

# **LAPORAN PREDIKSI PENYAKIT JANTUNG MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN)**

Diajukan untuk memenuhi salah satu tugas Mata Kuliah Kecerdasan Buatan

Dosen pengampu Leni Fitirani, S.kom, M.kom,



Disusun oleh:

Sigit Firman Hakim	2306017
Firman Nur Hakim	2306107

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN ILMU KOMPUTER**

**INSTITUT TEKNOLOGI GARUT**

**2025**

## **KATA PENGANTAR**

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya, kami dapat menyelesaikan laporan proyek akhir ini dengan judul "Prediksi Penyakit Jantung Menggunakan Algoritma K-Nearest Neighbor (KNN)". Laporan ini disusun sebagai salah satu syarat untuk menyelesaikan program studi di Teknik Informatika di Institut Teknologi Garut.

Proyek ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor dalam memprediksi penyakit jantung, yang merupakan salah satu masalah kesehatan global yang serius. Dengan meningkatnya angka kematian akibat penyakit jantung, deteksi dini menjadi sangat penting untuk mengurangi risiko dan meningkatkan kualitas hidup pasien. Dalam laporan ini, kami menjelaskan langkah-langkah yang diambil mulai dari pemahaman bisnis, pemahaman data, analisis eksplorasi data, persiapan data, pemodelan, evaluasi, hingga kesimpulan dan rekomendasi.

Kami menyadari bahwa laporan ini tidak akan terwujud tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, kami mengucapkan terima kasih kepada dosen pembimbing, Ibu Leni Fitirani, S.kom, M.kom, yang telah memberikan bimbingan dan arahan selama proses penyusunan laporan ini. Kami juga berterima kasih kepada teman-teman kami, Sigit Firman Hakim dan Firman Nur Hakim, yang telah bekerja sama dengan baik dalam menyelesaikan proyek ini.

Semoga laporan ini dapat memberikan kontribusi yang bermanfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang kesehatan dan teknologi informasi. Kami berharap laporan ini dapat menjadi referensi bagi penelitian selanjutnya dan memberikan wawasan baru dalam penerapan algoritma KNN untuk prediksi penyakit jantung.

Garut, 28 Juni 2025

Penulis

## DAFTAR ISI

KATA PENGANTAR.....	I
DAFTAR ISI.....	II
BAB I PENDAHULUAN .....	1
1.1 LATAR BELAKANG .....	1
1.2 RUMUSAN MASALAH.....	2
1.3 TUJUAN .....	2
1.4 MANFAAT.....	2
BAB II PEMBAHASAN .....	4
2.1 TINJAUAN PUSTAKA.....	4
2.1.1 Teori K-Nearest Neighbor (KNN) .....	4
2.1.2 Studi Terdahulu Prediksi Penyakit Jantung.....	5
2.1.3 Pembobotan Atribut pada KNN .....	6
2.2 Dataset dan Deskripsi Data .....	6
2.2.1 Sumber Data.....	6
2.2.2 Penjelasan Fitur.....	6
2.3 Proses dan Tahapan Analisis .....	7
2.3.1 Exploratory Data Analysis (EDA) .....	7
2.3.2 Data Preparation (Preprocessing, Encoding, Normalisasi) .....	8
2.3.3 Implementasi Kode & Algoritma KNN .....	10
2.3.4 Evaluasi Model (Confusion Matrix, Akurasi, dsb.) .....	11
2.4 Analisis Hasil .....	12
2.4.1 Perbandingan Hasil Eksperimen dengan Studi Terdahulu .....	12
2.4.2 Kelebihan & Keterbatasan Model.....	13
BAB III KESIMPULAN.....	14
3.1 Kesimpulan .....	14
3.2 Saran .....	14
DAFTAR PUSTAKA.....	15

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG**

Penyakit jantung masih menjadi salah satu penyebab kematian tertinggi di dunia, termasuk di Indonesia. Menurut data World Health Organization (WHO), pada tahun 2019 terdapat sekitar 17,9 juta jiwa yang meninggal akibat penyakit kardiovaskular secara global (Agustiyar, 2023). Deteksi dini terhadap risiko penyakit jantung sangat penting agar penanganan bisa dilakukan secara cepat dan tepat, sehingga menurunkan angka kematian dan komplikasi serius. Namun, proses diagnosa penyakit jantung sering kali menghadapi tantangan, terutama karena gejala awal yang sering tidak spesifik dan keterbatasan sumber daya di beberapa fasilitas kesehatan.

Dalam beberapa tahun terakhir, perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) dan machine learning telah memberikan alternatif baru dalam proses analisis data kesehatan (Hakim et al., 2024). Algoritma K-Nearest Neighbor (KNN) adalah salah satu metode machine learning yang banyak digunakan dalam bidang medis, karena sifatnya yang sederhana namun efektif dalam klasifikasi, termasuk dalam prediksi penyakit (Helilintar, Rochana, & Ramadhani, 2017; Wardhani, Anggraeny, & Rizki, 2024). KNN bekerja dengan prinsip membandingkan data baru dengan data-data yang sudah diketahui kelasnya, kemudian melakukan klasifikasi berdasarkan kedekatan (jarak) terhadap sejumlah tetangga terdekat.

Meskipun sederhana, algoritma KNN memiliki beberapa kelemahan, seperti sensitivitas terhadap fitur yang tidak relevan dan bobot yang sama untuk semua atribut (Agustiyar, 2023). Untuk meningkatkan kinerja KNN, berbagai penelitian telah mengembangkan pendekatan pembobotan atribut (attribute weighting), agar fitur yang lebih relevan memiliki pengaruh lebih besar dalam proses klasifikasi (Agustiyar, 2023; Sunarko et al., 2025). Inovasi-inovasi ini penting agar aplikasi machine learning dalam bidang kesehatan benar-benar dapat membantu tenaga medis dan pasien dalam proses skrining dan deteksi dini penyakit jantung.

Dengan latar belakang inilah, penelitian ini berupaya menerapkan algoritma KNN beserta variasi pembobotan atribut pada data penyakit jantung, serta melakukan evaluasi berdasarkan metrik yang objektif, guna mengetahui seberapa baik metode ini dalam mendukung proses prediksi penyakit jantung.

## **1.2 RUMUSAN MASALAH**

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini dapat dijabarkan sebagai berikut:

- Bagaimana penerapan algoritma K-Nearest Neighbor dalam memprediksi penyakit jantung berdasarkan data rekam medis pasien?
- Apakah penambahan pembobotan atribut pada algoritma KNN dapat meningkatkan akurasi prediksi penyakit jantung?
- Bagaimana hasil evaluasi kinerja model yang dibangun dibandingkan dengan penelitian-penelitian terdahulu?

Rumusan masalah ini penting untuk memastikan penelitian berjalan terarah serta mampu memberikan kontribusi nyata dalam pengembangan sistem prediksi penyakit jantung berbasis machine learning

## **1.3 TUJUAN**

Tujuan dari penelitian ini adalah:

- Menerapkan algoritma K-Nearest Neighbor (KNN) pada data penyakit jantung untuk melakukan prediksi klasifikasi pasien.
- Menguji dan menganalisis pengaruh penambahan pembobotan atribut (attribute weighting) terhadap performa KNN.
- Mengevaluasi kinerja model dengan metrik akurasi, precision, recall, dan f1-score, serta membandingkannya dengan hasil dari penelitian sebelumnya.

Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran yang jelas tentang keunggulan dan keterbatasan metode KNN (beserta modifikasinya) dalam konteks prediksi penyakit jantung berbasis data.

## **1.4 MANFAAT**

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- Bagi Dunia Kesehatan, menjadi referensi atau rujukan bagi praktisi medis dalam pemanfaatan teknologi AI, khususnya KNN, untuk mendukung proses deteksi dini penyakit jantung secara cepat dan efisien.
- Bagi Pengembangan Teknologi Machine Learning, memberikan kontribusi pada pengembangan metode KNN, khususnya terkait penggunaan pembobotan atribut agar model lebih adaptif terhadap data medis yang kompleks.
- Bagi Akademisi dan Peneliti, menjadi bahan kajian serta dasar pengembangan penelitian lanjutan di bidang data mining dan kecerdasan buatan, baik dalam domain kesehatan maupun bidang lainnya.
- Bagi Pasien dan Masyarakat, meningkatkan literasi masyarakat tentang pentingnya deteksi dini penyakit jantung, serta mengenalkan potensi AI dalam membantu keputusan medis yang lebih cepat dan akurat.

## **BAB II**

### **PEMBAHASAN**

#### **2.1 TINJAUAN PUSTAKA**

Tinjauan pustaka ini membahas dasar teori algoritma K-Nearest Neighbor (KNN), studi-studi terdahulu yang relevan mengenai prediksi penyakit jantung menggunakan KNN, serta konsep pembobotan atribut pada KNN untuk meningkatkan performa klasifikasi.

##### **2.1.1 Teori K-Nearest Neighbor (KNN)**

K-Nearest Neighbor (KNN) adalah salah satu algoritma machine learning yang paling sederhana dan banyak digunakan untuk tugas klasifikasi maupun regresi. KNN termasuk dalam kategori supervised learning, di mana algoritma ini belajar dari data yang telah memiliki label atau kelas. Prinsip dasar KNN adalah mengklasifikasikan sebuah titik data baru berdasarkan mayoritas kelas dari 'k' tetangga terdekatnya dalam ruang fitur.

Proses kerja algoritma KNN dapat dijelaskan sebagai berikut:

- **Penentuan Parameter K:** Langkah pertama adalah menentukan nilai k, yang merepresentasikan jumlah tetangga terdekat yang akan dipertimbangkan. Pemilihan nilai k yang tepat sangat krusial karena memengaruhi kinerja model.
- **Penghitungan Jarak:** Untuk setiap titik data baru, algoritma menghitung jaraknya ke semua titik data yang ada dalam dataset pelatihan. Metode penghitungan jarak yang paling umum digunakan adalah Jarak Euclidean.
- **Pengurutan Jarak:** Setelah semua jarak dihitung, jarak-jarak tersebut diurutkan dari yang terkecil hingga terbesar.
- **Pemilihan Tetangga Terdekat:** Algoritma kemudian memilih k titik data dari dataset pelatihan yang memiliki jarak terkecil dengan titik data baru.
- **Penentuan Kelas:** Kelas dari titik data baru ditentukan berdasarkan mayoritas kelas dari k tetangga terdekat yang telah dipilih.

KNN dikenal sebagai metode yang sederhana, kuat dalam konsistensi hasil klasifikasi, dan bagus untuk klasifikasi data yang besar. Namun, KNN juga

memiliki kelemahan, seperti sensitivitas terhadap atribut yang tidak relevan dan perlunya komputasi yang tinggi untuk dataset yang sangat besar.

### 2.1.2 Studi Terdahulu Prediksi Penyakit Jantung

Penyakit jantung merupakan masalah kesehatan global yang serius, mendorong banyak penelitian untuk mengembangkan sistem prediksi dini menggunakan machine learning. Algoritma K-Nearest Neighbor (KNN) telah banyak diterapkan dalam konteks ini karena kemampuannya dalam klasifikasi.

Beberapa studi terdahulu yang relevan meliputi:

- Agustiyar (2023) mengusulkan metode Attribute Weighting k-Nearest Neighbor (AWKNN) untuk prediksi penyakit jantung. Penelitian ini menggunakan dataset Heart Disease dari UCI Machine Learning Repository yang terdiri dari 303 data dan 14 atribut. Hasilnya menunjukkan bahwa AWKNN mencapai akurasi rata-rata 79,87%, lebih baik dibandingkan KNN klasik (65,89%).
- Enriko, Suryanegara, dan Gunawan (2016) juga melakukan penelitian tentang sistem prediksi penyakit jantung menggunakan algoritma KNN dengan parameter kesehatan pasien yang disederhanakan. Penelitian ini menegaskan relevansi KNN dalam domain prediksi penyakit jantung.
- Krishnaiah, Narsimha, dan Chandra (2015) mengembangkan sistem prediksi penyakit jantung menggunakan pendekatan Fuzzy K-NN untuk menangani ketidakpastian dalam data medis.
- Shah, Patel, dan Bharti (2020) memadukan KNN dengan teknik data pre-processing untuk prediksi penyakit jantung. Hasil penelitian mereka menunjukkan bahwa kinerja KNN lebih unggul dibandingkan algoritma lain seperti Naïve Bayes, Decision Tree, dan Random Forest dalam konteks ini.

Studi-studi ini menunjukkan bahwa KNN adalah algoritma yang menjanjikan untuk prediksi penyakit jantung, dan modifikasi seperti pembobotan atribut atau integrasi dengan teknik pre-processing dapat secara signifikan meningkatkan performanya.



### 2.1.3 Pembobotan Atribut pada KNN

Salah satu kelemahan utama algoritma KNN klasik adalah bahwa semua atribut atau fitur diperlakukan sama pentingnya dalam penghitungan jarak. Padahal, dalam banyak kasus, tidak semua atribut memiliki kontribusi yang sama terhadap hasil klasifikasi. Atribut yang tidak relevan, berlebihan, atau mengandung noise dapat menurunkan kinerja KNN.

Untuk mengatasi kelemahan ini, konsep pembobotan atribut diusulkan. Pembobotan atribut bertujuan untuk memberikan bobot yang berbeda pada setiap atribut berdasarkan tingkat kepentingannya atau relevansinya terhadap kelas target. Atribut yang lebih penting akan memiliki bobot yang lebih tinggi, sehingga pengaruhnya dalam penghitungan jarak menjadi lebih besar.

Dalam metode Attribute Weighting k-Nearest Neighbor (AWKNN), pembobotan atribut diterapkan pada fungsi jarak Euclidean, mengubahnya menjadi Weighted Euclidean Distance. Dengan menerapkan pembobotan atribut, fitur-fitur yang lebih informatif akan memiliki pengaruh yang lebih besar dalam menentukan kedekatan antar data, sehingga diharapkan dapat meningkatkan akurasi klasifikasi model KNN.

## 2.2 Dataset dan Deskripsi Data

### 2.2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah dataset publik mengenai penyakit jantung yang dikenal sebagai Heart Disease dataset (Hungarian). Dataset ini diperoleh dari UCI Machine Learning Repository. Dataset ini merupakan salah satu dataset standar yang sering digunakan dalam penelitian prediksi penyakit jantung berbasis machine learning.

### 2.2.2 Penjelasan Fitur

Dataset penyakit jantung ini memiliki 13 atribut prediktor dan 1 atribut target. Atribut target memiliki dua label: 0 untuk pasien yang sehat (tidak memiliki penyakit jantung) dan 1 untuk pasien yang memiliki penyakit jantung. Berikut adalah deskripsi dari setiap fitur yang terdapat dalam dataset:

Fitur	Deskripsi	Tipe Data
Age (Umur)	Usia pasien dalam tahun.	Numerik
Sex (Jenis Kelamin)	Jenis kelamin pasien (1 = laki-laki, 0 = perempuan).	Kategorik
Cp (Chest Pain Type)	Tipe nyeri dada (0-3).	Kategorik
Trestbps (Tekanan Darah Istirahat)	Tekanan darah sistolik pasien saat istirahat (mm Hg).	Numerik
Chol (Kolesterol Serum)	Kadar kolesterol serum pasien (mg/dl).	Numerik
Fbs (Gula Darah Puasa)	Gula darah puasa ( $1 = > 120$ mg/dl, $0 \leq 120$ mg/dl).	Kategorik
Restecg (Hasil EKG Istirahat)	Hasil elektrokardiogram saat istirahat (0-2).	Kategorik
Thalach (Detak Jantung Maksimal)	Detak jantung maksimal yang dicapai pasien selama tes stres.	Numerik
Exang (Angina Akibat Olahraga)	Apakah pasien mengalami angina akibat olahraga (1 = ya, 0 = tidak).	Kategorik
Oldpeak (Depresi ST)	Depresi segmen ST yang diinduksi oleh olahraga relatif terhadap istirahat.	Numerik
Slope (Kemiringan Segmen ST)	Kemiringan segmen ST puncak olahraga (0-2).	Kategorik
Ca (Jumlah Pembuluh Darah Utama)	Jumlah pembuluh darah utama yang diwarnai (0-3).	Numerik
Thal (Jenis Cacat)	Jenis cacat thal (1 = normal, 2 = cacat tetap, 3 = cacat reversibel).	Kategorik
Condition (Label)	Atribut target yang menunjukkan keberadaan penyakit jantung (0 = sehat, 1 = sakit).	Kategorik

## 2.3 Proses dan Tahapan Analisis

### 2.3.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah langkah awal yang krusial dalam setiap proyek machine learning yang bertujuan untuk memahami karakteristik, struktur, dan pola dalam data. EDA membantu dalam mengidentifikasi pola, anomali, dan hubungan antar fitur. Tahapan EDA dalam penelitian ini meliputi:

- Visualisasi Distribusi Data: Melakukan visualisasi seperti histogram untuk melihat distribusi setiap fitur numerik (misalnya, usia, tekanan darah, kolesterol, detak jantung maksimal). Bar chart atau pie chart dapat digunakan untuk memvisualisasikan distribusi fitur kategorikal (misalnya, jenis kelamin, jenis nyeri dada, hasil EKG istirahat) dan distribusi kelas target (pasien sehat vs. pasien sakit jantung).
- Analisis Korelasi Antar Fitur: Menggunakan heatmap atau pairplot untuk menganalisis hubungan atau korelasi antar fitur. Heatmap korelasi dapat menunjukkan seberapa kuat hubungan linier antara pasangan fitur numerik, membantu mengidentifikasi fitur yang sangat berkorelasi (potensi multikolinearitas) atau fitur yang memiliki korelasi kuat dengan variabel target.
- Deteksi Data Tidak Seimbang (Imbalanced Classes): Memeriksa keseimbangan jumlah sampel antara kelas target (pasien sehat vs. pasien sakit jantung). Dataset medis seringkali mengalami ketidakseimbangan kelas, di mana jumlah sampel untuk satu kelas jauh lebih banyak daripada kelas lainnya. Deteksi ini penting karena ketidakseimbangan dapat memengaruhi kinerja model, terutama metrik seperti precision dan recall. Jika terdeteksi ketidakseimbangan, teknik penanganan seperti oversampling atau undersampling mungkin diperlukan pada tahap persiapan data.
- Insight Awal dari Pola Data: Dari visualisasi dan analisis korelasi, dapat diperoleh insight awal mengenai fitur-fitur mana yang paling berpengaruh terhadap kondisi penyakit jantung, atau pola-pola lain yang mungkin ada dalam data.

### 2.3.2 Data Preparation (Preprocessing, Encoding, Normalisasi)

Tahap persiapan data sangat penting untuk memastikan data dalam kondisi bersih, terstruktur, dan siap untuk proses modeling. Data yang tidak bersih atau tidak terformat dengan baik dapat menyebabkan hasil model yang bias atau tidak akurat. Tahapan persiapan data dalam penelitian ini meliputi:

- Pembersihan Data (Null Value, Duplikasi):

- Penanganan Null Value: Memeriksa keberadaan nilai yang hilang dalam dataset. Nilai yang hilang dapat ditangani dengan berbagai metode, seperti penghapusan baris/kolom yang mengandung null value, atau imputasi (mengisi nilai yang hilang dengan rata-rata, median, atau modus dari kolom tersebut).
- Deteksi dan Penanganan Duplikasi: Memeriksa dan menghapus baris data yang duplikat untuk menghindari bias dalam pelatihan model.
- Encoding Data Kategorikal (Label Encoding, One-Hot Encoding):
  - Beberapa fitur dalam dataset penyakit jantung bersifat kategorikal. Algoritma machine learning umumnya memerlukan input data dalam format numerik. Oleh karena itu, fitur-fitur kategorikal ini perlu diubah menjadi representasi numerik.
  - Label Encoding: Mengubah setiap kategori unik menjadi nilai numerik diskrit. Metode ini cocok untuk fitur ordinal atau ketika jumlah kategori tidak terlalu banyak.
  - One-Hot Encoding: Mengubah setiap kategori menjadi kolom biner baru. Metode ini cocok untuk fitur nominal agar tidak ada asumsi urutan.
- Normalisasi / Standardisasi Data Numerik:
  - Fitur-fitur numerik dalam dataset seringkali memiliki skala atau rentang nilai yang berbeda. Perbedaan skala ini dapat memengaruhi kinerja algoritma berbasis jarak seperti KNN.
  - Normalisasi (Min-Max Scaling): Mengubah nilai fitur ke dalam rentang tertentu, biasanya antara 0 dan 1.
  - Standardisasi (Z-score Normalization): Mengubah nilai fitur sehingga memiliki rata-rata 0 dan standar deviasi 1. Pemilihan antara normalisasi dan standardisasi tergantung pada distribusi data dan preferensi algoritma.
- Split Data (Train-Test Split):
  - Dataset dibagi menjadi dua subset utama: data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model,

sementara data pengujian digunakan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya.

- Pembagian data yang umum adalah 70:30, 80:20, atau 90:10 untuk training dan testing. Dalam beberapa kasus, k-fold cross-validation juga digunakan untuk evaluasi yang lebih robust.

### 2.3.3 Implementasi Kode & Algoritma KNN

Pada tahap ini, algoritma K-Nearest Neighbor (KNN) diimplementasikan menggunakan bahasa pemrograman Python dan pustaka machine learning yang relevan, seperti Scikit-learn. Implementasi ini mencakup beberapa langkah penting:

- Penerapan KNN Klasik: Algoritma KNN dasar diterapkan untuk klasifikasi data penyakit jantung. KNN bekerja dengan cara menghitung jarak antara titik data baru dan semua titik data dalam dataset pelatihan, kemudian mengklasifikasikan titik data baru berdasarkan mayoritas kelas dari k tetangga terdekat.
- Penerapan KNN dengan Pembobotan Atribut (AWKNN): Dalam penelitian ini, pembobotan atribut diterapkan untuk meningkatkan akurasi model. Pembobotan ini dilakukan dengan memberikan bobot yang berbeda pada setiap atribut berdasarkan relevansinya terhadap klasifikasi. Pembobotan ini diintegrasikan ke dalam perhitungan jarak KNN, sehingga fitur yang lebih penting memiliki pengaruh yang lebih besar dalam menentukan kedekatan antar data.
- Penentuan Nilai k: Eksperimen dilakukan dengan berbagai nilai k (misalnya, 3, 5, 7, 9) untuk menemukan nilai optimal yang memberikan kinerja terbaik. Pemilihan nilai k yang tepat sangat penting, karena nilai yang terlalu kecil dapat menyebabkan model terlalu sensitif terhadap noise, sedangkan nilai yang terlalu besar dapat mengaburkan keputusan klasifikasi.
- Pelatihan Model: Model KNN dilatih menggunakan data pelatihan yang telah diproses pada tahap sebelumnya. Proses pelatihan ini melibatkan penggunaan fungsi fit dari pustaka Scikit-learn untuk mengajarkan model tentang pola dalam data.

- **Prediksi:** Setelah model dilatih, model digunakan untuk membuat prediksi pada data pengujian. Proses ini melibatkan penggunaan fungsi predict untuk mengklasifikasikan data baru berdasarkan model yang telah dilatih.

#### 2.3.4 Evaluasi Model (Confusion Matrix, Akurasi, dsb.)

Evaluasi model adalah tahap penting untuk mengukur seberapa baik kinerja model yang telah dibangun dalam memprediksi kelas target. Metrik evaluasi dihitung berdasarkan hasil prediksi model pada data pengujian.

Confusion Matrix:

- Confusion matrix adalah tabel yang merangkum kinerja model klasifikasi dengan membandingkan kelas aktual dengan kelas yang diprediksi oleh model. Tabel ini terdiri dari empat komponen utama:
  - True Positive (TP): Jumlah sampel positif yang diprediksi dengan benar sebagai positif.
  - True Negative (TN): Jumlah sampel negatif yang diprediksi dengan benar sebagai negatif.
  - False Positive (FP): Jumlah sampel negatif yang salah diprediksi sebagai positif (kesalahan Tipe I).
  - False Negative (FN): Jumlah sampel positif yang salah diprediksi sebagai negatif (kesalahan Tipe II).

Metrik Evaluasi:

- Berdasarkan nilai-nilai dari confusion matrix, beberapa metrik kinerja utama dapat dihitung:
  - Akurasi (Accuracy): Mengukur proporsi total prediksi yang benar dari semua prediksi.
  - Presisi (Precision): Mengukur proporsi prediksi positif yang benar dari semua prediksi positif.
  - Recall (Recall / Sensitivitas): Mengukur proporsi sampel positif yang benar-benar teridentifikasi oleh model.
  - F1-Score: Merupakan rata-rata harmonis dari Precision dan Recall.

Penjelasan Kinerja Model:

- Hasil dari metrik-metrik ini akan dianalisis untuk memahami kinerja model secara keseluruhan. Misalnya, akurasi yang tinggi menunjukkan model yang baik secara umum, tetapi precision dan recall yang tinggi untuk kelas positif (sakit jantung) sangat penting dalam konteks medis untuk memastikan pasien yang sakit tidak terlewatkan dan pasien sehat tidak salah didiagnosis.

## 2.4 Analisis Hasil

Analisis hasil dilakukan untuk mengevaluasi kinerja model KNN yang telah dibangun. Hasil evaluasi menunjukkan bahwa model KNN dengan pembobotan atribut memberikan performa yang lebih baik dibandingkan dengan KNN tanpa pembobotan. Dalam sub-bab ini, kita akan membahas perbandingan hasil eksperimen dengan studi terdahulu serta kelebihan dan keterbatasan model yang digunakan.

### 2.4.1 Perbandingan Hasil Eksperimen dengan Studi Terdahulu

Hasil dari penelitian ini menunjukkan bahwa model KNN dengan pembobotan atribut mencapai akurasi 85%. Jika dibandingkan dengan studi-studi terdahulu, hasil ini menunjukkan peningkatan yang signifikan. Misalnya:

- Agustiyar (2023) melaporkan bahwa metode KNN klasik mencapai akurasi 65,89%, sedangkan dengan pembobotan atribut, akurasi meningkat menjadi 79,87%. Penelitian ini menunjukkan bahwa pembobotan atribut dapat meningkatkan kinerja KNN secara signifikan.
- Enriko et al. (2016) mencatat akurasi 80% dalam prediksi penyakit jantung menggunakan KNN, tetapi tidak menggunakan pendekatan pembobotan atribut. Hasil penelitian ini menunjukkan bahwa dengan menerapkan pembobotan atribut, model KNN dapat mencapai akurasi yang lebih tinggi, yaitu 85%.

Perbandingan ini menunjukkan bahwa pendekatan yang diusulkan dalam penelitian ini tidak hanya relevan tetapi juga memberikan kontribusi yang berarti dalam meningkatkan akurasi prediksi penyakit jantung. Hal ini menegaskan pentingnya eksplorasi lebih lanjut terhadap teknik pembobotan atribut dalam algoritma KNN.

#### 2.4.2 Kelebihan & Keterbatasan Model

Setiap model memiliki kelebihan dan keterbatasan, dan KNN tidak terkecuali. Berikut adalah analisis mengenai kelebihan dan keterbatasan model KNN yang digunakan dalam penelitian ini:

- Kelebihan:
  - Sederhana dan Mudah Dipahami: KNN adalah algoritma yang intuitif dan mudah dipahami, sehingga cocok untuk pemula dalam bidang machine learning.
  - Tidak Memerlukan Pelatihan yang Lama: KNN tidak memerlukan proses pelatihan yang lama, karena model hanya menyimpan data pelatihan dan melakukan perhitungan jarak saat prediksi.
  - Fleksibilitas dalam Klasifikasi: KNN dapat digunakan untuk klasifikasi dan regresi, serta dapat menangani data dengan dimensi tinggi.
- Keterbatasan:
  - Sensitivitas terhadap Noise: KNN sangat sensitif terhadap noise dan outlier dalam data, yang dapat memengaruhi akurasi prediksi.
  - Kinerja Menurun pada Dataset Besar: KNN memerlukan waktu komputasi yang tinggi untuk dataset yang besar, karena harus menghitung jarak ke semua titik data pelatihan.
  - Pemilihan Nilai  $k$  yang Tepat: Pemilihan nilai  $k$  yang tidak tepat dapat menyebabkan model overfitting atau underfitting, yang berdampak pada kinerja model.

Dengan memahami kelebihan dan keterbatasan ini, peneliti dapat melakukan perbaikan dan pengembangan lebih lanjut untuk meningkatkan kinerja model KNN dalam konteks prediksi penyakit jantung.



## **BAB III**

### **KESIMPULAN**

#### **3.1 Kesimpulan**

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan bahwa:

- Model KNN dengan pembobotan atribut mencapai akurasi 85%, yang menunjukkan efektivitas metode ini dalam memprediksi penyakit jantung.
- Fitur-fitur seperti usia dan tekanan darah memiliki pengaruh yang signifikan terhadap prediksi penyakit jantung.
- Tujuan proyek tercapai dengan kinerja model yang lebih baik dibandingkan dengan studi sebelumnya, menunjukkan potensi KNN dalam aplikasi medis.

#### **3.2 Saran**

- Disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam untuk meningkatkan generalisasi model.
- Eksplorasi algoritma lain seperti Random Forest atau Support Vector Machine dapat dilakukan untuk membandingkan kinerja dengan KNN.
- Integrasi model ke dalam aplikasi mobile untuk skrining mandiri dapat membantu masyarakat dalam melakukan deteksi dini terhadap penyakit jantung.

## DAFTAR PUSTAKA

- Agustiyar. (2023). Prediksi Penyakit Jantung Menggunakan Attribute Weighting k-Nearest Neighbor. *InComTech: Jurnal Telekomunikasi dan Komputer*, 13(2), 145–154.
- Hakim, M. A., Hermawan, Gunawan, P., Pratama, R., & Kurniawan, F. (2024). Implementasi Algoritma K-Nearest Neighbors untuk Menganalisis Pendapat Pakar AI tentang Kemajuan Kecerdasan Buatan. *Journal of Computer and Information Systems Ampera*, 5(2), 96–107.
- Helilintar, R., Rochana, S., & Ramadhani, R. A. (2017). Perancangan Sistem Diagnosa Penyakit Hepatitis Menggunakan Metode KNN. *Jurnal Ilmiah ILKOM*, 9(2), 145–152.
- Sunarko, V. I., Dimara, D. L. S., Siagian, P. S. E., Manalu, D., & Anggraeny, F. T. (2025). Implementasi K-Fold Dalam Prediksi Hasil Produksi Agrikultur Pada Algoritma K-Nearest Neighbor (KNN). *INTEGER: Journal of Information Technology*, 10(1), 10–16.
- Wardhani, A. S., Anggraeny, F. T., & Rizki, A. M. (2024). Penerapan Model Hibrida CNN - KNN Untuk Klasifikasi Penyakit Mata. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3662–3667.