

BAB I

PENDAHULUAN

1.1. Latar Belakang

Klasifikasi merupakan sebuah metode yang digunakan untuk memasukkan data atau objek ke dalam suatu kelas. Klasifikasi termasuk dalam kelompok supervised learning, pada kelompok ini data belum diketahui label atau kelasnya. Klasifikasi adalah salah satu fungsi dasar dari data mining. Pada proses klasifikasi, data mining menggunakan informasi analisis yang sudah ada untuk melakukan proses prediksi atau pengelompokan tentang data baru. Data mining merupakan proses yang menggunakan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Salah satu metode klasifikasi data mining yang umum digunakan yaitu *Support Vector Machine*.

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi yang didapatkan dari tradisional machine learning. Selain SVM ada beberapa jenis metode tradisional machine learning antara lain decision tree, naïve bayes, find similar, akan tetapi dari beberapa metode tersebut SVM merupakan salah satu metode klasifikasi yang menghasilkan solusi terbaik dan nilai akurasi tertinggi yaitu 92% untuk 10 kategori (Resti & Selva Jumeilah, 2017). Sebagai perbandingan terdapat penelitian yang menerapkan metode tradisional machine learning yaitu penelitian mengenai perancangan system klasifikasi penyakit jantung menggunakan naïve bayes yang dilakukan oleh Mufti Ari Bianto dkk, menggunakan 303 data dengan 2 kelas menghasilkan nilai akurasi dengan rata-rata akurasi senilai 90,61% dengan nilai presisi 87,44% dan nilai recall 87,95% (Ari Bianto, 2019).

Terkait dengan penyakit, penyakit jantung merupakan organ tubuh yang berguna untuk memompa darah yang membawa oksigen dan nutrisi kepada seluruh sel tubuh guna kelangsungan hidup manusia, serta membantu proses metabolisme tubuh. Penyakit Jantung merupakan salah satu penyakit yang menjadi penyebab

kematian paling umum diseluruh dunia (Utomo & Mesran, 2020). Faktor utama yang mempengaruhi terjadinya penyakit jantung antara lain yaitu terjadinya tekanan darah tinggi, kolesterol tinggi, stress, konsumsi alkohol, gaya hidup, obesitas, diabetes.

Klasifikasi penyakit jantung merupakan proses pengelompokan data ke dalam kategori beresiko atau tidak beresiko mengalami penyakit jantung berdasar pada faktor-faktor Riwayat Kesehatan yang dimiliki. Pada penelitian ini, dataset yang digunakan yaitu Framingham Heart Study. Label dataset tersebut yaitu 1 menunjukkan bahwa seseorang beresiko mengalami penyakit jantung dalam kurun waktu 10 tahun, sedangkan nilai 0 menunjukkan bahwa seseorang tidak beresiko mengalami penyakit jantung. Klasifikasi ini memanfaatkan faktor beresiko yang berhubungan dengan penyakit jantung.

Menurut data dari WHO jumlah kematian karena penyakit jantung mencapai angka 26,6 juta dari populasi orang dewasa (Utomo & Mesran, 2020). Sedangkan di Indonesia menurut Riskesdas tahun 2018 pengidap penyakit jantung meningkat pada tiap tahunnya, yaitu 1,3 % pada pria dan 1,6% pada wanita. Jadi terdapat kurang lebih 2.784.064 jiwa di Indonesia yang memiliki riwayat penyakit jantung (Bahar & Moordiningsih, 2021). Diperlukan ketepatan akurasi yang baik dalam mengklasifikasikan pasien penyakit jantung sebagai deteksi dini agar tidak terjadi kesalahan dalam memprediksi suatu penyakit, khususnya penyakit jantung.

Beberapa penelitian sebelumnya telah dilakukan, contohnya Penelitian oleh Muhammad Febrian Dion Tino , Herliyan Hasanah, Tri Djoko Santoso membandingkan algoritma Support Vector Machine(SVM) dan Neural Network untuk Klasifikasi penyakit jantung. Penelitian ini menunjukkan dengan menggunakan metode Neural Network memiliki hasil akurasi yang lebih rendah yaitu 82% dan nilai AUC sebesar 0,823% (Dion et al., n.d.).

Klasifikasi penyakit kanker payudara menggunakan metode Support Vector Machine dengan Backward Elimination menghasilkan nilai akurasi 95,65% dan AUC sebesar 1,000 dan termasuk ke dalam kategori Excellent Classification. Sebelumnya nilai akurasi telah diuji tanpa menggunakan Backward Elimination dan AUC sebesar 0,700 menghasilkan nilai 65,22% sehingga dengan menerapkan

seleksi fitur Backward Elimination dapat meningkatkan nilai akurasi sebesar 30,43% (Resmiati & Arifin, n.d.).

Penelitian oleh Hilda Apriyani dan Kurniati membandingkan metode Naïve Bayes dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. Penelitian tersebut menunjukkan bahwa dengan metode Support Vector Machine menghasilkan nilai akurasi yang lebih tinggi yaitu 96,27% dan tingkat eror sebesar 3,72%. Pada algoritma SVM menggunakan dua kernel polynomial. Dengan ini dapat menggambarkan bahwa metode Support Vector Machine memiliki keunggulan dalam klasifikasi penyakit diabetes. Hal tersebut membuktikan bahwa dengan metode Support Vector Machine dapat digunakan dalam mendukung praktik medis dan pengambilan keputusan yang berkaitan dengan medis salah satunya yaitu diabetes militus (Apriyani, 2020).

Dari beberapa contoh penelitian diatas, penelitian ini bertujuan untuk melakukan klasifikasi penyakit jantung. Metode yang digunakan yaitu *Support Vector Machine* (SVM) dengan menambahkan seleksi fitur Information Gain serta penanganan ketidakseimbangan data guna meningkatkan nilai akurasi yang dihasilkan. Data yang digunakan berasal dari kaggle berjumlah 4.238 data.

1.2. Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang tersebut dapat disimpulkan bahwa pada sektor kesehatan juga memerlukan adanya sistem prediksi dengan menggunakan *machine learning*. Hal tersebut digunakan untuk mendeteksi resiko penyakit jantung. Maka dari itu perlu ditentukan metode *machine learning* serta pemilihan fitur yang tepat guna mendapatkan nilai keakuratan terbaik. Pemilihan fitur digunakan untuk memilih fitur yang memiliki nilai kontribusi yang signifikan dalam menentukan resiko penyakit jantung. Selain itu ditemukan adanya ketimpangan data yang ada pada dataset, maka dari itu selain adanya pemilihan fitur juga harus ada penanganan untuk ketimpangan data tersebut guna meningkatkan kinerja dari metode *machine learning*.

1.2.2 Metode Usulan

Berdasarkan uraian masalah diatas solusi yang diusulkan pada penelitian ini adalah menggunakan Metode *Support Vector Machine(SVM)* dengan menggunakan seleksi fitur *Information Gain* serta penanganan ketidakseimbangan data menggunakan Teknik *oversampling Synthetic Minority Oversampling Technique (SMOTE)*.

1.2.3 Pertanyaan Penelitian

1. Berapakah nilai akurasi yang dihasilkan oleh model yang dibangun menggunakan metode *Support Vector Machine* ?
2. Apakah dengan menggunakan seleksi fitur *Information Gain* dan Teknik *oversampling SMOTE* dapat memberikan pengaruh terhadap nilai akurasi klasifikasi penyakit jantung dengan metode SVM?

1.3. Batasan Masalah

Adapun batasan-batasan dalam penelitian ini sebagai berikut :

1. *Dataset* yang digunakan merupakan data yang diambil dari *Kaggle*, dengan jumlah 4.238 data. *Dataset* ini terdiri dari lima belas fitur dan satu fitur sebagai diagnosa.
2. Hasil klasifikasi penyakit jantung dibagi ke dalam dua kelas yaitu kelas Resiko dan kelas Tidak Resiko.
3. Metode usulan yang digunakan yaitu SVM.
4. Seleksi fitur yang digunakan yaitu *Information Gain*.
5. Teknik *oversampling* yang digunakan yaitu Teknik SMOTE.

1.4. Tujuan Dan Manfaat

1.4.1 Tujuan

Berdasarkan pertanyaan dan uraian rumusan masalah diatas, maka penelitian ini bertujuan untuk :

1. Mengetahui nilai akurasi model dengan metode *Support Vector Machine*.

2. Mengidentifikasi pengaruh nilai akurasi model dengan penggunaan seleksi fitur *Information Gain* dan Teknik *oversampling* SMOTE.

1.4.2 Manfaat

Manfaat dari penelitian ini adalah untuk menghasilkan model klasifikasi yang dapat mengidentifikasi penyakit jantung. Serta diharapkan bisa mendapatkan hasil akurasi yang terbaik. Selain itu diharapkan penelitian ini bisa berkontribusi penting untuk penelitian selanjutnya dan dapat memberikan dampak positif di bidang ini.

1.5. Sistematika Laporan

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang penelitian, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah penelitian, serta mengenai sistematika penulisan laporan penelitian.

BAB II TINJAUAN PUSTAKA

Pada bab ini menjelaskan mengenai landasan teori yang digunakan dalam penelitian serta penelitian terdahulu yang berkaitan dengan penelitian yang diajukan penulis.

BAB III METODE USULAN

Pada bab ini menjelaskan mengenai metode yang diusulkan penulis mencakup tentang arsitektur sistem, *dataset* yang akan digunakan, dan tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan mengenai implementasi sistem dengan menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain* dan Teknik SMOTE. Hasil ini juga mencakup lingkungan uji coba.

BAB V PENUTUP

Bagian ini mencakup kesimpulan dari penelitian, yang disesuaikan dengan rumusan yang telah dibuat dan saran untuk penelitian berikutnya.