

BAB I PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung merupakan salah satu penyebab kematian utama di dunia. Data dari WHO (2021) menunjukkan bahwa penyakit jantung menyebabkan lebih dari 17,9 juta kematian secara global[1]. Di Indonesia, peningkatan angka kematian akibat penyakit jantung meningkat dari 356,05 menjadi 412,46 kematian dari tahun 2000 hingga 2019[2]. Kondisi ini menegaskan pentingnya pengembangan teknologi yang mampu mendeteksi penyakit jantung secara dini guna mencegah kematian mendadak akibat serangan jantung. Salah satu pendekatan yang semakin mendapat perhatian adalah penerapan *machine learning*, yang terbukti mampu mengolah data medis secara efisien dan mengenali pola yang sulit dideteksi secara standar[3].

Dalam klasifikasi penyakit jantung, algoritma *Decision Tree* dan *Naïve Bayes* sering digunakan karena kemudahan implementasi, efisiensi komputasi, serta interpretasi hasil yang jelas[4]. *Decision Tree* berkerja dengan membagi data ke dalam beberapa subset berdasarkan aturan pohon keputusan yang diperoleh dari atribut data. Pemilihan atribut ini umumnya dilakukan menggunakan metrik seperti *entropy*, *information gain*, dan *gini index*[5]. Sementara itu, *Naïve Bayes* didasarkan pada asumsi bahwa setiap fitur dalam data bersifat independen, sehingga memungkinkan perhitungan probabilitas gabungan untuk menentukan kelas output. Meskipun asumsi ini tidak selalu sesuai dengan kondisi di dunia nyata, *Naïve Bayes* tetap menunjukan performa klasifikasi yang kompetitif[6].

Kedua algoritma ini banyak diterapkan dalam berbagai bidang, mulai dari klasifikasi teks hingga diagnosis medis dan analisis sentimen. *Naïve Bayes*, dengan efisiensi komputasinya yang tinggi, tetap dapat berkerja dengan baik meskipun terdapat penyimpangan dalam distribusi data[7]. Keberhasilan kedua algoritma ini sangat bergantung pada pemilihan *hyperparameter* yang tepat, di mana optimasi parameter dapat meningkatkan akurasi model serta mengurangi risiko *overfitting*[8]. Salah satu metode yang terbukti efisien adalah *Random Search*, yang lebih unggul dalam menemukan kombinasi parameter optimal

dibandingkan *Grid Search*, baik dari segi waktu maupun efektivitas pencarian [9], [10].

Berdasarkan latar belakang, penelitian ini bertujuan untuk membandingkan efektivitas *hyperparameter tuning* pada algoritma *Decision Tree* dan *Naïve Bayes* dalam memprediksi penyakit jantung. Dengan optimasi parameter yang tepat, kedua model diharapkan mampu meningkatkan akurasi dalam mendeteksi risiko penyakit jantung, sehingga dapat berkontribusi dalam upaya pencegahan dan penanganan dini penyakit jantung.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang sudah di paparkan diatas, maka dapat dirumuskan latar belakang masalah sebagai berikut:

1. Bagaimana perbandingan kinerja algoritma *Decision Tree* dan *Naive Bayes* setelah dilakukan *hyperparameter tuning* dalam memprediksi penyakit jantung?
2. Bagaimana hasil evaluasi berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score* setelah dilakukan *hyperparameter tuning* pada algoritma *Decision Tree* dan *Naïve Bayes*?

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Peneliti menggunakan algoritma *Decision Tree* dan *Naive Bayes*.
2. Dataset yang digunakan berupa dataset yang diambil dari *UCI Machine Learning Repository*, yang berjudul *Framingham Heart Study Dataset* tahun 2022 yang berisi 4240 jumlah data dan 16 kolom atribut, namun peneliti hanya menggunakan 7 kolom atribut dan tidak mencakup data medis dari rumah sakit atau institusi kesehatan lainnya.
3. Peneliti menggunakan seleksi fitur berdasarkan hubungan korelasi antara fitur dan variabel target, fitur yang digunakan meliputi: fitur *age*, *BPMeds*, *prevalentHyp*, kolesterol, *sysBP*, *diaBP*, *BMI* dan target.
4. Penelitian ini menggunakan metode evaluasi (*accuracy*, *presicion*, dan *recall*), *cross-validation*, dan *confusion matrix*.

5. Peneliti menggunakan metode *Random Search* untuk *hyperparameter tuning*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah penelitian ini bertujuan:

1. Menganalisis perbandingan kinerja *Decision Tree* dan *Naïve Bayes* setelah diterapkan *hyperparameter tuning* menggunakan Teknik *Random Search* untuk mengoptimalkan parameter model.
2. Mengevaluasi peningkatan performa model berdasarkan metrik *accuracy*, *precision*, *recall* dan *f1-score* setelah proses *hyperparameter tuning*.

1.5 Manfaat Penelitian

Penelitian ini memberikan berbagai manfaat, baik secara manfaat teoritis maupun manfaat praktis, sebagai berikut:

1. Manfaat teoritis

Penelitian ini diharapkan dapat memberikan referensi tentang kinerja *machine learning* seperti, penyetelan parameter yang optimal, penerapan algoritma *decision tree* dan *naïve bayes* dalam proses klasifikasi penyakit jantung. Dalam penelitian ini, peneliti memberikan wawasan tentang pengaturan serta pemilihan teknik *hyperparameter* untuk meningkatkan akurasi dan efisiensi model.

2. Manfaat praktis

Bagi lembaga masyarakat serta lainnya, diharapkan dapat memahami faktor - faktor risiko penyakit jantung seperti, hipertensi, kolesterol tinggi, atau diabetes. Dengan melakukan pencegahan yang baik, seperti perubahan gaya hidup lebih sehat.

1.6 Sistematika Penulisan

Sistematika penulisan yang disusun pada skripsi ini terdapat 5 bab dan pokok-pokok permasalahan diuraikan sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, tujuan dari penelitian, dan manfaat penelitian.

BAB II : TINJAUAN PUSTAKA

Pada bab ini berisi studi literatur dan dasar teori yang digunakan pada penelitian.

BAB III : METODE PENELITIAN

Pada bab ini berisi tentang alur penelitian, metode, serta alat dan bahan yang digunakan.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang tahapan penelitian, implementasi algoritma, pengaturan parameter, dan hasil evaluasi kinerja model.

BAB V : PENUTUP

Pada bab ini berisi kesimpulan dan saran guna untuk penelitian selanjutnya.

