

BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi merupakan salah satu elemen penting dalam *Machine Learning* yang memegang peran *vital* dalam analisis dan pemahaman data. Melalui teknik klasifikasi, berbagai masalah dapat diselesaikan, seperti klasifikasi data, kategorisasi teks, pengklasifikasian gambar, prediksi struktur protein, dan lain lain (Arian et al., 2020). Proses klasifikasi bertujuan untuk mencari model terbaik yang dapat merepresentasikan pola hubungan antara data dengan label yang sesuai (Otero et al., 2012). Tantangan utama dalam klasifikasi adalah bagaimana mengembangkan model yang mampu memisahkan data berdasarkan kelas atau labelnya dengan akurat. Ada berbagai metode yang digunakan untuk menyelesaikan permasalahan klasifikasi ini, seperti *K-Nearest Neighbor (KNN)*, *XGBoost*, *naive Bayes*, *decision tree*, dan lain sebagainya. Salah satu algoritma yang telah terbukti memiliki kinerja yang baik dalam klasifikasi adalah KNN, beberapa penelitian mencatat keunggulannya dalam hal akurasi (Bansal et al., 2022).

K-Nearest Neighbors (KNN) dapat digunakan baik untuk masalah klasifikasi maupun regresi dalam *Machine Learning*. Keberhasilan KNN dalam menangani permasalahan klasifikasi telah menjadikannya sebagai salah satu pendekatan yang sangat diminati dalam komunitas *Machine Learning* (Yin et al., 2017). Konsep dasar dari algoritma KNN adalah bahwa suatu *data point* akan diklasifikasikan atau

diprediksi berdasarkan mayoritas tetangga terdekatnya. Algoritma KNN mencari k-tetangga terdekat dari *data point* yang akan diprediksi dan menentukan kelas mayoritas dari tetangga-tetangga tersebut sebagai kelas yang akan diberikan kepada data point tersebut (Subbarayalu et al., 2019). Namun, terdapat masalah yang cukup signifikan dalam Algoritma KNN yaitu menentukan nilai *hyperparameter* K yang berarti jumlah tetangga yang ideal saat mengklasifikasikan sebuah *data point* dimana nilai K memiliki pengaruh signifikan terhadap akurasi sebuah model (Gauthama Raman et al., 2020).

Hyperparameter merupakan parameter-parameter yang tidak dapat diubah selama proses pelatihan model dalam *Machine Learning* (Yu & Zhu, 2020). Untuk menentukan *hyperparameter* yang optimal, umumnya dilakukan proses *tuning* secara manual. Akan tetapi, pendekatan manual tersebut memerlukan pemahaman yang mendalam mengenai parameter-parameter yang cocok untuk model terhadap dataset tertentu, sehingga proses pencarian parameter yang optimal membutuhkan banyak waktu (Syarif et al., 2016). Pendekatan lain yang lebih efektif untuk menemukan *hyperparameter* K terbaik adalah dengan menggunakan metode *Hyperparameter Tuning* yang memanfaatkan algoritma pencarian dalam proses pencarian nilai K terbaik. Namun, pemilihan metode *hyperparameter* yang tepat juga menjadi faktor penentu dalam keberhasilan pencarian nilai K karena setiap metode memiliki karakteristik dan kinerja yang berbeda. Oleh karena itu, penting untuk mempertimbangkan dengan cermat metode mana yang paling cocok untuk kasus penggunaan tertentu, agar hasil pencarian nilai K dapat lebih optimal yang tidak hanya menghasilkan model dengan akurasi tinggi tetapi juga efisien dalam

penggunaan sumber daya komputasi sesuai dengan kebutuhan aplikasi yang dihadapi.

Dalam penelitian ini, akan dilakukan perbandingan pada dua metode yang digunakan untuk mencari *hyperparameter* terbaik pada nilai K dalam Algoritma KNN. *Grid Search* dan *Random Search* menjadi dua algoritma pencarian yang dibandingkan. Penggunaan *Grid Search* dan *Random Search* sebagai metode yang dibandingkan dilakukan karena keduanya merupakan Teknik yang paling umum digunakan untuk optimasi parameter dalam *Machine Learning*. Keduanya juga memiliki karakteristik dan cara kerja yang mirip yaitu dengan mencoba berbagai kemungkinan baik secara acak maupun terstruktur. Evaluasi dilakukan berdasarkan nilai akurasi hasil prediksi (*accuracy*), waktu proses (*running time*) dan penggunaan memori (*memory usage*) dari masing-masing algoritma pencarian.

Beberapa penelitian terkait perbandingan optimasi *Hyperparameter* salah satunya adalah penelitian yang dilakukan oleh (Pramudhyta & Rohman, 2024a) yang membandingkan metode *Grid Search* dengan metode *Random Search* pada algoritma *XGBoost*. Hasil dari penelitian tersebut menyimpulkan bahwa *Grid Search* menunjukkan performa dan waktu proses yang lebih baik dibandingkan dengan *Random Search* untuk dataset dalam skala yang tidak terlalu besar. Selain itu, terdapat penelitian lain yang dilakukan oleh (Fajri & Primajaya, 2023a) yang membandingkan teknik *Grid Search* dengan *Random Search* pada algoritma SVM. Dalam penelitian tersebut hasil dari *Random Search* sedikit lebih unggul dibandingkan dengan *Grid Search*.

Diharapkan dengan mempertimbangkan temuan dari penelitian sebelumnya,

penelitian ini dapat memberikan wawasan yang lebih dalam mengenai metode hyperparameter tuning yang memiliki nilai akurasi yang lebih baik dan penggunaan sumber daya komputasi yang lebih rendah.

1.2 Rumusan Masalah

Berdasarkan penyusunan latar belakang di atas, dihasilkan rumusan masalah yang diangkat pada penelitian ini sebagai berikut.

1. Bagaimana perbandingan metode *Grid Search* dan *Random Search* berdasarkan kinerja akurasi pada algoritma KNN?
2. Bagaimana perbandingan metode *Grid Search* dan *Random Search* berdasarkan *memory usage* dan *running time* pada algoritma KNN?

1.3 Tujuan Penelitian

Tujuan penelitian ini dispesifikasikan berdasarkan rumusan masalah yang telah dibuat sebagai berikut.

1. Membandingkan metode *Grid Search* dan *Random Search* berdasarkan kinerja akurasi pada algoritma KNN.
2. Membandingkan metode *Grid Search* dan *Random Search* berdasarkan *memory usage* dan *running time* pada algoritma KNN.

1.4 Manfaat Penelitian

Penelitian ini dapat memberikan manfaat kepada pengguna *Machine Learning* khususnya bagi yang menggunakan Algoritma KNN. Manfaat tersebut di antaranya sebagai berikut.

1. Penelitian ini memberikan panduan praktis bagi pengguna dan peneliti *machine learning* untuk memilih metode *Hyperparameter Tuning* pada algoritma KNN.
2. Hasil penelitian dapat meningkatkan akurasi model KNN dalam memecahkan masalah klasifikasi dan regresi.
3. Penelitian ini membantu dalam mengoptimalkan penggunaan sumber daya komputasi dan waktu yang digunakan dalam proses *Hyperparameter Tuning*.

1.5 Batasan Masalah

Penelitian ini memiliki batasan-batasan masalah yang digunakan untuk membuat proses penelitian menjadi lebih spesifik. Batasan-batasan tersebut adalah sebagai berikut.

1. Evaluasi kinerja akan dilakukan pada sejumlah dataset representative dari *University of California Irvine (UCI) repository* namun tidak mencakup seluruh bidang dataset yang mungkin digunakan.
2. Pengujian dilakukan pada proses klasifikasi dan regresi tidak mencakup seluruh bidang *Machine Learning* yang bisa dilakukan seperti *clustering* dan *decision making*.

1.6 Sistematika Penulisan

Sistematika Penulisan digunakan dengan tujuan agar laporan penelitian dapat disusun dan diarahkan sesuai dengan tahapan penelitian. Adapun sistematika yang diterapkan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat latar belakang yang menjadi dasar dilaksanakannya penelitian, perumusan masalah yang akan diteliti, tujuan penelitian, manfaat yang diharapkan dari penelitian, metodologi penelitian, serta struktur penulisan laporan penelitian.

BAB II LANDASAN TEORI

Bab ini berisi pembahasan mengenai teori-teori yang relevan dengan penelitian, termasuk konsep, metode, dan algoritma yang digunakan. Selain itu, bab ini juga menjelaskan penelitian-penelitian sebelumnya yang berkaitan serta menjelaskan kebaruan dari penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, mulai dari waktu dan tempat penelitian, objek penelitian, variabel penelitian, matriks penelitian, hingga tahapan-tahapan yang dilalui dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil penelitian serta membahas rancangan yang telah dijelaskan pada bab sebelumnya. Pembahasan mencakup pembuatan model dan algoritma yang akan digunakan, serta eksperimen yang dilakukan bersama model lainnya.

BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil eksperimen yang dilakukan serta memberikan saran untuk penelitian selanjutnya berdasarkan batasan dan hasil penelitian yang terkait dengan topik yang dibahas.