

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Jantung merupakan salah satu organ vital yang menjadi pusat sistem peredaran darah pada tubuh manusia, yang melibatkan pembuluh darah sebagai salurannya. Disamping perannya yang penting, jantung juga menjadi salah satu organ tubuh yang paling rentan terkena penyakit(Prabowo & Kurniadi, 2023).

Penyakit jantung merupakan kondisi dimana pembuluh darah utama yang menyuplai darah ke jantung mengalami kerusakan dan tidak dapat bekerja sebagaimana mestinya(Anggriawan & Nugroho, 2023). Penyakit jantung atau penyakit kardiovaskular merupakan penyebab utama kematian manusia di seluruh dunia, terutama di negara berkembang termasuk Indonesia.

Sekitar 17,9 juta kematian per tahun akibat penyakit kardiovaskular dan merupakan 32 persen dari seluruh total kematian di dunia. Angka ini akan terus meningkat hingga 23 juta kematian per tahun di 2030.(WHO, 2021)

Penderita penyakit jantung di Indonesia semakin meningkat setiap tahunnya. Penyakit ini biasanya diawali dengan rasa sakit yang tidak jelas, rasa tidak nyaman yang samar-samar, atau rasa sesak pada bagian dada(KEMENKES, 2018).

Kemajuan teknologi telah membuat perkembangan dalam ilmu pengetahuan dan teknologi informasi telah sangat pesat bahkan menambah banyak bidang dan salah satu manfaatnya adalah terciptanya sistem prediksi. Dunia medis salah satu yang mendapat manfaat perkembangan teknologi dalam penerapan sistem prediksi(Anggriawan & Nugroho, 2023).

Berbagai metode klasifikasi banyak diterapkan untuk melakukan prediksi penyakit. Sistem prediksi sangat berguna dalam memperkirakan sesuatu yang kemungkinan akan terjadi berdasarkan data yang ada sehingga kerugian yang timbul dapat ditanggulangi atau keuntungan yang kemungkinan terjadi dapat dimaksimalkan hasilnya. Penelitian berjudul “Perbandingan Algoritma *Naive Bayes* dan C4.5 Untuk Klasifikasi Penyakit Anak”(Bahri et al., 2018). Penelitian yang dilakukan dengan

aplikasi *RapidMiner* versi 8.1 serta metode *K-fold Cross Validation* didapatkan hasil bahwa algoritma *Decision Tree C4.5* merupakan algoritma terbaik dalam melakukan klasifikasi penyakit anak dengan tingkat akurasi sebesar 90% sedangkan algoritma *Naive Bayes* memperoleh tingkat akurasi 89.58%.

Penelitian berjudul “Perbandingan Akurasi Algoritma C4.5 dan *Naive Bayes* untuk Deteksi Dini Gangguan Autisme Pada Anak”(Sugara, Bayu; Dedi Adidarma, 2019), algoritma C4.5 menunjukkan hasil nilai akurasi sebesar 72% sedangkan algoritma *Naive Bayes* menunjukkan nilai akurasi yang lebih besar dan lebih baik yaitu sebesar 73.33%.

Penelitian yang berjudul “Komparasi Dalam Prediksi Gagal Jantung Dengan Menggunakan Metode C4.5 dan *Naive Bayes*”(Julia, 2023). Kesimpulan yang didapat dari hasil analisis evaluasi model berdasarkan *cross validation* bahwa Algoritma C4,5 dengan hasil prediksi tertinggi adalah data tanpa *outliers* dengan nilai *accuracy* sebesar 85.02%, begitupun dengan algoritma *Naive Bayes* dengan hasil prediski tertinggi adalah data tanpa *outliers* dengan nilai *accuracy* sebesar 85.57%.

Penelitian yang dilakukan oleh (Khodijah & Sriyanto, 2023) yang berjudul “Perbandingan Kinerja Algoritma C4.5, *Naive Bayes*, dan *Random Forest* Dalam Prediksi Penyakit Jantung” yang menyimpulkan bahwa algoritma *Random Forest* menghasilkan nilai akurasi terbaik sebesar 98.60%, kemudian algoritma C4.5 menghasilkan nilai akurasi sebesar 88.34%, dan algoritma *Naive Bayes* sebesar 87.25%.

Penelitian berjudul “Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung” (Prabowo & Kurniadi, 2023), saat melakukan uji, hasil yang diperoleh adalah algoritma *Random Forest*, *SVM*, *AdaBoost* menjadi algoritma yang memiliki nilai terbaik dan bernilai sama dalam hasil ujinya. Untuk nilai akurasi 0,985364, presisi 0,985714, *recall* 0,985437, dan *f1-score* 0,985364. Sehingga ketiga algoritma tersebut dapat dijadikan rekomendasi untuk memuat prediksi penyakit jantung.

Penelitian berjudul “Deteksi Penyakit Jantung Menggunakan Metode Kalsifikasi *Decission Tree* dan *Regresi Logistik*” (Bukhari et al., 2024), dimana *decission tree* memiliki akurasi terhadap data uji sebesar 75 %, dan model regresi logistik memiliki

nilai akurasi terhadap data uji sebesar 87 %. Berdasarkan akurasi dari kedua model, regresi logistik lebih akurat untuk mendeteksi adanya penyakit jantung dibandingkan model *decision tree*.

Dalam penelitian ini memiliki tujuan untuk melakukan analisis terhadap algoritma dan fitur apa saja yang dapat memberikan hasil yang optimal. Terdapat banyak peneliti yang menggunakan algoritma klasifikasi data mining dalam memprediksi berbagai penyakit, akan tetapi masih sulit mengetahui metode atau algoritma yang paling akurat.

Berdasarkan penjelasan diatas, pada penelitian ini akan dilakukan proses perbandingan algoritma atau metode *Decission Tree C4.5* dan *Naive Bayes* untuk memprediksi penyakit jantung pada pasien. Peneliti memilih metode *C4.5* dan *Naive Bayes* untuk memprediksi penyalit jantung serta mengoptimalkan atribut yang berasal dari dataset terpercaya untuk memprediksi penyakit jantung pada pasien dengan tujuan agar akurasi menjadi meningkat.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka yang menjadi rumusan masalah dalam penelitian ini adalah bagaimana memprediksi serta mengukur tingkat akurasi penyakit jantung pada pasien dengan menggunakan metode *C4.5* dan *Naive Bayes*.

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang terjadi maka penelitian ini bertujuan untuk:

1. Mengetahui bagaimana algoritma klasifikasi *C4.5* dan *Naive Bayes* dalam memprediksi penyakit jantung pada pasien dan tingkat keakuratan prediksi penyakit jantung pada pasien.
2. Mengetahui perbandingan antara algoritma antara algoritma klasifikasi *C4.5* dan *Naive Bayes* dalam mengukur tingkat akurasi prediksi penyakit jantung pada pasien.

1.4. Manfaat Penelitian

Adapun manfaat yang terdapat dalam penelitian ini adalah:

1. Penulis dapat menambah pengetahuan terkait bagaimana cara mencari informasi yang penting dalam suatu data dengan menggunakan metode algoritma klasifikasi C4.5 dan *Naive Bayes*.
2. Pengembangan Data Mining dalam dunia kesehatan
3. Menjadi bahan pertimbangan kesehatan dalam perawatan pasien dengan penyakit jantung.
4. Dapat digunakan sebagai bahan acuan untuk melakukan penelitian selanjutnya.

1.5. Ruang Lingkup Penelitian

Agar penelitian ini lebih terarah, maka perlu adanya batasan masalah pada ruang lingkup bahasan antara lain sebagai berikut :

1. Menggunakan dan membahas algoritma klasifikasi C4.5 dan *Naive Bayes* dalam memprediksi penyakit jantung pada pasien.
2. Aplikasi yang akan digunakan untuk memproses data adalah tools *Google Colab* dan juga menggunakan bahasa pemrograman *python*.
3. Dataset prediksi penyakit jantung yang akan digunakan dalam penelitian ini diambil dari laman <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.