

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Machine Learning (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk meniru dan menggantikan peran manusia untuk memecahkan masalah[1]. Dalam konsep tradisional pemrograman, akan memberikan aturan dan instruksi kepada komputer untuk melakukan suatu tugas tertentu[2]. ML memiliki fokus pada pengembangan sebuah sistem yang mampu belajar sendiri untuk memutuskan sesuatu, tanpa harus berulang kali diprogram oleh manusia. Dengan metode tersebut, mesin tidak hanya bisa menemukan aturan untuk perilaku optimal dalam pengambilan keputusan, namun juga bisa beradaptasi dengan perubahan yang terjadi[3]. Selain itu, dalam proses ML ini menggunakan metode klasifikasi.

Klasifikasi adalah proses menemukan sekumpulan pola atau fungsi yang mendeskripsikan dan memisahkan suatu kelas data dengan kelas data lainnya dan digunakan untuk memprediksi kelas suatu objek yang labelnya tidak diketahui [4], [5]. Klasifikasi termasuk dalam metode pembelajaran atau supervised karena memerlukan pembelajaran dari data sebelumnya untuk mengetahui hasil dari data baru [6]. Dalam proses ini, algoritma pembelajaran mengidentifikasi pola yang menghubungkan atribut data dengan kelas yang sesuai, sehingga memungkinkan untuk membuat prediksi berdasarkan pola-pola ini.

Supervised Learning adalah Algoritma pembelajaran yang diawasi membangun model dilatih menggunakan dataset yang berisi masukan dan keluaran yang diinginkan[7]. Terdapat banyak metode yang ada dalam klasifikasi supervised learning diantaranya adalah K-NN[8]. Seiring dengan kemajuan teknologi dan peningkatan dalam penggunaan data, semakin banyak metode pembelajaran supervised seperti K-NN digunakan untuk mengatasi berbagai tantangan analisis data.

K-Nearest Neighbor (K-NN) merupakan metode supervised learning yang memerlukan informasi training untuk mengklasifikasikan objek yang jaraknya sangat dekat[9]. Prinsip pengoperasian K-NN adalah menemukan k tetangga terdekat dan mengklasifikasikan rekaman data numerik dengan menghitung jarak antara sampel uji dan seluruh sampel pelatihan menggunakan jarak Euclidian yang dapat digunakan pada dataset penyakit jantung (*heart disease*)[10]. K-NN adalah salah satu metode yang sederhana tetapi efektif dalam klasifikasi, dan sering digunakan dalam berbagai penyakit, termasuk klasifikasi penyakit stroke, paru-paru, dan jantung.

Dataset penyakit jantung didapat dari *Heart Disease Dataset* (kaggle.com) dan *Heart Disease Dataset* (kaggle.com) data pertama mempunyai 4238 data, 16 atribut serta 2 kelas dan data yang kedua mempunyai 1026 data, 14 atribut serta 2 kelas yaitu 1 (terkena serangan jantung) dan 0 (tidak terkena serangan jantung). Jantung adalah organ vital yang berfungsi sebagai pemompa darah untuk memenuhi kebutuhan oksigen dan nutrisi ke seluruh tubuh. Apabila jantung mengalami gangguan, peredaran darah dalam tubuh dapat terganggu sehingga menjaga kesehatan jantung sangatlah penting agar terhindar dari berbagai jenis penyakit jantung[11]. Analisis data pada dataset ini dapat membantu karena kurangnya akses untuk mencari informasi tentang penyakit jantung ini menyebabkan peningkatan angka kematian setiap tahunnya. Karena itu dibutuhkan sebuah sistem klasifikasi yang dapat memberikan

informasi tentang penyakit jantung serta dapat melakukan pengecekan klasifikasi secara dini.

Menurut Yofi Pratama dalam memprediksi penyakit gagal jantung menggunakan algoritma Knn berdasarkan 918 data dan terdiri dari 12 atribut dengan nilai $k=9$ menghasilkan akurasi 70,65%, nilai presisi 75% dan nilai recall menghasilkan 70,73%[12].

Menurut Hanna Willa Dhany dalam memprediksi penyakit jantung menggunakan algoritma Knn berdasarkan 303 data dan 13 atribut dengan 2 kelas 0 dan 1 menghasilkan akurasi sebesar 81,31% dengan Classification Error sebesar 18,68%[13].

Menurut Muhammad Fakhri Rizqullah dalam Penerapan Algoritma Knn untuk Keberlangsungan Pasien Gagal Jantung dengan 299 data dan 13 atribut dengan 2 kelas yaitu yes atau no dengan akurasi 63,54% dengan standar deviasi kurang lebih 5,57% dengan total true yes sebanyak 19 dan true no sebanyak 171[14].

Sehubungan dengan latar belakang diatas, maka penulis tertarik untuk melakukan penelitian untuk mendapatkan akurasi yang lebih baik lagi dari penelitian sebelumnya dengan berjudul **“Klasifikasi Penyakit Jantung Menggunakan Algoritma K-Nearest Neighbors”**.

1.2. RUMUSAN MASALAH

Berdasarkan dari latar belakang yang telah diuraikan di atas, maka di dapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan machine laerning dengan algoritma K-NN dalam mengklasifikasi penyakit jantung?
2. Seberapa besar tingkat akurasi algoritma K-NN apabila digunakan untuk memprediksi kelas penyakit jantung?

1.3. BATASAN MASALAH

Untuk menghindari permasalahan yang meluas, maka penulis membatasi pembahasan

permasalahan sebagai berikut :

1. Data yang digunakan bersumber dari repository public yaitu *kaggle.com* dengan nama *heart disease dataset*.
2. Metode yang digunakan adalah metode klasifikasi serta algoritma yang digunakan yaitu K-NN.
3. Kinerja algoritma K-NN diukur dengan menggunakan nilai *accuracy*, *precision*, dan *recall*.
4. Aplikasi yang digunakan adalah *Google Collaboratory*.

1.4. TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Untuk menerapkan *machine learning* dengan algoritma K-NN dalam mengklasifikasi penyakit jantung.
2. Untuk mengetahui seberapa akurat tingkat akurasi algoritma K-NN dalam memprediksi kelas penyakit jantung.

1.4.2. Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan Penelitian ini, yaitu :

1. Dapat membantu mengetahui atribut yang mempunyai korelasi baik dan tidak pada dataset penyakit jantung.

2. Dapat berkontribusi pada peningkatan pengetahuan medis terkait faktor risiko dan karakteristik pasien yang rentan terhadap penyakit jantung.
3. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *machine learning* khususnya untuk klasifikasi penyakit jantung.
4. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya

1.5. SISTEMATIKA PENULISAN

Sistematika penulisan menggambarkan secara umum tentang apa yang akan dibahas dalam setiap bab dari laporan penelitian ilmiah ini. Adapun isi pokok bahasan masing-masing bab sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang digunakan dalam penulisan laporan tugas akhir ini.

BAB II : LANDASAN TEORI

Pada bab ini menjelaskan teori-teori yang mendasari pembahasan laporan secara khusus berisi definisi- definisi yang di dapat melalui studi pustaka seperti kutipan buku, jurnal dan lain-lain sebagai dasar dalam pembuatan laporan. Pada bab ini juga memuat tinjauan pustaka yang berisi penelitian-penelitian sebelumnya yang berhubungan dengan penelitian ini.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang penjelasan dari metode-metode penelitian yang

digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian.

BAB IV : ANALISIS

Pada bab ini menjelaskan tentang gambaran umum pemilihan dataset pada kaggle.com. Kemudian menerapkan algoritma K-Nearest Neighbor (K-NN) untuk mengetahui tingkat akurasi data menggunakan aplikasi RapidMiner Studio.

BAB V : KESIMPULAN

Pada bab ini berisikan kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian

