BAB I
PENDAHULUAN

1.1 Latar Belakang

Penyakit yang mempengaruhi jantung (kardio) dan pembuluh darah (vaskular) tubuh disebut penyakit kardiovaskular. Sedangkan penyakit jantung lebih spesifik dan mengacu pada kondisi yang mempengaruhi fungsi jantung secara langsung, seperti penyakit jantung koroner, aritmia, atau gagal jantung [1].Penyakit jantung masih menjadi masalah kesehatan yang serius secara global [2]. Menurut data Organisasi Kesehatan Dunia (WHO), pada tahun 2023 menyebutkan bahwa setiap tahunnya kematian akibat penyakit jantung mencapai lebih dari 17,8 juta. Selain itu, data dari Kemenkes Republik Indonesia pada tahun 2023 di Indonesia angka kematian akibat penyakit jantung mencapai 650.000 penduduk per tahun [3]. Fakta ini menunjukkan bahwa penyakit jantung adalah masalah kesehatan serius dengan angka kejadian tinggi, yang memerlukan adanya upaya deteksi dini untuk menurunkan angka kematian dan meningkatkan kualitas pengelolaan penyakit ini.

Deteksi dini penyakit jantung tidak hanya menghabiskan waktu dan biaya, tetapi juga kompleks karena adanya gejala dan faktor risiko beragam, seperti tekanan darah, Kolesterol, gula darah, dan riwayat kesehatan. Manual diagnosis seringkali rumit dan rentan terhadap kesalahan. Oleh karena itu, machine learning (ML) hadir untuk menjadi solusi yang efektif dalam deteksi dini penyakit jantung, dengan kemampuan untuk menganalisis data besar dan mengidentifikasi pola yang kompleks [4]. Dengan mengolah berbagai parameter kesehatan pasien, machine learning dapat mempercepat proses diagnosis dan memberikan hasil yang konsisten.

Dalam konteks deteksi penyakit jantung, klasifikasi menggunakan metode supervised learning menjadi pendekatan utama karena menggunakan data berlabel dari data pasien dan status penyakit jantungnya. Penelitian ini menggunakan beberapa algoritma antara lain Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbors (KNN), dan Logistic Regression. Setiap algoritma memiliki keunggulan tersendiri tergantung pada karakteristik data dan tujuan prediksi.

Adapun keunggulan algoritma KNN efektif untuk kumpulan data kecil dan menengah, tetapi sensitif terhadap noise dan pemilihan nilai k yang sesuai. Di sisi lain, algoritma Naive Bayes memiliki kecepatan komputasi dan efisiensi yang baik untuk kumpulan data dengan fitur independen, namun kurang optimal untuk hubungan fitur yang kompleks. Sementara itu, algoritma SVM sangat efektif dalam klasifikasi biner dan pemprosesan data berdimensi tinggi, tetapi sensitif terhadap pemilihan parameter kernel dan regularisasi. Dan algoritma logistic regression mudah dilakukan dan diinterpretasikan tetapi terbatas pada hubungan linier antara fitur. Implementasi setiap algoritma memiliki tantangan dalam pemilihan parameter optimal karena sensitivitas hyperparameter yang berbeda-beda, yang sangat mempengaruhi model akurasi terutama pada data kesehatan yang sering kali tidak seimbang [5].

Untuk mengatasi tantangan tersebut, optimasi hyperparameter menggunakan GridSearchCV adalah solusi yang relevan. GridSearchCV secara sistematis mencari kombinasi parameter terbaik untuk setiap algoritma dengan cross-validation [6]. Misalnya untuk SVM pada kernel dan nilai C, Naive Bayes pada parameter smoothing , KNN dapat dioptimalkan pada nilai k dan jarak metric, dan Logistic Regression pada parameter regularisasi.

Berdasarkan uraian diatas, penelitian ini bertujuan untuk mengoptimalkan pemilihan hyperparameter pada algoritma machine learning untuk klasifikasi penyakit jantung, sehingga dapat meningkatkan akurasi model untuk deteksi dini dan memberikan solusi yang lebih efektif dalam bidang kesehatan.

1.2 Rumusan Masalah

Berdasarkan uraian yang dijelaskan dalam latar belakang, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa algoritma Support Vector Machine, Naive Bayes, K-Nearest Neighbors, dan Logistic Regression dalam klasifikasi penyakit jantung setelah optimasi hyperparameter menggunakan GridSearchCV?
2. Berapa tingkat akurasi model klasifikasi penyakit jantung yang dioptimalkan melalui GridSearchCV dibandingkan dengan model tanpa optimasi?

3. Seberapa efektifkah teknik GridSearchCV dalam meningkatkan performa berbagai algoritma machine learning untuk mendeteksi dini penyakit jantung?

1.3 Batasan Masalah

Dalam penelitian ini, batasan masalah ditetapkan untuk menghindari meluasnya permasalahan yang akan dibahas yaitu:

1. Penelitian ini hanya fokus pada empat algoritma Machine Learning, yaitu Support Vector Machine, Naive Bayes, K-Nearest Neighbors, dan Logistic Regression,
2. Optimasi hyperparameter dalam penelitian ini hanya menggunakan metode GridSearchCV.
3. Penelitian ini mengutamakan akurasi sebagai indikator utama untuk menunjukkan hasil model yang dioptimasi. Metrik lain, seperti presisi, recall, dan F1-score dihitung sebagai tambahan untuk memberikan gambaran yang lebih komprehensif tentang kinerja model, namun tidak akan menjadi fokus utama.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui kinerja dan efektivitas empat algoritma Machine Learning (Support Vector Machine, Naive Bayes, K-Nearest Neighbors, dan Logistic Regression) dalam klasifikasi penyakit jantung setelah optimasi hyperparameter menggunakan GridSearchCV.
2. Menganalisis tingkat akurasi model klasifikasi penyakit jantung yang dioptimalkan melalui GridSearchCV dan membandingkannya dengan model tanpa optimasi untuk menilai peningkatan Kinerja.
3. Mengevaluasi efektivitas teknik GridSearchCV dalam meningkatkan kinerja berbagai algoritma Machine Learning dalam mendeteksi dini penyakit jantung.

1.5 Manfaat Penelitian

Dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi pengetahuan dalam pengembangan model klasifikasi penyakit jantung menggunakan optimasi hyperparameter GridSearchCV pada machine learning yang dapat menjadi referensi untuk penelitian selanjutnya.
2. Membantu dokter dan tenaga medis dalam melakukan deteksi awal penyakit jantung secara lebih efisien dengan klasifikasi model yang telah dioptimasi, sehingga dapat mendukung pengambilan keputusan klinis.
3. Menyediakan alternatif deteksi dini penyakit jantung yang lebih terjangkau dan efisien, sehingga dapat meningkatkan akses masyarakat terhadap layanan pemeriksaan penyakit jantung secara mudah dan terjangkau.

1.6 Sistematika Penulisan

Pada penelitian kali ini penulis memuat uraian secara garis besar isi Skripsi untuk tiap-tiap bab. Pada bagian ini mendeskripsikan apa saja isi masing-masing bab yang akan disusun dan akan dijelaskan secara singkat dari Bab 1. Bab II, Bab III, Bab IV Bab V. sebagai berikut:

Bab 1 PENDAHULUAN: Latar Belakang, Rumusan Masalah, Batasan masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan.

Bab 2 TINJAUAN PUSTAKA: Studi Literature, Tabel Keaslian Penelitian, Data Mining, Machine Learning, Klasifikasi, Penyakit Jantung, SMOTE, Feature Selection, Korelasi Pearson, Hyperparameter GridsearchCV, Algoritma Support Vector Machine, Algoritma Naive Bayes, Algoritma K-Nearest Neighbors dan Algoritma Logistic Regression, dan Confusion Matrix.

Bab 3 METODE PENELITIAN: Objek Penelitian, Alur Penelitian, Studi Literature, Memuat Dataset, Preprocessing, Feature Selection, Normalisasi Data, Split Data, Modelling, Hyperparameter GridsearchCV, Evaluasi Model, Alat dan Bahan, Data Penelitian, dan Alat Penelitian.

Bab 4 HASIL DAN PEMBAHASAN: Memuat Dataset, Preprocessing Data, Korelasi Pearson, Normalisasi Data, Split Data, Modelling, Hyperparameter GridsearchCV, Algoritma Support Vector Machine, Algoritma Naive Bayes, Algoritma K-Nearest Neighbors, dan Algoritma Logistic Regression, Evaluasi dan Hasil.

Bab 5 PENUTUP: Kesimpulan dan saran.