

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Jantung merupakan organ penting dalam tubuh manusia yang berperan dalam memompa darah melalui pembuluh darah dalam sistem peredaran darah. Darah ini membawa oksigen yang sangat dibutuhkan untuk mendukung fungsi setiap sel dalam tubuh. Setiap harinya, jantung berdetak sekitar 100.000 kali. Penyakit jantung, juga dikenal sebagai penyakit kardiovaskular (CVDs), merupakan salah satu penyakit tidak menular yang sering menyerang individu di usia produktif dan menjadi penyebab kematian paling umum di seluruh dunia. Di Indonesia, penyakit jantung menjadi salah satu fokus utama dalam penanganan dan alokasi dana oleh BPJS Kesehatan. Dalam upaya pencegahan, prediksi penyakit jantung pada pasien menjadi langkah yang sangat penting [1]. Berbagai metode dapat digunakan untuk menganalisis data kesehatan, dan salah satu metode yang efektif adalah *Data Mining*.

Data mining menerapkan berbagai metode dan algoritma untuk menemukan dan mengekstrak pola data yang disimpan. Aplikasi data mining dan penemuan pengetahuan telah mendapat fokus yang kaya karena signifikansinya dalam pengambilan keputusan dan telah menjadi komponen penting dalam berbagai organisasi. Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang [2]. Bidang data mining telah berkembang dan

ditempatkan ke dalam bidang kehidupan manusia yang baru dengan berbagai integrasi dan kemajuan di bidang Statistik, Basis Data, Pembelajaran Mesin, Reorganisasi Pola, dan perawatan kesehatan. Data mining dalam perawatan kesehatan dianggap sebagai tugas penting namun rumit yang perlu dilakukan secara akurat dan efisien. Data mining mencoba untuk memecahkan masalah kesehatan dunia nyata dalam diagnosis dan pengobatan penyakit. Salah satu tugas terpenting dalam data mining adalah melakukan klasifikasi pada data.

Klasifikasi merupakan teknik yang berfungsi untuk memetakan data ke dalam kelompok atau kelas yang sudah ditetapkan. Dalam praktiknya, klasifikasi bisa dilakukan secara manual atau menggunakan bantuan teknologi. Klasifikasi sebuah gambar dalam sebuah label merupakan hal yang sangat mudah bagi manusia karena manusia merupakan makhluk visual yang bisa melihat, mengidentifikasi dan mengklasifikasi dengan cepat [3]. Klasifikasi manual dilakukan oleh manusia tanpa melibatkan algoritma cerdas dari komputer. Sementara itu, klasifikasi yang dibantu teknologi memanfaatkan berbagai algoritma, seperti Naïve Bayes, Support Vector Machine, Decision Tree, dan K-Nearest Neighbor [4]. Dalam konteks data mining, klasifikasi adalah salah satu tugas yang paling umum dilakukan, di mana algoritma belajar dari data pelatihan yang sudah diberi label untuk mengklasifikasikan data baru. Penerapan klasifikasi meliputi berbagai hal, seperti pengenalan pola, diagnosis medis, deteksi spam email, dan analisis sentimen teks. Dengan teknik klasifikasi yang tepat, sistem dapat melakukan prediksi yang akurat serta mendukung pengambilan keputusan di berbagai bidang. Algoritma diperlukan agar proses klasifikasi berjalan optimal. Beberapa algoritma yang sering digunakan

untuk klasifikasi adalah k-nearest neighbor (KNN), dan lainnya. Pada penelitian ini, peneliti memilih Algoritma K-nearest Neighbor.

Beberapa penelitian [5], yang telah menerapkan K-Nearest Neighbor, diantaranya dilakukan oleh Ikhsan Wisnuadji Gamadarenda melakukan proses klasifikasi terhadap sebuah dataset penyakit *ginjal kronis* dengan menggunakan menggunakan Algoritma k-Nearest Neighbor menghasilkan nilai akurasi sebesar 99,25%, sensitivity sebesar 99,5%, dan specificity sebesar 98,745%. Penelitian lainnya [6], yang dilakukan oleh *Indrayanti Indrayanti* proses klasifikasi terhadap sebuah dataset penyakit *diabetes mellitus* dengan menggunakan menggunakan Algoritma k-Nearest Neighbor Penelitian ini menghasilkan K terbaik pada percobaan K=13 dengan akurasi 75,14%. K=13 merupakan nilai k paling optimal diantara percobaan klasifikasi KNN menggunakan nilai K=1 sampai dengan K=49.

Algoritma K-nearest Neighbor (KNN) adalah algoritma yang banyak digunakan dalam klasifikasi penyakit, termasuk Jantung. KNN mengklasifikasikan data berdasarkan kedekatan jarak antara data pelatihan dengan objek yang akan diklasifikasikan. Kelebihan KNN mencakup ketahanannya terhadap noise dan efektivitasnya pada data pelatihan dalam jumlah besar. Namun, KNN juga memiliki kelemahan, seperti perlunya menentukan nilai parameter K (jumlah tetangga terdekat) dan ketidakjelasan dalam memilih jenis jarak serta atribut yang optimal untuk hasil terbaik[7]. Selain itu algoritma ini dapat digunakan untuk mengklasifikasi data medis yang mencakup informasi temporal, seperti riwayat penyakit diabetes, penyakit kanker paru paru dan penyakit jantung.

Berdasarkan latar belakang tersebut maka peneliti tertarik untuk melakukan penelitian dalam bentuk skripsi yang berjudul : “**PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER MENGGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)**”.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah diatas maka penulis dapat merumuskan masalah dari pembahasan tersebut yaitu:

1. Bagaimana penerapan algoritma KNN dalam mengklasifikasikan penyakit jantung dari data pasien?
2. Seberapa akurasi algoritma KNN dalam mengidentifikasi kasus positif (penyakit jantung) yang sebenarnya?

1.3 BATASAN MASALAH

Agar tidak terjadinya pembahasan yang diluar permasalahan dan topik penulisan ini, maka penulisan menetapkan batasan masalah yang akan dibahas. Adapun batasan masalah tersebut adalah sebagai berikut:

1. **Metode yang digunakan:** Penelitian ini hanya menggunakan algoritma K-nearest Neighbor sebagai metode klasifikasi.
2. **Sumber data:** Dataset yang digunakan adalah penyakit jantung dari Kaggle: (<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset/data>)
3. **Fokus analisis:** Penelitian ini hanya berfokus pada mengklasifikasi penyakit jantung.

1.4 TUJUAN DAN MANFAAAT PENELITIAN

Adapun tujuan dan manfaat dari penelitian yang dilakukan akan dijelaskan sebagai berikut:

14.1 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu untuk:

1. Menerapkan Algoritma KNN dalam melakukan klasifikasi dan memperoleh akurasi yang tepat dalam mendeteksi penyakit jantung koroner.
2. Melakukan optimasi akurasi terhadap akurasi, presisi, dan recall algoritma KNN dalam mengetahui tingkat akurasi yang lebih baik dalam mendeteksi penyakit jantung koroner.

1.4.2 Manfaat Penelitian

Adapun manfaat penelitian ini yaitu sebagai berikut:

1. Dengan menggunakan algoritma KNN, sistem dapat membantu mendeteksi pasien yang berpotensi menderita penyakit jantung lebih dini, sehingga tindakan pencegahan atau penanganan dapat segera dilakukan oleh tenaga medis.
2. Evaluasi akurasi dari algoritma KNN dapat memberikan wawasan yang lebih mendalam tentang ketepatan klasifikasi penyakit jantung. Ini membantu dokter dalam membuat keputusan klinis yang lebih tepat, terutama dalam mengidentifikasi kasus positif yang mungkin memerlukan perhatian segera.

3. Dengan kemampuan KNN untuk mengklasifikasikan penyakit secara lebih efektif, fasilitas kesehatan dapat mengalokasikan sumber daya, seperti waktu dokter dan alat diagnostik, secara lebih efisien.

1.1 SISTEMATIKA PENULISAN

Untuk mempermudah dan memahami lebih jelas materi yang akan disusun dalam skripsi ini, maka penulis memberikan suatu gambaran umum mengenai sistematika penulisan yang dapat dilihat sebagai berikut:

BAB I : PENDAHULUAN

Bab ini merupakan bab pendahuluan yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini memberikan landasan teori yang menjadi dasar untuk membahas laporan tertentu, dan memberikan definisi yang mendasari penelitian yang dilakukan melalui tinjauan literatur sebagai dasar untuk melakukan analisis.

BAB III : METODOLOGI PENELITIAN

Bab ini merujuk pada kerangka penelitian, metode atau pendekatan yang digunakan, dan alat yang digunakan untuk menjawab pertanyaan penelitian untuk mencapai tujuannya.

BAB IV : ANALISIS DAN HASIL

Pada bab ini, menjelaskan tentang gambaran umum objek penelitian, serta menganalisis data. Hasil analisis disajikan dalam bentuk visualisasi dari tools yang digunakan yaitu google collaboration.

BAB V : KESIMPULAN

Kesimpulan dan pedoman yang disajikan dalam bab ini berasal dari hasil analisis yang mencakup semua informasi yang diperoleh.