

**PERBANDINGAN ALGORITMA *DECISION TREE* DAN
NAÏVE BAYES DALAM MENGLASIFIKASIKAN PENYAKIT
JANTUNG**

SKRIPSI

OLEH:

MUHAMMAD ADILISYAH RIYADI

218160043



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2025**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 19/1/26

Access From (repositori.uma.ac.id)19/1/26

Perbandingan Algoritma *Decision Tree* dan *Naive Bayes* Dalam Mengklasifikasikan Pnyekait Jantung

SKRIPSI

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area**

Oleh:

MUHAMMAD ADILISYAH RIYADI

218160043

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2025**

HALAMAN PENGESAHAN

Judul Skripsi : Perbandingan Algoritma *Decision Tree* dan *Naive Bayes* Dalam Mengklasifikasikan Penyakit Jantung

Nama : Muhammad Adilisyah Riyadi

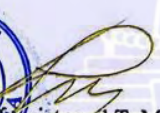
NPM : 218160043

Fakultas : Teknik

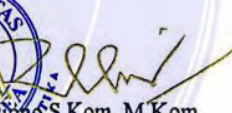
Disetujui Oleh
Komisi Pembimbing


Dr. Hartono S.Kom., M.Kom.
Pembimbing




Dr. Priatno, S.T., M.T
Dekan Fakultas Teknik




M. Mulhono S.Kom, M.Kom.
Ka. Prodi

Tanggal Lulus: 17 September 2025

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar serjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 17 September 2025



Muhammad Adilisyah Riyadi
218160043

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Adilisyah Riyadi
NPM : 218160043
Program Studi : Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

Perbandingan Algoritma *Decision Tree* Dan *Naive Bayes* Dalam Mengklasifikasi Penyakit Jantung.

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 17 September 2025



(Muhammad Adilisyah riyadi)
218160043

RIWAYAT HIDUP

Muhammad Adilisyah Riyadi lahir di Kota Tebing Tinggi pada tanggal 01 Juli 2003. Penulis ini merupakan anak kedua dari dua bersaudara dari ayah ALM Selamat Riyadi dan Ibunda Tengku Erlina. Penulis memulai Pendidikan di TK Inti Nusantara Tebing Tinggi. Kemudian melanjutkan ke SD Inti Nusantara Tebing Tinggi hingga kelas 6 dan menyelesaikan Pendidikan dasar di sekolah tersebut di tahun 2015.

Pada tahun yang sama, penulis melanjutkan Pendidikan di tingkat menengah pertama, yaitu di SMP Negeri 6 Tebing Tinggi, dan menyelesaikannya pada tahun 2018. Selanjutnya, penulis melanjutkan ke jenjang menengah atas di SMAN 4 Tebing Tinggi sejak tahun 2018 hingga lulus pada tahun 2021. Selepas menamatkan sekolah menengah atas, penulis melanjutkan Pendidikan ke perguruan tinggi di Universitas Medan Area pada tahun 2021 hingga tamat di tahun 2025.

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah SWT, yang telah melimpahkan Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Perbandingan Algoritma Decision Tree dan Naïve Bayes Dalam Mengklasifikasikan Penyakit Jantung”** skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Pendidikan Program Sarjana pada fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area. Dalam penyusunan skripsi ini, penulis telah menerima banyak bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
2. Bapak Dr.Eng. Supriatno, S.T., M.T. selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Kepala Program Studi Teknik Informatika.
4. Bapak Dr. Hartono S.Kom, M.Kom., selaku dosen pembimbing yang telah memberikan arahan dan bimbingan selama proses penyusunan skripsi.
5. Seluruh Dosen Teknik Informatika Universitas Medan Area yang telah memberikan penulis saran yang bermanfaat.
6. Ibunda dan sosok pendukung penulis yang selalu menyemangati dan memberikan dukungan selama ini.

7. IT Support Teknik Informatika bang Robby Kurniawan Sari Damanik, ST yang telah banyak membantu penulis dalam menyiapkan pemberkasan yang diperlukan dari awal sampai selesai.
8. Teman-teman kelas penulis yang telah memberikan dukungan dan bantuan selama proses penyusunan skripsi ini.



Medan, 17 September 2025

Muhammad Adilisyah Riyadi
218160043

ABSTRAK

Penyakit Jantung Koroner(PJK) adalah salah satu penyebab kematian paling umum di dunia, dan deteksi dini sangat penting untuk mengurangi efek fatalnya. Dalam penelitian ini, dua algoritma *Machine Learning*, *Decision Tree* dan *Naïve Bayes*, dievaluasi untuk mengklasifikasikan penyakit jantung menggunakan dataset *Kaggle*. Algoritma ini dievaluasi berdasarkan akurasi, presisi, *recall*, *fi-score* nya melalui tiga proporsi data latih dan uji (85:15, 75:25, dan 60:40). Hasil penelitian menunjukkan bahwa *Decision Tree* secara konsisten memiliki kinerja yang lebih baik daripada *Naïve Bayes*. Keunggulan *Decision Tree* juga ditunjuk oleh metrik seperti presisi dan *recall*. Keunggulan utama *Decision Tree* Adalah kemampuannya untuk menangani data numerik dan kategorikal dan membuat hasil klasifikasi lebih mudah dipahami. Sebaliknya, *Naïve Bayes* menunjukkan keterbatasan dalam mengidentifikasi pola kompleks dalam data medis. Meskipun dia berhasil. Hasil evaluasi menunjukkan bahwa algoritma *Decision Tree* lebih baik untuk klasifikasi penyakit jantung karena lebih stabil dan akurat. Penelitian ini memberikan wawasan untuk pengembang sistem deteksi dini penyakit jantung yang lebih efisien yang menggunakan *Machine Learning*.

Kata Kunci: *Decision Tree*, *Naïve Bayes*, Klasifikasi, Penyakit Jantung, *Machine Learning*, Evaluasi Model.

ABSTRACT

Corony Heart Diease (CHD) is one of the most common causes of death worldwide, and early detection is crucial to reduce its fatal effects, In this study, two Machine Learning algorithms, Decision Tree and Naïve Bayes, were evaluated to classify heart disease using the Kaggle dataset. These algorithms were evaluated based on their accuracy, precision, recall, and fi-score across three training and testing data proportions (85:15, 75:25, and 60:40). The result showed that Decision Tree consistently outperformed Naïve Bayes. The superiority of Decision Tree was also demonstrated by metrics such as precision dan recall. The main advantage of Decision Tree is its ability to handle both numerical and categorical data and make classification results easier to understand. In contracts, Naïve Bayes showed limitations in indentifying complex patterns in medical data. Despite its success, the evaluation results showed that the Decision Tree algorithm is better for heart disease classification because it is more stable and accurate. This study provides insights for developers of more efficient early detection systems for heart disease using Machine Learning.

Keywords: *Decision Tree, Naïve Bayes, Classification, Heart Disease, Machine Learning, Model Evaluation*

DAFTAR ISI

HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	Error! Bookmark not defined.
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
 BAB I PENDAHULUAN	 1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
 BAB II TINJAUAN PUSTAKA	 7
2.1 Penyakit Jantung.....	7
2.2 Klasifikasi.....	7
2.3 Algoritma <i>Decision Tree</i>	8
2.4 Algoritma <i>Naïve Bayes</i>	8
2.5 <i>Confusion Matrix</i>	9
2.6 Penelitian Terdahulu.....	10
 BAB III METODOLOGI PENELITIAN	 12
3.1 Kebutuhan Sistem, Perangkat Keras dan Perangkat Lunak	12
3.2 Tahapan Penelitian	12
3.2.1 Tinjauan Literatur	14
3.2.2 Pengumpulan Data.....	14
3.2.3 Pra-Pemrosesan Data	16
3.2.4 Implementasi	18
3.2.5 Evaluasi	18
3.3 Perhitungan Manual	19
3.3.1 Perhitungan Manual Untuk <i>Decision Tree</i>	20
3.3.2 Perhitungan Manual Untuk <i>Naïve Bayes</i>	22

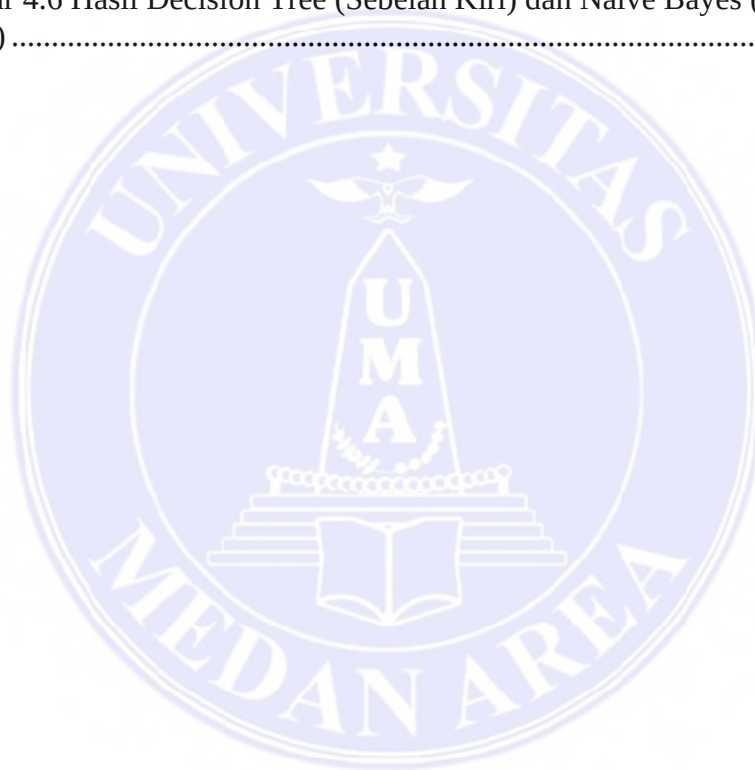
BAB IV HASIL DAN PEMBAHASAN.....	26
4.1 Data Penelitian	26
4.2 Pra-Pemrosesan	28
4.2.1 Memeriksa Data <i>Null</i>	30
4.2.2 Data Normalisasi	30
4.3 Exploratory Data Analysis(EDA)	32
4.3.1 Korelasi Atribut Terhadap Kelas	33
4.3.2 Korelasi Antar Fitur/Atribut	34
4.3.2 Box Plot	35
4.4 Pembagian Data.....	37
4.5 Implementasi	38
4.5.1 Pengujian Menggunakan Proporsi Data 85:15	38
4.5.2 Pengujian Menggunakan Proporsi Data 75:25	40
4.5.3 Pengujian Menggunakan Proporsi Data 60:40	41
 BAB V KESIMPULAN DAN SARAN	 42
5.1 Kesimpulan.....	42
5.2 Saran.....	43
 DAFTAR PUSTAKA	 45
LAMPIRAN.....	47

DAFTAR TABEL

	Halaman
Tabel 2.1 Rangkuman Penelitian Terdahulu	10
Tabel 3.1 Rangkuman dataset	15
Tabel 3.2 Dataset Yang Sudah Di Pra-Pemrosesan	17
Tabel 3.3 Data Latih Dummy	19
Tabel 3.4 Data Uji Dummy	19
Tabel 3.5 Data Atribut Gender	20
Tabel 3.6 Data Rangkuman Information Gain	22
Tabel 3.7 Data Kelas Positive	23
Tabel 3.8 Data Kelas Negative	23
Tabel 4.1 Rangkuman Dataset	27
Tabel 4.2 Rangkuman Pra-Pemrosesan	29
Tabel 4.3. Data Normalisasi	31
Tabel 4.4 Atribut Kelas	33
Tabel 4.5 Pembagian Data Latih dan Data Uji	38
Tabel 4.6 Hasil Pengujian dengan Proporsi Data 85:15	39
Tabel 4.7 Hasil Pengujian dengan Proporsi Data 75:25	40
Tabel 4.8 Hasil Pengujian dengan Proporsi Data 60:40	41

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Alur Tahapan Penelitian	13
Gambar 4.1 Hasil Mencari Data Null	30
Gambar 4.2 korelasi Antar Fitur	35
Gambar 4.3 Gambar Box Plot	37
Gambar 4.4 Hasil Decision Tree (Sebelah Kiri), Naive Bayes (Sebelah Kanan)	39
Gambar 4.5 Hasil Decision Tree (Sebelah Kiri) dan Naïve Bayes (Sebelah Kanan)	40
Gambar 4.6 Hasil Decision Tree (Sebelah Kiri) dan Naive Bayes (Sebelah Kanan)	41



DAFTAR LAMPIRAN

	Halaman
Lampiran1 Kode Program.....	47



BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit Jantung Koroner adalah penyakit yang disebabkan oleh aterosklerosis, proses yang menyuplai aliran darah ke jantung, dan penumpukan lemak pada arteri *coroner*, yang menyumbat aliran darah ke jantung(Bachtiar dkk., 2023). Menurut (Haganta Depari dkk., 2022)Berdasarkan data riset Kesehatan dasar, angka kejadian penyakit jantung di Indonesia terus meningkat setiap tahunnya. Pada tahun 2018, 15 dari 1000 orang, atau sekitar 4,2 juta orang di Indonesia , menderita penyakit jantung. Dokter harus mendiagnosa penyakit jantung melalui berbagai pemeriksaan fisik.

Seiring berkembangnya kemajuan teknologi yang sangat pesat ini menimbulkan beberapa masalah yang perlu diperhatikan(Anisa Nurhasanah dkk., 2024). Salah satunya adalah fakta bahwa, sebagai akibat dari perkembangan industri, akan ada jumlah data yang signifikan yang dikumpulkan untuk menangani informasi medis. Hal yang perlu diperhatikan adalah bagaimana kita dapat menguraikan dan menganalisis informasi ini sehingga dapat digunakan untuk memberikan layanan Kesehatan. Selanjutnya, *Mechine Learning* memainkan peran penting dalam proses penanganan dan analisis data medis. Dengan *Mechine Learning*, kita dapat mencoba berbagai metode yang tersedia untuk menemukan pola atau informasi yang penting dan bermanfaat yang diharapkan akan membantu meningkatkan kualitas layanan Kesehatan saat ini.(Sodikin, 2023)

Dalam hal ini, teknologi ini memungkinkan sistem untuk belajar dari data yang tersedia dan menemukan pola penting yang mungkin tidak terdeteksi oleh metode konvensional. Dengan menggunakan algoritma *Mechine Learning*, sistem dapat membantu dalam deteksi dini, diagnosa yang lebih akurat, dan perencanaan perawatan yang lebih efektif untuk pasien yang menderita penyakit jantung coroner.(Yulianto & Murdianto, 2024)

Setiap algoritma umumnya akan bekerja secara optimal. Karena itu, penelitian akan dilakukan dengan mengumpulkan data, mengolahnya sesuai dengan karakteristik yang relevan, dan kemudian mendapatkan hasil yang tepat untuk setiap metode model prediktif yang ideal dengan menggunakan *Decision Tree* dan *Naïve Bayes*.(Rosyadi dkk., 2025)

Menurut (Shiri Harzevili & Alizadeh, 2018) Keunggulan Dari *Naïve Bayes*, *Naïve Bayes* memiliki struktur yang sangat sederhana, menjadikannya mudah diimplementasikan dan dipahami. Efisien dalam Estimasi Parameter dibandingkan dengan banyak algoritma Klasifikasi lainnya, *Naïve Bayes* membutuhkan jumlah parameter yang sedikit untuk estimasi, sehingga cocok untuk dataset dengan jumlah fitur besar. Tidak memerlukan pembelajaran struktur(*Structural Learning*) berbeda dengan jaringan Bayesian umum, INaive Bayes tidak perlu mempelajari struktur dari data, karena struktur klsifikasinya sudah ditentukan sebelumnya, yaitu asumsi bahwa semua atribut saling berbeda bersyarat (*Conditionally Independent*) terhadap kelas. Kinerja prediktif yang baik meskipun asumsi tidak sepenuhnya benar walaupun asumsi conditional independence sering tidak benar dalam dunia nyata, *Naïve Bayes* tetap menunjukkan performa klasifikasi yang mengesankan di berbagai domain. Sangat cocok untuk dataset besar penelitian terdahulu

menyebutkan kinerja *Naïve Bayes* tetap stabil bahkan pada dataset dengan ukuran sampai yang besar.

Menurut (Matzavela & Alepis, 2021) Keunggulan dari *Decision Tree*, mudah dipahami dan diinterpretasi *Decision Tree* menyajikan hasil dalam bentuk aturan "if-then-else" yang sederhana, menjadikannya sangat mudah dipahami, bahkan oleh pengguna non-teknis. Visualisasi yang jelas struktur berbentuk pohon memudahkan representasi grafis, yang membuat hasil klasifikasi dapat divisualisasikan dengan sangat jelas dan intuitif. Dengan menangani data numerik dan kategorikal algoritma ini mampu memproses dan mengklasifikasikan data numerik maupun kategorikal secara bersamaan tanpa perlu banyak transformasi data sebelumnya. Sedikit persiapan data dibutuhkan berbeda dengan algoritma lain yang sering memerlukan normalisasi atau skala ulang data, *Decision Tree* bekerja cukup baik dengan data mentah (raw draw). Performa baik untuk dataset besar *Decision Tree* cocok untuk aplikasi dengan volume data besar dan kompleksitas tinggi, tanpa mengorbankan efisiensi dan kecepatan proses.

Beberapa penelitian terdahulu, yang berhasil menerapkan algoritma *Decision Tree* dan *Naive Bayes* yaitu "Perbandingan Algoritma *Naïve Bayes* dan *Decision tree* (c4.5) dalam Klasifikasi Dosen Berprestasi" oleh Andy Supriadi 2023, tahapan penelitian diawali dengan pengumpulan data dosen yang diambil dari wawancara dan *database* sistem kepegawaian Universitas Sebelas Maret dalam bentuk excel. Kemudian dilakukan pengolahan awal data dengan menghapus data yang tidak lengkap dan memberikan label pada data *training*, pertama pemodelan dengan *Naive Bayes*, kedua dengan *Decision Tree*, kemudian dilakukan pengujian dan evaluasi model dengan *tools Rapidminer versi 9.1*.

Studi terdahulu yang ditulis oleh Adi Sholihin Fauzan, Ade Irma Purnama Sari, Irfan Ali 2024 dengan judul “ Analisis Perbandingan Algoritma Decision Tree dan Naïve Bayes untuk Mengevaluasi Prestasi Belajar Siswa Studi Kasus: SMK AL-MUSYAWIRIN “ Penelitian yang berfokus pada data yang memberikan informasi yang berguna untuk menyimpulkan hasil. Metode penelitian kuantitatif dijelaskan sebagai suatu bentuk kajian sistematis terhadap suatu fenomena yang melibatkan observasi langsung dilapangan, serta pengumpulan data yang telah didapatkan selanjutnya dianalisis memanfaatkan Teknik statistik untuk menemukan korelasi antar variable yang diteliti dan juga membandingkan dua algoritma sebagai penilaian algoritma terbaik . Analisis yang diterapkan untuk memprediksi prestasi belajar siswa SMK AL-Musyawirin dengan metode algoritma *Naïve Bayes* dan algoritma *Decision Tree* dengan mengaplikasikan *tools rapidminer*.

Kemudian penelitian terdahulu yang menggunakan algoritma *Naïve Bayes* dan *Decision Tree* yaitu, Gifthera Dwilestari dan Turfa Azmi Afifah 2025 meneliti tentang “ Perbandingan kinerja Algoritma *Naïve Bayes* dan *Decision Tree* dalam klasifikasi kanker paru-paru “ yang digunakan dalam studi ini dirancang untuk secara sistematis mengeksplorasi dan menganalisis kinerja dua algoritma klasifikasi yaitu, *Naïve Bayes* dan *Decision Tree*, dalam konteks klasifikasi kanker paru-paru. Penelitian ini mengikuti pendekatan Knowledge Discovery in Database(KDD), yang mencakup tahapan-tahapan kritis mulai dari pemilihan data, praproses data, transformasi, penambangan data, hingga interpretasi hasil.

Ketiga penelitian diatas memiliki keunggulan yang saling melengkapi dalam mendukung penggunaan algoritma *Decision Tree* dan *Naïve Bayes*. Penelitian Andy Supriadi (2023) unggul dalam penggunaan data yang valid dari sistem

kepegawaian, tahapan *preprocessing* data yang teliti, dan evaluasi sistematis menggunakan alat professional seperti *Rapidminer*. Penelitian Adi Sholihin Fauzan dkk(2024) menonjol dengan pendekatan kuantitatif berbasis lapangan menggabungkan observasi langsung. Selain itu, penelitian yang dilakukan oleh Gifthera dan Turfa Azmi Afifah (2025) menunjukkan manfaat penerapan metodologi KDD, penggunaan fitur medis nyata, dan evaluasi model klasifikasi, ketepatan, recall, dan skor F1. Kombinasi keunggulan tersebut menunjukkan bahwa algoritma *Naïve Bayes* dan *Decision Tree* sangat fleksibel dalam penerapan dan dapat digunakan di berbagai bidang. Mereka juga memberikan hasil yang akurat jika didukung dengan metodologi yang tepat dan pemrosesan data yang optimal.

1.2 Rumusan Masalah

Berdasarkan penjelasan dari latar belakang diatas maka dapat disimpulkan beberapa rumusan masalah, antara lain sebagai berikut :

1. Bagaimana caranya membandingkan kedua algoritma tersebut dan memilih algoritma terbaik?
2. Bagaimana kinerja algoritma *Decision Tree* dan *Naïve Bayes* dalam menganalisis data medis untuk mendukung deteksi dan diagnose penyakit jantung coroner?

1.3 Batasan Masalah

Batasan masalah diperlukan untuk mempermudah penulisan skripsi ini dan agar pembahasan lebih terarah, dan sesuai dengan latar belakang yang sudah diuraikan:

Penelitian ini hanya difokuskan pada hal-hal berikut:

1. Algoritma yang digunakan dalam penelitian ini adalah *Decision Tree* dan *Naïve Bayes*, tanpa membandingkan dengan algoritma lain.
2. Data yang digunakan dalam penelitian ini berasal dari dataset yang telah dikumpulkan dan tersedia dalam bentuk data sekunder.

1.4 Tujuan Penelitian

Berikut beberapa tujuan penelitian yang dapat disimpulkan sebagai berikut:

1. Membandingkan kinerja algoritma *Decision Tree* dan *Naïve Bayes* dalam klasifikasi penyakit jantung coroner.
2. Menentukan algoritma terbaik di antara *Decision Tree* dan *Naïve Bayes* untuk digunakan dalam diagnose penyakit jantung coroner.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Meningkatkan pengetahuan tentang *mechine learning*, terutama dengan menggunakan algoritma *Decision Tree* dan *Naïve Bayes* untuk analisis data medis.
2. Meningkatkan efisiensi dalam pengolahan data medis dengan memanfaatkan algoritma *Decision Tree* dan *Naïve Bayes* untuk mengolah data dalam jumlah besar dan menemukan pola yang relevan.
3. Informasi, hasil penelitian dapat digunakan oleh penelitian lain yang tertarik dengan algoritma *Decision Tree* dan *Naïve Bayes* di bidang teknologi dalam dunia keseshatan.

BAB II

TINJAUAN PUSTAKA

2.1 Penyakit Jantung

Penyakit jantung coroner (PJK) adalah salah satu jenis penyakit jantung yang menjadi penyebab kematian nomor satu di dunia. PJK adalah penyakit yang terkait dengan gaya hidup dan Sosial ekonomi Masyarakat, dan merupakan masalah Kesehatan utama di negara maju(Hadi dkk., 2017).

Menurut(Zahrawardani dkk., 2013) Faktor risiko banyak ditemukan pada penderita PJK. Faktor risiko lipida, termasuk kadar kolestrol dan trigliserida, sangat penting karena sifat substansi ini yang mendorong timbulnya plak di arteri coroner.

2.2 Klasifikasi

Klasifikasi adalah proses menilai objek data sehingga dapat dimasukkan ke dalam kelas tertentu dari jumlah kelas yang tersedia. Klasifikasi terdiri dari dua tugas utama yaitu, membuat prototipe model untuk disimpan dalam memori, dan menggunakan model tersebut untuk menemukan, mengklasifikasikan, atau memprediksi objek data lain untuk dikelas mana objek data tersebut dalam model yang sudah disimpannya(Putro dkk., 2020).

Metode Klasifikasi menggunakan pola dari data latih yang diberi label untuk menerapkannya pada data baru yang kelasnya belum diketahui. *Decision Tree*, *Naïve Bayes*, adalah beberapa algortima klasifikasi populer.(Damanik dkk, 2024)

Menurut (Fabiyanto & Pratama Putra, 2024)Untuk mendeteksi penyakit jantung, klasifikasi memungkinkan proses diagnosa yang lebih cepat dan efisien.

Selain itu, klasifikasi multikelas, yang terdiri dari dua kelas, seperti “positif” dan “negatif”, dapat digunakan untuk menentukan Jenis penyakit atau Tingkat keparahan.

2.3 Algoritma *Decision Tree*

Decision Tree atau pohon Keputusan adalah struktur diagram alur berbentuk pohon dengan setiap simpul internal (node non-daun) menunjukkan pengujian terhadap atribut tertentu, setiap cabang pada pohon menunjukkan hasil dari pengujian tersebut, dan setiap simpul daun (daun terminal) menunjukkan informasi label kelas. Simpul akar adalah titik tertinggi dari struktur pohon (Antika dkk., 2023).

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2.1)$$

Disini, S adalah Kumpulan kasusnya, n adalah jumlah kasus partisi atribut A, A adalah atribut, S_i adalah jumlah kasus ke I, S adalah jumlah kasus S. (Setiawan dkk., 2024)

2.4 Algoritma *Naïve Bayes*

Menurut (Felicia Watratan dkk., 2020) Berdasarkan *teorema Bayes*, metode klasifikasi *Naïve Bayes* menggunakan metode probabilitas dan *statistic* dan berusaha memprediksi peluang di masa depan berdasarkan pengalaman masa lalu. Metode ini dikenal sebagai *teorema Bayes*. Ciri utama dari *Naïve Bayes* ini adalah keyakinan naif yang kuat bahwa setiap situasi atau peristiwa berdiri sendiri.

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (2.2)$$

Di sini, X adalah data dengan kelas yang belum diketahui, H adalah hipotesis bahwa X termasuk dalam kelas tertentu, $P(H|X)$ adalah probabilitas H berdasarkan data X, $P(X|H)$ adalah probabilitas data X muncul jika Hipotesis H benar, $P(H)$ adalah probabilitas awal dari hipotesis H, dan $P(X)$ adalah probabilitas keseluruhan dari data X (Mulyani dkk, 2022)

2.5 Confusion Matrix

Menurut (Setiawan dkk., 2024) Terdapat empat komponen utama, *Confusion Matrix* adalah alat yang digunakan untuk mengukur kinerja model klasifikasi dalam memprediksi berbagai kelas atau label. *True Positive*(TP) menunjukkan Ketika model benar memprediksi kelas positif, *True Negative*(TN) menunjukkan Ketika model benar memprediksi kelas negatif; *False Positive* (FP) menunjukkan ketika model salah memprediksi kelas positif; dan *False Negative* (FN) menunjukkan Ketika model salah memprediksi kelas negatif.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (2.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.5)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2.6)$$

2.6 Penelitian terdahulu

Penelitian terdahulu dibutuhkan sebagai acuan untuk melihat sejauh mana topik yang diteliti telah dikajikan sebelumnya. Berikut rangkuman dari penelitian terdahulu yang dapat dilihat pada tabel di bawah ini:

Tabel 2. 1 Rangkuman Penelitian terdahulu

Penulis(Tahun)	Judul	Metode	Hasil
(Ifani hariyanti dkk., 2024)	Perbandingan <i>algoritma Deciion Tree</i> dan <i>Naïve Bayes</i> dalam klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa	<i>Decision Tree</i> dan <i>Naïve Bayes</i>	Penelitian ini menunjukkan akurasi <i>Decision Tree</i> tertinggi sebesar 86,67% dibandingkan <i>Naïve Bayes</i> sebesar 70,00%.
Nur Syifa Fauziah dan Raditya Danar Dana,2023)	Implementasi Algoritma <i>Naïve Bayes</i> dalam klasifikasi Status Kesejahteran Masyarakat Desa Gunungsari	<i>Naïve Bayes</i>	Peneltian ini mendapatkan hasil akurasi dari data training mencapai 93,69%, presisi 97,27%, dan <i>recall</i> 95,23%.
Gifthera Dwilestari dan Turfa Azmi Afifah,2025)	Perbandingan Kinerja Algoritma <i>Naïve Bayes</i> dan <i>Decision Tree</i> dalam Klasifikasi Kanker Paru-Paru.	<i>Naïve Bayes Dan Decision Tree</i>	Penelitian ini menunjukkan bahwa akurasi <i>Naïve bayes</i> sebesar 92,47%, presisi 93,02%, dan <i>recall</i> 98,77%. Sementara itu, <i>Decision Tree</i> memiliki akurasi sebesar 88,17%, presisi 94,87%, dan <i>recall</i> 66,67%.
Bayu Wicaksono dan Nuri Cahyono,2024)	Analisis Sentimen Komentar Instagram pada Program Kampus Merdeka dengan Algoritma <i>Naïve Bayes</i> dan <i>Decision Tree</i>	<i>Naïve Bayes dan Decision Tree</i>	Penelitian ini menunjukkan bahwa <i>decision tree</i> memiliki akurasi terbaik sebesar 85%, presisi 85%, <i>recall</i> 85% dan <i>F1-score</i> 85% dibandingkan <i>Naïve Bayes</i> memiliki akurasi sebesar 83%, presisi 83%, <i>recall</i> 83% dan <i>f1-score</i> 83%.
Salmawati dkk,2022)	Klasifikasi Mahasiswa Berpotensi <i>Drop Out</i> menggunakan Algoritma <i>Naïve Bayes</i> dan <i>Decision Tree</i>	<i>Naïve Bayes dan Decision Tree</i>	Penelitian ini menunjukkan akurasi <i>Naive Bayes</i> sebesar 97,83% dan <i>Decision tree</i> sebesar 99,13%, presisi untuk <i>Naïve Bayes</i> 61,54% dan <i>Decision Tree</i> 100%, <i>recall</i> untuk <i>Naïve Bayes</i> 84,61% dan <i>Decision Tree</i> 100%, <i>F1-</i>

			score untuk <i>Naïve Bayes</i> 76,19% dan <i>Decision Tree</i> 91,67%.
--	--	--	--



BAB III

METODOLOGI PENELITIAN

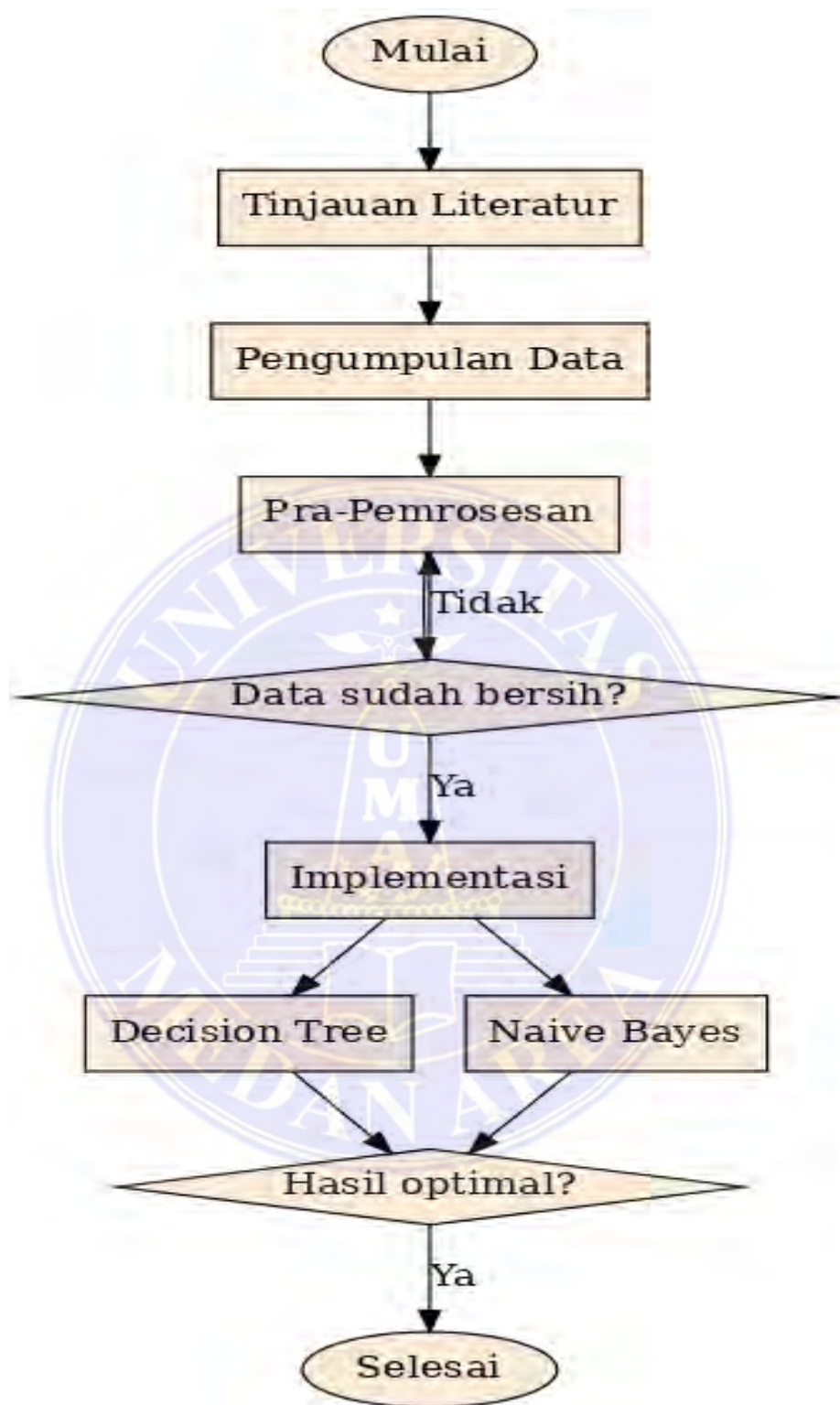
3.1 Kebutuhan Sistem, Perangkat Keras dan Perangkat Lunak

Penelitian ini Membutuhkan perangkat keras serta perangkat lunak yang spesifik guna membantu jalannya penelitian. Spesifikasi tersebut juga diharapkan dapat memaksimalkan dari temuan penelitian. Adapun kebutuhan perangkat keras dan perangkat lunak yang diperlukan dalam pembuatan sistem ini terdiri dari :

- 1) Untuk perangkat keras ialah sebagai berikut :
 - a. Laptop,
 - b. *Processor Intel Core i7* generasi kke-11,Ram 8GB, dan SSD 512GB.
- 2) Untuk perangkat lunak ialah sebagai berikut :
 - a. Bahasa Pemograman *python*, dan
 - b. Spyder.

3.2 Tahapan Penelitian

Pada tahapan penelitian menjelaskan tentang alur dari penelitian dimana algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasi penyakit jantung. Untuk penjelasan lebih lanjut dapat dilihat pada gambar 3.1



gambar 3.1 alur tahapan penelitian

3.2.1 Tinjauan Literatur

Pada tahap ini, peneliti melakukan kajian literatur yang bertujuan untuk memperoleh pemahaman mendalam mengenai konsep dan pendekatan yang berkaitan dengan perbandingan algoritma *Decision Tree* dan *Naïve Bayes* dalam proses klasifikasi data. Analisis terhadap penelitian terdahulu dilakukan untuk mengetahui bagaimana algoritma tersebut digunakan, keunggulan dan keterbatasan masing-masing varian, serta Teknik evaluasi yang diterapkan seperti akurasi, presisi, recall, dan f1-score. Selain itu, kajian ini juga membantu peneliti dalam mengidentifikasi variabel-variabel penting yang berperan dalam proses klasifikasi dan menentukan pendekatan paling tepat yang dapat diadopsi dalam penelitian ini.

3.2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh melalui platform *kaggle* pada tautan berikut. (<https://www.kaggle.com/datasets/bharath011/heart-disease-classification-dataset>) Dataset yang digunakan terdiri dari 9 atribut atau fitur yang menjelaskan mengenai gejala-gejala penyakit jantung seperti usia, jenis kelamin, tekanan darah kadar kolesterol, detak jantung, dan atribut medis lainnya yang relevan. Berikut merupakan rangkuman dari data pada penelitian ini yang dapat dilihat di bawah ini :

tabel 3.1 Rangkuman dataset

age	gender	impulse	pressurehigh	pressurelow	glucose	kcm	troponin	class
64	1	66	160	83	160	1.8	0.012	negative
21	1	94	98	46	296	6.75	1.06	positive
55	1	64	160	77	270	1.99	0.003	negative
64	1	70	120	55	270	13.87	0.122	positive
55	1	64	112	65	300	1.08	0.003	negative
58	0	61	112	58	87	1.83	0.004	negative
32	0	40	179	68	102	0.71	0.003	negative
63	1	60	214	82	87	300	2.37	positive
44	0	60	154	81	135	2.35	0.004	negative
67	1	61	160	95	100	2.84	0.011	negative
44	0	60	166	90	102	2.39	0.006	negative
63	0	60	150	83	198	2.39	0.013	negative
64	1	60	199	99	92	3.43	5.37	positive
54	0	94	122	67	97	1.42	0.012	negative
47	1	76	120	70	319	2.57	0.003	negative
61	1	81	118	66	134	1.49	0.017	positive
86	0	73	114	68	87	1.11	0.776	positive
45	0	70	100	68	96	0.606	0.004	negative
37	0	72	107	86	274	2.89	0.003	negative
45	1	60	109	65	89	1.6	0.02	positive
60	1	92	151	78	301	1.6	0.005	negative

3.2.3 Pra-Pemrosesan data

Pra-Pemrosesan data dalam perbandingan algoritma *Decision Tree* dan *Naïve Bayes* mencakup sejumlah tahapan yang bertujuan untuk memastikan bahwa data yang digunakan berada dalam format yang sesuai dan layak untuk masing-masing algoritma *Decision Tree* dan *Naïve Bayes*. Pada dataset yang digunakan ditemukan bahwa seluruh nilai dalam dataset tidak mengandung data kosong (missing value), sehingga tidak perlu proses imputasi. Namun, terdapat atribut class yang berisi nilai kategori *positive* dan *negative* harus diubah menjadi format numerik, yaitu *positive* = 1 dan *negative* = 0. Hal ini dilakukan agar dapat dikenali oleh algoritma klasifikasi yang hanya menerima input dalam bentuk numerik.

Adapun ringkasan data yang telah melalui tahap pra-pemrosesan untuk memastikan bahwa fitur-fitur yang digunakan dalam Klasifikasi penyakit jantung relevan ditampilkan pada tabel berikut :

tabel 3. 2 Dataset Yang Sudah Di Pra-pemrosesan

age	gender	impulse	pressurehigh	pressurelow	glucose	kcm	troponin	class
64	1	66	160	83	160	1.8	0.012	0
21	1	94	98	46	296	6.75	1.06	1
55	1	64	160	77	270	1.99	0.003	0
64	1	70	120	55	270	13.87	0.122	1
55	1	64	112	65	300	1.08	0.003	0
58	0	61	112	58	87	1.83	0.004	0
32	0	40	179	68	102	0.71	0.003	0
63	1	60	214	82	87	300	2.37	1
44	0	60	154	81	135	2.35	0.004	0
67	1	61	160	95	100	2.84	0.011	0
44	0	60	166	90	102	2.39	0.006	0
63	0	60	150	83	198	2.39	0.013	0
64	1	60	199	99	92	3.43	5.37	1
54	0	94	122	67	97	1.42	0.012	0
47	1	76	120	70	319	2.57	0.003	0
61	1	81	118	66	134	1.49	0.017	1
86	0	73	114	68	87	1.11	0.776	1
45	0	70	100	68	96	0.606	0.004	0
37	0	72	107	86	274	2.89	0.003	0
45	1	60	109	65	89	1.6	0.02	1
60	1	92	151	78	301	1.6	0.005	0

3.2.4 Implementasi

Pada tahap ini, penelitian akan menguji algoritma *Decision Tree* dan *Naïve Bayes* pada data yang telah diproses. Implementasi ini akan dilakukan menggunakan Bahasa pemrograman *Python* dan program *Spyder*. Masing-masing algoritma *Decision Tree* dan *Naive Bayes* diterapkan untuk mempelajari pola dari data yang telah melalui tahap pra-pemrosesan, dengan tujuan membandingkan efektivitas klasifikasi antar algoritma. Implementasi dari kedua algoritma ini akan dilakukan secara terpisah, dan hasil dari klasifikasi masing-masing algoritma akan dievaluasi menggunakan *Confusion Matrix*.

3.2.5 Evaluasi

Tahap evaluasi bertujuan untuk menilai kinerja masing-masing algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan penyakit jantung. Evaluasi dilakukan menggunakan dataset pengujian yang telah diproses selama tahap pra-pemrosesan. *Confusion Matrix* digunakan sebagai evaluasi, dengan metrik utama berupa akurasi, presisi, *recall*, dan *f1-score*. Semakin tinggi nilai dari metrik evaluasi tersebut, maka semakin baik pula performa algoritma. Setelah seluruh metrik dihitung, hasilnya akan dibandingkan antar algoritma untuk menentukan algoritma mana yang memberikan hasil klasifikasi paling optimal. Temuan dari tahap evaluasi ini akan digunakan sebagai dasar dalam menarik Kesimpulan terkait efektivitas masing-masing algoritma *Decision Tree* dan *Naïve Bayes* dalam klasifikasi penyakit jantung.

3.3 Perhitungan Manual

Pada penelitian ini, perhitungan manual digunakan untuk memberikan Gambaran tentang bagaimana proses klasifikasi dilakukan. Perhitungan manual ini akan menggunakan data dummy atau data contoh yang terdiri dari lima(5) fitur yaitu *age*, *gender*, *pressurehigh*, *glucose*, *troponin*, serta satu kolom kelas (*class*) yang bernilai 0 untuk *negative* dan 1 untuk *positive*. Adapun perhitungan manual ini akan dibagi menjadi dua bagian, yaitu perhitungan untuk algoritma *Decision Tree* dan perhitungan manual untuk *Naïve Bayes*. Berikut ini adalah Langkah-langkah perhitungannya :

Tabel 3. 3 Data Latih Dummy

age	gender	pressurehigh	glucose	troponin	class
60	1	160	280	0.012	positive
45	0	120	130	0.004	negative
50	1	140	150	0.011	positive
35	0	110	110	0.003	negative
65	1	170	300	0.020	positive

Tabel 3. 4 Data Uji Dummy

age	gender	pressurehigh	glucose	troponin
55	1	150	200	0.010
40	0	115	100	0.004

3.3.1 Perhitungan Manual untuk *Decision Tree*

Perhitungan manual dilakukan untuk menunjukkan bagaimana algoritma *Decision Tree* bekerja untuk menentukan atribut terbaik yang akan digunakan sebagai akar pohon (root node) berdasarkan nilai *Information Gain*. Dataset yang digunakan adalah data dummy, yang menunjukkan data asli dalam bentuk yang lebih sederhana.

Jumlah data: 5

Kelas:

- Positive = 3
- Negative = 2

1. Entropy Awal (S)

$$\text{Entropy}(S) = -\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right) = -(0.6\log_2 0.6 + 0.4\log_2 0.4) \\ = -(0.6 \times -0.737 + 0.4 \times -1.322) = 0.971$$

2. *Information Gain* untuk setiap Atribut

a. Atribut Gender

Tabel 3. 5 Data Atribut Gender

gender	class	jumlah
1	positive	3
0	negative	2

- gender=1 (3 positive): Entropy = 0

- gender=0 (2 negative): Entropy = 0

$$\text{Gain}(\text{gender}) = 0.971 - \left(\frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0\right) = 0.971$$

b. Atribut Pressurehigh

Dikategorikan menjadi:

- ≤ 140 : baris 2, 3, 4 $\rightarrow$ 1 positive, 2 negative $\rightarrow$ Entropy = 0,918

- 140: baris 1, 5 $\rightarrow$ 2 positive $\rightarrow$ Entropy = 0

$$\text{Gain(Pressurehigh)} = 0,971 - \left(\frac{3}{5} \cdot 0,918 + \frac{2}{5} \cdot 0 \right) = 0,421$$

c. Atribut Age

Dikategorikan menjadi:

- ≤ 50 : baris 2, 3, 4 $\rightarrow$ 1 positive, 2 negative $\rightarrow$ Entropy = 0.918

- 50: baris 1, 5 $\rightarrow$ 2 positive $\rightarrow$ Entropy = 0

$$\text{Gain(age)} = 0,971 - \left(\frac{3}{5} \cdot 0,918 \right) = 0,421$$

d. Atribut Glucose

Dikategorikan menjadi:

- ≤ 200 : baris 2, 3, 4 $\rightarrow$ 1 positive, 2 negative $\rightarrow$ Entropy = 0.918

- 200: baris 1, 5 $\rightarrow$ 2 positive $\rightarrow$ Entropy = 0

$$\text{Gain(glucose)} = 0,971 - \left(\frac{3}{5} \cdot 0,918 \right) = 0,421$$

e. Atribut Troponin

Dikategorikan menjadi:

- ≤ 0.01 : baris 2, 3, 4 1 positive, 2 negative $\rightarrow$ Entropy = 0.918

- 0.1: baris 1,5 $\rightarrow$ 2 positive $\rightarrow$ Entropy = 0

$$\text{Gain(troponin)} = 0,971 - \left(\frac{3}{5} \cdot 0,918 \right) = 0,421$$

3. Rangkuman Information Gain

Tabel 3. 6 Data Rangkuman Information Gain

Atribut	Information Gain
gender	0.971
pressurehigh	0.421
age	0.421
glucose	0.421
troponin	0.421

3.3.2 Perhitungan Manual untuk Naïve Bayes

Perhitungan manual dengan data dummy dilakukan untuk memberikan pemahaman tentang cara algoritma *Naïve Bayes* mengklasifikasikan data. Algoritma ini menggunakan *Teorema Bayes*, yang menyatakan bahwa fitur-fitur saling independent satu sama lain.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana:

- H = hipotesis atau kelas target (positive / negative),
- X = fitur dari data uji,
- P(H) = probabilitas awal (prior probability)
- P(X|H) = probabilitas fitur terjadi dalam kelas H (likelihood),

$P(H|X)$ = probabilitas prediksi setelah mempertimbangkan data (posterior).

Langkah 1: Hitung Mean dan Standar Deviasi

Berikut adalah nilai mean (μ) dan standar deviasi (σ) dari setiap fitur berdasarkan kelas:

a.Kelas Positive:**Tabel 3. 7** Data Kelas Positive

Fitur	Mean (μ)	Standar Deviasi (σ)
age	58.33	6.23
gender	1.00	0.00 Δ
pressurehigh	156.67	12.47
glucose	243.33	62.36
troponin	0.0143	0.00368

b.Kelas Negative:**Tabel 3. 8** Data Kelas Negative

Fitur	Mean (μ)	Standar Deviasi (σ)
age	40.0	7.07
gender	0.00	0.00 Δ
pressurehigh	115.0	5.0
glucose	120.0	14.14
troponin	0.0035	0.0005

Catatan: Karena fitur gender bersifat konstan dalam masing-masing kelas, standar deviasi = 0. Untuk menghindari pembagian dengan nol, σ disubstitusi dengan nilai kecil $\varepsilon=0.0001$ $\varepsilon=0.0001$.

Langkah 2: Hitung Likelihood (P(X|H))

Dengan menggunakan rumus distribusi *Gaussian*:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Dihitung untuk setiap fitur terhadap dua kelas, lalu dikalikan untuk menghasilkan total likelihood

Misalnya:

- Untuk kelas positive:

$$P(\text{age} = 55 | \text{positive}) = \text{Gaussian}(55; \mu = 58.33, \sigma = 6.23)$$

Dan begitu juga untuk fitur lainnya.

Langkah 3: Hitung Posterior Probability

$$P(\text{positive}|X) \propto P(\text{age}|\text{pos}) \cdot P(\text{gender}|\text{pos}) \cdot \dots \cdot P(\text{pos})$$

$$P(\text{negative}|X) \propto P(\text{age}|\text{neg}) \cdot P(\text{gender}|\text{neg}) \cdot \dots \cdot P(\text{neg})$$

Jika hasil:

- $P(\text{positive}|X) > P(\text{negative}|X)$, maka prediksi = positive
- Sebaliknya, jika lebih kecil, prediksi = negative

Hasil akhir

Setelah perhitungan dilakukan:

- $P(\text{positive}|X) \approx$ tidak dapat dihitung secara valid (karena $\sigma = 0$)
- $P(\text{negative}|X)$ juga mengalami kendala serupa.

Namun, untuk keperluan ilustrasi apabila pendekatan numerik diberikan dengan nilai $\varepsilon = 0.0001$, hasil prediksi adalah

Prediksi kelas: positive



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan evaluasi algoritma *Decision Tree* dan *Naïve Bayes* dalam Klasifikasi penyakit jantung menggunakan dataset Kaggle, Untuk menguji, tiga skenario pembagian dapat digunakan (85:15, 75:25, 60:40). Dan *Decision Tree* secara konsisten memperoleh nilai akurasi yang tinggi dalam ketiga skenario tersebut, dengan nilai rata-rata di atas 97%. Nilai *f1-score*, *recall*, dan presisi juga menunjukkan hasil yang sangat memuaskan dan stabil.

Sebaliknya, algoritma *Naïve Bayes* menunjukkan kinerja yang lebih buruk dan kurang konsisten, dengan akurasi rata-rata di bawah 70%. Hasil evaluasi ini menunjukkan bahwa *Naïve Bayes* memiliki keterbatasan dalam mengidentifikasi pola kompleks yang muncul dari data medis, terutama karena sifatnya yang menganggap setiap fitur independent. Banyak karakteristik saling berkaitan, seperti usia, tekanan darah, kadar glukosa, dan kadar troponin, membuat klasifikasi penyakit jantung tidak ideal.

Selain memiliki kelebihan dalam hal akurasi dan stabilitas, Algoritma *Decision Tree* juga memiliki kelebihan dalam hal keterjelasan. Model ini memiliki kemampuan untuk memvisualisasikan proses klasifikasi melalui struktur pohon Keputusan sehingga hasilnya mudah dipahami bahkan oleh orang-orang yang tidak terlalu teknis. Selain itu, *Decision Tree* dapat berfungsi dengan baik tanpa perlu melakukan banyak transformasi atau normalisasi data.

Dengan mempertimbangkan semua hasil implementasi dan evaluasi, dapat disimpulkan bahwa *Decision Tree* Adalah algoritma yang paling cocok dan disarankan untuk digunakan dalam klasifikasi penyakit jantung. Algoritma ini dinilai baik dalam mendeteksi pola, lebih stabil dalam jumlah data yang dilatih dan diuji, dan memberikan hasil klasifikasi yang lebih akurat daripada *Naïve Bayes*.

5.2 Saran

Hasil dari penelitian ini memberikan rekomendasi yang diharapkan dapat digunakan dalam penelitian lanjutan dan penerapan sistem klasifikasi penyakit jantung. Adapun saran untuk penelitian lanjutan sebagai berikut:

Jika dataset yang digunakan memiliki kompleksitas dan keterkaitan antar fitur yang tinggi, algoritma *Decision Tree* dapat digunakan. Selain itu, penelitian lebih lanjut dapat mencoba menggabungkan algoritma *Decision Tree* dengan algoritma lain untuk meningkatkan akurasi dan ketahanan model terhadap ketidakseimbangan performa antara data latih dan data uji.

Disarankan untuk menggunakan dataset yang lebih besar dan bervariasi, serta memasukkan fitur tambahan yang berkaitan dengan kondisi riwayat kesehatan pasien. Semakin kaya data, semakin baik model menangkap pola dan membuat prediksi yang lebih akurat.

Disarankan untuk penelitian selanjutnya memakai aplikasi sederhana seperti kalkulator prediksi penyakit jantung berdasarkan fitur yang digunakan pada penelitian ini.

Penting untuk memperhatikan aspek interpretasi dan kemudahan pemahaman dari model klasifikasi yang digunakan sepanjang waktu. Model *Decision Tree* sangat transparan, tetapi pengembang harus membuat antarmuka

atau penjelasan hasil yang mudah dipahami oleh pengguna akhir terutama untuk sistem yang digunakan dalam praktik medis atau layanan Kesehatan Masyarakat.



DAFTAR PUSTAKA

- Anisa Nurhasanah, Haldini Reygita, & Salsa Nabila Marcella Kalalo. (2024). Pengaruh Teknologi Modern Terhadap Moralitas Dan Tanggung Jawab Siswa Sekolah Dasar. *Student Scientific Creativity Journal*, 2(1), 175–186. <https://doi.org/10.55606/sscj-amik.v2i1.2701>
- Antika, R., Rifa, A., Dikananda, F., Indriya Efendi, D., & Narasati, R. (2023). Penerapan Algoritma *Decision Tree* Berbasis Pohon Keputusan Dalam Klasifikasi Penyakit Jantung. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).
- Bachtiar, L., Gustaman, R. A., Maywati, S., Masyarakat, P. K., Kesehatan, I., & Siliwangi, U. (2023). Faktor Risiko Yang Berhubungan Dengan Kejadian Penyakit Jantung Koroner (PJK) (Analisis Data Sekunder di Rumah Sakit Umum Daerah Kabupaten Subang). In *Jurnal Kesehatan komunitas Indonesia* (Vol. 19).
- Damanik, A. R., Gunawan, I., Sormin, R. K., Tambunan, H. S., Sumarno, S., & Darma, S. (2024). Analisis Metode C4.5 Dalam Mengukur Akurasi Kepuasan Mahasiswa/i Terhadap Layanan Perpustakaan. *Digital Transformation Technology*, 4(1), 699–704. <https://doi.org/10.47709/digitech.v4i1.4498>
- Fabiyanto, D., & Pratama Putra, Z. (2024). Validasi Efektivitas Logistic Regression untuk Diagnosa Penyakit Jantung melalui Pendekatan Machine Learning. *Jurnal Ilmiah FIFO*, 16(2), 158. <https://doi.org/10.22441/fifo.2024.v16i2.006>
- Felicia Watratan, A., Puspita, A. B., Moeis, D., Informasi, S., & Profesional Makassar, S. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. In *Journal Of Aoolied Computer Science And Technology (JACOST)* (Vol. 1, Issue 1). <http://journal.isas.or.id/index.php/JACOST>
- Hadi, A., Gizi, J., Kesehatan Kemenkes Aceh, P., & Soekarno-Hatta, J. (2017). Faktor Risiko Terjadinya Penyakit Jantung Koroner Pada Pasien Rumah Sakit Umum Meuraxa Banda Aceh (Risk factors of coronary heart disease in Meuraxa hospital of Banda Aceh). In *AcTion Journal* (Vol. 2, Issue 1).
- Haganta Depari, D., Widiastiwi, Y., Mega Santoni, M., Ilmu Komputer, F., Pembangunan Nasional Veteran Jakarta, U., Fatmawati Raya, J. R., & Labu, P. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *JURNAL INFORMATIK Edisi Ke, 18*, 2022.
- Matzavela, V., & Alepis, E. (2021). Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, 2. <https://doi.org/10.1016/j.caeai.2021.100035>

- Mulyani, A., Kurniadi, D., Nashrulloh, M. R., Julianto, I. T., & Regita, M. (2022). THE Prediction Of Ppa And Kip-Kuliah Scholarship Recipients Using *Naïve Bayes Algorithm*. *Jurnal Teknik Informatika (Jutif)*, 3(4), 821–827. <https://doi.org/10.20884/1.jutif.2022.3.4.297>
- Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKoSIN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Rosyadi, R., Wibowo, A., & Teknologi Machine Learning Dalam Perencanaan, P. (2025). *under a Creative Commons Attribution-ShareAlike 4.0 International License[CC BY SA] Pemanfaatan Teknologi Machine Learning Dalam Perencanaan Dan Rehabilitasi Rumah Layak Huni: Perspektif Arsitektur*.
- Setiawan, A., Febrio Waleska, R., Adji Purnama, M., & Efrizoni, L. (2024). Sidomulyo barat, Sidomulyo Bar. In *Jurnal Informatika & Rekayasa Elektronika* (Vol. 7, Issue 1). <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- Shiri Harzevili, N., & Alizadeh, S. H. (2018). Mixture of latent multinomial naive Bayes classifier. *Applied Soft Computing Journal*, 69, 516–527. <https://doi.org/10.1016/j.asoc.2018.04.020>
- Sodikin, M. I. (2023). Penerapan dan Manfaat Machine Learning di Rumah Sakit. *Multiverse: Open Multidisciplinary Journal*, 2(2), 262–265. <https://doi.org/10.57251/multiverse.v2i2.1207>
- Yulianto, E., & Murdianto, T. (2024). *Peran Artificial Intelligence (AI) dalam Manajemen Arsip dan Dokumen* (Vol. 1, Issue 6). Elektronik.
- Zahrawardani, D., Sri Herlambang, K., & Anggraheny, H. D. (2013). Analisis Faktor Risiko Kejadian Penyakit Jantung Koroner di RSUP Dr Kariadi Semarang. In *Jurnal Kedokteran Muhammadiyah* (Vol. 1).

LAMPIRAN

Lampiran 1 Kode Program

```
# Import library
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Baca dataset
df = pd.read_csv('Heart Attack.csv')

# 2. Pisahkan fitur dan target
X = df.drop(columns='class')
y = df['class']

# 3. Normalisasi dengan Min-Max
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

# 4. Pembagian data (85:15)
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.15, random_state=42, stratify=y
)

# 5. Decision Tree
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

# 6. Naive Bayes
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)

# 7. Evaluasi Decision Tree
print("=== Decision Tree ===")
print("Akurasi:", accuracy_score(y_test, y_pred_dt))
print("Classification Report:\n",
      classification_report(y_test, y_pred_dt))
print("Confusion Matrix:\n",
      confusion_matrix(y_test,
                       y_pred_dt))

# 8. Evaluasi Naive Bayes
```

```

print("\n=== Naive Bayes ===")
print("Akurasi:", accuracy_score(y_test, y_pred_nb))
print("Classification Report:\n",
      classification_report(y_test, y_pred_nb))
print("Confusion Matrix:\n",      confusion_matrix(y_test,
y_pred_nb))

# 9. Visualisasi Decision Tree
plt.figure(figsize=(20, 10))
plot_tree(dt_model,
          filled=True,
          feature_names=X.columns,
          class_names=["Tidak", "Ya"],
          rounded=True,
          fontsize=10)
plt.title("Visualisasi Decision Tree (85:15)")
plt.show()

# 10. Visualisasi Naive Bayes - Confusion Matrix
plt.figure(figsize=(7, 5))
cm_nb = confusion_matrix(y_test, y_pred_nb)
sns.heatmap(cm_nb, annot=True, fmt="d", cmap="Blues",
            xticklabels=["Tidak", "Ya"], yticklabels=["Tidak", "Ya"])
plt.title("Confusion Matrix - Naive Bayes")
plt.xlabel("Prediksi")
plt.ylabel("Aktual")
plt.show()

# Import library
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Baca dataset
df = pd.read_csv('Heart Attack.csv') # Ganti jika nama file
berbeda

# 2. Pisahkan fitur dan target
X = df.drop(columns='class')
y = df['class']

# 3. Normalisasi dengan Min-Max
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

```



```

# 4. Pembagian data (75% latih, 25% uji)
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.25, random_state=42, stratify=y
)

# 5. Decision Tree
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

# 6. Naive Bayes
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)

# 7. Evaluasi Decision Tree
print("=== Decision Tree ===")
print("Akurasi:", accuracy_score(y_test, y_pred_dt))
print("Classification Report:\n",
      classification_report(y_test, y_pred_dt))
print("Confusion Matrix:\n",
      confusion_matrix(y_test, y_pred_dt))

# 8. Evaluasi Naive Bayes
print("\n=== Naive Bayes ===")
print("Akurasi:", accuracy_score(y_test, y_pred_nb))
print("Classification Report:\n",
      classification_report(y_test, y_pred_nb))
print("Confusion Matrix:\n",
      confusion_matrix(y_test, y_pred_nb))

# 9. Visualisasi Decision Tree
plt.figure(figsize=(20, 10))
plot_tree(dt_model,
          filled=True,
          feature_names=X.columns,
          class_names=["Tidak", "Ya"],
          rounded=True,
          fontsize=10)
plt.title("Visualisasi Decision Tree (75:25)")
plt.show()

# 10. Visualisasi Naive Bayes - Confusion Matrix
plt.figure(figsize=(7, 5))
cm_nb = confusion_matrix(y_test, y_pred_nb)
sns.heatmap(cm_nb, annot=True, fmt="d", cmap="Blues",
            xticklabels=["Tidak", "Ya"], yticklabels=["Tidak", "Ya"])
plt.title("Confusion Matrix - Naive Bayes (75:25)")
plt.xlabel("Prediksi")
plt.ylabel("Aktual")
plt.show()

```



```

# Import library
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Baca dataset
df = pd.read_csv('Heart Attack.csv') # Ganti jika nama file
berbeda

# 2. Pisahkan fitur dan target
X = df.drop(columns='class')
y = df['class']

# 3. Normalisasi dengan Min-Max
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

# 4. Pembagian data (60% latih, 40% uji)
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.40, random_state=42, stratify=y
)

# 5. Decision Tree
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

# 6. Naive Bayes
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)

# 7. Evaluasi Decision Tree
print("=== Decision Tree ===")
print("Akurasi:", accuracy_score(y_test, y_pred_dt))
print("Classification Report:\n",
      classification_report(y_test, y_pred_dt))
print("Confusion Matrix:\n",
      confusion_matrix(y_test,
                       y_pred_dt))

# 8. Evaluasi Naive Bayes
print("\n=== Naive Bayes ===")
print("Akurasi:", accuracy_score(y_test, y_pred_nb))

```

```

print("Classification Report:\n",
      classification_report(y_test, y_pred_nb))
print("Confusion Matrix:\n",      confusion_matrix(y_test,
y_pred_nb))

# 9. Visualisasi Decision Tree
plt.figure(figsize=(20, 10))
plot_tree(dt_model,
          filled=True,
          feature_names=X.columns,
          class_names=["Tidak", "Ya"],
          rounded=True,
          fontsize=10)
plt.title("Visualisasi Decision Tree (60:40)")
plt.show()

# 10. Visualisasi Naive Bayes - Confusion Matrix
plt.figure(figsize=(7, 5))
cm_nb = confusion_matrix(y_test, y_pred_nb)
sns.heatmap(cm_nb, annot=True, fmt="d", cmap="Blues",
            xticklabels=["Tidak", "Ya"], yticklabels=["Tidak", "Ya"])
plt.title("Confusion Matrix - Naive Bayes (60:40)")
plt.xlabel("Prediksi")
plt.ylabel("Aktual")
plt.show()

```