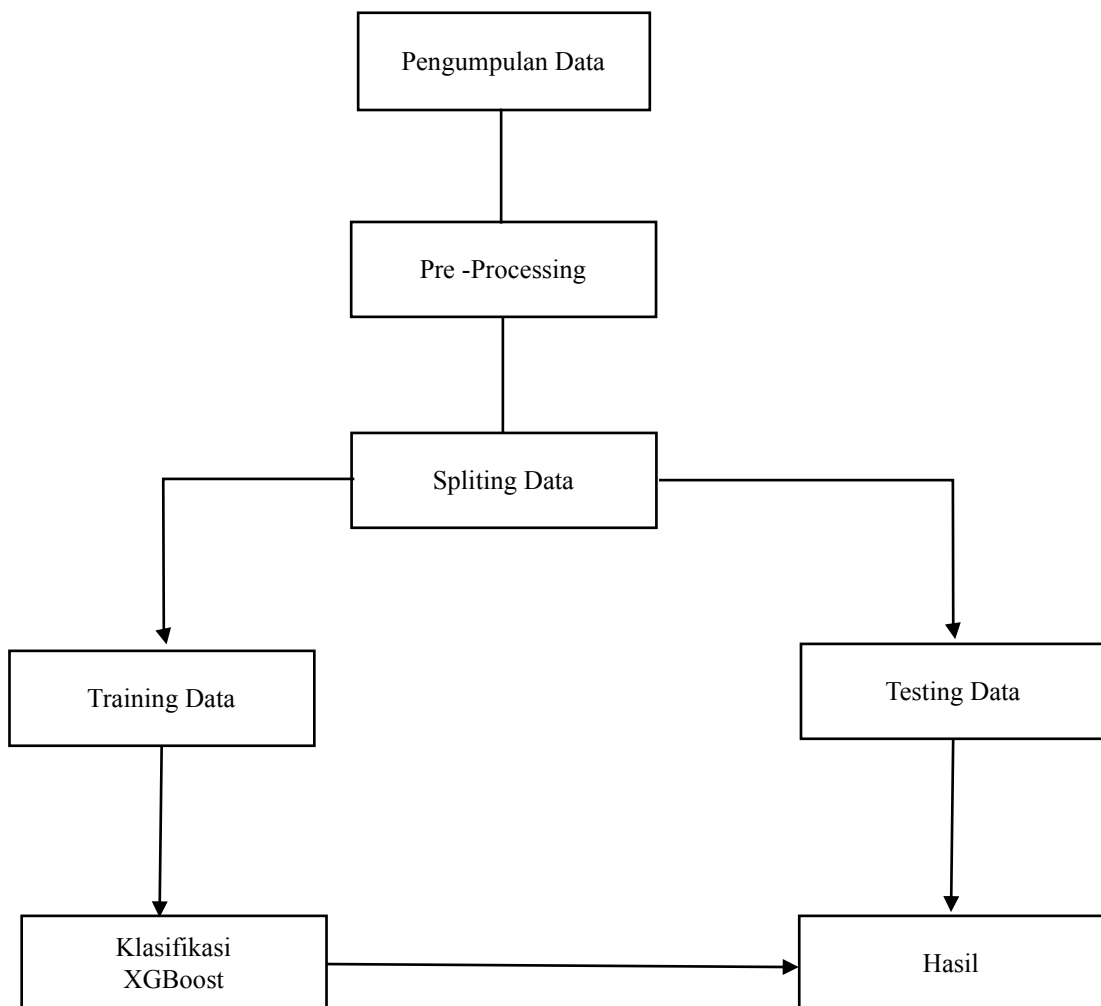
Proses penanganan dan diagnosis penyakit jantung di negara-negara berkembang menghadapi tantangan karena keterbatasan alat diagnostik, tenaga medis, dan sumber daya lainnya, yang mempersulit klasifikasi dan pengobatan pasien jantung. Machine learning dapat membantu mengatasi masalah ini dengan memprediksi dan mengklasifikasi jenis-jenis penyakit jantung berdasarkan karakteristik tertentu (Kurnia et al., 2023). Integrasi machine learning dalam bidang kesehatan menjadi pendekatan inovatif untuk prediksi dan diagnosis penyakit. Salah satu algoritma yang menjanjikan



dalam penelitian ini adalah Extreme Gradient Boosting (XGBoost). XGBoost telah menunjukkan potensi besar di berbagai bidang, termasuk medis (Murdiansyah, 2024).

Penelitian ini bertujuan untuk mengevaluasi efektivitas XGBoost dalam klasifikasi penyakit jantung, khususnya dari segi akurasi dan efisiensi. Hasil penelitian ini menunjukkan bahwa XGBoost berpotensi mencapai akurasi hingga 91,8% dan menjadi acuan dalam pengembangan sistem deteksi penyakit jantung berbasis teknologi yang lebih andal (Handika Permana, Rakhmat Umbara, dan Kasyidi, 2024).

2. METODOLOGI PENELITIAN



2.1 Pengumpulan data

Pengumpulan data ialah langkah paling awal di dalam penelitian ini karena dataset yang dipilih akan menjadi dasar dari analisis dan temuan yang akan dihasilkan (Handika Permana, Rakhmat Umbara and Kasyidi, 2024). Diawali dengan memilah parameter yang akan digunakan. Dengan korelasi dari masing-masing parameter tersebut dibangun berdasarkan hipotesis awal yang dibangun pada penelitian ini merujuk pada kondisi dimana tingkat kasus penyakit berbanding lurus dengan aspek yang akan dituju (Wardhana, Wang and Sibuea, 2023).

2.2 Pre- processing

Tahap selanjutnya yaitu preprocessing. Pada tahap ini peneliti melakukan proses, pengolahan data untuk membuang data yang bernilai kosong/ tidak lengkap (Dava Maulana, Id Hadiana and Rakhmat Umbara, 2024).



2.3 Proses Klasifikasi

Setelah menyelesaikan fase preprocessing, data dibagi menjadi dua yaitu set pelatihan dan pengujian. Dengan tujuan membagi data menjadi data training dan data testing. Dan pada tahap ini dilakukan pembagian menjadi data training untuk membentuk model/ pola/ *knowledge* dan data testing dilakukan untuk pengujian (Dava Maulana, Id Hadiana and Rakhmat Umbara, 2024).

2.4 XGBoost

Extreme Gradient Boosting (XGBoost) adalah metode machine learning berbasis ensemble yang menggunakan teknik boosting untuk menghasilkan prediksi yang kuat. XGBoost merupakan pengembangan dari Gradient Boosting Machine (GBM) klasik dengan menggabungkan sejumlah pohon keputusan (decision trees) untuk meningkatkan akurasi prediksi (Abdurrahman, Oktavianto, dan Sintawati, 2022). Proses kerja XGBoost terdiri dari beberapa langkah. Langkah pertama adalah melakukan prediksi awal sederhana dengan menghitung rata-rata target. Kedua, menghitung error sebagai selisih antara prediksi awal dan nilai aktual. Ketiga, membangun pohon keputusan untuk meminimalkan error, biasanya menggunakan metode Gradient Descent. Keempat, menerapkan metode ensemble yang melibatkan prediksi dari pohon sebelumnya, learning rate, dan pohon baru, menghasilkan prediksi ensemble.

Langkah kelima adalah memperbarui prediksi, dilakukan secara iteratif sesuai dengan jumlah pohon yang ditetapkan atau kriteria penghentian. Keenam, regularisasi untuk mengontrol kompleksitas model dan mencegah overfitting. Ketujuh, mengulangi langkah kedua hingga keenam sampai mencapai jumlah pohon atau kriteria penghentian yang ditentukan. Terakhir, hasil prediksi akhir untuk setiap sampel diperoleh dengan menggabungkan seluruh prediksi dari pohon-pohon dalam ensemble (Murdiansyah, 2024).

2.5 Hasil

Pada bagian ini akan diberikan penjelasan rinci mengenai temuan dan analisis dari setiap fase penelitian “Klasifikasi Penyakit Jantung Kardiovaskular Menggunakan Algoritma XGBoost”. Penjelasan meliputi proses pengumpulan data, pre-processing, proses klasifikasi, serta pengujian dan evaluasi kinerja algoritma (Handika Permana, Rakhmat Umbara and Kasyidi, 2024).

2.6 Studi Literatur

Penelitian yang dilakukan oleh Abdurrahman, Oktavianto, dan Sintawati (2022) berjudul *Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search* menggunakan data penderita diabetes dari situs UCI Machine Learning. Penelitian ini dimulai dengan mengatasi nilai yang hilang dalam dataset karena XGBoost tidak dapat bekerja dengan data yang tidak lengkap. Nilai-nilai yang hilang diimputasi dengan nilai bermakna, lalu data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Tiga pendekatan diterapkan pada dataset: tanpa hyperparameter tuning, tuning menggunakan GridSearch, dan Random Search. Hasil menunjukkan bahwa tanpa tuning, log loss negatif sebesar 25% dengan akurasi 75%, sedangkan dengan GridSearch dan Random Search, log loss negatif mencapai 5%, dengan akurasi 95%, yang menegaskan bahwa tuning hyperparameter dapat meningkatkan akurasi secara signifikan.

Penelitian oleh Sukmawati et al. (2024) berjudul *Efektivitas Algoritma AdaBoost dan XGBoost pada Dataset Obesitas Populasi Dewasa* mengevaluasi kinerja algoritma AdaBoost dan XGBoost pada kasus obesitas. Kedua algoritma ini dipilih karena mampu meningkatkan akurasi prediksi pada data yang kompleks, di mana AdaBoost menggunakan pendekatan iteratif dengan model sederhana, sementara XGBoost unggul dalam kecepatan dan efisiensi, ideal untuk dataset besar dengan banyak fitur.

Dalam penelitian *Seleksi Fitur dengan Particle Swarm Optimization pada Klasifikasi Penyakit Parkinson Menggunakan XGBoost* oleh Kurnia et al. (2023), metode Particle Swarm Optimization digunakan untuk seleksi fitur, sementara XGBoost diterapkan untuk klasifikasi. Penelitian ini juga menerapkan SMOTE untuk mengatasi ketidakseimbangan kelas serta tuning hyperparameter pada XGBoost. Hasil menunjukkan bahwa model dengan seleksi fitur dan tuning hyperparameter memiliki AUC sebesar 0,9325, lebih tinggi dibandingkan model tanpa seleksi fitur yang memperoleh AUC 0,9250.



Penelitian *Klasifikasi Penyakit Jantung Tipe Kardiovaskular Menggunakan Adaptive Synthetic Sampling dan Algoritma Extreme Gradient Boosting* menggunakan dataset penyakit jantung dari Kaggle dengan 15 fitur dan 4238 data. Tahapan penelitian meliputi imputasi data, transformasi dengan LabelEncoder, penyeimbangan data menggunakan ADASYN, pembagian data (80% pelatihan, 20% pengujian), serta tuning hyperparameter dengan Bayesian Optimization. Model XGBoost dengan ADASYN mencapai ROC-AUC sebesar 0,971 dan akurasi 0,916, lebih baik daripada model tanpa ADASYN yang memiliki ROC-AUC sebesar 0,698 dan akurasi 0,841. Metode ADASYN terbukti efektif meningkatkan performa pada dataset tidak seimbang, dan Bayesian Optimization membantu menemukan parameter optimal.

Penelitian oleh Murdiansyah (2024) berjudul *Prediksi Stroke Menggunakan Extreme Gradient Boosting* menerapkan algoritma XGBoost untuk memprediksi risiko stroke. Model XGBoost ini dibandingkan dengan model lain seperti Stacking, Random Forest, dan Majority Voting. Hasil menunjukkan bahwa XGBoost mencapai akurasi 95,4%, namun performanya masih di bawah Stacking dan Random Forest, di mana Stacking memberikan akurasi tertinggi sebesar 98%.

3. ANALISA DAN PEMBAHASAN

Tabel 1. Hasil Pembahasan

| No | Nama Peneliti dan Tahun | Metode Yang Dibahas | Tujuan Penelitiannya | Hasil yang didapat |
|----|---|--|--|---|
| 1 | (Abdurrahman, Oktavianto and Sintawati, 2022) | algoritma XGBoost sebagai teknik klasifikasi utama untuk mendeteksi penyakit diabetes. Algoritma ini merupakan model <i>ensemble</i> yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi klasifikasi. Selain itu, metode optimasi Gridsearch dan Random Search digunakan sebagai teknik <i>hyperparameter tuning</i> untuk meningkatkan kinerja XGBoost. Gridsearch melakukan pencarian secara ekstensif di seluruh ruang parameter yang memungkinkan, sementara Random Search memilih kombinasi parameter secara acak untuk efisiensi waktu dan sumber daya | Penelitian ini bertujuan untuk mengoptimalkan akurasi klasifikasi penyakit diabetes dengan menggunakan algoritma XGBoost yang ditingkatkan melalui <i>hyperparameter tuning</i> . Tujuan akhirnya adalah untuk mengetahui apakah optimasi <i>hyperparameter</i> ini dapat meningkatkan akurasi klasifikasi secara signifikan, sehingga lebih efektif untuk digunakan dalam diagnosis dini penyakit diabetes. | Penelitian menunjukkan bahwa XGBoost tanpa <i>hyperparameter tuning</i> menghasilkan akurasi 75% dengan nilai <i>negative log loss</i> sebesar 25%. Namun, setelah dilakukan optimasi menggunakan Gridsearch dan Random Search, akurasi meningkat menjadi 95% dengan nilai <i>negative log loss</i> sebesar 5%. Hal ini menegaskan bahwa baik Gridsearch maupun Random Search mampu meningkatkan akurasi XGBoost secara signifikan. Meskipun demikian, Gridsearch membutuhkan waktu eksekusi yang lebih lama |



| | | | | |
|---|-----------------------------------|--|--|---|
| | | | | karena melakukan pencarian di seluruh ruang parameter, sedangkan Random Search lebih efisien dalam waktu eksekusi |
| 2 | (Sukmawati <i>et al.</i> , 2024). | Metode yang diimplementasikan dalam penelitian ini adalah algoritma AdaBoost dan XGBoost. AdaBoost adalah teknik dalam data mining yang dirancang untuk meningkatkan tingkat akurasi pada metode klasifikasi. Selain itu, AdaBoost dikatakan sebagai ensemble learning yang sering digunakan pada algoritma boosting | Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma AdaBoost dan XGBoost dalam menganalisis pola makan, aktivitas fisik, serta faktor-faktor risiko lain yang berkontribusi terhadap obesitas pada populasi dewasa | Hasil penelitian menunjukkan bahwa penggunaan metode XGBoost cenderung lebih unggul dibandingkan dengan AdaBoost. Adapun nilai akurasi, precision dan recall dari XGBoost sebesar 92%. Sedangkan nilai akurasi dan recall dari AdaBoost sebesar 40% serta precision sebesar 39%. Hal tersebut dikarenakan metode AdaBoost cenderung menggunakan pohon keputusan yang dangkal dan cenderung mengalami kesulitan dengan dataset yang sangat kompleks serta menangani overfitting. Sedangkan, metode Xgboost tersebut cenderung cocok dengan dataset yang digunakan pada penelitian ini. |
| 3 | (Shafila, 2021) | Penelitian ini menggunakan metode Extreme Gradient Boosting (XGBoost) untuk klasifikasi data bioinformatika pada penyakit Ebola. Selain itu, tuning | Penelitian bertujuan untuk mengklasifikasikan data status kekebalan pasien Ebola menggunakan metode XGBoost dan meningkatkan akurasi | Hasil penelitian menunjukkan bahwa akurasi awal XGBoost mencapai 69,23%, yang kemudian meningkat menjadi 76,92% |



| | | | | |
|---|--|---|--|---|
| | | hyperparameter dilakukan menggunakan grid search untuk meningkatkan akurasi model. | model dengan optimasi hyperparameter. | setelah tuning hyperparameter. Nilai Kappa tertinggi yang dicapai adalah 0,675, dan Area Under Curve (AUC) sebesar 0,85 |
| 4 | (Kurnia <i>et al.</i> , 2023) | Penelitian ini menggunakan metode Particle Swarm Optimization (PSO) sebagai teknik seleksi fitur, dikombinasikan dengan XGBoost sebagai algoritma klasifikasi utama. Selain itu, penelitian ini juga menggunakan SMOTE untuk mengatasi ketidakseimbangan kelas pada dataset, dan menerapkan <i>hyperparameter tuning</i> dengan metode Random Search untuk meningkatkan performa XGBoost lebih lanjut. PSO dipilih untuk mengurangi jumlah fitur yang tidak relevan, sehingga meningkatkan akurasi dan efisiensi waktu dalam pengklasifikasian | Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi penyakit Parkinson menggunakan XGBoost dengan bantuan seleksi fitur dari PSO. Dengan menerapkan seleksi fitur, SMOTE, dan hyperparameter tuning, penelitian ini berharap dapat mengoptimalkan hasil klasifikasi, terutama untuk meningkatkan akurasi dan nilai AUC pada dataset Parkinson yang didapat dari rekaman suara pasien | Hasil penelitian menunjukkan bahwa model XGBoost dengan seleksi fitur PSO mencapai nilai AUC sebesar 0,9483 ketika dikombinasikan dengan SMOTE dan hyperparameter tuning. Sebagai perbandingan, model tanpa seleksi fitur hanya mencapai nilai AUC 0,9366 dalam kondisi yang sama. Hal ini menunjukkan bahwa penggunaan PSO sebagai seleksi fitur memberikan peningkatan akurasi klasifikasi yang signifikan. Selain itu, hasil evaluasi menunjukkan bahwa kombinasi seleksi fitur PSO, SMOTE, dan Random Search memberikan performa terbaik dibandingkan metode-metode lain yang hanya menerapkan sebagian teknik tersebut |
| 5 | (Muhammad Dava Maulana, Asep Id Hadiana, | Penelitian ini menggunakan algoritma XGBoost untuk klasifikasi | Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja algoritma XGBoost | Penelitian ini menunjukkan bahwa algoritma XGBoost yang |



| | | | | |
|--|------------------------------|--|---|--|
| | Fajri Rakhmat Umbara. 2023) | kualitas air minum, yang merupakan metode <i>Gradient Boosting</i> yang dikembangkan untuk meningkatkan akurasi dan mengurangi risiko <i>overfitting</i> . Dalam penelitian ini, data kualitas air diperoleh dari situs Kaggle, dan setelah dilakukan proses <i>preprocessing</i> termasuk <i>data cleaning</i> dan <i>class balancing</i> menggunakan teknik <i>oversampling</i> , data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Penelitian ini juga menggunakan <i>hyperparameter tuning</i> dengan metode Grid Search untuk menemukan nilai parameter optimal bagi XGBoost, seperti jumlah estimasi (<i>n_estimators</i>), kedalaman maksimum pohon (<i>max_depth</i>), dan <i>learning rate</i> | dalam mengklasifikasikan kualitas air minum berdasarkan dataset yang telah diolah. Dengan melakukan optimasi parameter, penelitian ini berupaya untuk memperoleh akurasi dan performa yang lebih tinggi dalam menentukan apakah air layak atau tidak untuk dikonsumsi | diterapkan pada dataset air minum menghasilkan akurasi sebesar 82,29% . Selain itu, nilai precision yang diperoleh adalah 78,62% , recall 85,90% , dan F1-score 82,09% . Hasil ini menunjukkan bahwa XGBoost dengan teknik <i>oversampling</i> dan <i>hyperparameter tuning</i> adalah metode klasifikasi yang efektif untuk kasus klasifikasi kualitas air minum, memberikan model dengan tingkat akurasi dan performa yang baik dalam membedakan air layak minum dari yang tidak |
|--|------------------------------|--|---|--|

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Extreme Gradient Boosting (XGBoost) terbukti efektif dalam mengklasifikasikan penyakit jantung. XGBoost mengungguli algoritma lain seperti Random Forest dan Decision Tree dalam beberapa aspek, yaitu memberikan akurasi prediksi yang lebih tinggi, efisiensi komputasi yang lebih baik, serta kemampuan menghindari overfitting yang lebih baik. Penggunaan teknik Adaptive Synthetic Sampling Approach (ADASYN) dan optimasi parameter menggunakan Bayesian Optimization semakin meningkatkan performa model XGBoost dalam mengatasi masalah ketidakseimbangan data dan menghasilkan model yang andal. Temuan dari penelitian ini menjadi rujukan berharga dalam pengembangan sistem pendeteksian penyakit jantung berbasis teknologi yang lebih handal, sehingga dapat membantu tenaga medis dalam pengambilan keputusan diagnosis yang lebih baik dan tepat waktu.

Kesuksesan penelitian ini diharapkan dapat mendorong adopsi dan pengembangan lebih lanjut dari algoritma XGBoost dalam bidang kesehatan, khususnya untuk deteksi dini dan prediksi penyakit jantung, sehingga dapat berkontribusi pada peningkatan akurasi diagnosis, efisiensi alokasi sumber daya kesehatan, dan pengembangan intervensi medis yang lebih tepat sasaran.

Extreme Gradient Boosting (XGBoost) merupakan algoritma *machine learning* berbasis *boosting* yang sangat efektif untuk masalah klasifikasi dan regresi, termasuk dalam prediksi penyakit jantung. Berdasarkan penelitian yang ada, XGBoost terbukti unggul dalam beberapa aspek:



1. **Akurasi Tinggi:** XGBoost mampu menghasilkan akurasi yang lebih tinggi dibandingkan metode *machine learning* lainnya, seperti *Random Forest* atau *Decision Tree*. Hal ini berkat teknik *gradient boosting* yang menggabungkan beberapa model lemah menjadi model yang lebih kuat.
2. **Efisiensi Komputasi:** XGBoost dirancang untuk efisiensi dan kecepatan, sehingga dapat memproses data besar dan kompleks dalam waktu yang relatif singkat. Optimalisasi memori yang efisien menjadikan XGBoost pilihan yang baik untuk menangani data berdimensi tinggi.
3. **Kemampuan Mengatasi *Overfitting*:** Dengan adanya fitur *regularization* yang canggih, XGBoost mampu mengurangi risiko *overfitting*, terutama pada data medis yang memiliki banyak variabel independen yang saling berkorelasi.

DAFTAR PUSTAKA

- Abdurrahman, G., Oktavianto, H. and Sintawati, M. (2022) 'Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search Pada Klasifikasi Penyakit Diabetes', *Informatics* [Preprint].
- Dava Maulana, M., Id Hadiana, A. and Rakhmat Umbara, F. (2024) 'Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum', *Teknik Informatika* [Preprint]. Available at: <https://doi.org/10.36040/jati.v7i5.7308>.
- Fadlillah, S., Sucipto, A. and Amestiasih, T. (2021) 'USIA, JENIS KELAMIN, PERILAKU MEROKOK, DAN IMT BERHUBUNGAN DENGAN RESIKO PENYAKIT KARDIOVASKULER', *Jurnal keperawatan* [Preprint].
- Handika Permana, A., Rakhmat Umbara, F. and Kasyidi, F. (2024) 'Klasifikasi Penyakit Jantung Tipe Kardiovaskular Menggunakan Adaptive Synthetic Sampling dan Algoritma Extreme Gradient Boosting', *Building of Informatics, Technology and Science (BITS)* [Preprint].
- Kurnia, D. et al. (2023) 'SELEKSI FITUR DENGAN PARTICLE SWARM OPTIMIZATION PADA KLASIFIKASI PENYAKIT PARKINSON MENGGUNAKAN XGBOOST Deni', *Jurnal Teknologi Informasi dan Ilmu Komputer* [Preprint].
- Murdiansyah, D.T. (2024) 'Prediksi Stroke Menggunakan Extreme Gradient Boosting', *Informatika dan Komputer* [Preprint].
- Rika Widianita, D. (2023) 'ANALISIS PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK KLASIFIKASI PENYAKIT KARDIOVASKULAR', *AT-TAWASSUTH: Jurnal Ekonomi Islam*, VIII(I), pp. 1–19.
- Shafila, G.A. (2021) 'Implementasi Metode Extreme Gradient Boosting (Xgboost) untuk Klasifikasi pada Data Bioinformatika', *statistika* [Preprint].
- Sukmawati, C.E. et al. (2024) 'Efektivitas algoritma AdaBoost dan XGBoost pada dataset obesitas populasi dewasa', *Informatics* [Preprint].
- Wardhana, R.G., Wang, G. and Sibuea, F. (2023) 'Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia', *Journal of Information System Management (JOISM)*, 5(1), pp. 40–45. Available at: <https://doi.org/10.24076/joism.2023v5i1.1136>.