

BAB I PENDAHULUAN

1.1 Latar Belakang

Jantung adalah salah satu organ tubuh yang harus kita jaga karena kesehatan sangat penting bagi manusia karena dengan sehat kita dapat melakukan berbagai hal dan berpikir dengan baik. Jantung adalah bagian tubuh manusia yang melakukan banyak fungsi penting, dan kelainan kecil pada jantung dapat berdampak besar pada bagaimana tubuh kita bekerja. Penyakit jantung merupakan penyebab kematian nomor satu di dunia[1].

Jantung adalah organ otot yang berfungsi untuk memompa darah melalui pembuluh darah dengan ritme yang teratur dan berkelanjutan. Organ ini sangat penting dalam sistem peredaran darah karena mendistribusikan oksigen dan nutrisi ke seluruh tubuh sekaligus membantu mengatur proses metabolisme. Jantung terletak di sisi kiri rongga dada. Sebagai salah satu organ yang paling krusial, jantung berpotensi menyebabkan kematian jika terjadi masalah[2]

Penyakit Jantung merupakan gangguan terhadap fungsi kerja jantung dalam tubuh. Salah satu penyakit yang paling sering terjadi di masyarakat adalah penyakit jantung, yang dapat menyerang siapa saja tanpa memandang usia, jenis kelamin, atau gaya hidup. Di negara Inggris, Amerika Serikat, Kanada, dan Australia, penyakit jantung adalah penyebab kematian utama, menurut CDC dan WHO. Penyakit jantung didiagnosis pada 26,6 juta atau 11,3% dari populasi orang dewasa. Lebih dari 12 juta orang di seluruh dunia meninggal karena penyakit jantung, salah satu penyakit yang memiliki angka kematian yang paling tinggi[3].

Dalam era digital saat ini, Kemajuan teknologi memberikan manfaat teknologi terciptanya sistem prediksi. Untuk memprediksi penyakit, banyak jenis klasifikasi digunakan. Sistem prediksi sangat berguna dalam memperkirakan apa yang mungkin terjadi berdasarkan data yang ada sehingga kita dapat memperbaiki kerugian atau memaksimalkan keuntungan yang mungkin terjadi[4].

Dalam penelitian untuk mendapatkan hasil yang akurat dan dapat

diandalkan, sangat penting untuk memilih algoritma yang tepat. *Decision Tree* dan *Random Forest* adalah dua algoritma yang sering digunakan untuk klasifikasi medis, tetapi masing-masing memiliki kelebihan dan kekurangan. Oleh karena itu, perlu dilakukan perbandingan untuk mengetahui mana yang lebih baik dalam prediksi penyakit jantung.

Decision Tree digunakan untuk klasifikasi dan pola prediksi data. Selain itu, variabel atribut x dan variabel target y digambarkan dalam hubungan antara satu sama lain [5]. Di sisi lain, *Random Forest* (RF) pertama kali diperkenalkan oleh Leo Breiman (2001). *Random Forest* (RF) algoritma yang berbasis pohon keputusan dan mampu menangani data yang lebih kompleks. Algoritma ini dikenal karena akurasi yang tinggi dan kemampuannya dalam mengatasi *overfitting*. *Overfitting* adalah suatu kondisi dalam pembelajaran mesin (*machine learning*) di mana model belajar terlalu baik pada data pelatihan, termasuk noise dan fluktuasi yang tidak relevan, sehingga kinerjanya buruk pada data yang belum pernah dilihat sebelumnya (data uji) [6].

Dengan membandingkan kedua model ini menggunakan dataset dari Kaggle. Penelitian ini bertujuan untuk mengetahui model mana yang lebih unggul dalam konteks klasifikasi dalam prediksi penyakit jantung.

Sebagai solusi diatas, permasalahan yang dihadapi yang lembaga kesehatan atau medis dalam memprediksi risiko penyakit jantung, penerapan machine learning, khususnya melalui algoritma *Decision Tree* dan *Random Forest*, diharapkan dapat menghasilkan hasil akurasi dalam proses penilaian. Dengan menggunakan kedua model ini, koperasi dapat mengoptimalkan pengambilan keputusan meminimalisir risiko penyakit jantung.

1.2 Rumusan Masalah

Rumusan masalah untuk penelitian yang berkaitan dengan penerapan algoritma *Decision Tree* dan *Random Forest* dalam prediksi risiko penyakit jantung:

1. Bagaimana kinerja dari model *Decision Tree* dalam memprediksi risiko penyakit jantung?
2. Bagaimana kinerja dari model *Random Forest* dalam memprediksi risiko penyakit jantung?
3. Bagaimana hasil dari perbandingan dari kedua model klasifikasi dalam kasus memprediksi risiko penyakit jantung?

1.3 Batasan Masalah

Penelitian ini memiliki batasan masalah yang berkaitan dengan penerapan algoritma *Decision Tree* dan *Random Forest* dalam prediksi risiko penyakit jantung.

1. Dataset yang digunakan merupakan dataset kaggle yang dapat diakses secara publik. Dataset ini mencakup berbagai faktor seperti *age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target*.
2. Model *machine learning* yang akan diuji pada penelitian ini yaitu *Decision Tree* dan *Random Forest*.
3. Metode Evaluasi yang digunakan untuk kedua model *machine learning* dengan akurasi, *precision, recall, F1-Score* dan AUC-ROC.
4. Hasil Prediksi dari model yang dihasilkan akan diinterpretasikan dalam konteks data yang digunakan. Penelitian ini tidak akan membahas generalisasi hasil ke populasi yang lebih luas di luar dataset yang dianalisis.
5. Aspek Geografis tidak akan membahas perbedaan risiko penyakit jantung berdasarkan faktor geografis atau budaya, meskipun kemungkinan faktor-faktor tersebut dapat mempengaruhi prevalensi dan risiko penyakit jantung.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah untuk menganalisis dan mengetahui model machine learning terbaik yang telah diuji dengan metode evaluasi dengan akurasi, *presisi*, *recall*, *F1-Score* dan AUC-ROC dan dataset yang telah ditetapkan yang nantinya diharapkan mampu menjadi opsi yang dapat digunakan dalam kasus lembaga kesehatan untuk mengatasi permasalahannya.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat penelitian untuk penerapan algoritma *Decision Tree* dan *Random Forest* dalam prediksi risiko penyakit jantung:

1. Penelitian ini dapat mengetahui hasil akurasi dalam prediksi risiko penyakit jantung dari 2 algoritma yaitu *Decision Tree* dan *Random Forest*.
2. Hasil dari penelitian ini dapat menjadi dasar untuk penelitian lebih lanjut di bidang yang sama, baik dalam konteks penyakit jantung maupun dalam penerapan algoritma lain untuk prediksi risiko penyakit lainnya.

1.6 Sistematika Penulisan

Adapun sistematika yang digunakan untuk menyusun skripsi ini terdapat 5(lima) bab. Setiap bab saling berhubungan satu sama lain, sehingga pembaca dapat dengan mudah memahami skripsi ini. Untuk sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan yang digunakan dalam penyusunan skripsi ini.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi Studi Literatur dan dasar teori yang digunakan dalam mendukung penelitian ini.

BAB III METODE PENELITIAN

Pada bab ini berisikan objek penelitian, alur penelitian serta alat dan bahan yang akan di buat.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dan pembahasan setelah uji coba.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian serta saran yang dapat digunakan penelitian di masa yang akan datang.

