

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berdasarkan informasi dari World Health Organization (WHO) *cardiovascular disease* (CVD) adalah salah satu penyebab utama kematian secara global. Diperkirakan 17,9 juta orang meninggal karena CVD pada 2019, yaitu 32% dari semua kematian global. Faktor penyebab CVD adalah pola hidup yang tidak sehat, sehingga dapat mengindikasikan meningkatnya risiko penyakit jantung [1]. Jantung adalah organ vital pada tubuh manusia sehingga penting bagi manusia untuk merawatnya. Sebagian besar penyakit berhubungan dengan jantung, serangan jantung (*heart attack*), stroke, dan gagal jantung (*heart failure*) merupakan beberapa jenis penyakit jantung. Banyak pasien atau penderita meninggal karena baru dikenali pada saat tahap akhir karena, sehingga prediksi tentang penyakit jantung perlu diterapkan dalam menangani hal tersebut [2].

Penyakit jantung menyebabkan angka kematian yang mencolok di seluruh dunia dan merupakan ancaman terhadap kesehatan untuk banyak orang. Saat ini prediksi terhadap penyakit jantung dapat membantu meminimalkan jumlah kematian daripada pasien. Pendeteksian kardiovaskular seperti penyakit jantung menjadi tantangan kritis oleh analis data klinis [3]. Pendekatan penelitian dalam upaya memberikan solusi dalam memprediksi penyakit jantung telah dilakukan [2] [3]. Tiap penelitian memiliki teknik tersendiri dalam prediksi terhadap penyakit jantung pada pasien. Metodologi *machine learning* dapat meninjau banyaknya data dan identifikasi terhadap manusia. Proses menyelesaikan tugasnya dengan menyediakan pelabelan terhadap data dan *output* pada sistem. Penggunaan metode seperti *Naive Bayes*, *Logistic Regression*, *Random Forest*, *SVM*, *KNN*, dan *Decision Tree* digunakan untuk memprediksi penyakit [4]. *Machine learning* mempelajari dari pola yang diberikan oleh sampel *training* dan kemudian digunakan untuk mengembangkan prediksi yang berguna. Metode klasifikasi berkembang luas di ranah medis untuk identifikasi dan memprediksi penyakit secara akurat [5].

Penelitian yang ada dengan menggunakan beberapa klasifikasi seperti *Naive Bayes*, *Logistic Regression*, *Random Forest*, *SVM*, *KNN*, dan *Decision Tree* memiliki keunggulan dan kekurangan tersendiri dalam pengolahan pada dataset dalam memprediksi penyakit [4]. Namun tingkat akurasi yang dihasilkan akan berpengaruh terhadap hasil prediksi yang akan didapat. Sehingga metode yang digunakan pada penelitian ini adalah *Decision Tree* untuk mengklasifikasi fitur-fitur yang ada pada dataset yang ada. Tingkat akurasi yang dimiliki oleh *Decision Tree* (73%) lebih baik dari pada metode lainnya. *Logistic regression* dan *SVM* (72%), *Random forests* (71%), *KNN* (66%) dan *Naive bayes* (60%) [4]. Penerapan metode *Decision Tree* dalam klasifikasi penyakit jantung mampu memperoleh tingkat akurasi yang terbaik dan dengan metode tersebut diharapkan dapat membantu kinerja medis dalam prediksi dini terhadap penyakit jantung.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Dalam mengklasifikasi dataset *heart failure* terdapat beberapa fitur yang berjenis kontinu dan kategorikal. Terdapat metode yang hanya efektif dalam memproses dataset dengan fitur yang berjenis kategorikal saja namun kurang efektif dengan fitur yang berjenis kontinu, maka dibutuhkan penerapan metode yang dapat memproses fitur dengan jenis kontinu dan kategorikal. Sehingga dapat mengklasifikasi fitur-fitur yang ada pada dataset *heart failure* untuk memperoleh keputusan berdasarkan target kelas yaitu *heart disease* atau *not heart disease*.

1.2.2 Metode Usulan

Pemanfaatan teknik *Decision Tree* metode *CART* menjadi usulan untuk mengklasifikasi dataset *heart failure* yang memiliki keunggulan dalam memproses fitur yang berjenis kontinu dan kategorikal untuk memperoleh klasifikasi berdasarkan target kelas yaitu *heart disease* atau *not heart disease*. Teknik *Decision Tree* metode *CART* menghasilkan *rule model* dalam bentuk pohon keputusan sehingga dapat dijadikan sebagai acuan dalam mengklasifikasi dan memperoleh hasil keputusan yang tepat.

1.2.3 Pertanyaan Penelitian

Pertanyaan pada penelitian ini yaitu berapa tingkat akurasi yang dihasilkan dari penggunaan metode *Decision Tree CART* dalam mengklasifikasi dataset *heart failure*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan

Tujuan dari penelitian ini yaitu mengukur tingkat akurasi klasifikasi dataset *heart failure* menggunakan metode *Decision Tree CART*.

1.3.2 Manfaat

Memperoleh hasil prediksi penyakit jantung untuk membantu tenaga medis dalam menangani pasien. Dan bisa juga dijadikan sebagai acuan untuk membangun aplikasi (yang siap diproduksi) untuk pendeteksian penyakit jantung.

1.4 Batasan Masalah

Batasan-batasan masalah pada penelitian ini, sebagai berikut :

1. *Raw material* data yang digunakan berasal dari data internasional yang disediakan oleh <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/> di Kaggle.
2. *Raw material* dataset *heart failure* berjumlah 918 data.
3. Jumlah fitur dataset *heart failure* berjumlah 11 fitur dan 1 target kelas.

1.5 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini sebagai berikut :

BAB I PENDAHULUAN

Bab I dipaparkan mulai dari (a) latar belakang penelitian, (b) rumusan masalah mencakup permasalahan, usulan solusi, pertanyaan penelitian, (c) tujuan dan manfaat dari penelitian, (d) batasan masalah, (e) serta sistematika penulisan penelitian.

BAB II STUDI PUSTAKA

Bab II dipaparkan dasar teori yang akan menjadi pijakan dari penelitian ini, yang mana di atasnya penelitian ini akan ditegakkan. Dimulai dari (a) landasan teori yang mencakup bahasan tentang penyakit jantung, klasifikasi, *Decision Tree*, metode *Decision Tree CART*, dan (b) penelitian terkait..

BAB III PERANCANGAN SISTEM

Bab III membahas mengenai gagasan rancangan sistem untuk setiap bagian mulai dari hulu ke hilir. Mulai dari (a) desain rancangan sistem, (b) bahasan tentang dataset, (c) proses *encoding*, (d) metode *Decision Tree Classifier*, (e) pembahasan tentang evaluasi dan validasi, (f) perhitungan manual yang mencakup tentang perhitungan menggunakan *Microsoft Excel* tentang metode *Decision Tree CART* dan akurasi menggunakan metode *Confusion Matrix*, (g) dan skenario uji coba.

BAB IV HASIL DAN PEMBAHASAN

Bab IV berisi tentang hasil dan pembahasan yang mencakup tentang hasil proses *encoding*, *split* dataset, hasil proses metode *Decision Tree Classifier*, hasil proses *Confusion Matrix*, implementasi uji coba, dan hasil skenario uji coba.

BAB V PENUTUP

Bab V berisi tentang (a) kesimpulan dari hasil penelitian dan (b) saran yang dengannya dapat dilakukan untuk penelitian lain selanjutnya yang lebih baik dan atau implementasi sistem yang lebih baik.