
ANALISIS PERBANDINGAN ALGORITMA LINEAR REGRESSION DAN RANDOM FOREST UNTUK PREDIKSI MINAT PEMBELIAN KONSUMEN PADA E-COMMERCE SHOPEE DAN TOKOPEDIA

Junior Mulatua S. Munthe¹, David Beckham Aritonang², Dio Kurnia Syahputra³,
Saut Parsaoran Tamba⁴

Universitas Prima Indonesia, Medan

Pusat Unggulan IPTEK (PUI-ITIK) Inovasi Teknologi Ilmu Komputer

e-mail: ¹juniorsmunthe@gmail.com, ²kaub75531@gmail.com,

³Kurniasyahputradio@gmail.com, ⁴sautparsaorantamba@unprimdn.ac.id

Email Penulis Korespondensi: ¹juniorsmunthe@gmail.com

Abstract: *The rapid growth of e-commerce in Indonesia, particularly on the Shopee and Tokopedia platforms, has increased the need for data-driven approaches to better understand consumer purchase intentions. This study aims to compare the performance of the Linear Regression and Random Forest algorithms in predicting consumer purchase intentions based on price perception, product ratings, consumer reviews, and trust in the platform. A quantitative approach was employed using primary data collected through questionnaires distributed to Shopee and Tokopedia users with a 1–5 Likert scale. The data were processed using Python through preprocessing, validity and reliability testing, and predictive model development. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R² Score. The results indicate that Random Forest achieved better predictive performance on the Shopee dataset, while Linear Regression produced better results on the Tokopedia dataset. These findings suggest that the effectiveness of a prediction model is influenced by the characteristics of the dataset, highlighting the importance of selecting an appropriate algorithm to achieve optimal prediction performance in e-commerce applications*

Keywords: *Purchase Intention, E-commerce, Linear Regression, Random Forest, Machine Learning*

Abstrak: Perkembangan e-commerce di Indonesia, khususnya pada platform Shopee dan Tokopedia, mendorong kebutuhan akan pemanfaatan analisis data untuk memahami minat pembelian konsumen secara lebih akurat. Penelitian ini bertujuan membandingkan kinerja algoritma Linear Regression dan Random Forest dalam memprediksi minat pembelian konsumen berdasarkan variabel persepsi harga, rating produk, ulasan konsumen, dan kepercayaan terhadap platform. Penelitian menggunakan pendekatan kuantitatif dengan data primer yang diperoleh melalui penyebaran kuesioner kepada pengguna Shopee dan Tokopedia menggunakan skala Likert 1–5. Data diolah menggunakan Python melalui tahapan preprocessing, uji validitas dan reliabilitas, serta pembangunan model prediksi. Evaluasi model dilakukan menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R² Score. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa yang lebih baik pada data Shopee, sedangkan Linear Regression menghasilkan performa yang lebih baik pada data Tokopedia. Temuan ini menunjukkan bahwa karakteristik data memengaruhi kinerja algoritma, sehingga pemilihan model prediksi perlu disesuaikan untuk memperoleh hasil yang optimal.

Kata kunci: Minat Pembelian, E-commerce, Linear Regression, Random Forest, Machine Learning

PENDAHULUAN

Perkembangan teknologi informasi dan internet telah mendorong pertumbuhan pesat sektor *electronic commerce (e-commerce)* di Indonesia. Platform seperti Shopee dan Tokopedia menjadi media utama dalam aktivitas transaksi digital yang melibatkan jutaan pengguna setiap hari. Kemudahan akses informasi, beragam pilihan produk, serta sistem pembayaran digital yang semakin berkembang menyebabkan masyarakat semakin bergantung pada platform e-commerce dalam memenuhi kebutuhan sehari-hari. Kondisi ini menyebabkan persaingan antar pelaku bisnis semakin ketat sehingga diperlukan strategi pemasaran berbasis data untuk memahami perilaku konsumen dan meningkatkan peluang pembelian. Pemanfaatan *machine learning* dalam analisis perilaku konsumen dan pengambilan keputusan bisnis terbukti mampu meningkatkan efektivitas prediksi serta mendukung strategi pemasaran yang lebih tepat sasaran [1], [2], [15], [20].

Berbagai penelitian telah menerapkan algoritma *machine learning* untuk memprediksi perilaku konsumen pada platform e-commerce. Saputra dan Fersellia [3] menggunakan algoritma Random Forest untuk memprediksi loyalitas pelanggan berdasarkan ulasan produk di Shopee dan memperoleh tingkat akurasi yang baik. Ruscikasani et al. [4] menerapkan metode berbasis perilaku sesi pengguna untuk memprediksi keputusan pembelian pada e-commerce. Penelitian lain menunjukkan bahwa Random Forest mampu menghasilkan performa yang baik dalam prediksi penjualan produk, pengelolaan persediaan, serta peramalan permintaan pada berbagai sektor bisnis [8]–[11]. Selain itu, perkembangan metode berbasis kecerdasan buatan menunjukkan bahwa model prediksi perilaku konsumen dapat terus ditingkatkan melalui pendekatan yang lebih adaptif terhadap perubahan pola pengguna [13], [14], [16]. Di sisi lain, Linear Regression masih banyak digunakan karena kemampuannya

dalam menjelaskan hubungan antar variabel secara sederhana, mudah diinterpretasikan, dan memiliki proses komputasi yang relatif efisien [5], [12].

Penelitian komparatif antara beberapa algoritma prediksi juga telah banyak dilakukan. Hasil penelitian Wijaya dan Fauziah [5] menunjukkan bahwa setiap algoritma memiliki karakteristik dan performa yang berbeda tergantung pada pola data yang digunakan. Penelitian lain menunjukkan bahwa Random Forest cenderung lebih unggul dalam menangani hubungan non-linear dan data yang kompleks dibandingkan metode regresi konvensional [6], [7], [12]. Keunggulan tersebut menjadikan Random Forest banyak digunakan dalam berbagai penelitian prediksi karena mampu mengurangi risiko *overfitting* serta menghasilkan akurasi yang lebih stabil pada data yang heterogen. Namun demikian, Linear Regression tetap menjadi metode yang relevan karena memberikan interpretasi yang lebih mudah terhadap hubungan antarvariabel dan sering digunakan sebagai model dasar dalam penelitian prediksi.

Selain aspek algoritma, berbagai penelitian menunjukkan bahwa faktor-faktor yang memengaruhi minat pembelian konsumen juga menjadi perhatian penting dalam lingkungan e-commerce. Persepsi harga, ulasan konsumen, rating produk, dan kepercayaan terhadap platform merupakan beberapa faktor yang terbukti memiliki pengaruh terhadap keputusan dan minat pembelian konsumen [17]–[19]. Informasi yang tersedia pada halaman produk, seperti ulasan dan rating, sering kali menjadi dasar pertimbangan konsumen sebelum melakukan transaksi. Di sisi lain, tingkat kepercayaan terhadap platform dan persepsi terhadap harga juga berperan dalam membentuk keyakinan konsumen terhadap produk yang ditawarkan. Oleh karena itu, faktor-faktor tersebut dipandang relevan untuk digunakan sebagai variabel prediktor dalam penelitian ini.

Meskipun demikian, sebagian besar

penelitian terdahulu masih berfokus pada prediksi transaksi aktual, penjualan produk, atau loyalitas pelanggan menggunakan data historis perusahaan [3], [8]–[11], [15].

Penelitian yang secara khusus membandingkan kinerja Linear Regression dan Random Forest dalam memprediksi minat pembelian konsumen sebelum terjadinya transaksi masih relatif terbatas. Selain itu, mayoritas penelitian menggunakan data sekunder berupa data transaksi atau aktivitas pengguna pada platform e-commerce, sedangkan penggunaan data primer berbasis kuesioner yang mengukur persepsi harga, rating produk, ulasan konsumen, dan kepercayaan terhadap platform masih jarang dilakukan. Kondisi ini menunjukkan adanya *research gap* pada aspek prediksi minat pembelian konsumen menggunakan pendekatan *machine learning* berbasis data persepsi konsumen.

Berdasarkan kesenjangan tersebut, penelitian ini menawarkan kontribusi berupa perbandingan kinerja algoritma Linear Regression dan Random Forest dalam memprediksi minat pembelian konsumen pada platform Shopee dan Tokopedia

menggunakan data primer hasil kuesioner. Evaluasi model dilakukan menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R^2 Score untuk mengukur tingkat akurasi dan kemampuan model dalam menjelaskan variasi data. Kebaruan (*novelty*) penelitian ini terletak pada penggunaan data persepsi konsumen sebagai variabel prediktor serta analisis komparatif dua algoritma yang banyak digunakan dalam konteks e-commerce Indonesia. Dengan demikian, penelitian ini tidak hanya berfokus pada hasil prediksi, tetapi juga memberikan gambaran mengenai algoritma yang lebih sesuai untuk digunakan dalam memprediksi minat pembelian konsumen berdasarkan karakteristik data yang digunakan.

Tujuan penelitian ini adalah menganalisis dan membandingkan performa algoritma Linear Regression dan Random Forest dalam memprediksi minat

pembelian konsumen berdasarkan persepsi harga, rating produk, ulasan konsumen, dan kepercayaan terhadap platform pada e-commerce Shopee dan Tokopedia. Hasil penelitian diharapkan dapat memberikan kontribusi akademis dalam pengembangan model prediksi berbasis *machine learning* serta menjadi bahan pertimbangan bagi pelaku bisnis dalam menyusun strategi pemasaran berbasis data yang lebih efektif dan tepat sasaran.

METODE

Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif untuk membandingkan performa algoritma Linear Regression dan Random Forest dalam memprediksi minat pembelian konsumen pada platform Shopee dan Tokopedia.

Data penelitian diperoleh melalui penyebaran kuesioner kepada responden yang pernah melakukan transaksi pada kedua platform tersebut. Variabel yang digunakan meliputi persepsi harga, rating produk, ulasan konsumen, dan kepercayaan terhadap platform sebagai variabel independen, serta minat pembelian konsumen sebagai variabel dependen.

Tahapan penelitian dimulai dari pengumpulan data melalui kuesioner, dilanjutkan dengan proses *preprocessing* data yang meliputi pembersihan data, transformasi skala Likert ke bentuk numerik, serta pembagian data menjadi data pelatihan (*training*) dan data pengujian (*testing*).

Selanjutnya dilakukan pembangunan model menggunakan algoritma Linear Regression dan Random Forest. Tahap terakhir adalah evaluasi model menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R^2 Score untuk menentukan algoritma yang memiliki performa terbaik.

Objek Penelitian

Objek dalam penelitian ini adalah

minat pembelian konsumen pada dua platform e-commerce yang banyak digunakan di Indonesia, yaitu Shopee dan Tokopedia.

Pemilihan kedua platform tersebut didasarkan pada kesamaan karakteristik layanan yang menyediakan berbagai fitur pendukung keputusan pembelian, seperti sistem penilaian produk, ulasan konsumen, informasi harga, dan tingkat kepercayaan terhadap penjual maupun platform. Faktor-faktor tersebut diketahui memiliki pengaruh terhadap keputusan dan minat pembelian konsumen dalam lingkungan perdagangan elektronik.

Populasi dan Sampel

Populasi penelitian mencakup seluruh pengguna Shopee dan Tokopedia yang pernah melakukan transaksi pembelian setidaknya satu kali dalam enam bulan terakhir. Teknik pengambilan sampel dilakukan menggunakan purposive sampling, yaitu metode penentuan sampel berdasarkan kriteria tertentu yang disesuaikan dengan tujuan penelitian. Jumlah responden yang digunakan dalam penelitian ini minimal sebanyak 400 orang agar data yang diperoleh memadai untuk proses pembangunan dan evaluasi model prediksi.

Variabel Penelitian

Penelitian ini menggunakan dua jenis variabel, yaitu variabel independen dan variabel dependen.

Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui penyebaran kuesioner secara daring kepada responden yang memenuhi kriteria penelitian.

Instrumen penelitian menggunakan skala Likert lima tingkat, dengan rentang nilai 1 sampai 5 untuk mengukur persepsi responden terhadap setiap indikator variabel. Penggunaan kuesioner memungkinkan peneliti memperoleh informasi mengenai persepsi, sikap, dan penilaian konsumen yang tidak dapat diamati secara langsung melalui data transaksi.

Teknik Pengolahan dan Analisis Data

Pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python yang dijalankan melalui Visual Studio Code sebagai lingkungan pengembangan utama. Platform ini dipilih karena mendukung integrasi berbagai pustaka analisis data dan machine learning yang dibutuhkan dalam penelitian.

Tahapan analisis meliputi proses persiapan data, pembangunan model prediksi, evaluasi model, serta visualisasi hasil penelitian.

Uji Instrumen Penelitian

Sebelum dilakukan analisis utama, instrumen penelitian terlebih dahulu diuji untuk memastikan kualitas data yang diperoleh, yaitu uji validitas dan reliabilitas.

Uji Asumsi Klasik

Karena salah satu model yang digunakan adalah Linear Regression, maka dilakukan pengujian asumsi klasik untuk memastikan model memenuhi persyaratan analisis statistik.

Pengujian yang dilakukan meliputi:

1. Uji Normalitas
2. Uji Multikolinearitas
3. Uji Heteroskedastisitas

Implementasi Model Prediksi

Penelitian ini membandingkan dua algoritma prediksi yang berbeda, yaitu linear regression dan Random forest.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat akurasi prediksi yang dihasilkan oleh masing-masing algoritma.

Pengukuran performa model menggunakan tiga metrik evaluasi, sebagai berikut:

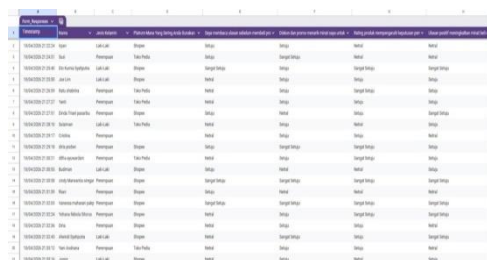
1. Coefficient of Determination (R^2 Score)
2. Mean Absolute Error (MAE)
3. Root Mean Square Error (RMSE)

HASIL DAN PEMBAHASAN

Data Penelitian

Penelitian ini menggunakan data primer yang diperoleh melalui penyebaran kuesioner kepada pengguna platform e-commerce Shopee dan Tokopedia. Pengumpulan data dilakukan secara daring menggunakan Google Form dengan tujuan memperoleh informasi mengenai faktor-faktor yang memengaruhi minat pembelian konsumen. Variabel yang digunakan dalam penelitian meliputi persepsi harga, rating produk, ulasan konsumen, kepercayaan terhadap platform, serta minat pembelian sebagai variabel dependen.

Data yang berhasil dikumpulkan kemudian diseleksi untuk memastikan kualitas dan kelengkapan informasi yang diberikan oleh responden. Proses seleksi dilakukan dengan memeriksa data yang tidak lengkap, duplikasi jawaban, maupun data yang tidak sesuai dengan kriteria penelitian. Hasil pengumpulan data menunjukkan bahwa seluruh responden merupakan pengguna aktif e-commerce yang pernah melakukan transaksi pembelian dalam periode enam bulan



No	Platform	Persepsi Harga	Rating Produk	Ulasan Konsumen	Kepercayaan Terhadap Platform	Minat Pembelian
1	Shopee	4.5	4.2	4.3	4.4	4.6
2	Tokopedia	4.3	4.1	4.2	4.3	4.5
3	Shopee	4.6	4.4	4.5	4.6	4.7
4	Tokopedia	4.2	4.0	4.1	4.2	4.4
5	Shopee	4.4	4.3	4.4	4.5	4.6
6	Tokopedia	4.1	3.9	4.0	4.1	4.3
7	Shopee	4.7	4.5	4.6	4.7	4.8
8	Tokopedia	4.0	3.8	3.9	4.0	4.2
9	Shopee	4.3	4.2	4.3	4.4	4.5
10	Tokopedia	4.4	4.3	4.4	4.5	4.6
11	Shopee	4.5	4.4	4.5	4.6	4.7
12	Tokopedia	4.2	4.1	4.2	4.3	4.4
13	Shopee	4.6	4.5	4.6	4.7	4.8
14	Tokopedia	4.1	3.9	4.0	4.1	4.3
15	Shopee	4.4	4.3	4.4	4.5	4.6
16	Tokopedia	4.3	4.2	4.3	4.4	4.5
17	Shopee	4.5	4.4	4.5	4.6	4.7
18	Tokopedia	4.2	4.1	4.2	4.3	4.4
19	Shopee	4.6	4.5	4.6	4.7	4.8
20	Tokopedia	4.1	3.9	4.0	4.1	4.3

terakhir.

Gambar 1 Data responden hasil kuisisioner pengguna Shopee dan Tokopedia.

Pengolahan Data

Sebelum dilakukan proses pemodelan, data terlebih dahulu melalui tahap preprocessing agar dapat digunakan secara optimal dalam algoritma machine learning. Tahapan yang dilakukan meliputi:

1. Pembersihan data (data cleaning) dengan menghapus data kosong dan data duplikat.
2. Transformasi data dari bentuk kategorikal menjadi numerik menggunakan skala Likert.

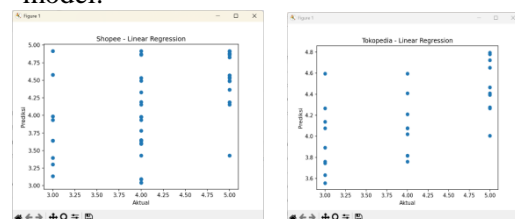
3. Pengelompokan data berdasarkan variabel penelitian.
4. Analisis statistik deskriptif untuk mengetahui karakteristik data.
5. Pembagian data menjadi data latih (training data) dan data uji (testing data).
6. Implementasi algoritma Linear Regression dan Random Forest.
7. Evaluasi performa model menggunakan metrik R^2 , MAE, dan RMSE.

Konversi skala Likert dilakukan dengan memberikan nilai 1 sampai 5 pada setiap jawaban responden. Proses ini bertujuan agar data dapat diolah menggunakan metode komputasi dan analisis statistik.

Implementasi Algoritma Linear Regression

Linear Regression digunakan untuk membangun model prediksi berdasarkan hubungan linier antara variabel independen dan variabel dependen. Model ini mengasumsikan bahwa perubahan pada variabel input akan menghasilkan perubahan yang proporsional terhadap variabel output.

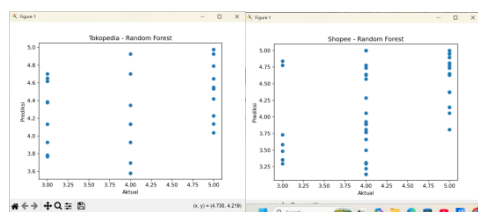
Pada penelitian ini, variabel persepsi harga, rating produk, ulasan konsumen, dan kepercayaan terhadap platform digunakan sebagai variabel prediktor untuk memprediksi minat pembelian konsumen. Setelah proses pelatihan model selesai, dilakukan pengujian menggunakan data testing untuk mengetahui kemampuan generalisasi model.



Shopee dan Tokopedia dengan mengikuti pola nilai aktual, meskipun masih terdapat beberapa titik prediksi yang menyebar. Hasil evaluasi menunjukkan bahwa pada dataset Shopee model memperoleh nilai R^2 sebesar 0,225, MAE sebesar 0,452, dan RMSE sebesar 0,603. Sementara itu, pada dataset Tokopedia diperoleh nilai R^2 sebesar 0,299, MAE sebesar 0,550, dan RMSE sebesar 0,678. Nilai R^2 yang lebih tinggi pada Tokopedia menunjukkan bahwa model Linear Regression lebih mampu menjelaskan variasi data dibandingkan pada Shopee. Secara umum, hasil pengujian menunjukkan bahwa Linear Regression mampu menangkap hubungan dasar antarvariabel, namun memiliki keterbatasan ketika menghadapi pola data yang lebih kompleks atau bersifat nonlinier.

Implementasi Algoritma Random Forest

Random Forest merupakan algoritma ensemble yang dibangun dari kombinasi beberapa decision tree. Metode ini bekerja dengan cara menggabungkan hasil prediksi dari banyak pohon keputusan sehingga mampu meningkatkan stabilitas dan akurasi mod



Gambar 3 Scatter Plot Random Forest Shoope dan Tokopedia

Berdasarkan Gambar 1.3, algoritma Random Forest mampu menghasilkan prediksi yang cukup mendekati nilai aktual pada data Shopee dan Tokopedia. Pada dataset Shopee, model memperoleh nilai R^2 sebesar 0,274, MAE sebesar 0,441, dan RMSE sebesar 0,586. Sementara itu, pada dataset Tokopedia, model menghasilkan nilai R^2 sebesar -0,086, MAE sebesar 0,679, dan RMSE sebesar 0,842. Hasil tersebut menunjukkan bahwa Random Forest memiliki performa yang lebih baik

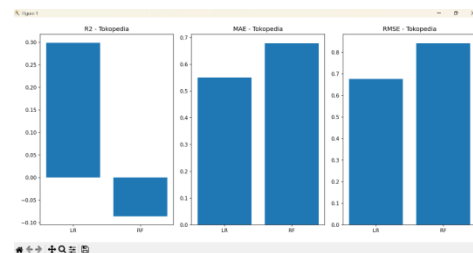
pada dataset Shopee karena mampu menjelaskan variasi data dengan lebih baik dan menghasilkan tingkat kesalahan prediksi yang lebih rendah. Sebaliknya, pada dataset Tokopedia, nilai R^2 yang negatif menunjukkan bahwa model kurang mampu menangkap pola data sehingga performanya lebih rendah dibandingkan pada dataset Shopee. Dengan demikian, algoritma Random Forest lebih sesuai digunakan untuk memprediksi minat pembelian konsumen pada data Shopee dibandingkan data Tokopedia. Implementasi Algoritma Random Forest.

Random Forest merupakan algoritma ensemble yang dibangun dari kombinasi beberapa decision tree. Metode ini bekerja dengan cara menggabungkan hasil prediksi dari banyak pohon keputusan sehingga mampu meningkatkan stabilitas dan akurasi model.

Hasil Evaluasi Model

Evaluasi performa model dilakukan menggunakan tiga metrik, yaitu R^2 Score, Mean Absolute Error (MAE), dan Root Mean Square Error (RMSE).

Gambar 4 Hasil Perbandingan Evaluasi Algoritma Shopee dan Tokopedia



Tokopedia

Evaluasi Dataset Shopee

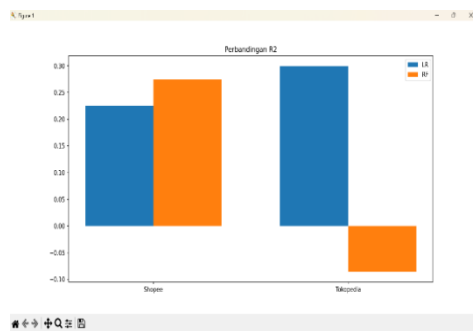
Hasil evaluasi menunjukkan bahwa Linear Regression memperoleh nilai R^2 sebesar 0,225, MAE sebesar 0,452, dan RMSE sebesar 0,603. Sementara itu, Random Forest memperoleh nilai R^2 sebesar 0,274, MAE sebesar 0,441, dan RMSE sebesar 0,586.

Nilai R^2 yang lebih tinggi pada Random Forest menunjukkan bahwa model mampu menjelaskan variasi data lebih baik dibandingkan Linear Regression. Selain itu, nilai MAE dan RMSE yang lebih rendah menunjukkan bahwa tingkat kesalahan prediksi Random Forest lebih kecil.

Evaluasi Dataset Tokopedia

Pada dataset Tokopedia, Linear Regression menghasilkan nilai R^2 sebesar 0,299, MAE sebesar 0,550, dan RMSE sebesar 0,678. Sebaliknya, Random Forest memperoleh nilai R^2 sebesar -0,086, MAE sebesar 0,679, dan RMSE sebesar 0,842.

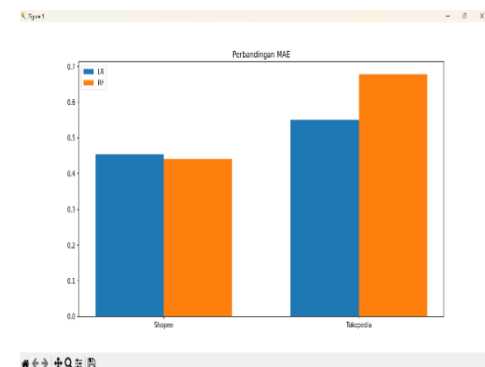
Nilai R^2 negatif pada Random Forest mengindikasikan bahwa model tidak mampu menjelaskan variasi data secara optimal. Kondisi ini menunjukkan



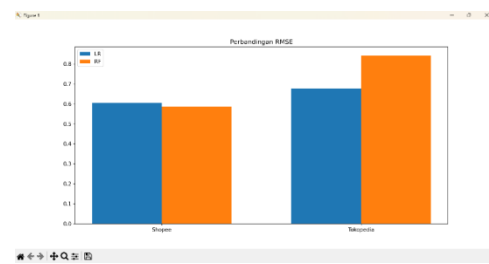
bahwa pola data Tokopedia lebih sesuai dimodelkan menggunakan pendekatan linier dibandingkan pendekatan ensemble.

Analisis Perbandingan Model

Perbandingan hasil evaluasi menunjukkan bahwa performa algoritma berbeda pada setiap platform e-commerce.



Gambar 5 Perbandingan Nilai R^2 , MAE dan RMSE



Pada Shopee, Random Forest menunjukkan performa terbaik dengan nilai R^2 yang lebih tinggi dan tingkat kesalahan yang lebih rendah. Hal ini mengindikasikan bahwa pola hubungan antarvariabel pada Shopee cenderung lebih kompleks sehingga lebih sesuai dimodelkan menggunakan algoritma berbasis ensemble.

Sebaliknya, pada Tokopedia, Linear Regression menghasilkan performa yang lebih baik. Nilai R^2 yang lebih tinggi serta nilai MAE dan RMSE yang lebih rendah menunjukkan bahwa hubungan antarvariabel pada dataset Tokopedia cenderung lebih linier.

Perbedaan hasil ini menunjukkan bahwa karakteristik data memiliki pengaruh yang signifikan terhadap kinerja algoritma machine learning.

Pembahasan

Hasil penelitian menunjukkan bahwa performa algoritma berbeda pada setiap dataset. Pada data Shopee, algoritma Random Forest memberikan hasil yang lebih baik dengan nilai R^2 sebesar 0,274, MAE sebesar 0,441, dan RMSE sebesar 0,586. Hal ini menunjukkan bahwa Random Forest lebih mampu menangkap pola hubungan yang kompleks antarvariabel.

Sebaliknya, pada data Tokopedia, algoritma Linear Regression menghasilkan performa yang lebih baik dengan nilai R^2 sebesar 0,299, MAE sebesar 0,550, dan RMSE sebesar 0,678. Hasil tersebut mengindikasikan bahwa hubungan antarvariabel pada dataset Tokopedia cenderung bersifat linier sehingga lebih sesuai dimodelkan menggunakan Linear Regression.

Secara keseluruhan, hasil penelitian menunjukkan bahwa pemilihan algoritma harus disesuaikan dengan karakteristik data yang digunakan. Random Forest lebih optimal untuk dataset Shopee, sedangkan Linear Regression lebih efektif untuk dataset Tokopedia dalam memprediksi minat pembelian konsumen pada platform e-commerce.

SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Linear Regression dan Random Forest dapat digunakan untuk memprediksi minat pembelian konsumen pada platform e-commerce Shopee dan Tokopedia. Hasil evaluasi menggunakan metrik R^2 , MAE, dan RMSE menunjukkan bahwa performa kedua algoritma berbeda pada setiap dataset. Pada data Shopee, Random Forest memberikan hasil yang lebih baik dengan nilai R^2 sebesar 0,274, MAE sebesar 0,441, dan RMSE sebesar 0,586. Sementara itu, pada data Tokopedia, Linear Regression menunjukkan performa yang lebih baik dengan nilai R^2 sebesar 0,299, MAE sebesar 0,550, dan RMSE sebesar 0,678. Hasil penelitian ini menunjukkan bahwa tidak terdapat satu algoritma yang selalu unggul untuk seluruh data. Performa model sangat dipengaruhi oleh karakteristik dataset yang digunakan. Random Forest lebih efektif dalam menangani pola data yang kompleks pada Shopee, sedangkan Linear Regression lebih sesuai untuk data Tokopedia yang memiliki hubungan variabel yang cenderung linier. Dengan demikian, tujuan penelitian untuk membandingkan performa kedua algoritma telah tercapai, dan hasil penelitian ini dapat menjadi referensi dalam pemilihan metode prediksi minat pembelian konsumen pada platform e-commerce.

DAFTAR PUSTAKA

A. Balyemah, S. Weamie, J. Bin, K. V. Jarnda, and F. J. Joshua,

“Predicting Purchasing Behavior on E-Commerce Platforms: A Regression Model Approach for Understanding User Features that Lead to Purchasing,” *International Journal of Communications, Network and System Sciences*, vol. 17, no. 6, pp. 81–103, 2024.

- D. E. Kurniawan and T. Erwina, “Enhancing Consumer Purchasing Pattern Prediction on E-Commerce Platforms Using Random Forest Algorithm: A Simulation Approach,” in *Proc. International Conference on Applied Engineering (ICAE)*, 2025, pp. 45–53.
- M. Saputra and A. Fersellia, “Prediksi Loyalitas Pelanggan E-Commerce Menggunakan Algoritma Random Forest pada Platform Shopee,” *Jurnal Teknologi Informasi dan Komputer*, vol. 5, no. 2, pp. 120–128, 2022.
- A. Ruscikasani, R. Prasetyo, and D. Wijayanti, “Prediction of Online Shopper Purchase Intention Based on Session Behavior Analysis,” *Journal of Data Science and Information Systems*, vol. 4, no. 1, pp. 15–25, 2021.
- H. Wijaya and L. Fauziah, “Comparative Analysis of Machine Learning Algorithms for Consumer Purchase Prediction,” *International Journal of Computer Applications*, vol. 184, no. 42, pp. 12–20, 2023.
- L. Breiman, “Random Forest-Based Consumer Behavior Prediction in E-Commerce Applications,” *Machine Learning Applications*, vol. 8, pp. 101–112, 2021.
- Y. Zhang and Q. Yu, “User Purchase Intention Prediction Based on Improved Deep Forest,” *Computer Modeling in Engineering & Sciences*, vol. 137, no. 2, pp. 1431–1450, 2023.
- A. Syah, F. Nurdiyansyah, and A. Y. Rahman, “Analisis Sentimen Aplikasi Shopee, Tokopedia,

- Lazada dan Blibli Menggunakan Leksikon dan Random Forest,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, pp. 1–10, 2024.
- E. S. Barus and D. Darmanto. (2024). Implementasi Metode Random Forest untuk Memprediksi Penjualan Produk. *Jurnal Teknik Informasi dan Komputer (Tekinkom)*.
- S. Ahmed and M. Ibrahim, “Machine Learning-Based Random Forest Prediction of Online Shopping Behavior,” in *Proc. ACM International Conference on Data Analytics*, 2025, pp. 221–230.
- M. Hossain and T. Rahman, “Random Forest Approach for E-Commerce Demand Forecasting and Purchase Prediction,” *Journal of Artificial Intelligence Research*, vol. 18, no. 4, pp. 211–223, 2022.
- W. Chen and Y. Li, “Comparative Study of Linear Regression and Ensemble Learning Methods for Consumer Purchase Prediction,” *Expert Systems with Applications*, vol. 201, pp. 117–126, 2022.
- A. Jain, “Predicting E-Commerce Purchase Behavior Using a DQN-Inspired Deep Learning Model for Enhanced Adaptability,” *arXiv preprint arXiv:2506.17543*, 2025.
- A. Verma, “Consumer Behaviour in Retail: Next Logical Purchase Using Deep Neural Network,” *arXiv preprint arXiv:2010.06952*, 2020.
- S. Gupta and P. Sharma, “E-Commerce Consumer Purchase Prediction Using Machine Learning Techniques,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 455–463, 2022.
- R. Kumar and V. Singh, “Machine Learning Models for Online Shopping Purchase Intention Analysis,” *Journal of Big Data Analytics*, vol. 7, no. 2, pp. 67–79, 2023.
- N. Sari and M. Kurniawan, “Pengaruh Persepsi Harga dan Ulasan Konsumen terhadap Minat Pembelian pada Marketplace Shopee,” *Jurnal Manajemen dan Bisnis Digital*, vol. 4, no. 2, pp. 89–98, 2022.
- F. Ramadhan and D. Putra, “Analisis Pengaruh Rating Produk terhadap Keputusan Pembelian Konsumen pada E-Commerce Tokopedia,” *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 6, no. 1, pp. 45–53, 2023.
- A. Pratama and R. Lestari, “Kepercayaan Platform dan Minat Pembelian Konsumen pada Marketplace Indonesia,” *Jurnal Ekonomi Digital*, vol. 5, no. 3, pp. 120–129, 2024.
- I. Maulana and S. Hidayat, “Machine Learning Approaches for Consumer Purchase Intention Prediction in Indonesian E-Commerce,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 33–44, 2024.