

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan meningkatnya penggunaan internet telah mendorong pertumbuhan pesat sektor perdagangan digital di Indonesia. *Platform e-commerce* seperti Shopee dan Tokopedia menjadi marketplace dominan yang memfasilitasi transaksi dalam skala besar setiap harinya. Dalam menghadapi persaingan yang semakin ketat, perusahaan dituntut untuk menerapkan strategi pemasaran berbasis data guna memahami perilaku konsumen secara lebih akurat. Penerapan machine learning dalam optimalisasi strategi pemasaran dan prediksi konversi pada *e-commerce* telah terbukti meningkatkan efektivitas pengambilan keputusan berbasis data [1], [2].

Seiring dengan berkembangnya analisis prediktif, berbagai algoritma machine learning telah digunakan untuk memodelkan perilaku konsumen dan meningkatkan akurasi prediksi. Random Forest menjadi salah satu algoritma yang banyak diterapkan dalam konteks *e-commerce*, seperti pada prediksi loyalitas pelanggan berbasis ulasan di Shopee [3] serta prediksi pembelian berbasis perilaku sesi pengguna [4]. Selain itu, studi komparatif menunjukkan bahwa evaluasi beberapa model prediksi diperlukan untuk menentukan algoritma yang paling optimal dalam konteks *e-commerce* [13]. Hal ini menunjukkan pentingnya pendekatan komparatif dalam penelitian prediksi.

Dalam konteks pemodelan data, *Linear Regression* dikenal sebagai metode yang sederhana dan mudah diinterpretasikan dalam menjelaskan hubungan antar variabel [5]. Namun, metode ini memiliki keterbatasan dalam menangani pola hubungan yang kompleks dan non-linear. Sebaliknya, *Random Forest* sebagai metode *ensemble learning* mampu meningkatkan akurasi melalui kombinasi banyak decision tree dan memiliki kemampuan generalisasi yang baik [6]. Beberapa penelitian komparatif menunjukkan bahwa *Random Forest* sering memberikan performa lebih baik dibandingkan *Linear Regression* dalam berbagai kasus prediksi, seperti prediksi harga dan komparasi model regresi [7], serta peramalan tren *e-commerce* dan harga komoditas [8].

Implementasi Random Forest dalam prediksi penjualan juga telah banyak dilakukan. Penelitian oleh Barus dan Darmanto [9] menunjukkan bahwa *Random Forest* efektif dalam memprediksi penjualan produk dengan tingkat akurasi yang baik. Hal serupa ditunjukkan dalam prediksi produksi supermarket menggunakan *Random Forest Regression* [10], serta pada prediksi penjualan dan persediaan produk usaha ritel [11] dan prediksi penjualan pada sektor parfum [12]. Implementasi *Random Forest* dalam sistem *point of sales* untuk prediksi penjualan harian juga menunjukkan hasil yang optimal [14], demikian pula pada prediksi pendapatan penjualan ritel modern seperti Indomaret [15]. Secara umum, hasil-hasil penelitian tersebut memperlihatkan keunggulan Random Forest dalam menangani data penjualan yang bersifat kompleks dan dinamis.

Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada prediksi transaksi aktual atau nilai penjualan yang telah terjadi. Penelitian yang secara khusus membahas prediksi **minat pembelian konsumen**, yaitu kecenderungan psikologis sebelum terjadinya transaksi, masih relatif terbatas. Selain itu, mayoritas studi menggunakan data sekunder berupa data transaksi perusahaan, sedangkan penggunaan data kuisioner untuk mengukur persepsi harga, kepercayaan terhadap platform, serta pengaruh rating dan ulasan masih jarang dilakukan. Padahal, variabel-variabel tersebut secara teoritis memiliki peran penting dalam membentuk minat pembelian konsumen pada platform *e-commerce*.

Berdasarkan kesenjangan penelitian (*research gap*) tersebut, diperlukan penelitian yang secara komprehensif menganalisis dan membandingkan kinerja algoritma **Linear Regression** dan **Random Forest** dalam memprediksi minat pembelian konsumen pada Shopee dan Tokopedia dengan menggunakan data kuisioner sebagai sumber data utama. Evaluasi model dilakukan menggunakan metrik MAE, RMSE, dan R^2 Score untuk mengukur tingkat akurasi serta kemampuan generalisasi model. Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan model prediksi berbasis machine learning pada konteks *e-commerce* serta memberikan implikasi praktis bagi pelaku bisnis dalam merancang strategi pemasaran berbasis data yang lebih efektif dan proaktif.

1.2 Identifikasi Masalah

Masalah yang teridentifikasi:

1. Persaingan yang semakin ketat pada platform e-commerce Shopee dan Tokopedia menuntut pelaku usaha untuk mampu memahami perilaku dan kecenderungan minat pembelian konsumen secara lebih akurat.
2. Sebagian besar pelaku usaha masih mengandalkan data historis penjualan sebagai dasar pengambilan keputusan, sehingga strategi yang diterapkan cenderung bersifat reaktif dan belum berbasis prediksi minat pembelian.
3. Minat pembelian sebagai indikator awal sebelum terjadinya transaksi belum banyak dimodelkan menggunakan pendekatan machine learning berbasis data kuisisioner yang menggambarkan persepsi dan preferensi konsumen.
4. Terdapat berbagai algoritma machine learning yang dapat digunakan untuk prediksi, namun belum diketahui secara pasti algoritma mana yang lebih optimal dalam memprediksi minat pembelian konsumen pada platform Shopee dan Tokopedia.
5. Penelitian terdahulu lebih banyak berfokus pada prediksi penjualan atau transaksi aktual, sehingga masih terdapat kesenjangan penelitian dalam konteks perbandingan algoritma *Linear Regression* dan *Random Forest* untuk prediksi minat pembelian berbasis data kuisisioner.

1.3 Batasan Masalah

Agar penelitian ini lebih terfokus, maka batasan masalah dalam penelitian ini adalah:

1. Penelitian ini hanya dilakukan pada pengguna platform e-commerce **Shopee dan Tokopedia** yang pernah melakukan transaksi pembelian minimal satu kali dalam enam bulan terakhir.
2. Data penelitian diperoleh melalui **kuisisioner (Google Form)** sebagai data primer dengan menggunakan skala Likert 1–5 untuk mengukur persepsi responden.
3. Variabel independen (X) yang digunakan dalam penelitian ini meliputi
 - Persepsi Harga
 - Rating Produk
 - Jumlah/Ulasan Konsumen
 - Kepercayaan terhadap Platform
4. Variabel dependen (Y) dalam penelitian ini adalah **Minat Pembelian Konsumen**.
5. Algoritma yang digunakan untuk pemodelan prediksi hanya:

- *Linear Regression*
 - *Random Forest*
6. Proses evaluasi model dilakukan menggunakan metrik:
- *Mean Absolute Error (MAE)*
 - *Root Mean Square Error (RMSE)*
 - *R² Score*
7. Penelitian ini tidak membahas aspek teknis internal sistem Shopee dan Tokopedia, serta tidak menggunakan data transaksi aktual dari pihak platform.
8. Penelitian ini tidak membandingkan algoritma lain di luar *Linear Regression* dan *Random Forest*.

1.4 Rumusan Masalah

Berdasarkan latar belakang di atas, terdapat rumusan masalah yaitu :

1. Bagaimana pengaruh variabel persepsi harga, rating produk, jumlah ulasan, dan kepercayaan terhadap platform terhadap minat pembelian konsumen pada Shopee dan Tokopedia berdasarkan data kuisisioner?
2. Bagaimana performa algoritma **Linear Regression** dalam memprediksi minat pembelian konsumen berdasarkan variabel-variabel tersebut?
3. Bagaimana performa algoritma **Random Forest** dalam memprediksi minat pembelian konsumen berdasarkan variabel-variabel tersebut?
4. Algoritma manakah yang memiliki tingkat akurasi dan kemampuan prediksi yang lebih baik berdasarkan nilai MAE, RMSE, dan R² Score?

1.5 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Untuk menganalisis pengaruh persepsi harga, rating produk, jumlah ulasan, dan kepercayaan terhadap platform terhadap minat pembelian konsumen pada Shopee dan Tokopedia berdasarkan data kuisisioner.
2. Untuk menganalisis performa algoritma **Linear Regression** dalam memprediksi minat pembelian konsumen berdasarkan variabel persepsi harga, rating produk, jumlah ulasan, dan kepercayaan terhadap platform.

3. Untuk menganalisis performa algoritma **Random Forest** dalam memprediksi minat pembelian konsumen berdasarkan variabel persepsi harga, rating produk, jumlah ulasan, dan kepercayaan terhadap platform.
4. Untuk membandingkan tingkat akurasi dan kemampuan prediksi antara algoritma Linear Regression dan Random Forest berdasarkan nilai MAE, RMSE, dan R^2 Score guna menentukan algoritma yang lebih optimal.

1.6 Manfaat Penelitian

1.6. 1 Manfaat Teoritis

1. Memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang Data Mining dan Machine Learning, khususnya dalam penerapan algoritma Linear Regression dan Random Forest untuk prediksi minat pembelian konsumen.
2. Menjadi referensi ilmiah bagi penelitian selanjutnya yang membahas perbandingan algoritma prediksi dalam konteks e-commerce.
3. Memperkaya kajian akademik terkait analisis perilaku konsumen berbasis pendekatan kuantitatif dan komputasional.

1.6. 2 Manfaat Praktis

1. Bagi pelaku *e-commerce* (Shopee dan Tokopedia), hasil penelitian ini dapat menjadi bahan evaluasi dalam memahami faktor-faktor yang mempengaruhi minat pembelian konsumen.
2. Memberikan rekomendasi algoritma yang lebih optimal dalam melakukan prediksi minat pembelian berdasarkan tingkat akurasi model.
3. Membantu pengambilan keputusan berbasis data (*driven decision making*) dalam strategi pemasaran digital.