

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Penyakit jantung dan pembuluh darah atau penyakit kardiovaskular masih menjadi penyebab kematian tertinggi di dunia. Data Organisasi Kesehatan Dunia (WHO) menyatakan bahwa setiap tahunnya kematian akibat penyakit kardiovaskular mencapai lebih dari 17,8 juta [1]. Sedangkan data dari Kemenkes RI pada tahun 2023 di Indonesia angka kematian akibat penyakit ini mencapai 650.000 penduduk per tahun [2].

Prevalensi penyakit ini terus meningkat karena adanya kerusakan jantung yang berpotensi berkembang dari gagal jantung akut menjadi gagal jantung kronis. [3] menyatakan sekitar 1 dari 4 orang akan mengalami gagal jantung seumur hidup, yang memberikan dampak pada penurunan kondisi kesehatan, social dan ekonomi. Penyebab utama gagal jantung atau *Congestive Hearth Failure* (CHF) adalah berbagai penyakit kardiovaskuler, seperti hipertensi, kardiomiopati, dan penyakit jantung coroner, yang menyebabkan gangguan fungsi ventrikel kiri dalam bentuk penurunan kemampuan kontraksi, relaksasi jantung, yang berdampak pada penurunan curah jantung.

Pengambilan keputusan serta proses klasifikasi atau prediksi merupakan bagian penting dalam tindakan medis, terutama dalam mendeteksi penyakit. Keputusan yang didasarkan pada data yang akurat dapat menghasilkan prediksi penyakit yang lebih tepat sasaran. Deteksi dini dan prediksi risiko penyakit jantung

dapat membantu menurunkan tingkat kematian melalui intervensi yang lebih cepat dan tepat. Bagi tenaga kesehatan, mendiagnosis pasien dengan tepat merupakan suatu tantangan besar, karena prediksi yang tidak akurat dapat menimbulkan konsekuensi yang tidak diharapkan.

Seiring dengan perkembangan teknologi kecerdasan buatan, Machine learning, sebagai salah satu cabang kecerdasan buatan (AI), merupakan salah satu solusi yang efektif untuk mengoptimalkan analisis data yang kompleks. Khususnya algoritma Random Forest dengan kemampuannya mempelajari pola-pola dalam dataset besar, dapat mengidentifikasi kombinasi faktor risiko seperti tekanan darah, kadar kolesterol, dan hasil elektrokardiogram yang paling signifikan dalam memprediksi kejadian penyakit jantung. Algoritma ini juga mampu mengatasi kompleksitas data medis yang melibatkan banyak variabel untuk membangun model prediksi yang lebih akurat. Hal ini memungkinkan dokter untuk mengidentifikasi pasien yang berisiko tinggi lebih dini dan memberikan intervensi yang tepat waktu, sehingga meningkatkan peluang untuk mencegah serangan gagal jantung.

Beberapa Penelitian terkait dengan machine learning seperti yang dilakukan oleh [39] dengan mengevaluasi empat algoritma pembelajaran machine learning *Random Forest*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) dan Regresi Logistik menghasilkan akurasi terbaik menggunakan Random Forest dengan akurasi 87.7%. Sedangkan [40] menggunakan algoritma Random Forest mengalami peningkatan dari 82,6% menggunakan Teknik optimasi K-Fold dan GridSearchCV menjadi 85%. Berdasarkan beberapa penelitian diatas terlihat machine learning dapat digunakan dalam prediksi gagal jantung dengan tingkat

akurasi yang baik. Namun, belum ada yang melakukan penelitian prediksi gagal jantung menggunakan metode Feature Selection.

Berdasarkan tingginya angka kematian akibat penyakit jantung dan potensi besar dari penerapan kecerdasan buatan dalam bidang kesehatan, maka penulis mengambil penelitian dengan judul “Optimisasi Prediksi Gagal Jantung Menggunakan Algoritma Random Forest Classifier”. Penelitian ini bertujuan untuk mengoptimalkan prediksi gagal jantung menggunakan algoritma Random Forest. Dengan menggunakan sejumlah fitur dalam dataset penelitian ini berupaya membangun model prediksi yang akurat dan dapat diandalkan.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah di atas, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana Pengaruh Machine Learning dalam Prediksi Penyakit Jantung menggunakan Random Forest Classifier?
2. Bagaimana cara mengoptimasikan prediksi gagal jantung menggunakan algoritma Random Forest Classifier dengan metode Feature Selection?
3. Bagaimana efektifitas penerapan algoritma Random Forest Classifier dengan metode Feature Selection dalam mengklasifikasikan penyakit gagal jantung?

1.3 BATASAN MASALAH

Untuk lebih memfokuskan penelitian, maka perlu dilakukan pembatasan masalah sebagai berikut

1. Dataset terdiri dari 918 observasi dengan 12 atribut yang merupakan integrasi antar 11 fitur umum dari beberapa observasi, diantaranya: 303 observasi Cleveland, 294 observasi Hungarian, 123 observasi Switzerland, 200 observasi Long Beach VA, dan 270 observasi Stalog (Heart) Data Set. Total keseluruhan terdapat 1190 observasi namun diantaranya terdapat 272 data observasi duplikat / sama.
2. Bahasa pemrograman yang digunakan adalah bahasa Python dengan memanfaatkan beberapa library yang telah ada, yaitu: pandas, numpy, matplotlib, seaborn, matplotlib, sklearn.
3. Algoritma yang digunakan adalah Random Forest.
4. Hasil penelitian difokuskan pada penerapan algoritma Random Forest yang kemudian akan dilakukan pemilihan fitur berdasarkan fitur terpenting pada dataset. Data independen terdiri dari 11 kolom hasil integrasi fitur penyakit jantung yang umum.

1.4 TUJUAN DAN MANFAAT

1.4.1 TUJUAN PENELITIAN

Sesuai dengan rumusan masalah di atas, tujuan penelitian ini sebagai berikut :

1. Mengidentifikasi faktor-faktor signifikan dalam Prediksi terjadinya gagal jantung menggunakan Feature Selection berdasarkan data yang digunakan,
2. Menguji algoritma Random Forest dalam memprediksi terjadinya gagal Jantung.
3. Mengevaluasi keefektifan penggunaan Machine Learning algoritma random forest dengan metode Feature Selection dalam prediksi Gagal Jantung.

1.4.2 MANFAAT PENELITIAN

Dari hasil penelitian yang dilakukan, penulis berharap bahwa hasil penelitian ini dapat bermanfaat sebagai berikut :

1. Bagi Dokter, Hasil penelitian ini dapat membantu dokter dalam mengambil keputusan klinis yang lebih tepat terkait dengan pengobatan dan manajemen pasien gagal jantung.
2. Bagi Petugas Kesehatan, Hasil penelitian ini dapat membantu para petugas kesehatan untuk memudahkan dalam memprediksi pasien gagal jantung
3. Bagi peneliti selanjutnya, Penelitian ini dapat memberikan kontribusi pada pengembangan metode prediksi dini gagal jantung.

1.5 SISTEMATIKA PENULISAN

Secara garis besar, penulisan laporan proyek penelitian ini terdiri dari enam bab. Gambaran umum mengenai penulisan ilmiah dapat dilihat dalam sistematika penulisan sebagai berikut:

BAB I : PENDAHULUAN

Dalam bab pendahuluan ini berisi tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Dalam bab Landasan Teori ini berisi tentang materi terkait penelitian baik dari penyakit Gagal Jantung, Machine Learning, Algoritma dan pembahasan dari Dataset dan penelitian-penelitian Sebelumnya.

BAB III : METODOLOGI PENELITIAN

Dalam bab ini membahas tentang Metode Machine Learning yang digunakan dalam penelitian baik dari Algoritma yang digunakan dan Metode pre-processing data sampai pada Hasil Prediksi menggunakan dataset.

BAB IV : ANALISIS

Data diolah menggunakan algoritma random forest classifier dengan metode feature selection. Setelah pengolahan data dianalisis dari hasil penelitian baik dari hasil prediksi, Fitur-fitur yang berpengaruh, F1-Score, Recall dan lainnya menghasilkan hasil analisis.

BAB V : KESIMPULAN

Merupakan hasil penelitian secara keseluruhan terkait penelitian Optimasi Prediksi Gagal Jantung Menggunakan Algoritma Random Forest Classifier. Disertai dengan saran-saran diberikan untuk penelitian selanjutnya.