

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi telah mengubah banyak aspek kehidupan, termasuk bidang medis. Kemajuan utamanya adalah penggunaan teknologi untuk mengumpulkan dan menganalisis data guna mendukung pengambilan keputusan yang lebih baik[1]. Teknologi informasi berperan penting dalam memprediksi risiko kesehatan dan membantu meningkatkan upaya pencegahan dan pengobatan penyakit. Misalnya prediksi penyakit jantung yang dipengaruhi oleh faktor faktor klinis, kini dapat dianalisis menggunakan pendekatan berbasis data[2].

Penyakit jantung adalah salah satu penyakit paling mematikan di dunia. Menurut laporan data dari berbagai organisasi kesehatan dunia seperti WHO, penyakit jantung mengakibatkan kematian setiap tahunnya sekitar 17,9 juta setiap tahunnya, dengan prediksi tersebut akan terus meningkat menjadi 23 juta pada tahun 2030[3]. Faktor faktor seperti diabetes, hipertensi dan kolesterol tinggi sangat berkontribusi secara signifikan terhadap perkembangan risiko penyakit jantung, oleh karena itu penting untuk mengembangkan model *prediktif* yang dapat mengidentifikasi individu berisiko tinggi sehingga *intervensi* dini dapat dilaksanakan[4].

Dalam beberapa tahun terakhir, algoritma *machine learning* menjadi semakin populer dalam analisis data medis, terutama untuk tugas *prediktif*[5]. Algoritma seperti *Support Vector Machine* juga digunakan secara luas dalam bidang medis, *Support Vector Machine* memiliki kelebihan dalam membangun *hiperbidang* pemisah untuk memaksimalkan jarak antar dua kelas yang berbeda, sehingga efektif dalam mengklasifikasikan himpunan data kompleks tetapi *Support Vector Machine* memiliki tantangan dalam menangani data yang besar karena kompleksitas dan komputasinya[6].

Sedangkan *Gradient Boosting Machine* merupakan algoritma dengan cara kerja membangun model *prediktif* dengan bertahap, Dimana setiap modelnya berusaha untuk memperbaiki kesalahan dari model sebelumnya. Keunggulan dari algoritma *Gradient Boosting Machine* memiliki keunggulan utama menangani berbagai jenis jenis data terutama pada kasus *klasifikasi* sedangkan *Gradient Boosting Machine* juga memiliki kelemahan yang terletak pada

waktu komputasi yang tinggi dibandingkan algoritma sederhana lain. *Gradient Boosting Machine* juga sudah sangat banyak digunakan untuk prediksi kasus medis lainnya seperti diabetes dan cancer[7].

Pemahaman yang lebih baik tentang faktor penyakit jantung dari sudut pandang klinis dan gaya hidup dapat memberikan manfaat dalam meningkatkan kesehatan Masyarakat dan dapat membantu mengurangi angka kematian penyakit jantung[8]. Penelitian ini menggunakan algoritma *Support Vector Machine* dan *Gradient Boosting Machine* untuk memberikan analisis tentang penyakit jantung dan membuat rekomendasi berdasarkan data yang dapat digunakan untuk memberikan upaya pencegahan yang lebih efektif dan tepat sasaran.

Berdasarkan uraian di atas, penelitian ini berjudul "Analisis Kinerja Algoritma *Support Vector Machine* dan *Gradient Boosting Machine* untuk Prediksi Risiko Penyakit Jantung." Penelitian ini diharapkan dapat berkontribusi dalam memahami faktor penyakit jantung dan memperkuat upaya pencegahan melalui pendekatan berbasis data.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian ini maka permasalahan dapat dirumuskan adalah:

1. Bagaimana kinerja algoritma *Support Vector Machine* dan *Gradient Boosting Machine* dalam memprediksi penyakit jantung?
2. Bagaimana pengaruh *SMOTE*, *StandardScaler*, dan *Correlation Thresholding* terhadap performa algoritma *Gradient Boosting Machine* dan *Support Vector Machine* dalam memprediksi penyakit jantung?

1.3 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah yang sudah dipaparkan sebelumnya, ada batasan yang telah ditentukan untuk mempermudah dalam proses pembuatan penelitian ini, yaitu:

1. Penelitian tidak menggunakan data pasien dari rumah sakit tertentu.
2. Analisis yang dilakukan bersifat *deskriptif*.
3. Dataset yang digunakan terbatas pada data *Kaggle* yang berfokus pada penyakit jantung dan tidak menyertakan data penyakit lain.
4. Data yang dianalisis tidak memperhitungkan faktor genetik atau riwayat kesehatan lainnya, namun hanya mencakup faktor faktor klinis.

5. Penelitian membandingkan kinerja algoritma *Support Vector Machine* dan *Gradient Boosting Machine* dalam memprediksi penyakit jantung.
6. Bahasa pemrograman yang digunakan pada penelitian ini adalah *Python*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah menggunakan algoritma *machine learning* untuk memecahkan masalah terkait prediksi penyakit jantung. Tujuan yang akan dicapai adalah:

1. Membandingkan kinerja algoritma *Support Vector Machine* dan *Gradient Boosting Machine* dalam memprediksi penyakit jantung dan menentukan algoritma mana yang lebih efektif.
2. Membandingkan akurasi setiap algoritma untuk prediksi penyakit jantung berdasarkan dataset yang digunakan.

1.5 Manfaat Penelitian

Adapun manfaat dalam penyusunan penelitian skripsi ini adalah:

1. Manfaat teoritis:

- a. Perkembangan keilmuan: Penelitian ini diharapkan memberi kontribusi terhadap pengembangan keilmuan di bidang kesehatan, khususnya dalam penerapan algoritma *machine learning* untuk memprediksi penyakit jantung.
- b. Peningkatan pemahaman: Hasil dari penelitian ini diharapkan dapat meningkatkan pemahaman tentang hubungan antara faktor-faktor klinis dengan penyakit jantung serta memperbanyak literatur yang ada.
- c. Model Teoritis: Penelitian ini dapat memberikan dasar bagi penelitian lebih lanjut untuk mengembangkan model *machine learning* yang lebih kompleks tentang faktor-faktor yang mempengaruhi penyakit jantung.

2. Manfaat Praktis:

- a. Perencanaan Kebijakan Kesehatan: Hasil penelitian diharapkan dapat digunakan oleh pengambilan kebijakan dan organisasi kesehatan untuk mengembangkan strategi pencegahan penyakit jantung yang lebih efektif berdasarkan data dan analisis yang valid.
- b. Alat Prediksi Risiko: Penelitian ini diharapkan dapat berpotensi menciptakan alat prediksi risiko dini yang dapat digunakan oleh penyedia layanan kesehatan untuk mengidentifikasi pasien yang berisiko tinggi terkena penyakit jantung.

- c. Edukasi Masyarakat: Temuan dari penelitian dapat digunakan untuk mengedukasi masyarakat tentang faktor-faktor yang meningkatkan penyakit jantung sehingga masyarakat dapat lebih sadar melakukan tindakan pencegahan demi kesehatannya.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab I pendahuluan ini memberikan Gambaran umum tentang masalah yang akan dibahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab II tinjauan Pustaka tentang pembahasan penelitian sebelumnya yang sudah dilakukan peneliti lain sebagai referensi penelitian ini dijalankan dan berisikan dasar teori-teori yang digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab III menguraikan tentang alur penelitian dalam mendapatkan hasil dan juga pengujian yang dilakukan dengan menerapkan metode *Support Vector Machine* dan *Gradient Boosting Machine*.

BAB IV HASIL DAN PEMBAHASAN

Bab IV berisi tentang tahapan yang dilakukan dalam penerapan alur yang telah dirancang meliputi rancangan implementasi didalam *code* untuk mendapatkan hasil dari perbandingan kedua metode.

BAB V PENUTUP

Bab V berisi kesimpulan yang diperoleh dari perumusan masalah dan saran yang diharapkan dari penelitian ini metode *Support Vector Machine* dan *Gradient Boosting Machine* bisa membantu untuk menganalisis kasus yang serupa di masa yang akan datang.