JURNAL MANAJEMEN AKUNTANSI DAN ILMU EKONOMI

Volume 3, Nomor 1 (April, 2026), Page: 8-14
Homepage : https://jurnal.ananpublisher.com/index.php/jumali

PREDIKSI PEMBATALAN PESANAN E-COMMERCE INDONESIA MENGGUNAKAN RANDOM FOREST DAN SMOTE

Felixiana Koten
Institut Bisnis Dan Teknologi Indonesia
felixiana0206@gmail.com

I Wayan Sudiarsa
Institut Bisnis Dan Teknologi Indonesia
sudiarsa@instiki.ac.id

Ni Komang Trisnawati
Institut Bisnis Dan Teknologi Indonesia
ntrisnawati317@gmail.com

Magdalena Matildis Palo Pera
Institut Bisnis Dan Teknologi Indonesia
rodrikgues15@gmail.com

Maria Avilia Ndinin
Institut Bisnis Dan Teknologi Indonesia
afrilanjela0@gmail.com

Corresponding author: sudiarsa@instiki.ac.id

Received: 17 Januari 2026 | Revised: 21 Januari 2026 | Published: 06 April 2025

Abstract

The high rate of order cancellations poses a serious challenge to the operational efficiency and profitability of the e-commerce industry in Indonesia. This study aims to identify the determinants that trigger cancellations and develop a machine learning-based predictive model. The method used is quantitative with a data mining approach using the Random Forest algorithm. The data source comes from the secondary Kaggle dataset "Indonesia E-Commerce Sales and Shipping 2023-2025" which includes 19,189 transaction data. The research stages include data pre-processing, the use of the SMOTE (Synthetic Minority Over-sampling Technique) technique to address class imbalance, model development, and evaluation using a confusion matrix. The results show that the Random Forest model is able to predict cancellations with a very high accuracy of 99.94%. The main factors that most influence cancellation decisions are payment methods (especially Online Payment and COD) and transaction value. These findings provide managerial implications for platform providers to tighten payment verification and increase cost transparency to mitigate the risk of future cancellations.

Keywords**: **Order Cancellation, E-Commerce, Random Forest, SMOTE, Consumer Behavior

Abstrak

Tingginya angka pembatalan pesanan (order cancellation) menjadi tantangan serius bagi efisiensi operasional dan profitabilitas industri e-commerce di Indonesia. Penelitian ini bertujuan untuk mengidentifikasi variabel determinan yang memicu pembatalan serta membangun model prediktif berbasis machine learning. Metode yang digunakan adalah kuantitatif dengan pendekatan data mining menggunakan algoritma Random Forest. Sumber data berasal dari dataset sekunder Kaggle "Indonesia E-Commerce Sales and Shipping 2023-2025" yang mencakup
19.189 data transaksi. Tahapan penelitian meliputi pra-pemrosesan data, penggunaan teknik SMOTE (Synthetic Minority Over-sampling Technique) untuk menangani ketidakseimbangan kelas, pembangunan model, dan evaluasi menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model Random Forest mampu memprediksi pembatalan dengan akurasi sangat tinggi sebesar 99,94%. Faktor utama yang paling berpengaruh terhadap keputusan pembatalan adalah metode pembayaran (khususnya Online Payment dan COD) serta nilai transaksi. Temuan ini memberikan implikasi manajerial bagi penyedia platform untuk memperketat verifikasi pembayaran dan meningkatkan transparansi biaya guna memitigasi risiko pembatalan di masa depan.

Kata kunci: Pembatalan Pesanan, E-Commerce, Random Forest, SMOTE, Perilaku Konsumen.

PENDAHULUAN

Meskipun e-commerce telah berkembang menjadi komponen penting dalam ekonomi Indonesia sebagai hasil dari transformasi digital, kemajuan ini telah terhambat oleh pembatalan pesanan konsumen secara sepihak. Pembatalan pesanan dalam manajemen rantai pasok tidak hanya mengakibatkan penurunan angka penjualan, tetapi juga mengakibatkan biaya logistik "jarak terakhir" (delivery last mile), biaya pengemasan yang terbuang, dan ketidakpastian dalam manajemen inventaris. Ketika konsumen melihat biaya tambahan di akhir transaksi, perilaku pembelian impulsif atau masalah psikologis lainnya dapat menyebabkan fenomena ini terjadi.

Ketidakmampuan platform e-commerce untuk mengidentifikasi pesanan yang berisiko tinggi batal sebelum diproses menjadi tujuan penelitian ini. Kajian ini bertujuan untuk memetakan karakteristik konsumen dan transaksi yang menyebabkan pembatalan dengan menggunakan data historis transaksi dari tahun 2023–2025. Algoritma Random Forest diharapkan dapat membantu pelaku industri mengambil tindakan pencegahan yang lebih akurat, meningkatkan efisiensi arus kas dan kepuasan layanan secara keseluruhan.

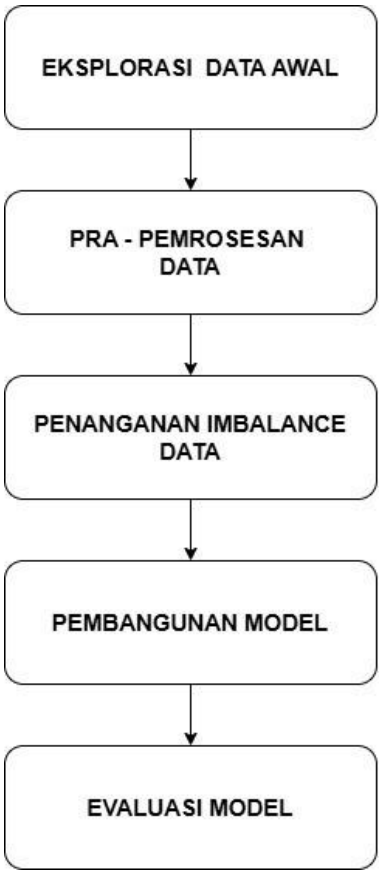
Kajian Teori

Teori disonansi kognitif menjelaskan pembatalan pesanan jika pelanggan merasa ragu setelah melakukan komitmen pembelian. Kemudahan akses, seperti fitur Cash on Delivery (COD), seringkali membuat komitmen keuangan konsumen lebih rendah, sehingga lebih mungkin untuk pembatalan dibandingkan

metode prabayar. Ketidakseimbangan data juga dikenal sebagai kelas yang tidak seimbang merupakan masalah utama dalam klasifikasi ini dari sudut pandang teknis pengolahan data; jumlah transaksi sukses jauh melampaui jumlah transaksi batal. Teknik SMOTE menyeimbangkan distribusi kelas secara sintetis, sehingga model memprediksi kesuksesan akurat dan peka terhadap pola pembatalan. Random Forest dipilih karena keunggulannya dalam mengolah data besar dan memberikan interpretasi fitur yang kuat.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan **kuantitatif** dengan jenis penelitian **deskriptif-kausalitas**. Data bersumber dari Kaggle *"Indonesia E-Commerce Sales and Shipping 2023-2025"* yang mencakup 19.189 observasi dengan 519 fitur setelah dilakukan transformasi data.



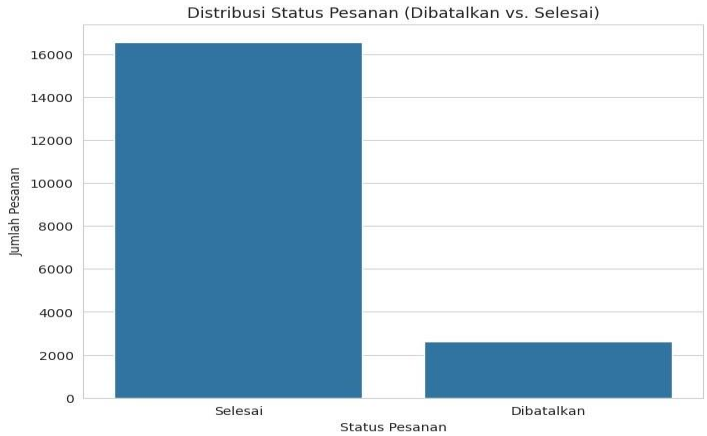
Gambar 1. Proses Klarifikasi

Langkah-langkah sistematis yang dilakukan adalah:

1. Eksplorasi Data Awal (EDA): Menemukan bahwa data awal sangat timpang (86,31% Selesai vs 13,69% Batal).
2. Pra-pemrosesan Data: Melakukan *one-hot encoding* pada variabel kategori seperti metode pembayaran dan lokasi geografis.
3. Penanganan Imbalance Data: Menerapkan SMOTE untuk menciptakan keseimbangan pada 16.562 sampel per kelas, sehingga total data latih menjadi 33.124 sampel.
4. Pembangunan Model: Menggunakan Random Forest untuk melatih data dengan pembagian *training* dan *testing* yang ketat.
5. Evaluasi: Menguji model melalui metrik akurasi, presisi, dan *recall* untuk memastikan validitas hasil.

HASIL DAN PEMBAHASAN

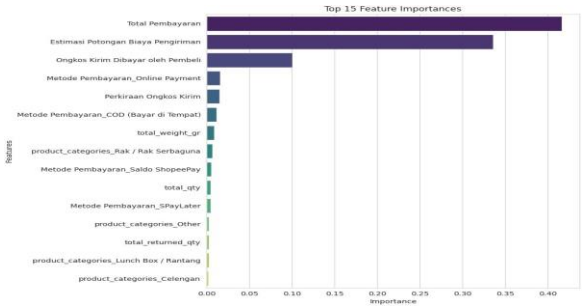
Berdasarkan pengolahan data menggunakan Google Colab, ditemukan pola-pola signifikan yang digambarkan dalam visualisasi berikut:



Gambar 2. Distribusi Status Pesanan Awal

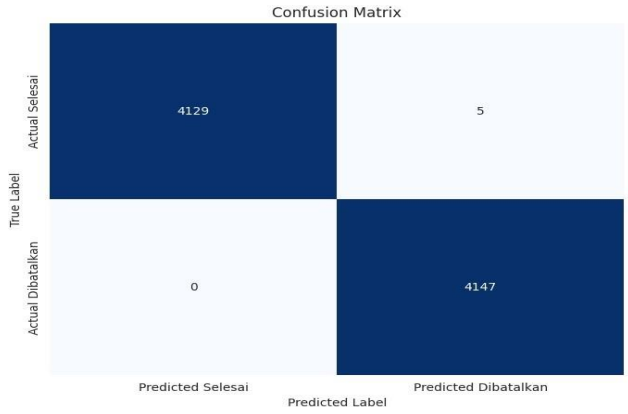
Sumber: Google Coleb (2026)

Gambar 2 menunjukkan bahwa tanpa penanganan SMOTE, model akan cenderung mengabaikan data pembatalan karena jumlahnya yang sedikit. Namun, setelah diseimbangkan, performa model meningkat tajam dalam mengenali ciri-ciri pembatalan.



Gambar 3. Feature Importance Model Random Forest
Sumber: Google Coleb (2026)

Penjelasan Gambar 3 mengungkap bahwa **Metode Pembayaran** adalah variabel paling krusial. Transaksi yang menggunakan *Online Payment* dan COD menunjukkan frekuensi pembatalan yang lebih tinggi. Hal ini menyarankan bahwa kemudahan memesan tanpa transfer di muka (pada COD) berkontribusi pada rendahnya komitmen pelanggan.



Gambar 4. Confusion Matrix Hasil Prediksi
Sumber: Google Coleb (2026)

Gambar 4 membuktikan efektivitas model dengan akurasi mencapai **99,94%**. Hasil ini menjawab permasalahan penelitian bahwa teknologi *machine learning* mampu memberikan peringatan dini bagi manajemen e-commerce sebelum mereka mengeluarkan biaya operasional untuk pesanan yang kemungkinan besar akan dibatalkan.

KESIMPULAN

Studi ini menemukan bahwa algoritma Random Forest dan teknik SMOTE dikombinasikan dengan sangat baik dapat memprediksi kemungkinan

pembatalan pesanan di e-commerce Indonesia dengan akurasi 99,94%. Nilai transaksi, serta metode pembayaran, terutama COD dan pembayaran online, adalah faktor utama yang menyebabkan pembatalan. Untuk mengurangi kerugian operasional, manajemen menyarankan penyedia platform untuk memperketat verifikasi pembayaran dan meningkatkan transparansi biaya. Penelitian lebih lanjut dapat mempelajari perilaku pembatalan pesanan di ekosistem digital yang lebih dinamis dengan melihat variabel demografi pelanggan atau algoritma deep learning.

DAFTAR PUSTAKA

Bakitacos. (2025). *Indonesia E-Commerce Sales and Shipping 2023-2025* [Data set]. Kaggle. https://www.kaggle.com/datasets/bakitacos/indonesia-e- commerce-sales-and-shipping-20232025

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14(1), 106. https://doi.org/10.1186/1471-2105-14-106

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

Davagdorj, K., Pham, V. H., Theera-Umpon, N., & Ryu, K. H. (2020). XGBoost- Based Framework for Smoking-Induced Noncommunicable Disease Prediction. *International Journal of Environmental Research and Public Health*, 17(18), 6513. https://doi.org/10.3390/ijerph17186513

Dewi, A. C., Hermawan, A., & Avianto, D. (2024). Classification of Customers' Repeat Order Probability Using Decision Tree, Naïve Bayes and Random Forest. *Pilar Nusa Mandiri: Journal of Computing and Information System*, 20(1), 43–50. https://doi.org/10.33480/pilar.v20i1.5243

Gao, X., Wen, J., & Zhang, C. (2019). An Improved Random Forest Algorithm for Predicting Employee Turnover. *Mathematical Problems in Engineering*, 2019, 1–12. https://doi.org/10.1155/2019/4140707

Hakim, R. N. S. (2023). Optimasi Algoritma Random Forest dengan Teknik Boosting dalam Prediksi Churn Pelanggan di Industri Telekomunikasi. *Jurnal Ilmiah Teknologi Informasi*, 21(1), 45–56.

Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., & Pentland, A. S. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, 7(1), 1–15. https://doi.org/10.1140/epjds/s13688-018-0165-5

Mohammadpour, S., Khedmati, M., & Zada, M. (2023). Classification of truck- involved crash severity: Dealing with missing, imbalanced, and high- dimensional safety data. *PLOS ONE*, 18(3), e0281901. https://doi.org/10.1371/journal.pone.0281901

Netayawijit, P. (2025). Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical

Decision Support. *Healthcare*, 13(20), 2588. https://doi.org/10.3390/healthcare13202588

Nugroho, A. (2025). The Influence of the Timeliness of Goods Delivery, The Speed of Goods Delivery Time, The Transparency of Goods Delivery Information on Customer Satisfaction and Company Performance Case Study at Posind Kendari Main Branch Office. *Apcore Online Journal*, 1(1), 160–165. https://doi.org/10.65232/xh5yav58

Rodan, A., Fayyoumi, A., Faris, H., Alsakran, J., & Al-Kadi, O. (2015). Negative correlation learning for customer churn prediction: A comparison study. *The Scientific World Journal*, 2015, 1–11. https://doi.org/10.1155/2015/347683

Starcke, J., Spadafora, J., Spadafora, P., & Toma, M. (2025). The Effect of Data Leakage and Feature Selection on Machine Learning Performance for Early Parkinson's Disease Detection. *Bioengineering*, 12(8), 845. https://doi.org/10.3390/bioengineering12080845

Suryanto, S., & Martias, M. (2021). Komparasi metode K-NN, support vector machine dan random forest pada e-commerce Shopee. *INSANtek*, 2(1), 15– 21.

Tan, T., Ngoc, K., Thanh, H., Thu, H., & Hoang, U. (2024). Enhancing Repurchase Intention on Digital Platforms Based on Shopping Well-Being Through Shopping Value, Trust and Impulsive Buying. *SAGE Open*, 14(3), 1–12. https://doi.org/10.1177/21582440241278454