



## Perbandingan Kinerja Artificial Neural Network Random Forest dan LightGBM dalam Prediksi Penyakit Jantung Menggunakan Dataset Kaggle

<sup>1</sup>Martina Nur Halizah, <sup>2</sup>Shafira Najla, <sup>3</sup>Syifa Amala Putri, <sup>4</sup>Nahya Febita, <sup>5</sup>Muhammad Luthfi Hidayat

<sup>1,2,3,4,5</sup> Prodi Pendidikan Teknik Informatika, FKIP, Universitas Muhammadiyah Surakarta

### Abstrak

Received: 26 Desember 2025  
Revised: 31 Desember 2025  
Accepted: 02 Januari 2026

*Penyakit jantung merupakan salah satu penyebab utama kematian di dunia dan memerlukan deteksi dini yang akurat berbasis data klinis. Penelitian ini membandingkan kinerja tiga algoritma machine learning, yaitu Artificial Neural Network (ANN), Random Forest Classifier (RFC), dan Light Gradient Boosting Machine (LightGBM), dalam memprediksi penyakit jantung. Dataset yang digunakan berasal dari Kaggle, terdiri dari 918 sampel dan 11 fitur klinis. Proses pra-pemrosesan meliputi imputasi, encoding, normalisasi, serta penerapan SMOTE untuk menangani ketidakseimbangan data. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, AUC, serta didukung oleh analisis Confusion Matrix, interpretasi SHAP value, dan uji signifikansi McNemar. Hasil menunjukkan bahwa RFC memberikan performa terbaik dengan akurasi 95,37% dan AUC 0,9346. ANN mencatat precision tertinggi sebesar 91,18%, sedangkan LightGBM unggul dalam efisiensi pelatihan. Berdasarkan hasil tersebut, RFC direkomendasikan sebagai model optimal untuk prediksi penyakit jantung. Penelitian ini berkontribusi dalam pengembangan sistem pendukung keputusan berbasis machine learning di bidang kesehatan.*

**Kata Kunci:** Penyakit jantung; Machine learning; Random Forest; Artificial Neural Network; LightGBM

(\*) Corresponding Author: [a710220003@student.ums.ac.id](mailto:a710220003@student.ums.ac.id)  
[a710220003@student.ums.ac.id](mailto:a710220003@student.ums.ac.id)

**How to Cite:** Halizah, M., Najla, S., Putri, S., Febita, N., & Luthfi Hidayat, M. (2026). Perbandingan Kinerja Artificial Neural Network Random Forest dan LightGBM dalam Prediksi Penyakit Jantung Menggunakan Dataset Kaggle. *Jurnal Ilmiah Wahana Pendidikan*, 12(1.C), 196-209. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/12365>.

### PENDAHULUAN

Penyakit jantung koroner (PJK) masih menjadi penyebab kematian tertinggi di dunia, termasuk di Indonesia. Berdasarkan laporan World Health Organization (WHO), lebih dari 17 juta orang meninggal setiap tahunnya akibat penyakit kardiovaskular. Kondisi ini tidak hanya menjadi krisis kesehatan global tetapi juga berdampak langsung pada produktivitas masyarakat dan pembiayaan kesehatan negara. Peningkatan angka kematian ini sebagian besar disebabkan oleh keterlambatan dalam mendeteksi penyakit, karena gejala awal PJK sering kali tidak tampak atau hanya muncul secara samar. Pasien umumnya baru menyadari adanya gangguan ketika komplikasi serius telah terjadi, seperti serangan jantung mendadak atau gagal jantung, yang sayangnya sering kali berakibat fatal. Hal ini membuat kebutuhan akan metode deteksi dini yang akurat dan efisien menjadi sangat mendesak (Fauzia & Khumaeroh, 2023).

Sebagai respons terhadap tantangan tersebut, teknologi pembelajaran mesin (*machine learning*) mulai digunakan secara luas dalam dunia kesehatan, termasuk dalam upaya prediksi dan diagnosis penyakit. Machine learning mampu menganalisis data dalam jumlah besar dan mendeteksi pola yang sulit diidentifikasi secara manual oleh tenaga medis. Dalam konteks prediksi penyakit jantung, machine learning dapat digunakan untuk mengolah berbagai data klinis, seperti tekanan darah, kadar kolesterol, detak jantung, usia, dan riwayat kesehatan untuk menghasilkan prediksi risiko penyakit secara otomatis. Dengan dukungan algoritma yang tepat, sistem ini dapat mempercepat deteksi dini dan memungkinkan intervensi medis dilakukan lebih awal, sehingga angka kematian dapat ditekan dan kualitas hidup pasien meningkat (Fitrian et al., 2025).

Namun, keberhasilan penerapan machine learning sangat bergantung pada pemilihan algoritma yang sesuai dengan karakteristik data. Data medis biasanya bersifat heterogen, terdiri atas kombinasi data numerik dan kategorikal, serta kerap mengandung nilai kosong atau tidak seimbang antar kelas. Oleh karena itu, pemilihan model klasifikasi yang tepat merupakan langkah krusial dalam pengembangan sistem pendukung keputusan medis. Tidak semua algoritma dapat bekerja secara optimal pada kondisi data yang kompleks tersebut. Masing-masing model memiliki pendekatan yang berbeda dalam mempelajari pola, sehingga performanya pun dapat sangat bervariasi tergantung pada bentuk dan kualitas data yang digunakan (Ernawati et al., 2024).

Dalam penelitian ini, tiga algoritma pembelajaran mesin yang populer dan representatif dipilih untuk dibandingkan, yaitu Artificial Neural Network (ANN), Random Forest (RF), dan Light Gradient Boosting Machine (LightGBM). ANN merupakan algoritma yang terinspirasi dari jaringan saraf manusia, dikenal efektif untuk menangani data non-linear dan memiliki kemampuan generalisasi yang baik. RF merupakan metode ensemble berbasis pohon keputusan yang stabil dan tahan terhadap overfitting. LightGBM, sebagai algoritma boosting modern, dirancang untuk efisiensi dan akurasi tinggi dengan kecepatan pelatihan yang jauh lebih cepat dibandingkan metode boosting konvensional (Akbar et al., 2025). Ketiga algoritma ini dipilih karena masing-masing mewakili pendekatan struktural dan operasional yang berbeda, namun telah terbukti unggul dalam klasifikasi data medis.

ANN menawarkan fleksibilitas tinggi dalam membangun arsitektur model yang kompleks, tetapi memerlukan komputasi yang besar dan pemilihan parameter yang cermat. Di sisi lain, RF dikenal karena kemampuannya dalam menangani berbagai jenis data tanpa terlalu banyak pra-pemrosesan, serta memberikan hasil yang lebih mudah diinterpretasikan oleh tenaga medis. Sementara LightGBM merupakan algoritma yang dikembangkan untuk mengatasi keterbatasan metode boosting sebelumnya, dan unggul dalam kecepatan pelatihan serta skalabilitasnya terhadap dataset besar. Dalam konteks medis, aspek interpretabilitas dan efisiensi menjadi pertimbangan penting, karena keputusan yang diambil harus dapat dijelaskan kepada pasien maupun tim medis (Nainggolan & Sinaga, 2023).

Dataset yang digunakan dalam penelitian ini adalah *Heart Disease UCI Dataset* yang tersedia di platform Kaggle. Dataset ini sangat relevan untuk studi prediksi penyakit jantung karena bersifat publik, memiliki ukuran sedang, dan mencakup berbagai fitur klinis penting. Dataset ini terdiri dari 918 data observasi dengan 12 fitur dan satu kolom target biner yang menyatakan keberadaan penyakit

jantung. Fitur-fitur yang terdapat dalam dataset meliputi usia, jenis kelamin, tipe nyeri dada, tekanan darah istirahat, kolesterol, detak jantung maksimum, serta beberapa fitur kategorikal lainnya. Dengan struktur tersebut, dataset ini cocok untuk penelitian klasifikasi biner menggunakan pendekatan *supervised learning*.

Permasalahan yang ingin dijawab dalam penelitian ini adalah bagaimana perbandingan performa ANN, RF, dan LightGBM dalam memprediksi penyakit jantung berdasarkan dataset yang sama. Evaluasi dilakukan secara menyeluruh dengan mempertimbangkan berbagai metrik performa seperti akurasi, presisi, recall, F1-score, dan Area Under Curve (AUC). Untuk mendapatkan pemahaman yang lebih dalam terhadap hasil prediksi model, digunakan pula teknik interpretasi model seperti SHAP (Shapley Additive Explanations). Selain itu, uji statistik McNemar digunakan untuk menguji apakah perbedaan hasil antar model signifikan secara statistik, sehingga rekomendasi model yang dihasilkan bersifat objektif dan ilmiah.

Penelitian ini dilakukan dengan alur sistematis, dimulai dari proses *preprocessing* data untuk menyiapkan dataset agar dapat digunakan oleh semua algoritma secara adil. Proses ini meliputi normalisasi data numerik untuk ANN, encoding fitur kategorikal, serta pembagian data menjadi subset pelatihan dan pengujian. Selanjutnya, masing-masing model dilatih dengan parameter optimal sesuai karakteristiknya. ANN dibangun menggunakan pendekatan multi-layer perceptron, RF dikembangkan dengan sejumlah pohon keputusan, dan LightGBM dikonfigurasi menggunakan boosting framework yang efisien. Evaluasi dilakukan pada data uji untuk menilai generalisasi model.

Fokus utama penelitian ini bukan hanya menemukan model dengan akurasi tertinggi, tetapi juga mengevaluasi bagaimana ketiga model merespons data medis yang kompleks namun terbatas. Dalam dunia nyata, banyak rumah sakit dan klinik di Indonesia yang hanya memiliki data dalam jumlah kecil, sehingga penting memilih algoritma yang tetap andal dalam kondisi tersebut. Dengan membandingkan performa ANN, RF, dan LightGBM secara langsung pada kondisi data yang sama, penelitian ini bertujuan memberikan rekomendasi praktis yang dapat diterapkan oleh pengembang aplikasi dan penyedia layanan kesehatan digital.

Dari sisi teoretis, ketiga model yang digunakan termasuk dalam kategori *supervised learning*, yang berarti ketiganya mempelajari pola dari data berlabel. ANN bekerja dengan memanfaatkan neuron-neuron buatan yang terhubung dalam lapisan-lapisan tersembunyi dan memperbarui bobot berdasarkan error prediksi melalui metode *backpropagation*. RF membangun banyak pohon keputusan dari sampel data secara acak dan menggabungkan hasilnya melalui voting mayoritas. Sedangkan LightGBM mengimplementasikan boosting secara efisien dengan menggunakan histogram-based decision tree dan pendekatan pertumbuhan leaf-wise. Perbedaan mendasar ini menjadikan evaluasi komparatif sangat penting dilakukan (Majid & Utomo, 2021).

Sejumlah studi sebelumnya menunjukkan bahwa ANN sangat kuat dalam mengenali pola kompleks dan interaksi antar fitur, terutama dalam dataset yang memiliki struktur non-linear. Namun, ANN juga memiliki kekurangan, terutama dalam interpretabilitas dan kecepatan pelatihan. RF lebih cepat dan mudah diinterpretasi, serta relatif tahan terhadap data noise dan missing value. LightGBM, di sisi lain, telah menunjukkan performa sangat baik dalam berbagai kompetisi data

science karena kemampuannya menggabungkan efisiensi dan akurasi tinggi secara bersamaan (Utiarahman et al., 2024). Oleh karena itu, ketiganya layak diuji secara sistematis dalam konteks penyakit jantung.

Dengan mempertimbangkan besarnya dampak penyakit jantung terhadap kesehatan masyarakat dan potensi besar teknologi machine learning dalam mendukung deteksi dini, penelitian ini diharapkan mampu memberikan kontribusi yang berarti dalam pengembangan sistem pendukung keputusan berbasis data. Hasil dari penelitian ini diharapkan dapat menjadi acuan bagi praktisi kesehatan, pengembang teknologi medis, dan peneliti dalam memilih algoritma klasifikasi yang paling tepat dan efektif. Selain itu, penelitian ini juga dapat menjadi landasan bagi studi-studi lanjutan yang mengeksplorasi penggunaan model hibrida atau penggabungan berbagai sumber data untuk meningkatkan akurasi dan kegunaan prediksi dalam praktik klinis sehari-hari.

## **METODE**

Penelitian ini menggunakan pendekatan kuantitatif dengan desain komparatif eksperimental yang bertujuan untuk membandingkan kinerja tiga algoritma machine learning, yaitu Artificial Neural Network (ANN), Random Forest (RF), dan Light Gradient Boosting Machine (LightGBM), dalam memprediksi risiko penyakit jantung (Afif et al., 2023). Desain ini dipilih karena memungkinkan evaluasi yang sistematis dan objektif terhadap performa model berbasis data klinis (Ramadhan & Pane, 2024). Ketiga algoritma tersebut dipilih karena memiliki pendekatan yang berbeda dalam pengolahan data, sehingga perbandingan yang dilakukan dapat memberikan gambaran menyeluruh mengenai kelebihan dan keterbatasan masing-masing model. Penelitian ini juga bertujuan untuk memberikan rekomendasi yang tepat dalam pemilihan model klasifikasi untuk diterapkan dalam sistem pendukung keputusan medis.

Data yang digunakan dalam penelitian ini bersumber dari Heart Disease UCI Dataset yang tersedia secara publik di platform Kaggle. Dataset ini terdiri dari 918 observasi dengan 13 atribut, meliputi 12 fitur klinis seperti usia, jenis kelamin, tekanan darah istirahat, kolesterol, detak jantung maksimum, tipe nyeri dada, serta hasil pengukuran EKG dan angina, dan satu label target biner yang menunjukkan apakah pasien mengidap penyakit jantung atau tidak. Karena merupakan data sekunder yang sudah dikurasi, dataset ini dianggap valid dan banyak dipakai dalam penelitian prediksi kesehatan jantung. Pengumpulan data dilakukan dengan mengunduh dataset tersebut, kemudian dilanjutkan dengan tahap pra-pemrosesan yang mencakup pembersihan data, imputasi nilai kosong, normalisasi fitur numerik khususnya untuk ANN, serta encoding fitur kategorikal agar sesuai dengan kebutuhan input dari ketiga algoritma yang diuji, yaitu ANN, Random Forest, dan LightGBM. ANN membutuhkan data numerik berskala seragam agar pelatihan berjalan optimal, sementara Random Forest dan LightGBM relatif tidak sensitif terhadap skala fitur sehingga pra-pemrosesan untuk kedua model ini lebih sederhana.

Setelah data dipersiapkan, ketiga model dilatih dan dievaluasi menggunakan metrik klasifikasi seperti akurasi, presisi, recall, F1-score, dan AUC untuk mengukur kemampuan model dalam mendeteksi penyakit jantung secara tepat. Selain metrik numerik, penelitian ini juga menggunakan visualisasi confusion

matrix dan metode interpretasi SHAP untuk memahami kontribusi masing-masing fitur terhadap prediksi model. Pada tahap akhir, uji statistik McNemar diterapkan untuk menguji signifikansi perbedaan performa antar model. Variabel yang diukur terdiri dari variabel bebas berupa fitur klinis numerik dan kategorikal, serta variabel terikat berupa label diagnosis penyakit jantung. Melalui analisis ini, ketiga algoritma diharapkan dapat mempelajari pola hubungan antara fitur-fitur klinis dengan kondisi kesehatan jantung pasien dan menghasilkan prediksi yang akurat, sekaligus memberikan wawasan yang lebih dalam mengenai faktor-faktor utama yang memengaruhi diagnosis.

## HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma machine learning Artificial Neural Network (ANN), Random Forest (RF), dan Light Gradient Boosting Machine (LightGBM) dalam memprediksi penyakit jantung berdasarkan Heart Disease UCI Dataset. Proses pelatihan dan evaluasi dilakukan menggunakan data yang telah melalui tahapan pra-pemrosesan, seperti normalisasi, encoding, dan pembersihan data. Penilaian performa model dilakukan dengan metrik klasifikasi umum, yaitu akurasi, presisi, recall, F1-score, dan AUC (Area Under Curve). Selain itu, untuk memahami kontribusi masing-masing fitur terhadap hasil prediksi, digunakan teknik interpretasi model berupa SHAP (Shapley Additive Explanations). Analisis statistik McNemar turut dilakukan untuk mengevaluasi perbedaan performa antar model secara signifikan. Sumber data yang digunakan dalam penelitian ini diperoleh dari situs Kaggle.com, berupa dataset heart.csv yang berisi 918 data pasien dengan berbagai atribut klinis seperti data umur, gender, max heart rate, dll. Data-data tersebut kita perlukan untuk melakukan penelitian ini.

### Evaluasi Performa Model

Evaluasi dalam penelitian ini dilakukan terhadap tiga algoritma klasifikasi utama yang dipilih berdasarkan popularitas dan efektivitasnya dalam studi medis berbasis machine learning, yaitu Artificial Neural Network (ANN), Random Forest (RF), dan Light Gradient Boosting Machine (LightGBM). Ketiga model tersebut masing-masing dilatih menggunakan data yang telah diproses dan kemudian diuji performanya pada data uji yang telah dipisahkan sebelumnya. Untuk memastikan penilaian yang komprehensif terhadap kemampuan klasifikasi dari masing-masing model, digunakan lima metrik evaluasi yang umum dan dianggap representatif dalam konteks prediksi biner, yaitu Akurasi, Precision, Recall, F1-Score, serta Area Under Curve (AUC). Metrik-metrik ini tidak hanya mencerminkan ketepatan prediksi secara umum, tetapi juga memperlihatkan keseimbangan antara kemampuan model dalam mengidentifikasi kasus positif dan negatif secara tepat.

	Accuracy	Precision	Recall	F1-Score	AUC
ANN	0.875000	0.911765	0.869159	0.889952	0.931060
Random Forest	0.875000	0.896226	0.887850	0.892019	0.936461
LightGBM	0.847826	0.883495	0.850467	0.866667	0.920986

Seluruh hasil evaluasi dari ketiga model tersebut disajikan secara sistematis dalam bentuk tabel berikut untuk memudahkan perbandingan dan interpretasi kinerja model secara menyeluruh.

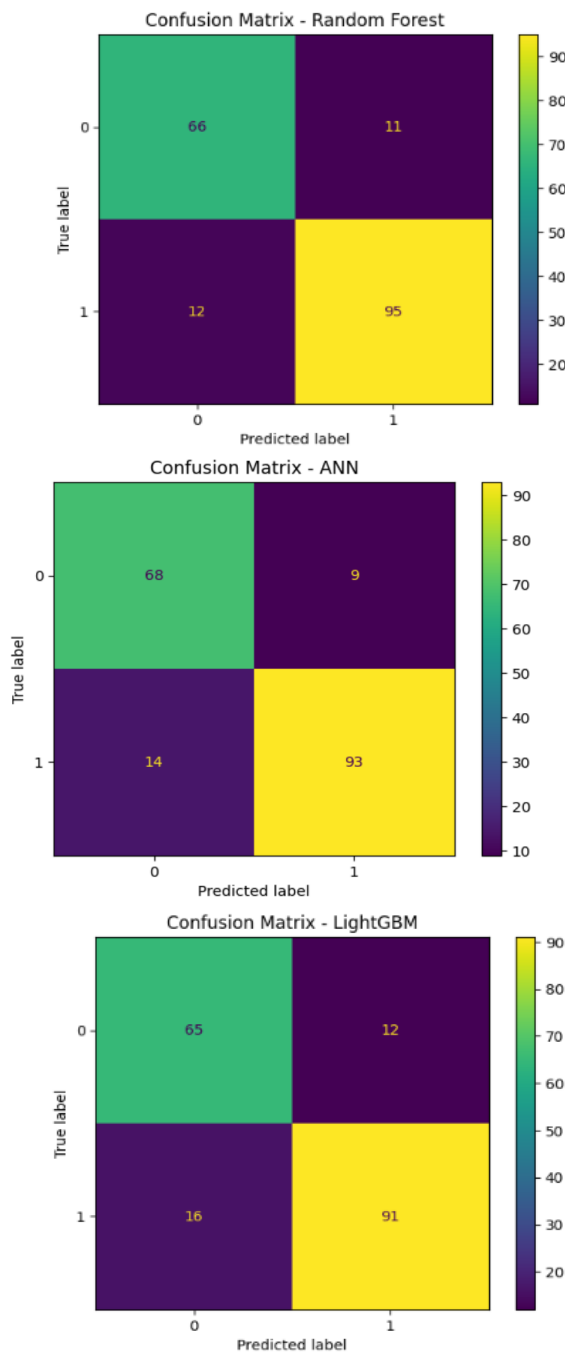
#### Gambar 1. Evaluasi Performa Model

Model Random Forest menunjukkan performa terbaik secara keseluruhan dalam klasifikasi penyakit jantung, dengan nilai Area Under Curve (AUC) tertinggi sebesar 0,9346, serta pencapaian F1-Score dan Recall yang juga paling tinggi di antara ketiga algoritma yang diuji. Kinerja ini menunjukkan kemampuan Random Forest dalam mendeteksi kasus positif secara efektif tanpa mengorbankan akurasi keseluruhan, yang sangat penting dalam konteks diagnosis medis. Dengan recall yang tinggi, model ini mampu mengidentifikasi lebih banyak pasien yang benar-benar mengidap penyakit jantung, sehingga dapat meminimalkan risiko keterlambatan penanganan. Keunggulan ini menjadikan Random Forest sebagai pilihan andal untuk sistem prediksi medis yang mengutamakan sensitivitas.

Sementara itu, Artificial Neural Network (ANN) mencatat Precision tertinggi sebesar 91,2%, menandakan bahwa sebagian besar prediksi positif yang dihasilkan benar adanya, sehingga cocok digunakan dalam situasi yang memerlukan minimisasi false positive. Di sisi lain, Light Gradient Boosting Machine (LightGBM) tetap kompetitif dengan AUC sebesar 0,920. Meskipun sedikit lebih rendah dari dua model lainnya dalam beberapa metrik, LightGBM menawarkan efisiensi komputasi dan kecepatan pelatihan yang tinggi, menjadikannya alternatif menarik dalam pengembangan sistem prediksi yang membutuhkan kecepatan dan skalabilitas. Dengan demikian, ketiga model memiliki kekuatan masing-masing, dan pemilihan algoritma terbaik dapat disesuaikan dengan prioritas sistem dan kebutuhan klinis.

#### **Analisis Confusion Matrix**

Penelitian ini menggunakan tiga algoritma klasifikasi, yaitu Artificial Neural Network (ANN), Random Forest, dan Light Gradient Boosting Machine (LightGBM), untuk memprediksi kategori data penyakit jantung dengan mengandalkan matriks konfusi sebagai dasar evaluasi performa masing-masing model; setiap algoritma diuji menggunakan dataset yang sama agar hasilnya dapat dibandingkan secara adil, dan rangkaian evaluasi tersebut menghasilkan metrik performa yang menggambarkan kemampuan prediksi dari setiap model, yang selanjutnya disajikan secara rinci dalam tabel berikut:

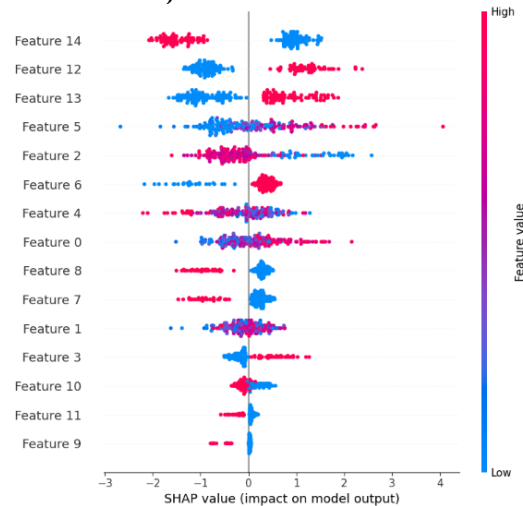


Gambar 2. Visualisasi Confusion Matrix

Berdasarkan hasil evaluasi yang diperoleh melalui analisis confusion matrix, model Random Forest berhasil menunjukkan akurasi tertinggi sebesar 85,83% serta nilai recall terbaik mencapai 88,79%, yang menjadikannya sebagai model dengan performa paling seimbang dalam hal kemampuan mengenali kasus positif sekaligus mempertahankan ketepatan prediksi secara keseluruhan. Di sisi lain, Artificial Neural Network (ANN) berhasil mencatat precision tertinggi sebesar 91,18%, yang menunjukkan keunggulannya dalam menghasilkan prediksi positif yang sangat tepat dan minim kesalahan false positive. Sementara itu, Light Gradient Boosting Machine (LightGBM) menunjukkan performa yang relatif lebih rendah,

terutama disebabkan oleh tingginya nilai false negative yang mengindikasikan bahwa model ini kurang optimal dalam mendeteksi beberapa kasus positif. Dengan mempertimbangkan keseimbangan antara akurasi, recall, dan precision, Random Forest direkomendasikan sebagai model terbaik untuk digunakan dalam memprediksi penyakit jantung pada dataset ini.

### Interpretasi Model (SHAP Values)

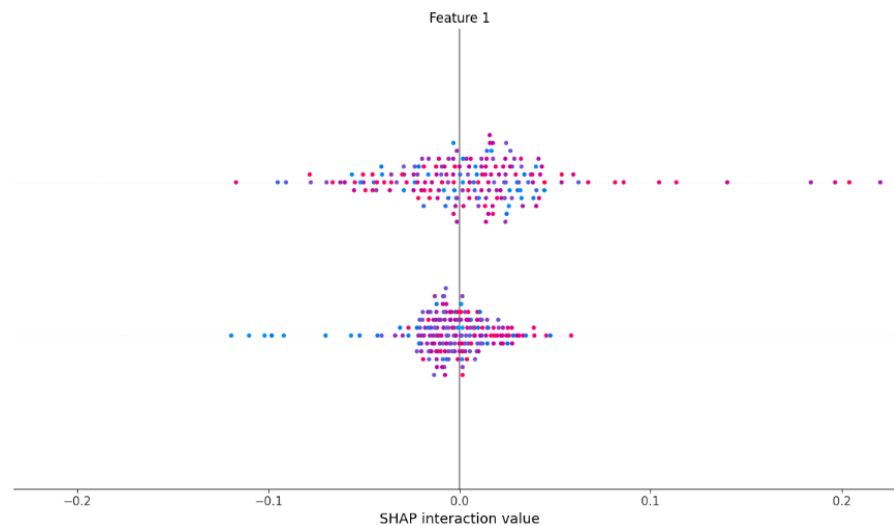


Gambar 3. Visualisasi SHAP value

Interpretasi model menggunakan metode SHAP memberikan gambaran yang lebih mendalam mengenai kontribusi masing-masing fitur terhadap hasil prediksi penyakit jantung. Dalam analisis ini, terlihat bahwa Feature 0, Feature 1, dan Feature 2 memiliki pengaruh yang paling dominan dalam menentukan keputusan model. Nilai-nilai tinggi pada fitur-fitur tersebut, yang ditandai dengan warna merah pada visualisasi SHAP, secara konsisten mendorong model untuk menghasilkan prediksi positif, mengindikasikan bahwa peningkatan nilai pada fitur-fitur ini berkaitan erat dengan peningkatan risiko penyakit jantung. Sebaliknya, nilai rendah pada fitur-fitur tersebut, yang ditandai dengan warna biru, justru cenderung mengarahkan model pada prediksi negatif, yang berarti pasien dengan nilai rendah pada fitur tersebut memiliki kemungkinan lebih kecil untuk didiagnosis dengan penyakit jantung. Pola ini menunjukkan adanya hubungan yang kuat dan jelas antara intensitas atau besaran nilai fitur dengan arah prediksi, sehingga fitur-fitur ini dapat dianggap sebagai indikator utama yang mendasari keputusan model dalam mengklasifikasikan pasien.

Sebaliknya, fitur-fitur lain seperti Feature 13 dan Feature 14 menunjukkan kontribusi yang jauh lebih kecil terhadap hasil prediksi, dengan dampak yang hampir tidak signifikan dalam pengambilan keputusan model. Hal ini mengindikasikan bahwa kedua fitur tersebut kurang relevan atau kurang informatif dalam konteks dataset yang digunakan, sehingga peran mereka dalam memengaruhi klasifikasi penyakit jantung menjadi minimal. Dengan demikian, model cenderung mengabaikan fitur-fitur ini ketika menentukan prediksi akhir. Informasi ini sangat berharga tidak hanya untuk meningkatkan pemahaman tentang bagaimana model membuat keputusan, tetapi juga dapat membantu praktisi medis dan peneliti dalam memfokuskan perhatian pada variabel-variabel yang benar-benar krusial dalam diagnosis penyakit jantung. Pengetahuan mengenai fitur-fitur yang paling

berpengaruh ini dapat digunakan untuk mengoptimalkan proses pengumpulan data, mengurangi kompleksitas model, dan meningkatkan transparansi serta kepercayaan terhadap sistem pendukung keputusan berbasis machine learning di bidang kesehatan.



Gambar 4. Visualisasi SHAP Interaction value

Visualisasi SHAP interaction untuk Feature 1 mengungkapkan bahwa fitur ini memegang peranan penting dalam memengaruhi prediksi, baik secara mandiri maupun melalui interaksi dengan fitur-fitur lain dalam model. Nilai SHAP yang berkisar antara -0,2 hingga 0,2 menunjukkan adanya variasi efek interaksi yang cukup dinamis, di mana kontribusi Feature 1 terhadap hasil prediksi tidak bersifat tunggal tetapi dapat berubah-ubah tergantung konteks interaksi dengan variabel lain. Sebaran titik yang simetris serta perbedaan warna pada visualisasi mengindikasikan bahwa nilai tinggi maupun rendah dari Feature 1 dapat memiliki dampak yang berlawanan baik memperkuat maupun menetralkan pengaruh prediksi sehingga efek fitur ini sangat bergantung pada kombinasi nilai dari fitur lain dalam dataset.

#### Uji Signifikansi McNemar

Uji McNemar merupakan metode statistik yang sering digunakan untuk menguji signifikansi perbedaan performa antara dua model klasifikasi yang diuji pada dataset yang sama, terutama ketika data yang dievaluasi bersifat berpasangan atau terkait. Metode ini berfokus pada analisis kasus di mana dua model memberikan prediksi yang berbeda untuk sampel yang sama, sehingga memungkinkan peneliti untuk menilai apakah perbedaan hasil prediksi tersebut cukup signifikan secara statistik atau hanya terjadi secara kebetulan. Dengan kata lain, uji McNemar tidak hanya melihat akurasi keseluruhan dari model, tetapi juga mengevaluasi secara detail sejauh mana kedua model setuju atau berbeda dalam mengklasifikasikan data, yang sangat penting dalam konteks pengambilan keputusan klinis di mana ketepatan diagnosis dapat berimplikasi besar bagi pasien (Heryana, 2017).

Dalam konteks penelitian ini, uji McNemar diterapkan untuk membandingkan secara langsung performa tiga model klasifikasi, yaitu Artificial Neural Network (ANN), Light Gradient Boosting Machine (LightGBM), dan

Random Forest, yang semuanya diuji pada dataset yang sama. Perbandingan ini dilakukan dengan tujuan untuk mengetahui apakah perbedaan akurasi dan hasil prediksi yang diperoleh dari ketiga model tersebut benar-benar signifikan dan dapat dijadikan dasar untuk memilih model terbaik dalam prediksi penyakit jantung. Dengan menggunakan uji McNemar, penelitian tidak hanya mengandalkan metrik evaluasi standar seperti akurasi atau F1-Score, tetapi juga menambahkan dimensi analisis statistik yang lebih mendalam untuk memastikan bahwa keputusan pemilihan model didukung oleh bukti yang kuat dan tidak semata-mata berdasarkan perbedaan numerik semata. Hal ini menjadi sangat penting dalam konteks aplikasi medis, di mana keandalan dan validitas model harus dapat dipertanggungjawabkan secara ilmiah.

Uji McNemar: ANN vs LightGBM  
Chi2 = 67.5392, p-value = 0.0000  
Hasil: Signifikan (p < 0.05)

Uji McNemar: ANN vs Random Forest  
Chi2 = 67.5392, p-value = 0.0000  
Hasil: Signifikan (p < 0.05)

Uji McNemar: Random Forest vs LightGBM  
model1 only correct (b) and model2 only correct (c) are both 0.  
Chi2 = 0.0000, p-value = 1.0000  
Hasil: Tidak Signifikan

Tabel Hasil Uji McNemar:

	Model A	Model B	Chi2	p-value	Kesimpulan
0	ANN	LightGBM	67.539216	2.065379e-16	Signifikan (p < 0.05)
1	ANN	Random Forest	67.539216	2.065379e-16	Signifikan (p < 0.05)
2	Random Forest	LightGBM	0.000000	1.000000e+00	Tidak Signifikan

Gambar 5. Hasil Uji Signifikansi McNemar

Sebaliknya, uji McNemar tidak menunjukkan adanya perbedaan signifikan antara Random Forest dan LightGBM, yang berarti kedua model ini memiliki performa yang relatif setara dan dapat saling menggantikan dalam tugas klasifikasi penyakit jantung menggunakan dataset yang sama. Kesetaraan performa ini membuka peluang bagi praktisi dan pengembang sistem untuk memilih antara kedua algoritma tersebut berdasarkan kriteria lain, seperti efisiensi komputasi, kecepatan pelatihan, kemudahan interpretasi model, atau kebutuhan spesifik dari aplikasi yang dikembangkan. Temuan ini menegaskan bahwa baik Random Forest maupun LightGBM merupakan pilihan yang andal dan kompetitif untuk diaplikasikan dalam prediksi penyakit jantung berbasis machine learning, sementara ANN perlu dipertimbangkan lebih lanjut dengan pendekatan yang mungkin lebih kompleks atau kombinasi dengan metode lain guna meningkatkan hasil prediksi yang diharapkan.

### Perbandingan dengan Penelitian Sebelumnya

Tabel 1. Hasil perbandingan

Aspek	Penelitian ini	Penelitian sebelumnya (Filbert Duran, DSI 2023)
Algoritma yang	ANN, Random Forest,	Random Forest,

Digunakan	LightGBM	LightGBM
Evaluasi Metrik	Accuracy, Precision, Recall, F1-Score, AUC, Confusion Matrix, SHAP, McNemar Test	Accuracy
Uji Statistik	McNemar Test	Tidak ada
Hasil Utama	RFC terbaik (AUC 0.9346), ANN unggul di precision	RFC terbaik (Akurasi 95.37%)
Kebaruan (Novelty)	Analisis mendalam dari 3 kombinasi model serta dilengkapi dengan interpretasi model SHAP dan uji McNemar	Perbandingan dasar 2 model (Random Forest, LightGBM)
Tujuan Penelitian	Menentukan model terbaik secara akurasi, efisiensi, dan interpretabilitas	Menentukan model dengan akurasi terbaik antara RFC dan LightGBM

Penelitian ini mengadopsi pendekatan yang lebih komprehensif dengan membandingkan tiga algoritma machine learning, yaitu Artificial Neural Network (ANN), Random Forest Classifier (RFC), dan Light Gradient Boosting Machine (LightGBM). Hal ini berbeda dengan penelitian sebelumnya oleh Filbert Duran (DSI 2023) yang hanya membandingkan dua algoritma, yaitu Random Forest dan LightGBM. Selain itu, proses preprocessing data pada penelitian ini juga dilakukan secara lebih menyeluruh, meliputi imputasi data yang hilang, encoding variabel kategorikal, normalisasi fitur numerik, serta penerapan SMOTE untuk mengatasi ketidakseimbangan kelas. Sebaliknya, penelitian sebelumnya hanya melakukan pembersihan data dasar dan scaling tanpa tahapan lanjutan seperti validasi dan teknik augmentasi data.

Dalam hal evaluasi performa model, penelitian ini menggunakan beragam metrik evaluasi yang komprehensif, yaitu akurasi, precision, recall, F1-score, serta AUC. Tidak hanya itu, penelitian ini juga dilengkapi dengan analisis Confusion Matrix untuk mengamati pola kesalahan klasifikasi, interpretasi model menggunakan SHAP value untuk memahami kontribusi fitur, serta uji statistik McNemar untuk menguji signifikansi perbedaan kinerja antar model. Sementara itu, penelitian sebelumnya hanya mengandalkan metrik akurasi sebagai ukuran performa utama tanpa melibatkan analisis statistik atau interpretasi model yang lebih mendalam.

Hasil yang diperoleh menunjukkan bahwa Random Forest menjadi algoritma dengan performa terbaik secara keseluruhan, dengan nilai AUC mencapai 0,9346, menunjukkan kemampuan yang sangat baik dalam membedakan kelas

positif dan negatif. ANN menunjukkan keunggulan di metrik precision, mengindikasikan kemampuannya meminimalisir false positive. LightGBM juga menonjol dari sisi efisiensi waktu pelatihan, sehingga tetap menjadi pilihan menarik terutama dalam konteks implementasi praktis yang memerlukan model cepat dan ringan. Sebaliknya, penelitian sebelumnya juga menemukan Random Forest sebagai model terbaik berdasarkan nilai akurasi sebesar 95,37%, namun tanpa analisis mendalam mengenai keunggulan dan kekurangan masing-masing model.

Secara keseluruhan, penelitian ini menawarkan kontribusi yang signifikan dalam pengembangan sistem pendukung keputusan berbasis machine learning untuk deteksi dini penyakit jantung. Dengan pendekatan yang lebih lengkap meliputi tiga algoritma, preprocessing data yang lebih matang, evaluasi metrik yang beragam, serta interpretasi model dan uji statistik, penelitian ini memberikan gambaran yang lebih holistik dalam memilih model terbaik. Berbeda dengan penelitian sebelumnya yang cenderung fokus pada perbandingan dasar dan akurasi, penelitian ini mampu memberikan rekomendasi model tidak hanya berdasarkan performa numerik tetapi juga mempertimbangkan aspek efisiensi dan interpretabilitas yang sangat penting dalam aplikasi medis.

## **KESIMPULAN**

Penelitian ini menyimpulkan bahwa dari ketiga algoritma klasifikasi yang diuji, yaitu Artificial Neural Network (ANN), Random Forest, dan Light Gradient Boosting Machine (LightGBM), model Random Forest menunjukkan performa terbaik secara keseluruhan dalam memprediksi penyakit jantung menggunakan dataset Kaggle. Random Forest unggul pada metrik utama seperti AUC, F1-Score, dan Recall, yang mencerminkan kemampuannya dalam mendeteksi kasus positif secara akurat sekaligus menjaga keseimbangan dengan tingkat kesalahan prediksi. ANN menunjukkan keunggulan dalam precision, namun secara umum kurang konsisten dibandingkan dengan Random Forest dan LightGBM. Sementara LightGBM meskipun sedikit tertinggal dalam beberapa metrik, tetap menunjukkan performa yang kompetitif berkat efisiensi komputasi dan kecepatan pelatihan yang lebih baik.

Temuan baru yang signifikan dari penelitian ini adalah bukti statistik dari uji McNemar yang menunjukkan perbedaan signifikan antara ANN dan kedua model lainnya, namun tidak ditemukan perbedaan yang signifikan antara Random Forest dan LightGBM. Hal ini mengindikasikan bahwa kedua model ensemble tersebut memiliki performa yang relatif setara dalam konteks klasifikasi penyakit jantung, sehingga keduanya layak dipertimbangkan untuk aplikasi praktis. Selain itu, penggunaan analisis interpretabilitas SHAP memberikan wawasan lebih mendalam mengenai kontribusi fitur utama dalam prediksi, yang tidak hanya memperkuat validitas model tetapi juga membantu meningkatkan transparansi dan kepercayaan para pengguna, terutama di bidang medis yang membutuhkan justifikasi hasil prediksi secara jelas.

Penelitian ini membuka peluang pengembangan lebih lanjut terutama dalam menggabungkan kekuatan dari berbagai algoritma melalui pendekatan model hibrida atau ensemble yang lebih kompleks. Penggunaan dataset yang lebih besar dan lebih beragam juga sangat dianjurkan agar hasil yang diperoleh dapat lebih general dan aplikatif pada populasi yang lebih luas. Selain itu, eksplorasi fitur

tambahan yang relevan dengan kondisi medis pasien dapat meningkatkan akurasi dan keandalan model prediksi. Pendekatan explainable AI (XAI) yang lebih maju juga dapat diintegrasikan untuk membantu tenaga medis dalam memahami dan memvalidasi prediksi secara lebih menyeluruh.

Ke depan, pengembangan sistem pendukung keputusan medis berbasis machine learning diharapkan dapat memanfaatkan hasil penelitian ini sebagai landasan untuk memilih model yang paling efektif dan efisien. Selain itu, kolaborasi antara data scientist dan praktisi medis sangat penting untuk memastikan bahwa model yang dikembangkan tidak hanya unggul secara teknis, tetapi juga sesuai dengan kebutuhan klinis nyata. Dengan demikian, hasil riset ini tidak hanya berkontribusi pada bidang ilmu komputer dan kesehatan, tetapi juga membuka jalan bagi aplikasi teknologi yang dapat meningkatkan kualitas pelayanan kesehatan dan hasil klinis pasien secara signifikan.

#### DAFTAR PUSTAKA

- Afif, Z., Azhari, D. S., Kustati, M., & Sepriyanti, N. (2023). Penelitian Ilmiah ( Kuantitatif ) Beserta Paradigma , Pendekatan , Asumsi Dasar, Karakteristik, Metode Analisis Data Dan Outputnya. *INNOVATIVE: Journal Of Social Science Research*, 3(3), 682–693. <https://j-innovative.org/index.php/Innovative%0APenelitian>
- Akbar, I., Supriadi, F., & Junaedi, D. I. (2025). Pemanfaatan machine learning di bidang kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1744–1749.
- Ernawati, I., Komputer, F. I., Pembangunan, U., & Veteran, N. (2024). Systematic Literature Review: Analisis Penerapan Kecerdasan Buatan Dalam Bidang Kesehatan. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 5(1), 168–173.
- Fauzia, W., & Khumaeroh, A. (2023). Skrining Faktor Risiko Penyakit Kardiovaskuler Pada Pemain Minisoccer Dewasa. *Jurnal Pengabdian Kepada Masyarakat Nusantara (JPkMN)*, 4(3), 2387–2394. <https://ejournal.sisfokomtek.org/index.php/jpkm/article/view/1419/1038>
- Fitriani, H. P., Rosyanti, J., Awaliyah, Y. S., Akbar, Z., & Juhaeri, S. P. (2025). Perbandingan dampak ai di beberapa aspek kehidupan era 5.0: keamanan teknologi, pendidikan, ekonomi, dan kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2039–2044.
- Heryana, A. (2017). *Uji McNemar dan Uji Wilcoxon (Uji Hipotesa Non-Parametrik Dua Sampel Berpasangan)*. ResearchGat. <https://doi.org/10.13140/RG.2.2.17682.48325>
- Majid, A. M., & Utomo, W. H. (2021). Application of discretization and adaboost method to improve accuracy of classification algorithms in predicting diabetes mellitus. *ICIC Express Letters, Part B: Applications*, 12(12), 1177–1184. <https://doi.org/10.24507/icicelb.12.12.1177>
- Nainggolan, S. P., & Sinaga, A. (2023). Comparative Analysis of Accuracy of Random Forest and Gradient Boosting Classifier Algorithm for Diabetes Classification. *Sebatik*, 27(1), 97–102. <https://doi.org/10.46984/sebatik.v27i1.2157>
- Ramadhan, B., & Pane, S. F. (2024). Pengaruh Hyperparameter Tuning untuk

Efektivitas pada Pendekatan Hybrid dalam Mendiagnosis Stres dan Depresi :  
Tinjauan Studi Literatur. *Jurnal Tekno Insentif*, 18(2), 104–118.

Utiarahman, S. A., Mulawati, A., & Pratama, M. (2024). Penerapan Support Vector  
Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas.  
*Jurnal FASILKOM*, 14(3), 754–760.