Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

Penerapan Regresi Logistik, K-NN, dan Naïve Bayes Berbasis Pendekatan CRISP-DM dalam Memprediksi Penyakit Jantung

Rayna Shera Chang[1], Natalie Grace Widjaja Kuswanto[1], Jessica Laurentia Tedja[1], Christopher Andreas[2,*]

[1] Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Ciputra, Surabaya, Indonesia
[2] Fakultas Teknologi Informasi, Program Studi Informatika, Universitas Ciputra, Surabaya, Indonesia
Email: [1]rsherachang@student.ciputra.ac.id, [2]nwidjaja01@student.ciputra.ac.id, [3]jlaurentia01@student.ciputra.ac.id, [4,*]christopher.andreas@ciputra.ac.id
Email Penulis Korespondensi: christopher.andreas@ciputra.ac.id
Submitted: **13/10/2025**; Accepted: **31/03/2026**; Published**: 31/03/2026**

Abstrak-Penyakit jantung menjadi penyebab mortalitas tertinggi secara global yang sebenarnya memiliki potensi untuk dikendalikan melalui deteksi dini dan manajemen faktor risiko yang efektif. Sebagai upaya meningkatkan akurasi dan efisiensi deteksi dini, teknologi *machine learning* dimanfaatkan untuk membangun model prediktif risiko penyakit jantung. Penelitian ini bertujuan untuk membandingkan performa tiga algoritma klasifikasi dalam memprediksi risiko penyakit jantung guna memperoleh model yang paling optimal. Penelitian ini menggunakan metodologi CRISP-DM untuk membangun dan membandingkan performa model prediktif terhadap risiko penyakit jantung dengan tiga algoritma *supervised learning*, yaitu *K-Nearest Neighbors* (K-NN), *Naïve Bayes*, dan *Logistic Regression*. Dataset yang digunakan merupakan data *heart disease* yang diambil dari platform *Kaggle*, dan terdiri dari 10.000 baris dengan sejumlah variabel seperti *Age*, *Blood Pressure*, *Smoking*, *Diabetes*, *Cholesterol*, *Triglyceride Level*, *Fasting Blood Sugar*, dan *CRP Level*. Pada model K-NN, dilakukan pengujian dengan tiga variasi nilai k (k=5, k=10, dan k=20) untuk melihat pengaruh jumlah tetangga pada kinerja model. Sementara itu, pada model *Naïve Bayes* dan *Logistic Regression*, digunakan parameter bawaan (*default parameters*) tanpa penyesuaian tambahan, agar hasil perbandingan performa setiap model tetap konsisten. Evaluasi performa model dievaluasi berdasarkan metrik *Accuracy* dan *F1-Score*. Hasil evaluasi yang diuji menyatakan bahwa model K-NN dengan k=5 memberikan hasil terbaik dengan akurasi 0.7203 dan *F1-Score* 0.7598, mengungguli model *Naïve Bayes* dan *Logistic Regression*.

Kata Kunci: CRISP-DM; K-Nearest Neighbors (K-NN); Naïve Bayes; Penyakit Jantung; Regresi Logistik

Abstract−Heart disease remains the leading cause of mortality globally, despite having significant potential to be controlled through early detection and effective risk-factor management. To improve the accuracy and efficiency of early detection, machine learning technology is employed to develop predictive models for heart disease risk. The research aims to compare the performance of three classification algorithms in predicting heart disease risk to identify the most optimal model. This research applies the CRISP-DM methodology to build and compare predictive models for heart disease risk using three supervised learning algorithms: K-Nearest Neighbors (K-NN), Naïve Bayes, and Logistic Regression. The dataset used is a heart disease dataset obtained from the Kaggle platform, consisting of 10,000 records with variables such as Age, Blood Pressure, Smoking, Diabetes, Cholesterol, Triglyceride Level, Fasting Blood Sugar, and CRP Level. For the K-NN model, experiments were conducted using three values of *k* (k = 5, k = 10, and k = 20) to examine the effect of the number of neighbors on model performance. Meanwhile, the Naïve Bayes and Logistic Regression models were implemented using default parameters without additional tuning to ensure a consistent performance comparison. Model performance was evaluated using Accuracy and F1-Score metrics. The evaluation results indicate that the K-NN model with k = 5 achieved the best performance, with an accuracy of 0.7203 and an F1-Score of 0.7598, outperforming the Naïve Bayes and Logistic Regression models.

Keywords: CRISP-DM; K-Nearest Neighbors (K-NN); Naïve Bayes; Heart Disease; Logistic Regression

1. PENDAHULUAN

Penyakit kardiovaskular masih menjadi salah satu penyebab dominan kematian di Indonesia dan mengakibatkan tantangan signifikan dalam sektor kesehatan [1]. Kondisi ini terjadi ketika jantung tidak mampu menjalankan fungsinya secara optimal dalam memompa darah, sehingga mengganggu distribusi oksigen dan nutrisi ke seluruh tubuh serta proses pembuangan sisa metabolisme. Gangguan pada fungsi vital ini dapat berdampak serius terhadap kesehatan secara menyeluruh. Berdasarkan data *Sample Registration System* (SRS) tahun 2014, penyakit jantung koroner berkontribusi terhadap 12,9% dari total kematian nasional. Tren ini menunjukkan peningkatan dari tahun ke tahun. Data *Institute for Health Metrics and Evaluation* (IHME) mencatat bahwa jumlah penderita penyakit jantung meningkat dari 12,93 juta orang pada tahun 2021 menjadi 15,5 juta pada tahun 2022. Pada tahun yang sama, hasil diagnosis dokter memperlihatkan prevalensi sebesar 1,5% atau sekitar 1,01 juta kasus, dengan DKI Jakarta menempati posisi kedua tertinggi dengan 1,9% atau sekitar 40.200 kasus [2]. Laporan lain dari Hidayat, Sunyoto, dan Al Fatta [3] mengungkapkan bahwa setidaknya 15 dari setiap 1.000 penduduk, atau sekitar 2,78 juta orang, mengidap penyakit jantung. Fakta-fakta ini menegaskan bahwa penyakit jantung merupakan isu kesehatan yang mendesak dan memerlukan strategi pencegahan serta penanganan yang lebih serius.

Beragam faktor dapat mempengaruhi kesehatan jantung seseorang, antara lain kebiasaan merokok, hipertensi, kadar kolesterol yang tinggi, diabetes, stres, pola makan tinggi lemak, serta obesitas [4]. Faktor-faktor tersebut sering kali tidak menunjukkan dampak secara langsung, sehingga deteksi dini berperan penting dalam memungkinkan intervensi yang lebih cepat dan efektif. Dengan mengenali potensi risiko sejak awal, langkah-langkah pencegahan dapat dirancang secara lebih tepat sasaran [5]. Sebagai bagian dari upaya untuk menanggulangi permasalahan ini, penelitian ini menerapkan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM) dalam

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

membangun model prediksi penyakit jantung. Model tersebut dikembangkan menggunakan tiga algoritma klasifikasi berbasis *supervised learning*, yaitu *K-Nearest Neighbor* (K-NN), *Naïve Bayes*, dan *Logistic Regression*. Ketiga algoritma ini dipilih karena dinilai mampu membentuk model prediktif dengan tingkat akurasi yang baik, dengan memanfaatkan data historis yang telah diberi label, seperti tekanan darah, kadar kolesterol, usia, indeks massa tubuh (BMI), serta sejumlah indikator medis lainnya.

Dalam algoritma *K-Nearest Neighbor (K-NN)*, proses klasifikasi dilakukan dengan mengukur jarak antara data baru dan sekumpulan *K* titik terdekat dari data pelatihan. Pendekatan ini memiliki mekanisme yang sederhana dan intuitif, namun tetap efektif dalam menghasilkan performa klasifikasi yang kompetitif [6]. Salah satu ukuran jarak yang umum digunakan dalam algoritma KNN adalah *Euclidean Distance*, yang menghitung jarak dengan mengambil akar kuadrat dari total selisih kuadrat setiap nilai fitur data [7]. Dalam konteks prediksi penyakit kardiovaskular, algoritma ini diterapkan untuk mengklasifikasikan pasien berlandaskan kemiripan karakteristik kesehatannya dengan data pasien lain yang hasil diagnosisnya telah diketahui. Namun, jumlah data rekam medis yang besar sering kali mengalami ketidakseimbangan kelas, sehingga diperlukan penerapan teknik *oversampling* seperti SMOTE guna meningkatkan akurasi algoritma klasifikasi, termasuk algoritma K-NN [8].

Sementara itu, algoritma *Naïve Bayes* dipilih karena efisiensinya dalam proses komputasi dan kemampuannya untuk menghitung probabilitas secara cepat. Meskipun metode ini menggunakan asumsi bahwa setiap fitur bersifat independen, hasil klasifikasi yang dihasilkan sering kali cukup akurat dan stabil, terutama untuk data kesehatan yang kompleks [9]. Dengan menggunakan algoritma ini, pasien dapat diklasifikasikan ke dalam kelompok risiko tertentu berdasarkan kombinasi faktor-faktor medis yang dimilikinya.

Selain itu, *Logistic Regression* juga digunakan dalam penelitian ini karena kemampuannya dalam memodelkan kemungkinan terjadinya suatu kondisi berdasarkan sejumlah variabel prediktor. Metode ini sangat cocok untuk klasifikasi biner, seperti memprediksi apakah seseorang memiliki risiko tinggi atau rendah terhadap penyakit jantung. Penerapan algoritma *Logistic Regression* memungkinkan identifikasi faktor-faktor risiko yang paling berpengaruh serta pengelompokan pasien ke dalam kategori risiko yang sesuai [10].

Dalam proses pengolahan dan analisis data, bahasa pemrograman *Python* dipilih sebagai bahasa pemrograman untuk analisis dan visualisasi data karena dapat dimanfaatkan untuk berbagai keperluan, mulai dari pemrograman statistik hingga *deep learning* [11]. Selain itu, *Python* juga menyediakan *library* lengkap seperti *NumPy*, *Pandas*, *Matplotlib*, dan *Scikit-learn* (*sklearn*) untuk mendukung pemrograman *machine learning* [12]. Adapun dataset yang digunakan dalam penelitian ini diperoleh dari platform *Kaggle*, yang disediakan oleh Ördekçi [13]. Dataset ini berisi sekitar 10.000 data pasien yang mencakup berbagai atribut penting, seperti usia, jenis kelamin, tekanan darah, kolesterol, BMI, kebiasaan olahraga, merokok, riwayat keluarga, diabetes, stres, jam tidur, konsumsi alkohol, kadar homosistein, serta *C-Reactive Protein* (CRP). Dengan banyaknya fitur yang tersedia, dataset ini sangat bermanfaat untuk proses klasifikasi dan prediksi risiko penyakit jantung secara lebih komprehensif dan akurat. Meskipun dataset yang digunakan menyediakan berbagai atribut kesehatan yang beragam, penelitian-penelitian terdahulu umumnya hanya memanfaatkan variabel klinis konvensional dalam proses pemodelan penyakit jantung, seperti usia, jenis kelamin, tekanan darah, kolesterol, serta hasil elektrokardiografi [15-17]. Variabel-variabel tersebut terbukti efektif dalam membangun model klasifikasi, namun belum sepenuhnya merepresentasikan faktor gaya hidup dan indikator inflamasi yang berperan penting dalam perkembangan penyakit jantung. Oleh karena itu, terdapat celah penelitian (*research gap*) dalam pengembangan model prediksi yang mengintegrasikan variabel metabolik, kebiasaan hidup, dan penanda peradangan secara bersamaan. Penelitian ini mengisi celah tersebut dengan menambahkan variabel seperti status merokok, diabetes, kadar trigliserida, kadar gula darah puasa, serta *C-Reactive Protein* (CRP) sebagai indikator inflamasi untuk meningkatkan representasi risiko penyakit jantung. Dengan pendekatan ini, penelitian diharapkan mampu menghasilkan model prediksi yang lebih komprehensif dan relevan secara klinis. Oleh karena itu, manfaat penelitian ini adalah menyediakan model prediksi risiko penyakit jantung yang lebih akurat dan seimbang, serta berkontribusi dalam pengembangan sistem pendukung keputusan medis untuk mendukung deteksi dini dan pengambilan keputusan klinis berbasis analisis data.

2. METODOLOGI PENELITIAN

2.1 Penyakit Jantung

Penyakit jantung merupakan kondisi yang dipengaruhi oleh berbagai faktor risiko, seperti usia, tekanan darah, kolesterol, diabetes, dan kebiasaan merokok. Faktor-faktor tersebut dapat menyebabkan gangguan pada sistem kardiovaskular yang berdampak pada kesehatan secara keseluruhan. Dalam penelitian ini, variabel-variabel tersebut digunakan sebagai indikator untuk memprediksi kemungkinan seseorang mengalami penyakit jantung, sehingga dapat mendukung proses deteksi dini melalui pendekatan berbasis data.

2.2 Metode Klasifikasi

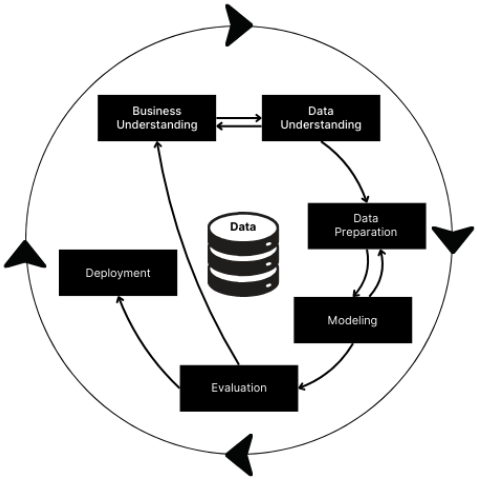
Penelitian ini menggunakan tiga algoritma klasifikasi berbasis *supervised learning*, yaitu *Logistic Regression, K-Nearest Neighbors (K-NN)*, dan *Naïve Bayes*. Ketiga metode ini diterapkan pada tahap *modeling* dalam pendekatan CRISP-DM untuk membandingkan performa dalam memprediksi risiko penyakit jantung. *Logistic Regression* digunakan untuk klasifikasi biner dengan memodelkan hubungan antara variabel prediktor dan probabilitas terjadinya

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

penyakit jantung. *K-Nearest Neighbors* (K-NN) mengklasifikasikan data berdasarkan kedekatan dengan data latih terdekat, dengan variasi nilai k untuk menentukan performa terbaik. *Naïve Bayes* digunakan untuk menghitung probabilitas klasifikasi berdasarkan distribusi data, dengan asumsi independensi antar fitur. Ketiga metode tersebut dibandingkan berdasarkan hasil evaluasi untuk menentukan model yang paling optimal.

2.3 Tahapan Penelitian

Dalam upaya memprediksi penyakit jantung, digunakan sebanyak 10.000 data dari *dataset Kaggle* tentang *Heart Disease Data for Health Research* dengan pendekatan model CRISP-DM serta penerapan tiga algoritma *machine learning*, yaitu *Logistic Regression, K-Nearest Neighbor* (K-NN), dan *Naïve Bayes*. Dataset ini dirancang untuk mengidentifikasi potensi risiko penyakit jantung pada pasien berdasarkan hasil pengukuran diagnostik. Proses analisis data dilakukan dengan mengacu pada standar tahapan *data mining* yang dikenal sebagai *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang mencakup enam fase utama: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* sebagaimana ditunjukkan di Gambar 1.



Gambar 1. Alur Tahapan Metodologi CRISP-DM

Adapun langkah-langkah yang dilakukan dalam proses analisis data sebagai berikut:

a. *Business Understanding*
   Tahap ini bertujuan untuk memahami permasalahan penelitian, yaitu tingginya risiko penyakit jantung yang sering terlambat terdeteksi. Oleh karena itu, penelitian ini bertujuan membangun model prediksi berbasis *machine learning* untuk mengidentifikasi kemungkinan risiko penyakit jantung berdasarkan data kesehatan pasien. Model ini diharapkan dapat membantu proses deteksi dini.
b. *Data Understanding*
   Dataset yang digunakan merupakan data *heart disease* dari Kaggle yang terdiri dari 10.000 data pasien dengan berbagai atribut kesehatan seperti usia, tekanan darah, kolesterol, diabetes, dan kebiasaan merokok. Pada tahap ini dilakukan identifikasi variabel prediktor (X) yang merupakan atribut kesehatan pasien dan Variabel target (y) yang merupakan *Heart Disease Status* (0 = tidak, 1 = ya). Selain itu dilakukan eksplorasi data untuk melihat distribusi data, mendeteksi *outlier*, serta memahami karakteristik dataset sebelum dilakukan pengolahan lebih lanjut.
c. *Data Preparation*
   Tahap ini merupakan proses pengolahan data sebelum dilakukan pemodelan agar dataset siap digunakan. Proses ini dimulai dengan pembersihan data (*data cleaning*) dengan menghapus nilai yang hilang (*missing value*) serta data duplikat untuk menjaga kualitas data. Selanjutnya, dilakukan transformasi data dengan mengubah variabel kategorikal menjadi numerik menggunakan *Label Encoding*. Untuk memastikan setiap fitur memiliki skala yang sama, dilakukan *feature scaling* menggunakan metode *Min-Max Scaling*, terutama karena algoritma seperti K-NN sensitif terhadap perbedaan skala. Selain itu, karena dataset memiliki ketidakseimbangan kelas, digunakan metode SMOTE untuk menyeimbangkan distribusi data antara kelas mayoritas dan minoritas [14]. Terakhir, dataset dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20% untuk keperluan pelatihan dan evaluasi model.
d. *Modeling*
   Selanjutnya, proses pembangunan model prediksi dilakukan menggunakan algoritma *machine learning*. Dalam penelitian ini digunakan tiga algoritma, yaitu *K-Nearest Neighbors* (K-NN*), Logistic Regression*, dan *Naïve Bayes*. Ketiga metode ini diterapkan pada tahap *Modeling* dalam pendekatan CRISP-DM untuk membandingkan performa dalam memprediksi risiko penyakit jantung. Pada implementasinya, algoritma K-NN diuji dengan beberapa variasi nilai parameter k, yaitu k = 5, 10, dan 20, untuk menentukan konfigurasi terbaik. Sementara itu, *Logistic Regression* dan *Naïve Bayes* digunakan dengan parameter *default* agar perbandingan antar model tetap konsisten. Setiap model dilatih menggunakan *data training* untuk mempelajari pola hubungan antar variabel, kemudian

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

digunakan untuk melakukan prediksi pada data *testing.*

e. *Evaluation*
Pada proses ini dilakukan pengukuran kinerja model dalam memprediksi data yang belum pernah dilihat sebelumnya. Evaluasi dilakukan menggunakan metrik *accuracy* untuk mengukur tingkat ketepatan prediksi, serta *F1-Score* untuk menilai keseimbangan antara *precision* dan *recall*, terutama pada dataset yang tidak seimbang. Selain itu, digunakan *confusion matrix* untuk memberikan gambaran yang lebih rinci mengenai hasil klasifikasi, termasuk jumlah prediksi benar dan salah pada masing-masing kelas. Hasil evaluasi dari ketiga model kemudian dibandingkan untuk menentukan model dengan performa terbaik.

f. *Deployment*
Setelah model analisis diimplementasikan, langkah selanjutnya adalah penerapannya untuk mengestimasi kemungkinan individu terdiagnosis penyakit jantung berdasarkan informasi medis yang ada. Adapun metode ini akan diuji terlebih dahulu dengan data pasien yang mencakup faktor-faktor seperti usia, tekanan darah, kebiasaan merokok, dan indikator kesehatan lainnya. Sebagai ilustrasi, misalkan terdapat seorang individu berusia 60 tahun dengan tekanan darah sebesar 150 mmHg, merupakan seorang perokok, tidak menderita diabetes, memiliki kondisi tekanan darah tinggi, kadar kolesterol HDL normal, kadar kolesterol LDL yang tinggi, kadar trigliserida sebesar 250 mg/dL, kadar gula darah puasa sebesar 130 mg/dL, serta kadar CRP sebesar 10 mg/L. Data individu ini kemudian dapat dimasukkan ke dalam model yang telah dibangun, untuk memprediksi apakah individu tersebut memiliki risiko terkena penyakit jantung atau tidak.

3. HASIL DAN PEMBAHASAN

3.1 Evaluasi Model Prediktif

Setelah melalui tahapan *data preparation* dan pemodelan sebagaimana dijelaskan pada bagian metodologi, masing-masing algoritma klasifikasi dilatih (*training*) menggunakan data latih untuk mempelajari pola hubungan antara variabel prediktor dan status penyakit jantung. Model yang telah dilatih kemudian diuji (*testing*) menggunakan data uji untuk mengevaluasi kemampuannya dalam melakukan klasifikasi secara objektif. Proses evaluasi ini bertujuan untuk menilai sejauh mana model mampu membedakan pasien yang menderita penyakit jantung dan yang tidak menderita penyakit jantung berdasarkan data yang tersedia. Hasil evaluasi kinerja model selanjutnya dianalisis menggunakan *Confusion Matrix* sebagai alat utama untuk menggambarkan distribusi prediksi benar dan salah dari setiap model.

3.1.1 Confusion Matrix

Confusion Matrix digunakan sebagai alat evaluasi utama untuk memahami performa setiap model dalam melakukan klasifikasi terhadap data pasien. Dalam konteks penelitian ini, *Confusion Matrix* membandingkan hasil prediksi yang dihasilkan oleh model dengan label aktual pada dataset, sehingga dapat diketahui seberapa tepat model membedakan pasien yang menderita penyakit jantung (*Heart Disease*) dan yang tidak menderita penyakit jantung (*Non-Heart Disease*). Keunggulan penggunaan *Confusion Matrix* terletak pada kemampuannya memberikan gambaran rinci mengenai jenis-jenis kesalahan yang dilakukan oleh model, bukan hanya sekadar tingkat akurasi secara keseluruhan. Melalui analisis komponen di dalamnya, peneliti dapat mengidentifikasi area kekuatan model dan titik-titik kelemahan yang perlu diperbaiki, terutama dalam konteks medis di mana kesalahan deteksi dapat berdampak signifikan terhadap kesehatan pasien.

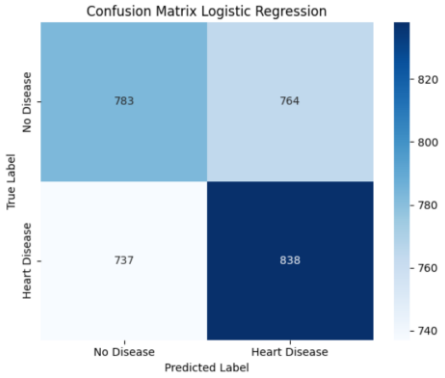
Secara umum, terdapat empat komponen utama dalam *Confusion Matrix* yang menjadi indikator penting untuk mengukur kinerja model. Keempat komponen ini memberikan informasi yang saling melengkapi terkait kualitas prediksi, baik untuk kasus positif maupun negatif, dan dijelaskan sebagai berikut:

a. *True Positive* (TP) adalah jumlah prediksi positif yang benar. Komponen ini menunjukkan banyaknya pasien yang memang menderita penyakit jantung dan berhasil diidentifikasi dengan tepat oleh model sebagai penderita. Nilai TP yang tinggi menjadi indikator bahwa model memiliki kemampuan yang baik dalam mendeteksi kasus yang benar-benar positif. Dalam aplikasi medis, tingginya TP berarti semakin sedikit pasien berisiko tinggi yang terlewatkan, sehingga potensi intervensi dini dapat dimaksimalkan.
b. *True Negative* (TN) adalah jumlah prediksi negatif yang benar, yaitu pasien yang tidak menderita penyakit jantung dan diprediksi tidak sakit oleh model. TN yang tinggi menunjukkan bahwa model mampu membedakan pasien sehat dari pasien sakit secara akurat. Hal ini penting karena kesalahan dalam bentuk prediksi positif palsu terhadap pasien sehat dapat menimbulkan kekhawatiran yang tidak perlu, biaya pemeriksaan tambahan, atau bahkan pengobatan yang tidak dibutuhkan.
c. *False Positive* (FP) adalah jumlah prediksi positif yang salah. Dalam kasus ini, pasien yang sebenarnya sehat justru diprediksi oleh model sebagai penderita penyakit jantung. FP yang tinggi dapat menjadi masalah karena menyebabkan over-diagnosis, di mana pasien harus menjalani pemeriksaan lanjutan atau intervensi medis padahal tidak diperlukan. Meskipun dalam dunia medis FP masih lebih dapat ditoleransi dibanding FN, tingkat FP yang terlalu tinggi tetap perlu diwaspadai karena dapat membebani sistem layanan kesehatan dan menimbulkan ketidaknyamanan psikologis bagi pasien.
d. *False Negative* (FN) adalah jumlah prediksi negatif yang salah, yaitu pasien yang sebenarnya menderita penyakit

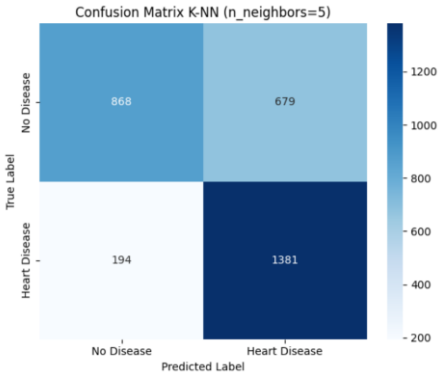
Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

jantung namun tidak terdeteksi oleh model dan diprediksi sebagai sehat. FN dianggap sebagai jenis kesalahan yang paling berbahaya dalam diagnosis medis, karena dapat mengakibatkan pasien tidak menerima penanganan yang seharusnya. Hal ini berpotensi memperburuk kondisi kesehatan pasien, meningkatkan risiko komplikasi serius, bahkan berujung pada kematian jika penyakit tidak segera ditangani. Oleh karena itu, salah satu tujuan utama dalam pengembangan model prediksi medis adalah meminimalkan nilai FN seminimal mungkin.

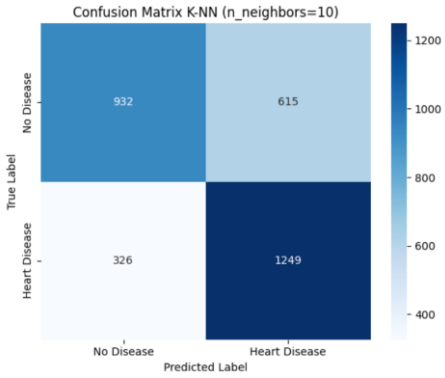
Selanjutnya, hasil evaluasi kinerja masing-masing model klasifikasi divisualisasikan menggunakan *Confusion Matrix* yang ditampilkan pada Gambar 2 hingga Gambar 6.



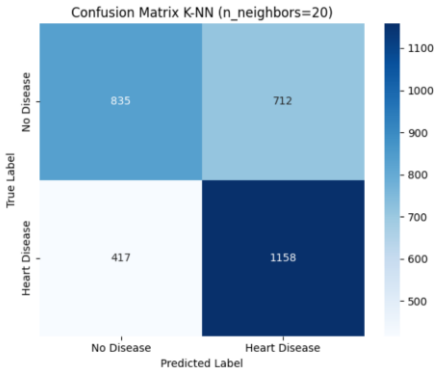
Gambar 2. *Confusion matrix Logistic Regression*



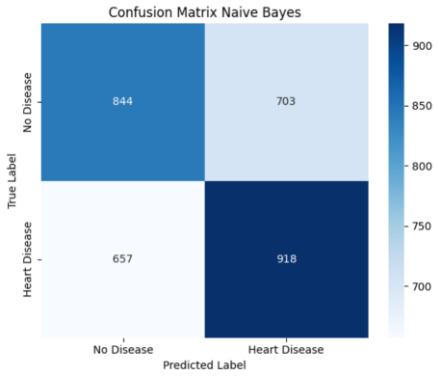
Gambar 3. *Confusion matrix* K-NN (k=5)



Gambar 4. *Confusion matrix* K-NN (k=10)



Gambar 5. *Confusion matrix* K-NN (k=20)



Gambar 6. *Confusion matrix Naïve Bayes*

Berdasarkan hasil analisis yang ditunjukkan pada Gambar 2, setiap model menunjukkan pola performa yang berbeda pada keempat komponen *Confusion Matrix*. Model *Logistic Regression*, misalnya, menampilkan jumlah *True Negative* (TN) dan True Positive (TP) yang cukup seimbang, menandakan bahwa model ini memiliki kemampuan moderat dalam mengenali kedua kelas. Namun, pada saat yang sama, jumlah *False Positive* (FP) dan *False Negative* (FN) yang dihasilkan tergolong tinggi. Tingginya FN pada Gambar 2 menunjukkan bahwa *Logistic Regression* sering gagal mengidentifikasi pasien yang sebenarnya menderita penyakit jantung, sebuah kelemahan signifikan dalam aplikasi medis karena dapat menyebabkan keterlambatan diagnosis dan pengobatan. Di sisi lain, FP yang cukup tinggi berarti model ini juga cenderung memberikan hasil positif palsu, yang dapat mengarah pada pemeriksaan tambahan dan biaya medis yang tidak diperlukan.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

Model K-Nearest Neighbors (K-NN) dengan k=5 sebagaimana ditampilkan pada Gambar 3 menunjukkan performa yang jauh lebih baik, terutama pada komponen TP yang lebih tinggi dan FN yang lebih rendah dibandingkan *Logistic Regression*. Hal ini mengindikasikan bahwa K-NN (k=5) lebih akurat dalam mendeteksi pasien berisiko tinggi, sehingga potensi kasus yang terlewat menjadi lebih kecil. Meskipun terdapat sejumlah FP, jumlahnya masih dalam batas yang dapat ditoleransi mengingat manfaat deteksi dini yang dihasilkan. Dalam dunia medis, sedikit over-diagnosis umumnya dapat diterima karena pasien yang teridentifikasi positif tetap akan menjalani pemeriksaan lanjutan sebelum diagnosis akhir diputuskan.

Ketika jumlah tetangga yang digunakan pada K-NN ditingkatkan menjadi k=10 dan k=20, seperti yang ditunjukkan pada Gambar 4 dan Gambar 5, pola hasil menunjukkan adanya penurunan performa. Walaupun nilai TP tetap relatif tinggi, FN mulai meningkat, menandakan bahwa beberapa kasus positif mulai terlewatkan. Selain itu, FP juga mengalami kenaikan, yang berarti model mulai kehilangan ketajaman dalam membedakan pasien sehat dari pasien sakit. Fenomena ini sejalan dengan teori dasar K-NN yang menyatakan bahwa semakin besar nilai k, semakin besar pula kecenderungan model untuk melakukan generalisasi berlebihan (*over-smoothing*), sehingga batas antara kelas positif dan negatif menjadi kurang jelas.

Sementara itu, model *Naïve Bayes* yang ditampilkan pada Gambar 6 menunjukkan hasil yang berada di tengah-tengah antara *Logistic Regression* dan K-NN. Nilai TP dan TN yang dihasilkan moderat, tetapi FN cenderung cukup tinggi. Salah satu penyebabnya adalah asumsi independensi antar fitur pada algoritma *Naïve Bayes* yang jarang terpenuhi dalam data medis. Misalnya, tekanan darah tinggi sering kali berkorelasi dengan kadar kolesterol yang tidak normal, sehingga asumsi bahwa variabel-variabel ini tidak saling memengaruhi dapat mengurangi akurasi perhitungan probabilitas.

Hasil perbandingan dari keempat komponen *Confusion Matrix* ini memberikan gambaran yang jelas bahwa K-NN dengan k=5 adalah model yang paling seimbang dalam mendeteksi penyakit jantung. Tingginya nilai TP dan rendahnya FN menjadikannya unggul untuk tujuan deteksi dini, sementara tingkat FP yang dihasilkan masih dapat diterima dalam konteks medis. Temuan ini menjadi dasar kuat untuk memilih K-NN k=5 sebagai model utama yang akan digunakan pada tahap implementasi dan pengujian lebih lanjut.

3.1.2 Evaluation Matrix

Hasil evaluasi model pada penelitian ini diukur menggunakan dua metrik utama, yaitu *accuracy* dan *F1-Score*, untuk memberikan gambaran menyeluruh mengenai kinerja setiap algoritma. *Accuracy* menunjukkan persentase prediksi yang tepat dibandingkan dengan seluruh data uji, sementara *F1-Score* merupakan ukuran keseimbangan antara *precision* dan *recall*, yang menjadi sangat penting ketika data yang digunakan memiliki distribusi kelas yang tidak seimbang. Dalam konteks deteksi penyakit jantung, *F1-Score* memegang peranan krusial karena mampu menunjukkan sejauh mana model dapat mendeteksi pasien yang benar-benar positif tanpa terlalu banyak menghasilkan kesalahan prediksi pada pasien negatif.

Berdasarkan hasil yang disajikan pada Tabel 1, terlihat jelas bahwa *K-Nearest Neighbors* (K-NN) dengan k=5 menghasilkan performa paling unggul di antara seluruh model yang diuji. Model ini mencapai *accuracy* sebesar 0,7203, jauh lebih tinggi dibandingkan *Logistic Regression* yang hanya memperoleh 0,5202, serta mengungguli K-NN dengan k=10 dan k=20 yang masing-masing memperoleh 0,6985 dan 0,6383. Bahkan dibandingkan *Naïve Bayes* yang hanya mencatat *accuracy* sebesar 0,5643, K-NN k=5 tetap menjadi pilihan terbaik. Angka ini menegaskan bahwa K-NN k=5 memiliki kemampuan yang lebih stabil dan presisi dalam mengklasifikasikan data pasien.

Tabel 1. Perbandingan *Evaluation Metrics*

	Logistic Regression	K-NN (k=5)	K-NN (k=10)	K-NN (k=20)	*Naïve Bayes*
Accuracy	0.519218	0.720372	0.698591	0.638373	0.564382
F1-Score	0.527542	0.759835	0.726374	0.672279	0.574468

Keunggulan K-NN k=5 tidak hanya terlihat dari *accuracy*, tetapi juga dari nilai *F1-Score* yang mencapai 0,7598. Nilai ini menunjukkan bahwa model mampu menjaga keseimbangan optimal antara *precision*, yakni kemampuan menghindari *false positive* dan *recall*, kemampuan mengurangi *false negative*. Hal ini sangat penting dalam dunia medis, di mana kesalahan melewatkan pasien yang sebenarnya berisiko (FN) dapat berakibat fatal, sementara kesalahan mendeteksi pasien sehat sebagai sakit (FP) meskipun tidak diinginkan, masih dapat ditangani melalui pemeriksaan lanjutan.

Sebagai perbandingan, *Logistic Regression* menunjukkan nilai *F1-Score* terendah, yaitu 0,5275. Hasil ini menandakan bahwa model linear tersebut kesulitan menangkap kompleksitas pola hubungan antarvariabel medis dalam dataset, yang kemungkinan besar bersifat non-linear dan memiliki interaksi antar fitur yang signifikan. K-NN dengan k=10 dan k=20 masing-masing memiliki *F1-Score* 0,7263 dan 0,6722, menunjukkan tren penurunan performa ketika nilai k diperbesar. Penurunan ini selaras dengan karakteristik algoritma K-NN, di mana penambahan jumlah tetangga cenderung menghaluskan batas klasifikasi hingga kehilangan sensitivitas terhadap kasus positif yang jelas berbeda dari mayoritas data.

Naïve Bayes mencatat *F1-Score* sebesar 0,5744, yang meskipun lebih baik dari *Logistic Regression*, tetap tertinggal dari seluruh variasi K-NN. Salah satu penyebabnya adalah asumsi independensi antar fitur yang mendasari

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

Naïve Bayes. Dalam data medis seperti ini, variabel-variabel seperti tekanan darah, kolesterol, gula darah, dan CRP seringkali memiliki korelasi yang kuat, sehingga pelanggaran asumsi independensi tersebut dapat mengurangi akurasi perhitungan probabilitas.

Secara keseluruhan, hasil ini menunjukkan bahwa K-NN dengan k=5 merupakan model yang memiliki kinerja terbaik dalam merepresentasikan keterkaitan antara variabel input dengan prediksi penyakit jantung pada dataset yang digunakan, dibandingkan dengan model lain. Dengan akurasi dan *F1-Score* yang lebih tinggi, model K-NN (k=5) dapat memberikan prediksi yang lebih akurat dan lebih seimbang dalam mengidentifikasi baik kasus positif (terkena penyakit jantung) maupun negatif (tidak terkena penyakit jantung).

3.1.3 Evaluasi Ketahanan (*Robustness*) Model

Untuk menguji ketahanan (*robustness*) model terhadap variasi parameter, dilakukan pengujian sensitivitas terhadap algoritma *K-Nearest Neighbors* (K-NN) dengan memodifikasi nilai parameter k, yaitu jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Parameter ini memiliki peranan penting dalam mempengaruhi kinerja prediksi, karena menentukan seberapa banyak tetangga yang dipertimbangkan saat mengklasifikasikan data uji. Pengujian dilakukan dengan tiga nilai berbeda, yaitu k = 5, k = 10, dan k = 20.

Berdasarkan hasilnya, terlihat bahwa nilai k yang lebih kecil (k = 5) menghasilkan performa yang lebih optimal dibandingkan nilai k yang lebih besar (k = 10 dan k = 20). Secara umum, tren penurunan performa pada metrik akurasi dan *F1-Score* menunjukkan bahwa peningkatan nilai k cenderung mengakibatkan menurunnya kemampuan model dalam melakukan klasifikasi yang akurat, khususnya terhadap kelas minoritas. Fenomena ini dapat dijelaskan melalui konsep *overgeneralization*, yaitu ketika jumlah tetangga yang terlalu banyak menyebabkan keputusan klasifikasi menjadi terlalu umum dan kurang sensitif terhadap pola-pola lokal dalam data. Dalam konteks dataset yang memiliki ketidakseimbangan kelas (*class imbalance*), hal ini sangat penting, karena model yang terlalu general berpotensi mengabaikan karakteristik unik dari kelas minoritas, yang justru krusial dalam deteksi dini penyakit jantung.

Dengan demikian, hasil ini menunjukkan bahwa model K-NN dengan nilai k yang terlalu besar dapat kehilangan presisi dalam membedakan antara kelas positif dan negatif. Sebaliknya, nilai k yang lebih kecil mampu mempertahankan sensitivitas terhadap variasi lokal dalam data, sehingga memberikan hasil klasifikasi yang lebih akurat dan seimbang. Secara keseluruhan, hasil evaluasi ini memperkuat pemahaman bahwa pemilihan parameter yang tepat merupakan bagian integral dari desain model *machine learning* yang *robust*.

Untuk menguji penerapan model K-NN (k=5) yang telah terbukti memberikan performa prediksi yang seimbang, dilakukan simulasi pada sebuah studi kasus. Data profil medis seorang individu berusia 60 tahun dengan tekanan darah 150 mmHg, status perokok, tanpa riwayat diabetes, memiliki tekanan darah tinggi, kadar kolesterol HDL normal, kadar kolesterol LDL tinggi, trigliserida sebesar 250 mg/dL, kadar gula darah puasa 130 mg/dL, serta kadar CRP sebesar 10 mg/L, dimasukkan ke dalam model. Hasil prediksi menunjukkan bahwa individu tersebut berpotensi mengalami penyakit jantung. Prediksi ini memperkuat potensi penggunaan model sebagai alat bantu dalam proses *screening* awal dan pengambilan keputusan medis. Selain itu, melalui penerapan kasus nyata ini, terlihat bahwa model mampu mengolah berbagai variabel medis secara efektif untuk memberikan estimasi risiko yang dapat dipertimbangkan oleh tenaga medis dalam merencanakan intervensi pencegahan yang lebih tepat.

3.2 Pembahasan

Berdasarkan hasil evaluasi yang dilakukan terhadap lima model klasifikasi yang diuji, ditemukan bahwa masing-masing algoritma menunjukkan karakteristik performa yang unik dalam mengidentifikasi risiko penyakit jantung. Dataset yang digunakan memiliki ketidakseimbangan kelas dengan perbandingan 80:20 antara pasien yang tidak menderita dan menderita penyakit jantung. Oleh karena itu, *F1-Score* menjadi metrik yang lebih tepat untuk mengevaluasi performa karena mampu menangkap keseimbangan antara *precision* dan *recall*. Model *K-Nearest Neighbors* (K-NN) dengan nilai k = 5 menunjukkan performa paling unggul secara konsisten pada metrik *accuracy* dan *F1-Score*. Hal ini mengindikasikan bahwa model tersebut mampu memberikan keseimbangan antara kemampuan mengenali pasien berisiko (*recall*) dan ketepatan dalam menghindari kesalahan prediksi (*precision*), yang sangat krusial dalam konteks deteksi penyakit.

Keunggulan K-NN juga diperkuat oleh temuan dari Hakim et al. [19], yang menyatakan bahwa K-NN merupakan metode yang sangat adaptif dalam mengenali pola-pola non-linear pada data medis, terutama saat digunakan untuk prediksi berdasarkan data klinis pasien seperti tekanan darah dan kadar kolesterol. Dalam penelitian tersebut, akurasi yang diperoleh mencapai angka yang kompetitif, mendukung efektivitas algoritma ini dalam konteks serupa.

Lebih lanjut, hasil penelitian oleh Napiah dan Heristian [15] juga menguatkan temuan ini. Dalam studi mereka, dilakukan perbandingan antara algoritma *Logistic Regressio*n, *Naïve Bayes*, dan *K-Nearest Neighbors* dalam klasifikasi penyakit jantung menggunakan pendekatan *supervised learning* berbasis dataset terbuka. Hasil evaluasi menunjukkan bahwa algoritma K-NN mencapai tingkat akurasi tertinggi sebesar 91%, disusul oleh *Logistic Regression* sebesar 87%, dan *Naïve Bayes* sebesar 83%. Meskipun terdapat perbedaan pada karakteristik dataset serta tahapan preprocessing yang digunakan, kesamaan pola performa ini memberikan dukungan empiris terhadap superioritas relatif K-NN dalam kasus klasifikasi risiko penyakit jantung. Hal ini memperkuat generalisasi bahwa pendekatan berbasis tetangga terdekat memiliki keunggulan dalam menangkap hubungan kompleks antar fitur klinis, terutama ketika distribusi data tidak linier dan kelas tidak seimbang.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

4. KESIMPULAN

Penelitian ini menggunakan pendekatan CRISP-DM untuk menganalisis dan memprediksi risiko penyakit jantung dengan memanfaatkan tiga algoritma *machine learning*: *Logistic Regression*, *K-Nearest Neighbors* (K-NN), dan *Naïve Bayes*. Proses dimulai dengan pemahaman masalah dan tujuan, yaitu untuk memprediksi potensi seseorang terkena penyakit jantung berdasarkan faktor risiko seperti usia, tekanan darah, kolesterol, dan kebiasaan merokok. Setelah itu, dataset yang terdiri dari 10.000 data pasien dengan 21 variabel medis dipersiapkan dengan mengatasi missing value, mengonversi data kategorik menggunakan LabelEncoder, dan menerapkan *Min-Max Scaling* untuk menyamakan skala fitur. Model kemudian dibangun menggunakan ketiga algoritma tersebut untuk mengklasifikasikan data ke dalam dua kelas: *Heart Disease* atau *Non-Heart Disease*. Hasil evaluasi menunjukkan bahwa model K-NN (k=5) memberikan performa terbaik dengan akurasi 0.7203 dan *F1-Score* 0.7598, yang menunjukkan keseimbangan terbaik antara *Precision* dan *Recall* dalam mendeteksi kasus penyakit jantung. Sementara *Logistic Regression* dan *Naïve Bayes* memberikan hasil yang lebih rendah, keduanya tetap memberikan kontribusi dalam prediksi risiko penyakit jantung. Model K-NN (k=5) terbukti sebagai pilihan optimal dalam konteks prediksi penyakit jantung pada dataset ini, memberikan prediksi yang lebih akurat dan lebih seimbang dalam mengidentifikasi baik kasus positif maupun negatif. Untuk hasil yang lebih baik, dapat dipertimbangkan untuk menambahkan variabel yang berkaitan dengan diabetes. Hal ini dilakukan agar dapat memperkaya dataset dan memberikan informasi yang lebih detail, serta meningkatkan akurasi prediksi. Mengumpulkan data yang lebih relevan, seperti tipe nyeri dada, pola makan atau faktor lingkungan sekitar, juga dapat membantu menghasilkan prediksi yang lebih realistis. Untuk meningkatkan kinerja model, eksplorasi lebih lanjut dengan algoritma lain seperti *Random Forest* atau model lainnya yang dapat memberikan perbandingan yang lebih luas dan membantu menemukan model yang juga dapat diandalkan dalam memprediksi penyakit jantung.

REFERENCES

[1] Kementerian Kesehatan Republik Indonesia, “Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer,” *Kementerian Kesehatan Republik Indonesia*, 2022. [Online]. Available: https://kemkes.go.id/eng/penyakit-jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer

[2] R. Fadil, “Gambaran Profil Lipid pada Pasien Penderita Jantung Koroner di RSPAD Gatot Soebroto,” *Bab 1 dalam Laporan Penelitian*, 2024.

[3] H. Hidayat, A. Sunyoto, and H. Al Fatta, “Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier,” *J. SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, 2023, doi: 10.47970/siskom-kb.v7i1.464.

[4] S. Yusuf, P. Joseph, S. Rangarajan, S. Islam, A. Mente, P. Hystad, M. Brauer, V. R. Kutty, R. Gupta, A. Wielgosz, K. F. AlHabib, A. Dans, P. Lopez-Jaramillo, A. Avezum, F. Lanas, A. Oguz, I. M. Kruger, R. Diaz, K. Yusoff, R. Kelishadi, P. Mony *et al.*, “Modifiable risk factors, cardiovascular disease and mortality in 155,722 individuals from 21 high-, middle-, and low-income countries,” *The Lancet*, vol. 395, no. 10226, pp. 795–808, 2020, doi: 10.1016/S0140-6736(19)32008-2.

[5] A. H. Anwar, “Sistematic Review Faktor Resiko Penyakit Jantung Koroner di Indonesia,” *Indonesian Journal of Health Research Innovation (IJHRI)*, vol. 2, no. 1, pp. 57-69, 2025, doi: https://doi.org/10.64094/fqanc998

[6] R. Naifa Saniy, Y. Tarida Sheevana Sitorus, N. Angela Meyana, S. Najwa, A. Apriliyanti Pravitasari, and F. Indrayatna, “Penerapan Algoritma K-Nearest Neighbor pada Klasifikasi Penyakit Jantung”, *BIAS*, vol. 2022, no. 1, pp. 222–229, 2023, doi: https://doi.org/10.1234/bias.v2022i1.192

[7] R. Helilintar, R. A. Ramadhani, and S. Rochana, *Data Mining: K-Nearest Neighbor*, Kediri: Fakultas Teknik Universitas Nusantara PGRI Kediri, 2017. [Online]. Available: https://www.researchgate.net/publication/321804055 DATA MINING K-Nearest Neighbor

[8] K. M. Mohi Uddin, R. Ripa, N. Yeasmin, N. Biswas, and S. K. Dey, “Machine learning-based approach to the diagnosis of cardiovascular disease using a combined dataset,” *Intelligent Medicine*, vol. 7, pp. 1–15, 2023, doi: https://doi.org/10.1016/j.ibmed.2023.100100

[9] A. A. Surya and Y. Yamasari, “Array,” *J. Informatics and Computer Science (JINACS)*, vol. 5, no. 3, pp. 447–455, 2024, doi: 10.26740/jinacs.v5n03.p447-455.

[10] I. Amal, “Analisis Deteksi Dini Penyakit Jantung dengan Pendekatan Regresi Logistik pada Data Pasien,” *Skripsi*, Universitas Muhammadiyah Makassar, 2024.

[11] I. S. Karima, “Penerapan Machine Learning untuk memprediksi Resiko Pengidap Penyakit Jantung menggunakan Algoritma Decision Tree,” *Jurnal Ilmiah Teknik Informatika*, vol. 14, no. 1, pp. 73–81, 2025, doi: http://dx.doi.org/10.22441/format.2025.v14.i1.007

[12] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python,” *Bina Insani ICT J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.

[13] O. Ördekçi, “Heart disease,” Kaggle, https://www.kaggle.com/datasets/oktayrdeki/heart-disease

[14] M. M. Baharuddin, H. Azis, and T. Hasanuddin, “Analisis performa metode K-Nearest Neighbor untuk identifikasi jenis kaca,” *Ilkom Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, 2019, doi: https://doi.org/10.33096/ilkom.v11i3.489.269-274

[15] S. Heristian, “Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung,” *J. Infortech*, vol. 6, no. 1, pp. 46–51, 2024, doi: 10.31294/infortech.v6i1.21888.

[16] A. Ratnasari, J. Wahidin, A. E. Setiawan, and P. Bintoro, “Machine Learning untuk Klasifikasi Penyakit Jantung,” *Aisyah J. Informatics and Electrical Engineering (A.J.I.E.E)*, vol. 6, no. 1, pp. 145–150, 2024, doi: 10.30604/jti.v6i1.272.

[17] D. Fabiyanto and Z. Pratama Putra, “Validasi Efektivitas Logistic Regression untuk Diagnosa Penyakit Jantung melalui Pendekatan Machine Learning,” *Jurnal Ilmiah FIFO*, vol. 16, no. 2, pp. 158–167, 2024, doi: 10.22441/fifo.2024.v16i2.006.

[18] A. Yulandari, S. K. Nur, and A. Hernita, “Perbandingan Metode Decision Tree, Naïve Bayes, dan K-Nearest Neighbor (KNN) untuk Meningkatkan Akurasi Algoritma Machine Learning dalam Memprediksi Heart Disease (Penyakit Jantung),” *Madani:*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2771−2779
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.8518

J. Ilmiah Multidisiplin, vol. 2, no. 11, pp. 529–536, 2024, doi: 10.5281/zenodo.14377870.

[19] L. Hakim, A. Sobri, L. Sunardi, and D. Nurdiansyah, "Prediksi Penyakit Jantung Berbasis Machine Learning dengan Menggunakan Metode K-NN," *J. Digital Teknol. Inform.*, vol. 7, no. 2, pp. 14–20, 2024, doi: https://doi.org/10.32502/digital.v7i2.9429