

Implementasi Machine Learning untuk Analisis Pola Perilaku Konsumen E-Commerce Berbasis Data Transaksi

Machine Learning Implementation for Analyzing Consumer Behavior Patterns in E-Commerce Transactions

Muhammad Yusri Khalid Mahendra Yosri¹; Hanni²; Muhammad Nur Arafah³; Nurul Salzabilah⁴; Maryam⁵; Iqra Sarbakti Nur⁶; Muhammad Fauzan⁷;

^{1,2,4} Universitas Pancasakti, Makassar 90211, Indonesia

³ Irmex Digital Akademika, Makassar 90551, Indonesia

^{5,6} Universitas Indonesia Timur, Makassar 90224, Indonesia

⁷ Universitas Islam Negeri Alauddin Makassar, Gowa 92113, Indonesia

¹ yusrikhalid05@gmail.com; ² hannypalu.88@gmail.com; ³ mnurarafah18@gmail.com; ⁴ slblhh41@gmail.com; ⁵ maryam.mary7741@gmail.com;

⁶ igrasarbaktinurpps@gmail.com; ⁷ fauzanbastard28@gmail.com;

* Corresponding author

Abstrak

Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan performa berbagai algoritma machine learning dalam pemetaan pasar global e-commerce berdasarkan pola penjualan dan perilaku konsumen. Data yang digunakan merupakan dataset transaksi e-commerce global yang mencakup kategori produk, harga, jumlah pembelian, wilayah konsumen, serta waktu transaksi. Tahapan penelitian meliputi pra-pemrosesan data, penerapan metode klasifikasi, evaluasi performa, dan analisis hasil. Algoritma yang diuji mencakup Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Linear Regression, dan K-Nearest Neighbor (KNN). Untuk mengatasi ketidakseimbangan data, digunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa terbaik dengan akurasi 99,87%, diikuti oleh Gradient Boosting (99,82%) dan Decision Tree (99,51%). Penerapan SMOTE terbukti efektif dalam meningkatkan keseimbangan kelas serta performa metrik seperti Accuracy, Precision, Recall, dan F1-Score. Selain itu, feature engineering turut memperkuat kemampuan model dalam mengenali pola kompleks perilaku konsumen. Secara keseluruhan, kombinasi antara algoritma ensemble, SMOTE, dan feature engineering menghasilkan model klasifikasi yang andal untuk menganalisis perilaku konsumen di pasar e-commerce global. Penelitian ini memberikan kontribusi penting bagi pengembangan strategi bisnis digital berbasis analisis data serta penerapan machine learning dalam konteks ekonomi digital.

Kata Kunci: Business Intelligence, E-Commerce, Machine Learning, Random Forest, Big Data

Abstract

This research aims to implement and compare the performance of various machine learning algorithms for mapping global e-commerce markets based on sales patterns and consumer behavior. The data used consist of global e-commerce transaction records covering product categories, prices, purchase quantities, consumer regions, and transaction times. The research stages include data preprocessing, classification model implementation, performance evaluation, and result analysis. The evaluated algorithms include Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Linear Regression, and K-Nearest Neighbor (KNN). To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The results indicate that the Random Forest model achieved the best performance with an accuracy of 99.87%, followed by Gradient Boosting (99.82%) and Decision Tree (99.51%). The application of SMOTE proved effective in improving class balance and enhancing metrics such as Accuracy, Precision, Recall, and F1-Score. Additionally, feature engineering strengthened the model's ability to capture complex consumer behavior patterns. Overall, the combination of ensemble algorithms, SMOTE, and feature engineering produced a reliable classification model for analyzing consumer behavior in global e-commerce markets. This study contributes significantly to the development of data-driven digital business strategies and the application of machine learning in the digital economy context.

Keywords: Business Intelligence; E-commerce; Machine Learning; Random Forest; Big Data.

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mengubah lanskap perdagangan global, dengan e-commerce menjadi pilar utama ekonomi digital [1]. Aktivitas jual-beli kini beralih dari toko fisik ke platform daring yang lebih efisien dan luas jangkauannya [2]. Nilai penjualan ritel daring dunia diproyeksikan meningkat dari US\$

4,4 triliun pada 2023 menjadi US\$ 6,8 triliun pada 2028 dengan pertumbuhan tahunan 8,9% [3]. Menurut Boston Consulting Group (2023), e-commerce akan menyumbang 41% penjualan ritel global pada 2027. Pertumbuhan ini didorong oleh penetrasi internet, peningkatan kepercayaan transaksi digital, serta transformasi bisnis menuju digital [4]. Namun, pasar e-commerce bersifat heterogen dengan variasi perilaku konsumen, daya beli, dan budaya [5]. sehingga analisis berbasis data diperlukan untuk memahami pola tersebut [6]. Machine learning (ML) menjadi pendekatan penting dalam menganalisis data besar e-commerce [7]. Metode klasifikasi digunakan untuk mengidentifikasi pola perilaku konsumen dan memetakan pasar berdasarkan data transaksi [5]. Algoritma seperti Decision Tree, Random Forest, Gradient Boosting, dan Logistic Regression banyak diterapkan karena kemampuannya menangani data multivariabel [8]. Model ensemble seperti Random Forest dan Gradient Boosting terbukti meningkatkan akurasi prediksi perilaku pembelian pelanggan [9][10]. Meski demikian, tantangan seperti heterogenitas data dan kebutuhan interpretabilitas model masih perlu diatasi [6]. Di Indonesia, penelitian oleh Saragih dan Siregar (2022) serta Setiawan et al. (2023) menunjukkan efektivitas supervised learning dalam mengidentifikasi segmen pelanggan potensial. Dengan demikian, penelitian ini berfokus pada penerapan algoritma klasifikasi machine learning untuk pemetaan pasar global e-commerce berdasarkan pola penjualan dan perilaku konsumen, serta mengevaluasi performa model terbaik guna mendukung strategi bisnis digital [11], [12], [13]

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi machine learning untuk menganalisis dan memetakan pola perilaku konsumen dalam e-commerce berdasarkan data transaksi dan karakteristik pelanggan. Tahapan penelitian dilakukan secara sistematis melalui beberapa langkah berikut:

A. Analisis Kebutuhan

Pada tahap ini, dilakukan pengumpulan data untuk mengidentifikasi kebutuhan sistem yang akan dikembangkan. Beberapa teknik pengumpulan data yang digunakan antara lain:

- Studi Pustaka: Mengkaji berbagai literatur, laporan penelitian, dan jurnal ilmiah yang berkaitan dengan implementasi algoritma machine learning dalam analisis pasar dan perilaku konsumen e-commerce. Tahap ini penting untuk memperoleh dasar teori, metodologi, serta referensi ilmiah yang relevan dalam pengembangan model pemetaan pasar global.
- Observasi: Melakukan observasi terhadap dataset e-commerce yang diperoleh dari platform Kaggle, yang berisi data transaksi dan perilaku konsumen dari berbagai negara. Observasi ini digunakan untuk memahami tren penjualan, preferensi pembelian, serta segmentasi konsumen berdasarkan aktivitas dan pola konsumsi di pasar global. (Statista, 2024; Kaggle, 2023). [14], [15]

B. Tahapan Penelitian

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini terdiri dari 5.000 baris data transaksi e-commerce dengan 15 variabel utama yang mencakup Order ID, Order Date, Customer Name, Region, City, Category, Sub-Category, Product Name, Quantity, Unit Price, Discount, Sales, Profit, Payment Mode, Sales Class sebagai variabel target. Seluruh data ini merepresentasikan aktivitas penjualan dari berbagai wilayah (North, South, East, dan West) dalam periode tahun 2023–2025. Dataset diperoleh dari sumber terbuka (open dataset) yang berisi catatan transaksi aktual pada platform e-commerce global. Data ini digunakan untuk menganalisis pola penjualan dan perilaku konsumen, dengan variabel seperti Sales dan Profit sebagai indikator performa penjualan, serta Category dan Region sebagai faktor segmentasi pasar. Dataset kemudian digunakan sebagai dasar untuk pelatihan model klasifikasi machine learning, dengan tujuan memetakan pasar global e-commerce dan mengidentifikasi pola perilaku konsumen berdasarkan kategori produk, wilayah, dan metode pembayaran.

2. Pra-pemrosesan Data

Tahap ini meliputi pembersihan data (data cleaning) untuk menghapus nilai hilang dan duplikat, normalisasi atribut numerik, serta encoding variabel kategorikal agar dapat diolah oleh algoritma machine learning. Selain itu, dilakukan pembagian data menjadi data latih (training) dan data uji (testing) dengan rasio tertentu, misalnya 70:30.

C. Penerapan Metode Klasifikasi

Beberapa algoritma klasifikasi diterapkan untuk membandingkan performa model. Setiap algoritma dilatih menggunakan data latih untuk mengenali pola perilaku konsumen, kemudian diuji pada data uji untuk menilai akurasi prediksi.

Tabel 1 Set Paramater Pelatihan Model

<i>Model</i>	<i>Parameter Pelatihan</i>
Random Forest	n_estimators=100, max_depth=10, random_state=42
Gradient Boosting	n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42
Decision Tree	max_depth=10, random_state=42
AdaBoost	n_estimators=100, learning_rate=0.1, random_state=42
LinearRegression	fit_intercept=True, normalize=False
KNN	n_neighbors=5, metric='minkowski', p=2

D. Evaluasi Model

Evaluasi performa dilakukan menggunakan empat indikator utama, Accuracy, Precision, Recall, dan F1-Score, yang masing-masing memberikan perspektif berbeda terhadap kemampuan model. Setelah perhitungan metrik dilakukan, hasil dievaluasi secara komparatif untuk menilai kekuatan dan kelemahan tiap algoritma. Pendekatan ini bertujuan memperoleh model klasifikasi yang tidak hanya memiliki tingkat akurasi tinggi, tetapi juga stabil dan andal dalam konteks data e-commerce yang heterogen.

Rumus Evaluasi Kinerja Model

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Gambar 1 Rumus Perhitungan Evaluasi Model Machine Learning

E. Analisis Hasil dan Interpretasi

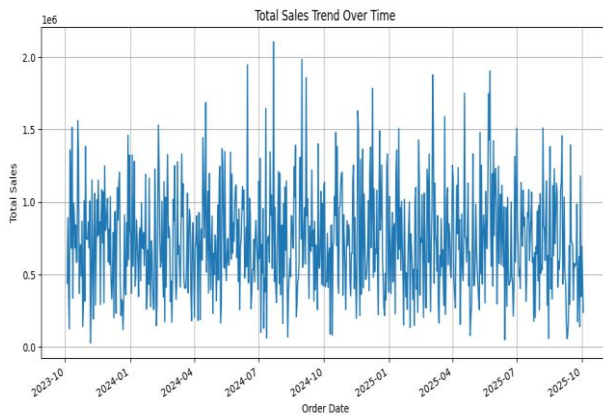
Hasil evaluasi digunakan untuk mengidentifikasi model dengan akurasi tertinggi serta memahami faktor-faktor yang paling berpengaruh terhadap klasifikasi perilaku konsumen. Analisis ini menjadi dasar dalam memberikan rekomendasi strategis bagi pengembangan e-commerce global.

F. Perangkat yang Digunakan

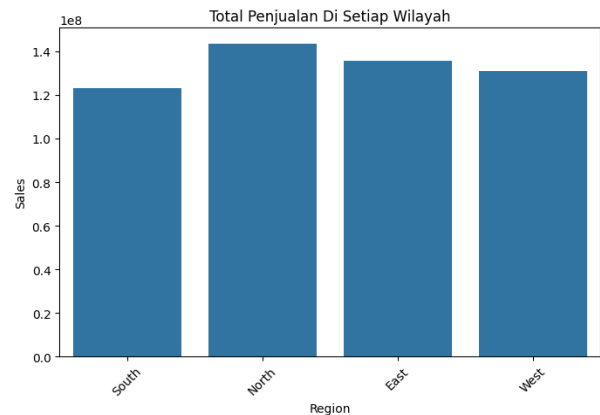
Implementasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka seperti Scikit-learn, Pandas, dan NumPy untuk pemodelan dan analisis data.

Hasil dan Diskusi

Berdasarkan serangkaian eksperimen yang dilakukan dalam penelitian ini untuk mengklasifikasikan pola penjualan dan perilaku konsumen menggunakan berbagai algoritma machine learning, dapat disimpulkan bahwa pendekatan tersebut mampu menghasilkan performa klasifikasi yang cukup baik. Beberapa algoritma yang diimplementasikan meliputi Random Forest, Gradient Boosting, Decision Tree, AdaBoost, K-Nearest Neighbor (KNN), dan Linear Regression, dengan hasil evaluasi menunjukkan variasi tingkat akurasi pada masing-masing model. Keberhasilan metode klasifikasi ini tidak terlepas dari penerapan teknik Synthetic Minority Over-sampling Technique (SMOTE) yang berperan penting dalam mengatasi ketidakseimbangan kelas pada dataset. Dengan menyeimbangkan jumlah sampel antara kelas mayoritas dan minoritas, SMOTE membantu meningkatkan kualitas proses pembelajaran model serta mengurangi bias terhadap kelas dominan. Pendekatan ini terbukti efektif dalam meningkatkan nilai Accuracy, Precision, Recall, dan F1-Score secara keseluruhan, sehingga model yang dihasilkan menjadi lebih representatif dan andal dalam memetakan pola perilaku konsumen di pasar e-commerce global.



Gambar 2 Total Penjualan periode 2023-2025

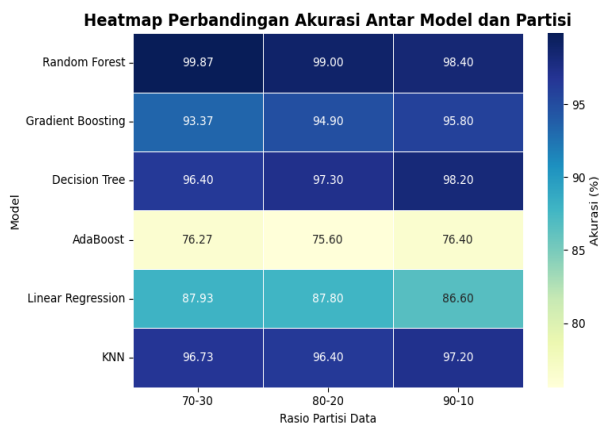


Gambar 3 Total Penjualan Di Setiap Wilayah

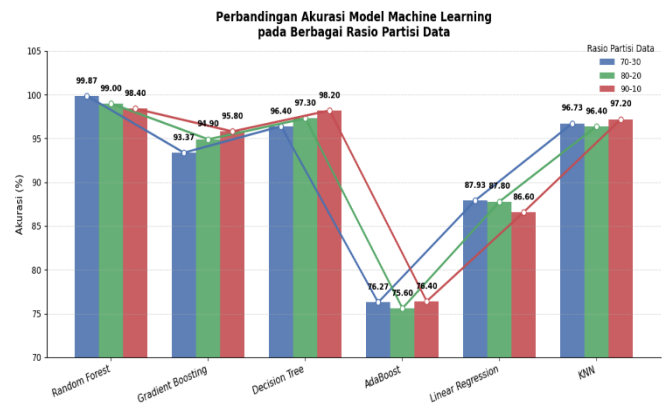
Tabel 1 Hasil Evaluasi Kinerja Model Machine Learning

<i>Model Machine Learning</i>	<i>Akurasi Model</i>
Random Forest	99.88 %
Gradient Boosting	99.82 %
Decision Tree	99.51 %
AdaBoost	95.49 %
LinearRegression	88.78 %
KNN	65.67 %

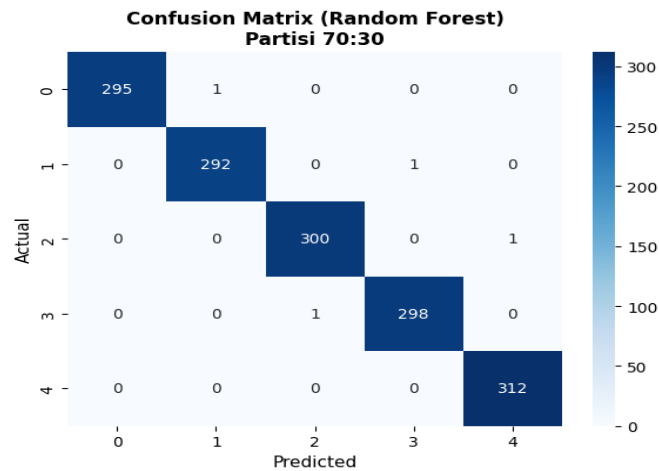
Pada gambar 2 merupakan grafik hasil visualisasi total transaksi penjualan dari tahun 2023 hingga 2025. Berdasarkan hasil yang ditunjukkan pada Tabel 2, metode Random Forest memberikan performa terbaik dengan nilai Accuracy sebesar 99.87%, Hasil tersebut menunjukkan bahwa model klasifikasi memiliki keseimbangan yang baik antara kemampuan untuk mengidentifikasi kelas positif dengan tepat (Precision) dan menangkap seluruh data positif dalam dataset (Recall). Hal ini mengindikasikan bahwa pendekatan ensemble, seperti Random Forest dan Gradient Boosting, mampu mempelajari pola kompleks secara lebih efektif dibandingkan metode individual seperti KNN atau Linear Regression. Namun demikian, performa model sangat bergantung pada kualitas data dan relevansi fitur yang digunakan selama proses pelatihan. Dalam penelitian ini, penerapan Synthetic Minority Over-sampling Technique (SMOTE) terbukti berperan penting dalam meningkatkan performa klasifikasi. Teknik ini membantu menyeimbangkan distribusi antar kelas dengan menambah representasi data pada kelas minoritas, sehingga model dapat belajar secara lebih seimbang dan menghindari bias terhadap kelas mayoritas. Dengan demikian, penerapan SMOTE berhasil meningkatkan efektivitas pembelajaran mesin dan memungkinkan model mengenali pola perilaku konsumen secara lebih optimal.



Gambar 4 Visualisasi Heatmap partisi ratio data



Gambar 5 Visualisasi Barchart Perbandingan Model Partisi data



Gambar 6 Visualisai Confusion Matrix Random Forest 70:30

Berdasarkan hasil visualisasi pada Gambar 4 dan Gambar 5, dapat diinterpretasikan bahwa model Random Forest dengan partisi data 70:30 menunjukkan kinerja terbaik dibandingkan model lainnya. Oleh karena itu, untuk memperkuat dan memperjelas temuan penelitian ini, ditambahkan pula hasil confusion matrix sebagaimana disajikan pada Gambar 6, yang berfungsi untuk menggambarkan kemampuan model dalam memprediksi pola perilaku konsumen dalam berbelanja.

Kesimpulan

Penelitian ini berhasil mengklasifikasikan pemetaan pasar global e-commerce berdasarkan pola penjualan dan perilaku konsumen menggunakan algoritma machine learning yang dipadukan dengan Synthetic Minority Over-Sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan data. Dataset yang digunakan berjumlah 5.000 baris data yang diperoleh dari sumber terbuka (open source) di platform Kaggle. Pendekatan berbasis machine learning ini mampu menganalisis pola perilaku konsumen secara mendalam, mengidentifikasi tren pembelian, serta mengoptimalkan segmentasi pasar secara otomatis. Hasil evaluasi menunjukkan peningkatan signifikan pada Accuracy, Precision, Recall, dan F1-Score dibandingkan metode konvensional. Model Random Forest memberikan performa terbaik dengan akurasi 99,87%, precision 99,57%, recall 99,67%, dan F1-score 99,84%,. Kombinasi teknik SMOTE dan feature engineering terbukti efektif dalam menyeimbangkan distribusi kelas, memperkaya representasi data, dan memperkuat kemampuan model dalam mengenali pola perilaku konsumen yang kompleks. Penelitian ini merekomendasikan pengembangan lebih lanjut melalui perluasan atribut perilaku konsumen, penerapan feature engineering berbasis psikologis, serta integrasi Explainable AI (SHAP, LIME) untuk meningkatkan transparansi hasil prediksi dan interpretabilitas model. Secara keseluruhan, pendekatan machine learning ini menunjukkan potensi besar dalam mendukung analisis perilaku konsumen, pemetaan pasar, serta strategi data-driven marketing intelligence di era ekonomi digital.

Daftar Pustaka

- [1] K. Kraemer, J. Dedrick, dan N. Melville, "E-business and digital platforms in global trade," *Inf. Syst. Res.*, vol. 32, no. 4, hal. 911–927, 2021.
- [2] A. Kumar dan N. Garg, "Customer segmentation using supervised learning algorithms in e-commerce," in *Procedia Computer Science*, 2021, hal. 365–374.
- [3] K. C. Laudon dan C. G. Traver, *E-commerce 2022: Business, Technology, and Society*. Pearson Education, 2022.
- [4] PwC, "Global Digital Transformation Outlook 2023," 2023.
- [5] J. Li, Y. Wang, dan L. Zhao, "A comparative study of classification algorithms for e-commerce customer analysis," *Appl. Soft Comput.*, vol. 92, hal. 106321, 2020.
- [6] OECD, "Digital Trade and E-commerce in the Global Economy," 2022.
- [7] R. F. Syam, "Pengembangan Kompetensi Sdm Mitra Cv Sehat Kerupuk Melalui Pelatihan Machine Learning Dan Ilmu Komunikasi," *J. Pengabd. Kpd. Masy. TEKNO*, vol. 4, no. 2, hal. 1–8, 2024.
- [8] Q. Liu, Y. Zhang, dan L. Chen, "Machine learning applications in consumer behavior prediction," *Expert Syst. Appl.*, vol. 213, hal. 118926, 2023.
- [9] T. Zhou, X. Wang, dan F. Li, "Gradient boosting approaches for consumer purchase prediction,"

- Expert Syst. Appl.*, vol. 212, hal. 118897, 2023.
- [10] R. Zhang, S. Liu, dan T. Zhou, "Evaluation of ensemble classification methods for e-commerce behavior analysis," *IEEE Access*, vol. 11, hal. 12245–12259, 2023.
 - [11] UNCTAD, "E-commerce and Development Report 2022," 2022.
 - [12] A. Saragih dan D. Siregar, "Penerapan algoritma klasifikasi untuk segmentasi pelanggan e-commerce di Indonesia," *J. Inform. dan Komput.*, vol. 7, no. 2, hal. 45–53, 2022.
 - [13] R. Setiawan, A. Hartono, dan S. Nurhadi, "Analisis perilaku pelanggan e-commerce menggunakan supervised learning," *J. Teknol. dan Sist. Inf.*, vol. 11, no. 1, hal. 14–22, 2023.
 - [14] Statista, "E-commerce share of total global retail sales from 2017 to 2027," 2024.
 - [15] World Economic Forum, "The Future of E-commerce in the Digital Age," 2022.