

BAB I PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan e-commerce yang signifikan dalam beberapa tahun terakhir, serta seiring dengan meningkatnya akses digital dan perubahan perilaku konsumen, menjadikan kemampuan untuk bisa memprediksi perilaku pembelian menjadi hal yang penting. Prediksi tersebut berguna untuk memberikan penawaran diskon maupun produk tertentu yang sesuai dengan minat konsumen. Berdasarkan data pada laman *statista*, nilai pasar pada e-commerce secara global pada tahun 2025 diproyeksikan akan mencapai USD 3,66 triliun. Dengan tingkat pertumbuhan tahunan (CAGR 2025-2030) sebesar 6,29%, volume pasar diperkirakan akan meningkat hingga USD 4,96 triliun pada tahun 2030. Jumlah pengguna e-commerce secara global diperkirakan mencapai 4 miliar pengguna pada tahun 2030 [1]. Dalam konteks ini, e-commerce dituntut untuk dapat mengakses data yang dapat digunakan untuk mengidentifikasi dan mengklasifikasikan niat konsumen agar data tersebut dapat digunakan untuk memprediksi tingkat penjualan [2]. Untuk meningkatkan prediksi tersebut, sebagian besar organisasi mempekerjakan tenaga ahli guna menganalisis data secara langsung [3]. Namun, tenaga manusia kurang efektif untuk melakukan analisis pada data yang sangat besar.

Machine learning menjadi salah satu solusi yang banyak digunakan dalam menganalisis data besar untuk menghasilkan prediksi yang bermanfaat. Dua algoritma yang digunakan dalam prediksi perilaku konsumen adalah Random Forest dan XGBoost [4], [5], [6], [7]. Random Forest dikenal karena memiliki kinerja yang bagus dalam menangani ketidakseimbangan data dan banyaknya variabel [8], sedangkan XGBoost memiliki kemampuan untuk menghindari *overfitting* dengan mengurangi kompleksitas model karena dibangun dengan struktur pohon regresi yang lebih teratur [9]. Namun, performa dari kedua algoritma ini sangat dipengaruhi oleh pemilihan parameter yang sesuai. Pemilihan *hyperparameter* yang tidak tepat dapat menyebabkan model menjadi kurang optimal dari segi akurasi dan waktu komputasi [10].

Terdapat beberapa metode yang umum digunakan untuk menyetel parameter algoritma. Teknik yang umum digunakan untuk menyetel parameter algoritma adalah *GridSearchCV*, yang memeriksa setiap kombinasi parameter secara keseluruhan [11]. Namun, metode ini memakan waktu dan sumber daya komputasi yang besar. Metode *RandomizedSearchCV* memungkinkan pencarian *hyperparameter* yang lebih menyeluruh dengan iterasi yang lebih sedikit [12]. *RandomizedSearchCV* menunjukkan hasil yang unggul dari *GridSearchCV* dengan waktu komputasi yang lebih rendah [13]. Penelitian ini berfokus pada penerapan *RandomizedSearchCV* untuk mengoptimalkan kinerja algoritma *Random Forest* dan *XGBoost* dalam memprediksi apakah perilaku pengguna pada platform e-commerce akan menghasilkan transaksi atau tidak.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa kedua algoritma yang telah dioptimasi, serta mengidentifikasi sejauh mana metode *tuning* dapat meningkatkan performa algoritma.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah Apakah dengan menerapkan teknik optimasi *hyperparameter* pada *Random Forest* dan *XGBoost* akan meningkatkan performa dari kedua model tersebut?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini meliputi:

1. *Dataset* yang digunakan dalam penelitian ini berasal dari *dataset* publik *UCI Machine Learning Repository* dengan nama *Online Shoppers Purchasing Intention Dataset*.
2. Algoritma yang digunakan dalam penelitian ini adalah *Random Forest* dan *XGBoost*.
3. Teknik optimasi yang digunakan dalam penelitian ini adalah *RandomSearchCV*.

1.4 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk menganalisis pengaruh optimasi *hyperparameter* terhadap performa algoritma *Random Forest* dan *XGBoost*.

1.5 Manfaat Penelitian

Adapun beberapa manfaat dari penelitian ini antara lain:

1. Menambah pengetahuan akan efektifitas dari penerapan optimasi *hyperparameter* pada algoritma *Random Forest* dan *XGBoost*.
2. Memberikan kontribusi terhadap penelitian prediksi perilaku konsumen pada platform *e-commerce*.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini terdiri dari 5 bab, dengan susunan sebagai berikut:

BAB I PENDAHULUAN, berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA, berisi pembahasan dasar teori yang berkaitan dengan penelitian serta penelitian-penelitian sebelumnya yang relevan.

BAB III METODE PENELITIAN, bab ini menguraikan tahapan yang akan dilaksanakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini berisi hasil dan pembahasan dari penerapan metode yang dijelaskan pada bab sebelumnya.

BAB V PENUTUP, berisi kesimpulan dan saran yang diperoleh dari penelitian dan saran untuk penelitian selanjutnya.