

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gaya hidup adalah pola kebiasaan dan preferensi yang dianut oleh individu atau kelompok dalam berbagai aspek kehidupan sehari-hari, termasuk pada pekerjaan, aktivitas fisik, hobi, pola tidur, interaksi sosial, serta kebiasaan makan dan minum. Gaya hidup juga suatu hal yang sangat berpengaruh terhadap kesehatan. Menurut World Health Organization (WHO), 60 persen faktor yang berhubungan dengan kesehatan individu dan kualitas hidup berhubungan dengan gaya hidup. Jutaan orang di dunia yang menerapkan gaya hidup tidak sehat diketahui berisiko atau sudah merasakan dampak negatifnya bagi kesehatan [1].

Gaya hidup tidak sehat mencakup beberapa aktivitas yaitu malas berolahraga, sering begadang, kebiasaan merokok, mengonsumsi minuman beralkohol secara berlebihan serta kebiasaan buruk lainnya [2]. Gaya hidup tidak sehat dapat menyebabkan berbagai dampak negatif pada kesehatan tubuh kita, salah satunya adalah penyakit degenerative. Penyakit degenerative yang sering terjadi yaitu diabetes tipe 2, penyakit jantung, osteoporosis dan penyakit kanker [3]. Kondisi ini meliputi tekanan darah tinggi, kadar gula darah yang tinggi, kelebihan lemak tubuh, dan kadar kolesterol yang tidak normal.

Penyakit jantung adalah sebuah kondisi kinerja jantung yang terganggu sehingga tidak bisa melaksanakan tugasnya dengan baik dalam memompa darah dan mengalirkan oksigen ke seluruh tubuh [4]. Kebanyakan orang dengan penyakit jantung tidak menyadari gejala awal dan sebagian besar orang meninggal karena adanya serangan jantung [5]. Saat ini, terdapat beberapa gangguan yang mempengaruhi fungsi jantung, termasuk penyakit jantung koroner, gangguan irama jantung, penyakit jantung bawaan, kelainan katup jantung, dan gagal jantung [6]. Berdasarkan data WHO (World Health Organization) pada tahun 2021, tercatat 17,8 juta kematian akibat penyakit jantung, dengan 12,934,931 kasus terjadi pada tahun yang sama [7].

Jumlah kasus penyakit jantung yang tinggi terjadi karena kurangnya metode yang akurat untuk mendeteksi apakah seseorang menderita penyakit jantung atau

tidak. Beberapa permasalahan dalam deteksi penyakit jantung meliputi ketidakseimbangan data, data yang hilang (missing value), dan salah satunya kesalahan dalam penyetelan parameter model. Oleh karena itu, untuk mempermudah ahli jantung dalam mendiagnosis, merencanakan pengobatan dan memberikan perawatan yang sesuai diperlukan adanya teknologi yang bisa mengatasi permasalahan tersebut [8]. Dalam era digital, teknik machine learning telah menjadi alat yang kuat dalam memprediksi dan mengklasifikasikan penyakit [9]. Machine learning merupakan pendekatan komputasional yang memungkinkan sistem untuk belajar secara mandiri dari data dan pengalaman, tanpa perlu pemrograman eksplisit untuk setiap tugas tertentu [10].

Machine learning biasanya mengonversikan tugasnya ke dalam bentuk optimasi dan menyelesaikannya dengan berbagai metode optimasi. Dalam konteks ini, pemberian parameter dilakukan melalui proses *hyperparameter tuning*, yang dilakukan sebelum fase pelatihan dimulai [11]. Proses ini bertujuan untuk menemukan kombinasi nilai *hyperparameter* yang optimal agar model dapat bekerja dengan performa terbaik. Akan tetapi, tidak ada hubungan langsung antara output algoritma machine learning dengan *hyperparameter*. Selain itu, pencarian nilai parameter dalam algoritma machine learning masih dilakukan secara manual. Oleh karena itu, penting untuk menyesuaikan *hyperparameter tuning* secara teratur dengan melatih sejumlah model dengan kombinasi nilai yang berbeda dan menguji hasil model untuk menentukan model yang memiliki performa terbaik.

Penelitian ini menggunakan beberapa model machine learning yang biasa digunakan untuk klasifikasi seperti Random Forest, K-Nearest Neighbors (K-NN), *XGBOOST* dan Decision Tree. Masing-masing algoritma tentunya memiliki keunggulan dan kelemahan yang berbeda-beda dalam hal akurasi, waktu pemrosesan dan kemampuan menginterpretasikan hasil. *Random Forest*, mampu mengatasi overfitting dengan jumlah pohon yang digunakan banyak [12]. Namun, algoritma ini dapat memakan waktu komputasi lebih lama ketika banyak menggunakan jumlah pohon pada dataset yang besar [13]. KNN, disisi lain, efektif dan cukup sederhana dalam menangani data non-linier tanpa memerlukan asumsi tertentu. Namun KNN cukup lambat dalam melakukan pencarian nilai k ketika

dataset yang digunakan sangat besar [14]. *XGBOOST*, meskipun cepat dan akurat, algoritma ini sangat bergantung pada pemilihan parameter yang tepat seperti learning rate dan max-depth [15]. Sedangkan Decision Tree, diunggulkan dalam menghasilkan model yang relatif mudah dipahami akan tetapi, algoritma ini cukup rentan terhadap overfitting [16].

Dalam proses membangun model machine learning, pemilihan parameter yang tepat sangat penting untuk meningkatkan performa algoritma. Setiap algoritma memiliki parameter spesifik yang dapat memengaruhi hasilnya. Seperti jumlah tetangga pada KNN, jumlah pohon keputusan pada Random Forest, learning rate pada XGBoost dan kedalaman maksimum pohon pada algoritma Decision Tree. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan kinerja model beberapa algoritma machine learning dengan mengimplementasikan teknik *hyperparameter tuning* Grid Search Cross Validation. Metode Grid Search CV dipilih karena metode ini dalam pencarian kombinasi parameter terbaik dilakukan secara sistematis dan menyeluruh [17].

Penelitian ini juga bertujuan untuk mengetahui algoritma machine learning mana yang memiliki performa terbaik ketika sebelum dilakukan *hyperparameter tuning* dan setelah dilakukan *hyperparameter tuning*. Hasil penelitian ini diharapkan tidak hanya dapat meningkatkan akurasi model, tetapi juga mengidentifikasi algoritma machine learning mana yang memiliki performa terbaik sebelum dan setelah *hyperparameter tuning*. Penelitian ini juga diharapkan dapat memberikan kontribusi yang baik kedepannya dalam bidang kesehatan mengenai klasifikasi penyakit khususnya jantung.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang ada dan dijelaskan pada bagian latar belakang, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana implementasi *hyperparameter tuning* Grid Search CV dalam meningkatkan kinerja beberapa algoritma machine learning pada klasifikasi penyakit jantung?

2. Algoritma machine learning apa yang memberikan performa terbaik sebelum dan setelah dioptimasi dengan Grid Search CV pada klasifikasi penyakit jantung?

1.3 Batasan Masalah

Batasan masalah dari penelitian ini yaitu:

1. Penelitian ini dilakukan untuk mengetahui tingkat akurasi, precision, recall dan F1-Score pada klasifikasi penyakit jantung.
2. Optimasi yang digunakan menggunakan *hyperparameter tuning* Grid Search CV (Cross Validation).
3. Data yang digunakan pada penelitian ini adalah dataset public yang didapat dari situs *Kaggle*.
4. Evaluasi performa pada akurasi, precision, recall, dan F1-score sebagai metrik evaluasi.
5. Pada penelitian ini menggunakan *google colab* sebagai platform utama untuk analisis dan pemrosesan data.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk meningkatkan kinerja model beberapa algoritma machine learning seperti Random Forest, K-Nearest Neighbors, *XGBOOST*, dan Decision Tree dengan mengimplementasikan *hyperparameter tuning* Grid Search Cross Validation (Grid Search CV) pada klasifikasi penyakit jantung. Penelitian ini juga bertujuan untuk mengetahui algoritma machine learning mana yang memiliki performa terbaik ketika sebelum dilakukan *hyperparameter tuning* dan setelah dilakukan *hyperparameter tuning*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah:

1. Mengetahui mengenai pemodelan data mining menggunakan beberapa algoritma machine learning pada dataset penyakit jantung. Penelitian ini juga memberikan pengetahuan tentang cara meningkatkan kinerja model suatu algoritma machine learning dengan *hyperparameter tuning* Grid Search CV.

2. Hasil dari penelitian ini diharapkan bisa membantu bidang kesehatan dalam mendeteksi dini penyakit jantung pada pasien secara akurat dan tepat dengan pemilihan algoritma yang memiliki performa terbaik setelah dilakukan tuning parameter menggunakan metode Grid Search CV (Cross Validation).
3. Membantu masyarakat yang terkena penyakit jantung dengan harapan dapat mendeteksi gejala penyakit sejak dini, sehingga dapat meningkatkan kualitas hidup mereka.
4. Menambahkan informasi dari segi algoritma dan teknik *hyperparameter tuning* yang digunakan beserta topik yang diteliti yang memungkinkan untuk dijadikan referensi pada penelitian selanjutnya.

1.6 Sistematika Penelitian

BAB I PENDAHULUAN

Bab I berisi Latar belakang masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Sistematika Penulisan.

BAB II TINJAUAN PUSTAKA

Bab II berisi Studi Literatur dan Landasan Teori. Pada bab ini akan dibahas mengenai penelitian terdahulu dan teori sesuai penelitian.

BAB III METODE PENELITIAN

Bab III didalamnya terdapat Object Penelitian, Tahapan Penelitian, serta Alat & Bahan.

BAB IV HASIL DAN PEMBAHASAN

Bab IV ini merupakan tahapan yang berisi hasil penelitian didalamnya terhadap analisis dan uji penelitian sesuai dengan metode yang digunakan.

BAB V PENUTUP

Bab V berisi kesimpulan dan saran yang peneliti rangkum selama proses penelitian.