

Literatur Review: Pendekatan Ensemble Learning untuk Klasifikasi Penyakit Jantung Koroner

Caesar Ariel Ramadan^{1*}, Febri Eka Fahriza², Fikri Hidayatullah³, Muhamad Tegar Amru⁴

¹⁻⁴Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}caesarariel04@gmail.com, ²fahrizaxcx@gmail.com, ³fikrihidayatullahv@gmail.com, ⁴tegaramru94@gmail.com

(* : coressponding author)

Abstrak– Penyakit Jantung Koroner adalah salah satu penyebab utama kematian global, terutama di negara berkembang. Berbagai metode machine learning telah diterapkan untuk membantu klasifikasi PJK, termasuk metode ensemble learning. Studi ini bertujuan untuk melakukan kajian literatur yang membandingkan metode ensemble, seperti random forest, artificial neural network (ANN), serta support vector machine (SVM), guna klasifikasi PJK. Berdasarkan literatur yang dianalisis, dilakukan perbandingan akurasi, kelebihan, dan kekurangan dari setiap metode, serta teknik preprocessing data yang digunakan. Studi ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai efektivitas metode ensemble dalam meningkatkan akurasi prediksi dan diagnosa dini PJK.

Kata Kunci: Penyakit Jantung Koroner, *Machine Learning*, Klasifikasi, *Ensemble Learning*, *Preprocessing Data*

Abstract– *Coronary Heart Disease* (CHD) is one of the leading causes of global death, especially in developing countries. Various machine learning methods have been applied to assist in the classification of CHD, including ensemble learning methods. This study aims to conduct a literature review comparing ensemble methods, such as random forest, artificial neural network (ANN), and support vector machine (SVM), for CHD classification. Based on the analyzed literature, a comparison is made on the accuracy, advantages, and disadvantages of each method, as well as the data preprocessing techniques used. This study is expected to provide deeper insights into the effectiveness of ensemble methods in improving prediction accuracy and early diagnosis of CHD.

Keywords: *Coronary Heart Disease*, *Machine Learning*, *Classification*, *Ensemble Learning*, *Data Preprocessing*

1. PENDAHULUAN

Gangguan kardiovaskular, seperti penyakit jantung paru, secara signifikan memengaruhi kesehatan manusia. Kondisi ini ditandai dengan gangguan pada fungsi jantung dan paru-paru yang dapat memengaruhi kemampuan tubuh dalam memompa darah dan mengoksidasi darah dengan baik. Diagnosis dini dan akurat sangat penting untuk memastikan penanganan yang tepat dan efektif dalam mengelola penyakit ini (Perani Rosyani, 2023).

Repositori Pembelajaran Mesin UCI memiliki dataset publik dari Cleveland, Hungaria, Swiss, VA Long Beach, dan Statlog, yang sering digunakan oleh akademisi untuk mencari model ML terbaik dalam memprediksi penyakit jantung. Meramalkan munculnya penyakit kardiovaskular adalah salah satu tantangan dalam biomedis komputer yang dapat diselesaikan dengan metode machine learning (ML). Variabel dalam lima kumpulan data tersebut memiliki ciri-ciri yang serupa, sehingga bisa digabungkan. (Apriyanto Alhamad, 2019).

Penanganan yang tepat dan efektif dalam mengelola penyakit ini. Pilihan gaya hidup yang tidak sehat, termasuk pola makan yang tinggi karbohidrat serta lemak, kelebihan berat badan, kurang berolahraga, merokok, serta mempunyai riwayat penyakit jantung dalam keluarga, merupakan faktor risiko untuk kondisi ini. Dengan total pembiayaan lebih dari 7,7 triliun rupiah, penyakit jantung merupakan penyakit yang paling banyak dibiayai oleh BPJS Kesehatan.

Guna membantu para profesional medis dengan cepat mengidentifikasi orang yang mempunyai penyakit jantung, ada beberapa hal yang dapat dilaksanakan guna menentukan apakah seseorang terindikasi penyakit jantung atau tidak, salah satunya ialah dengan penggunaan Machine learning, dengan penggunaan ini terbukti mampu menyelesaikan topik klasifikasi, dan optimasi dalam pembuatan sebuah system penyedia layanan kesehatan.

Dalam penelitian (Hidayat, 2023) dengan study kasus yang sama yaitu klasifikasi penyakit

jantung, dari penelitian yang ada menggunakan beberapa metode machine learning untuk dapat memprediksi seseorang terindikasi penyakit jantung yaitu random forest classifier, ann, svm, naïve bayes, support vector machine, decision tree, ensemble learning dll sebagainya, hasil penelitian sebelumnya memiliki hasil akurasi yang paling baik adalah 90% dengan menggunakan teknik seperti preprocessing data, penentuan hyperparameter, kombinasi metode balancing data dll sebagainya.

Tujuan literatur ini adalah menganalisis dan membandingkan berbagai metode machine learning yang digunakan dalam penelitian sebelumnya untuk klasifikasi penyakit jantung koroner, seperti random forest, artificial neural network (ANN), support vector machine (SVM), naïve bayes, dan decision tree. Hal ini untuk melihat kelebihan, kekurangan, serta tingkat akurasi masing-masing metode. Mengidentifikasi teknik-teknik preprocessing data dan penentuan hyperparameter yang diterapkan untuk meningkatkan performa model, serta memahami bagaimana metode balancing data dan kombinasi metode lain berkontribusi dalam meningkatkan akurasi model.

Riset ini diharapkan bisa memberikan pemahaman lebih dalam mengenai peran machine learning dalam membantu diagnosa dini penyakit jantung, serta tantangan-tantangan yang dihadapi dalam penerapannya di dunia nyata. Hal ini diharapkan bisa memberikan kontribusi bagi pengembangan sistem pendukung keputusan berbasis teknologi yang lebih baik di bidang kesehatan.

2. METODE PENELITIAN

2.1 Systematic Literatur Review (SLR)

Systematic literature review (SLR) ialah metode guna menemukan, meninjau, mengevaluasi, serta menginterpretasikan semua penelitian yang telah dilaksanakan pada topik atau fenomena tertentu, dengan bantuan pertanyaan penelitian yang telah didefinisikan dengan baik. Guna melaksanakan tinjauan sistematis serta mengidentifikasi jurnal-jurnal yang relevan, teknik SLR mengikuti serangkaian proses yang telah ditetapkan (E Triandini, 2019).

2.2 Pengumpulan Data

Data untuk studi literatur ini dikumpulkan dari lima jurnal ilmiah yang relevan dan terpercaya, yang diambil dari beberapa database akademik seperti Google Scholar, IEEE Xplore, ScienceDirect, SpringerLink, MDPI. Kata kunci yang digunakan dalam pencarian meliputi “coronary heart disease,” “machine learning,” “classification,” “prediction model,” dan “data preprocessing.” Jurnal-jurnal yang ditemukan kemudian dipilih berdasarkan kriteria inklusi serta eksklusi yang telah ditetapkan sebelumnya.

2.3 Klasifikasi

Metode klasifikasi banyak data membantu menentukan perilaku pola data karena seluruh kumpulan data dapat dianggap sebagai anggota class bergantung pada target atribut ataupun output. Singkatnya, ini adalah proses diskriminasi data. Pembentukan pohon, atau pohon, melalui pelatihan pada sampel data yang dimiliki digunakan untuk melaksanakan klasifikasi hutan acak. Random Forest melakukan klasifikasi dengan menggunakan metode yang mengambil hasil voting keputusan terbanyak berdasarkan pohon yang sudah dibentuk (Perani Rosyana, 2021).

2.4 Random Forest

Random Forest adalah kumpulan pohon pilihan yang dibangun dengan sampel yang dipilih secara acak, tetapi masing-masing pohon memiliki aturan yang membelah simpul. Model ini menggunakan subset fitur pada setiap pohon dan mencoba menemukan ambang batas terbaik dalam pembagian data. Jadi, banyak pohon akan dilatih dengan cara yang lebih lemah dan masing-masing akan membuat prediksi yang berbeda. (Deo Haganta Depari, 2022).

2.5 Kriteria Seleksi

Kriteria seleksi untuk jurnal-jurnal dilakukan dengan kriteria diterbitkan dalam lima tahun terakhir (2018-2023) untuk memastikan relevansi dengan perkembangan teknologi machine learning terbaru, Kemudian menggunakan teknik optimasi atau preprocessing data yang dapat meningkatkan performa model.

2.6 Analisis Hasil

Hasil dari analisis dan perbandingan setiap jurnal disajikan dalam bentuk tabel serta grafik untuk memberikan gambaran yang jelas mengenai efektivitas berbagai metode. Kesimpulan dari analisis ini digunakan untuk merumuskan rekomendasi mengenai metode machine learning yang paling efektif dan teknik-teknik optimasi yang dapat digunakan dalam penelitian masa depan.

3. ANALISA DAN PEMBAHASAN

Tabel 1. Hasil Temuan Artikel Relevan

2.1 Metode Pengumpulan Data

Metode yang digunakan pada pengumpulan data dalam program aplikasi ini adalah sebagai berikut:

No	Nama Peneliti dan Tahun	Metode yang dibahas	Tujuan Penelitiannya	Hasil yang didapat
1	(Ilham Aziz et al.,2023)	metode algoritma decision tree, Bootstrap Aggregating (Bagging), dan Adaptive Boosting (Adaboost)	menevaluasi Adaptive Boosting (AdaBoost) serta Bootstrap Aggregating (Bagging) untuk menemukan teknik kelompok terbaik untuk klasifikasi penyakit jantung serta meningkatkan akurasi klasifikasi data penyakit jantung.	menunjukkan bahwa baik Bagging maupun AdaBoost meningkatkan akurasi algoritma C4.5. Namun, Bagging menghasilkan akurasi tertinggi di antara ketiga dataset yang digunakan: Penyakit Jantung, Prediksi Penyakit Jantung, dan Prediksi resiko gagal jantung.
2	(Hidayatullah Harahap et al.,2021)	algoritma Random Forest	Untuk melakukan proses klasifikasi yang dapat digunakan dalam mendiagnosa penyakit jantung	menunjukkan akurasi 85,3% dalam klasifikasi penyakit jantung menggunakan algoritma Random Forest. Ini menandakan keberhasilan algoritma dalam klasifikasi.
3	(Halizah Alfajr & Defiyanti, 2024)	menerapkan Principal Component Analysis (PCA) untuk mengoptimalkan fitur	membuat model prediksi untuk penyakit jantung dengan menerapkan metode Random Forest	menunjukkan bahwa model prediksi penyakit jantung menggunakan metode Random Forest meraih akurasi 98%, presisi 100%, recall 96%, dan F1-score 98%. Penghapusan fitur fbs meningkatkan akurasi model dari 0.9617 menjadi 0.9823. Selain itu, penelitian ini juga mencatat bahwa fitur "thal" memiliki pengaruh
4	(Adnin Kamila et al., 2023)	k-fold cross validation, decision tree	melakukan prediksi penyakit jantung pada pasien dengan menggunakan metode	menunjukkan bahwa klasifikasi penyakit jantung menggunakan algoritma Random Forest lebih baik

			klasifikasi, yaitu decision tree dan random forest, serta membandingkan hasil klasifikasi untuk menemukan hasil prediksi terbaik untuk penyakit jantung	dibandingkan dengan Decision Tree. Random Forest menghasilkan akurasi 81.82%, precision 87.04%, recall 79.82%, dan F1 score 83.13%. Sementara itu, Decision Tree memiliki akurasi 77.44%, precision 89.42%, recall 74.78%, dan F1 score 80.92%
5	(Hidayat et al., 2023)	klasifikasi menggunakan Random Forest Classifier	untuk meningkatkan akurasi klasifikasi penyakit jantung menggunakan algoritma Random Forest.	dapat mencapai akurasi 94% pada pembagian data 80/20, yang lebih tinggi dibandingkan dengan akurasi 90% dari penelitian sebelumnya. Pada pembagian data 70/30, akurasi yang diperoleh adalah 92%. Confusion matrix menunjukkan bahwa pada pembagian 80/20 terdapat 11 kesalahan prediksi dari 205 data, sedangkan pada pembagian 70/30 terdapat 48 kesalahan dari 308 data.

4. KESIMPULAN

Berdasarkan hasil tinjauan literatur, dapat disimpulkan bahwa metode pembelajaran ensembel, seperti Random Forest, Jaringan Syaraf Tiruan (ANN), dan Support Vector Machine (SVM), memiliki potensi signifikan dalam meningkatkan akurasi prediksi penyakit jantung koroner (PJK). Dalam analisis penelitian yang dilakukan, metode Random Forest terbukti menjadi salah satu algoritma paling efektif dengan akurasi mencapai 98% setelah fitur dioptimalkan. Teknik Bootstrap Aggregating (Bagging) dan Adaptive Boosting (AdaBoost) juga menunjukkan perbaikan kinerja algoritma dasar seperti Decision Tree dalam pengklasifikasian data penyakit jantung.

Penggunaan teknik preprocessing data dan penentuan hyperparameter yang tepat berperan signifikan dalam meningkatkan performa model. Selain itu, metode balancing data dan kombinasi beberapa algoritma telah terbukti meningkatkan keakuratan klasifikasi, terutama dalam kasus dataset yang tidak seimbang.

Secara keseluruhan, studi ini memberikan wawasan bahwa penggunaan metode ensemble dalam klasifikasi PJK dapat membantu dalam diagnosa dini dan lebih akurat, sehingga dapat mendukung keputusan klinis yang lebih baik dalam penanganan penyakit jantung.

REFERENCES

- Akbar Hidayatullah Harahap, I. M. (2021). Klasifikasi Diagnosa Penyakit Jantung menggunakan Algoritma Random Forest. *Gunung Djati Conference Series*, 43-51. Diambil kembali dari <https://conferences.uinsgd.ac.id/index.php/gdcs/article/download/95/56>
- Apriyanto Alhamad, A. I. (2019). Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 352-360. Diambil kembali dari <https://jurnal.untan.ac.id/index.php/jepin/article/view/37188/75676584470>

- Deo Haganta Depari, Y. W. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *IFTK: JURNAL INFORMATIK*, 239-248. Diambil kembali dari <https://ejournal.upnvj.ac.id/informatik/article/view/4694/1852>
- E Triandini, S. J. (2019). Metode Systematic Literature Review untuk Identifikasi Platform dan Metode Pengembangan Sistem Informasi di Indonesia. *Indonesian Journal of Information Systems (IJIS)*, 63-77. Diambil kembali dari <https://ojs.uajy.ac.id/index.php/IJIS/article/view/1916/1309>
- Hidayat, A. S. (2023, September). Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier. *JurnalSistem Komputer dan Kecerdasan Buatan*, 31-40. Diambil kembali dari <https://jurnal.tau.ac.id/index.php/siskom-kb/article/download/464/341/2073>
- Mochammad Ilham Aziz, A. Z. (2023). Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 1-12. Diambil kembali dari <https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/5169/3200>
- Nur Halizah Alfajr, S. D. (2024). PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE RANDOM FOREST DAN PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA). *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*. Diambil kembali dari <https://journal.eng.unila.ac.id/index.php/jitet/article/download/5055/2083>
- Perani Rosyani, G. G. (2023). LITERATUR RIVIEW SISTEM PAKAR MENGINDENTIFIKASI PENYAKIT JANTUNG DAN PARU-PARU MENGGUNAKAN METODE FORWARD CHAINING. *CHIPSET : Jurnal Ilmu Komputer, Teknik, dan Multimedia*, 33-37. Diambil kembali dari <https://jurnal.publikasimahasiswa.id/index.php/chipset/article/view/5/5>
- Perani Rosyana, S. R. (2021). Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO). *justiN : Jurnal Sistem dan Teknologi Informasi*, 132-134. Diambil kembali dari <https://jurnal.untan.ac.id/index.php/justin/article/view/44120/75676589292>
- Sabrina Adnin Kamila, R. S. (2023). Klasifikasi Penyakit Jantung Menggunakan Decision Tree dan. *PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS*. Diambil kembali dari <https://proceeding.unpkediri.ac.id/index.php/stains/article/download/2816/1982>