

BAB I PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung adalah kondisi Ketika bagian jantung yang meliputi pembuluh darah jantung, selaput jantung, katup jantung, dan otot jantung mengalami gangguan yang dapat disebabkan oleh berbagai hal, seperti sumbatan pada pembuluh darah jantung, peradangan, infeksi, atau kelainan bawaan[1]. Penyakit jantung merupakan salah satu penyakit yang menjadi masalah kesehatan masyarakat di seluruh dunia. Menurut data dari World Health Organization (WHO), penyakit jantung menjadi penyebab kematian nomor satu di dunia, dengan lebih dari 17,9 juta kematian setiap tahunnya. Di Indonesia sendiri, penyakit jantung juga menjadi penyebab kematian utama, tren peningkatan penyakit jantung yakni 0,5% pada 2013 menjadi 1,5% pada 2018 dan terus meningkat setiap tahunnya[2]. Hal ini menunjukkan pentingnya upaya pencegahan dan deteksi dini penyakit jantung untuk mengurangi angka kematian akibat penyakit tersebut.

Dua teknik klasifikasi yang sering digunakan dalam pengolahan data medis adalah algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM)[3], [4], [5], [6]. Kedua algoritma ini memiliki keunggulan masing-masing dalam menangani dataset kompleks seperti data medis. *Random Forest* adalah algoritma pembelajaran berbasis pohon yang memanfaatkan kekuatan dari serangkaian pohon keputusan untuk membuat prediksi yang akurat. Di sisi lain, SVM merupakan metode klasifikasi yang membangun hyperplane optimal untuk memisahkan data ke dalam kelas-kelas yang berbeda.

Dalam dunia kedokteran, keakuratan dan kecepatan dalam mendiagnosis penyakit sangat krusial. *Random Forest*, sebagai salah satu algoritma pembelajaran mesin, telah menunjukkan efektivitasnya dalam mengolah data medis. Algoritma ini bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk mendapatkan prediksi yang lebih stabil dan akurat[5].

Sama halnya dengan *Random Forest* (RF), *Support Vector Machine* (SVM) juga memiliki peran penting dalam pengolahan data medis, dengan kemampuan

mengoptimalkan hyperplane antar kelas. *Support Vector Machine* (SVM) mampu menangani dataset besar dan kompleks, serta beradaptasi dengan berbagai jenis kernel untuk meningkatkan performa model[7].

Namun, salah satu tantangan utama dalam pengolahan data medis adalah ketidakseimbangan data. Ketidakseimbangan ini terjadi ketika jumlah data pada kelas mayoritas jauh lebih besar dibandingkan kelas minoritas. Ketidakseimbangan data ini dapat menyebabkan model lebih fokus pada kelas mayoritas dan kurang akurat dalam mengenali kelas minoritas[8]. Untuk mengatasi masalah tersebut, penelitian menerapkan SMOTE dengan harapan untuk meningkatkan performa algoritma klasifikasi, khususnya dalam mengenali pasien dengan risiko penyakit jantung.

SMOTE, atau *Synthetic Minority Oversampling Technique*, adalah metode yang digunakan untuk mengatasi ketidakseimbangan kelas dalam kumpulan data dengan menghasilkan sampel sintetis untuk kelas minoritas[9].

Oleh karena itu, tujuan dari penelitian ini adalah untuk melakukan perbandingan antara algoritma *Random Forest* dan *Support Vector Machine* dalam klasifikasi penyakit jantung dengan bantuan SMOTE. Dengan pemahaman yang lebih mendalam dari masing-masing algoritma, diharapkan dapat memberikan pengetahuan untuk pengembangan sistem klasifikasi penyakit jantung yang lebih efektif dan handal.

1.2 Rumusan Masalah

Dalam konteks ini, tujuan penelitian ini adalah untuk membandingkan hasil klasifikasi antara *Random Forest* (RF) dan *Support Vector Machine* (SVM).

Beberapa pertanyaan yang akan dijawab dalam penelitian ini meliputi:

- Bagaimana perbandingan performa klasifikasi penyakit jantung antara algoritma RF dan SVM?
- Apakah teknik *oversampling* mempengaruhi hasil kinerja model?
- Bagaimana algoritma RF dan SVM dapat diterapkan untuk menganalisis faktor risiko penyakit jantung?
- Seberapa efektif penggunaan algoritma RF dan SVM dalam menganalisis risiko penyakit jantung?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

- Penelitian ini akan menggunakan dataset UCI Heart Disease Data, Heart Attack Analysis & Prediction Dataset, dan Heart Disease Dataset sebagai sumber data untuk prediksi penyakit jantung. Data lainnya tidak akan dipertimbangkan.
- Analisis kinerja algoritma klasifikasi SVM dan Random Forest akan difokuskan pada akurasi prediksi, serta evaluasi lainnya seperti presisi, recall, dan F1-score.
- Variabel-variabel yang digunakan dalam dataset UCI Heart Disease Data akan dipilih secara cermat berdasarkan relevansinya dalam prediksi penyakit jantung.
- Pengujian kinerja algoritma klasifikasi akan dilakukan dengan menggunakan metode validasi silang untuk menghindari overfitting.
- Penelitian ini tidak akan memperhitungkan faktor-faktor eksternal seperti faktor lingkungan atau genetik yang dapat mempengaruhi prediksi penyakit jantung.
- Analisis hasil akan dibatasi pada perbandingan langsung antara kinerja SVM dan Random Forest dalam memprediksi penyakit jantung, tanpa mempertimbangkan faktor-faktor eksternal lainnya.
- Penelitian ini tidak akan membahas implementasi atau optimisasi teknis dari algoritma klasifikasi SVM dan Random Forest, fokusnya adalah pada perbandingan kinerja keduanya dalam konteks prediksi penyakit jantung.
- Penelitian ini tidak akan memprediksi tingkatan penyakit jantung.
- Penggunaan SMOTE pada penelitian ini hanya untuk mengetahui apakah hasil evaluasi algoritma akan lebih baik jika kelas data seimbang.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

- Membandingkan performa algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM) dalam klasifikasi penyakit jantung.
- Mengidentifikasi algoritma yang memberikan hasil paling akurat untuk mendukung proses diagnosis dan pengambilan keputusan medis.
- Mengevaluasi kelebihan dan kekurangan masing-masing algoritma berdasarkan metrik performa, seperti akurasi, presisi, recall, dan F1-score.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang dapat dihasilkan dari penelitian ini meliputi:

- Memberikan pemahaman yang lebih mendalam tentang kinerja algoritma klasifikasi *Random Forest* (RF) dan *Support Vector Machine* (SVM) dalam klasifikasi penyakit jantung.
- Memberikan rekomendasi mengenai algoritma klasifikasi yang paling efektif dalam prediksi penyakit jantung.
- Membantu mendorong pengembangan metode prediksi penyakit jantung yang lebih akurat dan efisien.

1.6 Sistematika Penulisan

Penelitian ini terdapat susunan sistematika penulisan sebagai berikut:

1.6.1 BAB I

Membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian tentang evaluasi performa algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM) pada data medis.

1.6.2 BAB II

Membahas tentang studi literatur dan dasar teori yang berhubungan dengan evaluasi performa algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM).

1.6.3 BAB III

Membahas tentang objek penelitian yaitu dataset UCI Heart Disease Data, Heart Attack Analysis & Prediction Dataset, dan Heart Disease Dataset. Kemudian alur penelitian atau *flowchart* mulai dari input dataset, pre-processing dataset, split dataset, train model, dan evaluate model.

1.6.4 BAB IV

Membahas hasil yang didapatkan dari analisis *Random Forest* (RF) dan *Support Vector Machine* (SVM) dengan bantuan SMOTE.

1.6.5 BAB V

Membahas tentang kesimpulan dan saran dari penelitian yang akan diberikan baik kepada pengembang selanjutnya maupun peneliti yang akan mengambil penelitian ini sebagai referensi penelitiannya.

