

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Penyakit jantung atau penyakit kardiovaskuler adalah penyakit yang disebabkan oleh gangguan fungsi jantung dan pembuluh darah, beberapa contoh dari penyakit tersebut adalah penyakit jantung koroner, penyakit serebrovaskular hipertensi dan stroke[1]. Menurut WHO, setiap tahunnya penyakit jantung adalah salah satu penyebab utama kematian secara global. Diperkirakan 17.9 juta orang meninggal dikarenakan penyakit jantung pada tahun 2019, angka ini merepresentasikan 32% dari semua kematian secara global. Dari semua kematian tersebut, 85% disebabkan oleh serangan jantung dan stroke. Hal ini diperparah dengan fakta yang menyatakan bahwa 1/3 penyakit jantung terjadi di negara yang berpenghasilan rendah dan menengah[2]. Di Indonesia sendiri, menurut data Riset Kesehatan Dasar (Riskesdas) pada tahun 2018, angka prevalensi penyakit jantung sudah mencapai 1.5% atau 15 dari 1000 orang di Indonesia menderita penyakit jantung[3].

Dengan tingginya angka penderita penyakit jantung dan kematian yang disebabkan oleh penyakit jantung, maka perlu dilakukan langkah preventif dini agar resiko terjadinya penyakit jantung dapat diminalisir dan efek-efek yang fatal dan mematikan dapat dicegah serta proses pengobatan dapat dimulai[4]. Beberapa metode umum untuk melakukan deteksi penyakit jantung secara akurat adalah melalui *Electrocardiogram (ECG)*, *Cardiac Magnetic Resonance Imaging (MRI)*, *Echocardiogram (ultrasound)*, *Stress test*, dan *Angiography*[4]. Akan tetapi, salah satu kelemahan dari metode diagnostik tersebut adalah biaya dan aksesibilitas yang tinggi sehingga tidak terjangkau untuk seluruh masyarakat [5, 6, 7].

Oleh karena itu, penggunaan pembelajaran mesin dapat digunakan sebagai salah satu metode untuk mendeteksi penyakit jantung. Selain dapat mengurangi biaya diagnostik, implementasi pembelajaran mesin juga dapat digunakan untuk membantu meningkatkan akurasi serta keyakinan dokter atau ahli medis dalam memberikan diagnosis.

Algoritma pembelajaran mesin yang digunakan dalam penelitian ini adalah algoritma *Support Vector Machine (SVM)*. *Support Vector Machine* merupakan model pembelajaran mesin berbasis kernel untuk menyelesaikan masalah klasifikasi

dan regresi[8]. Kapabilitas dalam melakukan generalisasi dan kemampuan untuk memecahkan masalah klasifikasi biner membuat SVM lebih unggul dari model *supervised* lainnya [9, 10, 11]. Kekuatan SVM berasal dari kemampuannya untuk mempelajari pola klasifikasi data dengan akurasi dan reproduktifitas yang seimbang [12]. Selain itu, pemilihan algoritma tersebut juga didasarkan oleh penelitian terdahulu yang menggunakan dataset yang sama yaitu *Heart Disease UCI* menunjukkan algoritma *Support Vector Machine* terlihat unggul atau memiliki akurasi tertinggi dibandingkan dengan algoritma lain ketika mendeteksi penyakit jantung dengan tingkat akurasi 85 - 98% [13, 14].

Dalam implementasinya, algoritma SVM menggunakan bantuan kernel untuk memecahkan masalah nonlinear dengan memetakannya ke dimensi yang lebih tinggi terutama untuk data dengan jumlah fitur yang besar[15]. Salah satu kernel tersebut adalah kernel RBF atau *Radial Basis Function*. Berdasarkan penelitian terdahulu kernel RBF dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan kernel lain dengan menggunakan dataset yang digunakan di dalam penelitian ini [16, 17]. Hal ini disebabkan oleh dataset yang digunakan mengandung lebih dari satu fitur, di mana variabel hasil yang sesuai dalam ruang linier dua dimensi biasanya tidak dapat dipisahkan secara linier.[16].

Algoritma genetika merupakan salah satu algoritma metaheuristik yang dapat digunakan untuk menyelesaikan permasalahan *Combinatorial Optimization Problem* [18] yang merupakan jenis masalah dari proses seleksi fitur [19]. Berdasarkan penelitian terdahulu, penggunaan algoritma genetika untuk seleksi fitur terlihat dapat meningkatkan akurasi model secara cukup signifikan dengan peningkatan sekitar 5 - 20% [20, 21, 22]. Namun pada penelitian sebelumnya algoritma pembelajaran mesin yang digunakan adalah algoritma *Naïve Bayes* dengan dataset yang berbeda dengan yang digunakan di dalam penelitian ini. Oleh karena itu, algoritma pembelajaran mesin yang digunakan dalam penelitian ini adalah algoritma *Support Vector Machine* dengan algoritma genetika untuk melakukan seleksi fitur atau mereduksi fitur-fitur yang tidak terlalu signifikan sehingga dapat meningkatkan tingkat akurasi model.

1.2 Rumusan Masalah

Rumusan masalah yang ada dalam penelitian ini adalah sebagai berikut:

- Bagaimana cara mengimplementasi algoritma *Support Vector Machine* dengan seleksi fitur algoritma genetika untuk melakukan prediksi penyakit jantung?
- Bagaimana tingkat akurasi dan f-1 score dari algoritma *Support Vector Machine* dengan seleksi fitur algoritma genetika dalam melakukan prediksi penyakit jantung?

1.3 Batasan Permasalahan

Batasan-batasan masalah yang ada dalam penelitian ini adalah sebagai berikut:

- Dataset yang digunakan adalah data penyakit jantung (*Heart Disease UCI*) yang diperoleh dari website UCI Machine Learning Repository [23].
- Penelitian dibatasi oleh dua kelas klasifikasi yaitu klasifikasi positif dan klasifikasi negatif dan tidak mengklasifikasikan jenis penyakit jantung.
- Kernel yang digunakan algoritma SVM adalah kernel RBF.

1.4 Tujuan Penelitian

Beberapa tujuan dari penelitian yang dilakukan adalah sebagai berikut:

- Melakukan implementasi algoritma *Support Vector Machine* dengan seleksi fitur algoritma genetika untuk melakukan prediksi penyakit jantung.
- Mengukur hasil akurasi dan nilai f-1 score dari algoritma *Support Vector Machine* dengan seleksi fitur algoritma genetika untuk melakukan prediksi penyakit jantung.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah dengan menggunakan model pembelajaran mesin yang disertai dengan metode diagnosis yang mudah diakses dan relatif terjangkau seperti mesin EKG dapat mengurangi biaya yang dibutuhkan dan meningkatkan aksesibilitas dibandingkan dengan melakukan proses diagnostik konvensional yang umumnya dilakukan untuk mendeteksi penyakit jantung dengan asumsi tingkat akurasi deteksi yang baik kedepannya. Selain itu, model pembelajaran mesin juga dapat membantu meningkatkan akurasi serta keyakinan dokter atau ahli medis dalam memberikan diagnosis terutama pada fasilitas kesehatan dengan peralatan diagnostik yang kurang memadai.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab 1 berisikan latar belakang dari masalah yang diteliti, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian dan sistematika penulisan .
- Bab 2 LANDASAN TEORI
Bab 2 berisikan teori-teori yang melandasi penelitian yang meliputi penjelasan untuk penyakit jantung, *Support Vector Machine*, kernel, algoritma genetika dan *confusion matrix*.
- Bab 3 METODOLOGI PENELITIAN
Bab 3 berisikan metodologi yang digunakan di dalam penelitian dalam merancang model pembelajaran mesin untuk memecahkan masalah yang sudah dijabarkan.
- Bab 4 HASIL DAN DISKUSI
Bab 4 berisikan hasil implementasi dari penelitian dan hasil pengujian dari model pembelajaran mesin yang dihasilkan.
- Bab 5 SIMPULAN DAN SARAN
Bab 5 berisikan kesimpulan dari penelitian dan saran yang dapat diimplementasi untuk penelitian selanjutnya.