Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Perbandingan XGBoost dan Random Forest Untuk Prediksi Pembatalan Pesanan Marketplace Berbasis SHAP

Comparison of XGBoost and Random Forest for Marketplace Order Cancellation Prediction Based on SHAP

Dian Pertiwi*, Usman Sudibyo

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202214602@mhs.dinus.ac.id, [2]usman.sudibyo@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202214602@mhs.dinus.ac.id
Received: **09/07/2026**, *Revised*: **21/07/2026**, *Accepted*: **01/09/2026**, *Published*: **08/09/2026**

Abstrak−Tingginya tingkat pembatalan pesanan pada marketplace menjadi salah satu tantangan yang dapat memengaruhi efisiensi operasional, pengelolaan inventaris, serta kualitas layanan kepada pelanggan. Kemampuan untuk mengidentifikasi pesanan yang berpotensi dibatalkan sejak dini menjadi penting agar pelaku bisnis dapat mengambil langkah mitigasi yang lebih tepat. Tujuan Penelitian ini adalah membandingkan kinerja algoritma Random Forest dan XGBoost dalam proses prediksi pembatalan pesanan marketplace serta menginterpretasikan faktor-faktor yang memengaruhi prediksi menggunakan Explainable Machine Learning berbasis SHAP (SHapley Additive exPlanations). Penelitian menggunakan dataset transaksi marketplace yang telah melalui tahapan pembersihan data, transformasi fitur, dan penanganan ketidakseimbangan kelas melalui penerapan metode Synthetic Minority Over-sampling Technique (SMOTE). Selanjutnya, kinerja model dievaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score. Empat skenario model diuji, yaitu Random Forest + SMOTE, Random Forest Tuned + SMOTE, XGBoost + SMOTE, dan XGBoost Tuned + SMOTE. Berdasarkan Hasil penelitian mengindikasikan bahwa model XGBoost Tuned + SMOTE menghasilkan kinerja terbaik dengan nilai accuracy mencapai 86,62%, precision mencapai 51,43%, recall mencapai 34,95%, dan F1-Score mencapai 41,62%. Sementara itu, model XGBoost + SMOTE menghasilkan recall tertinggi sebesar 37,09% dengan F1-Score sebesar 41,48%. Mengingat tujuan penelitian adalah mendeteksi pesanan yang berpotensi dibatalkan pada data yang tidak seimbang, model XGBoost + SMOTE dipilih sebagai model terbaik karena menghasilkan recall tertinggi sebesar 37,09% dengan F1-Score yang tetap kompetitif sebesar 41,48%, sehingga lebih mampu mendeteksi kelas minoritas dibandingkan model lainnya. Analisis SHAP memperlihatkan bahwa beberapa fitur transaksi berkontribusi lebih dominan terhadap keputusan hasil prediksi yang dihasilkan oleh model. Temuan ini dapat dimanfaatkan sebagai dasar pengambilan keputusan untuk mengurangi risiko pembatalan pesanan dan meningkatkan efektivitas operasional marketplace.

Kata Kunci: Explainable Machine Learning; Marketplace; Pembatalan Pesanan; SHAP; SMOTE; XGBoost

Abstract−The high rate of order cancellations in marketplaces is a challenge that can impact operational efficiency, inventory management, and customer service quality. The ability to identify orders potentially canceled early is crucial so businesses can take more appropriate mitigation measures. The purpose of this study is to compare the performance of the Random Forest, and XGBoost algorithms in the marketplace order cancellations prediction process and to interpret the factors that influence predictions using SHAP (SHapley Additive exPlanations)-based explainable machine learning. The study used a marketplace transaction dataset that had undergone data cleaning, feature transformations, and class through the application of the Synthetic Minority Oversampling Technique (SMOTE) method. Next, the model performance is evaluated using the Accuracy, Precision, Recall, and F1-Score metrics. Four model scenarios were tested: Random Forest + SMOTE, Random Forest Tuned + SMOTE, XGBoost + SMOTE, and XGBoost Tuned + SMOTE. Based on the research results, it is indicated that the XGBoost Tuned + SMOTE model produces the best performance with accuracy values reaching 86.62%, precision reaching 51.43%, recall reaching 34.95%, and F1-Score reaching 41.62%. Meanwhile, the XGBoost + SMOTE model produced the highest recall of 37.09% with an F1-Score of 41.48%. Considering that the research objective is to detect potentially canceled orders in imbalance data, the XGBoost + SMOTE model was chosen as the best model because it produced the highest recall of 37,09% with a competitive F1-Score of 41,48%, making it more capable of detecting minority classes compared to other models. SHAP analysis showed that several transaction features contribute more dominantly to the prediction results produced by the model. These findings can be used as a basis for decision-making to reduce the risk of order cancellations and improve the effectiveness of marketplace operations.

Keywords: Explainable Machine Learning; Marketplace; Order Cancellation; SHAP; SMOTE; XGBoost

1. PENDAHULUAN

Perkembangan *electronic commerce (e-commerce)* telah mengubah cara masyarakat melakukan transaksi. Kemudahan akses *marketplace*, beragam metode pembayaran, serta dukungan layanan logistik yang semakin baik mendorong peningkatan aktivitas jual beli secara daring. Meskipun demikian, peningkatan jumlah transaksi tidak selalu diikuti dengan keberhasilan penyelesaian pesanan. Salah satu permasalahan yang masih sering ditemukan pada *platform marketplace* adalah pembatalan pesanan sebelum transaksi selesai. Kondisi ini dapat menimbulkan kerugian bagi penjual maupun penyedia *platform* karena berdampak pada pengelolaan persediaan, efisiensi operasional, dan kualitas layanan kepada pelanggan[1],[2].

Pembatalan pesanan berkaitan erat dengan perilaku konsumen dalam berbelanja secara daring. Wang et al [1] menjelaskan bahwa banyak pelanggan menambahkan produk ke dalam keranjang belanja, tetapi tidak melanjutkan proses pembelian hingga tahap pembayaran. Fenomena tersebut dikenal sebagai *shopping cart abandonment* dan menjadi salah satu penyebab tingginya transaksi yang tidak terselesaikan. Selain itu, keputusan pelanggan dalam

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

menyelesaikan transaksi juga dipengaruhi oleh faktor kepercayaan, nilai belanja yang dirasakan, serta perilaku pembelian impulsif [2]. Oleh karena itu, diperlukan pendekatan yang mampu mengidentifikasi pola transaksi yang berpotensi mengalami pembatalan sehingga dapat dilakukan tindakan pencegahan lebih awal.

Perkembangan *machine learning* memberikan peluang untuk memprediksi perilaku transaksi berdasarkan data historis. Melalui proses pembelajaran dari data sebelumnya, model dapat digunakan untuk memperkirakan kemungkinan suatu pesanan akan diselesaikan atau dibatalkan. Penelitian Nuryana et al [3] menunjukkan bahwa pendekatan *machine learning* mampu digunakan untuk memprediksi penyelesain transaksi *e-commerce* dengan performa yang baik. Informasi hasil prediksi tersebut dapat dimanfaatkan sejumlah *decision tree* yang dikombinasikan menggunakan pendekatan *ensemble learning* sehingga mampu menghasilkan model yang konsisten serta memiliki kemampuan generalisasi yang optimal [4]. Selain Random Forest, algoritma XGBoost juga banyak digunakan karena memiliki mekanisme *gradient boosting* yang mampu meningkatkan performa prediksi pada berbagai kasus klasifikasi[5]. Kedua algoritma tersebut dikenal mampu memberikan kinerja yang baik dalam mengelola data berskala besar serta digunakan dalam berbagai penelitian berbasis data transaksi.

Dalam penerapannya, data transaksi *marketplace* umumnya memiliki karakteristik tidak seimbang *(class imbalance)*, yaitu jumlah transaksi berhasil jauh lebih banyak daripada transaksi yang dibatalkan. Kondisi tersebut dapat mengakibatkan model lebih sering memprediksi kelas mayoritas sehingga kemampuan mendeteksi transaksi yang berpotensi dibatalkan menjadi rendah. Untuk mengatasi masalah tersebut, metode *Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk membentuk sampel sintetis pada kelas minoritas sehingga distribusi antar kelas menjadi lebih seimbang [6]. Penelitian terbaru menunjukkan bahwa kombinasi SMOTE dengan algoritma klasifikasi dapat meningkatkan kinerja model dalam mengenali kelas minoritas secara efektif [7].

Selain ketidakseimbangan data, pembangunan model *machine learning* juga perlu memperhatikan potensi *data leakage. Data leakage* terjadi ketika informasi yang sebenarnya tidak tersedia pada saat prediksi ikut digunakan dalam proses pelatihan model sehingga menghasilkan performa yang tampak tinggi, namun tidak mencerminkan kondisi sebenarnya [8]. Pada data transaksi *marketplace*, beberapa atribut yang berkaitan langsung dengan hasil akhir transaksi berpotensi menyebabkan *data leakage*. Oleh karena itu, identifikasi dan penghapusan atribut tersebut menjadi langkah penting agar model yang dihasilkan lebih valid dan dapat diterapkan pada kondisi nyata.

Beberapa penelitian terdahulu telah membahas pemanfaatan *machine learning* dalam bidang *e-commerce.* Koten et al [9] menerapkan Random Forest dan SMOTE untuk memprediksi pembatalan pesanan pada *e-commerce* indonesia. Meskipun demikian, penelitian tersebut hanya menggunakan satu algoritma sehingga belum menunjukkan perbandingan performa dengan pendekatan lain. Nuryana et al [3] menggunakan *machine learning* untuk memprediksi penyelesaian transaksi *e-commerce*, tetapi belum membahas pengaruh *class imbalance* maupun interpretasi model. Dewi et al [10] memanfaatkan Random Forest untuk mengklasifikasikan probabilitas pelanggan melakukan pemesanan ulang, namun fokus penelitian berada pada perilaku pelanggan, bukan pembatalan pesanan. Mohammadpour et al [11] menunjukkan pentingnya penanganan data tidak seimbang dalam meningkatkan kualitas model klasifikasi, tetapi penelitian tersebut tidak diterapkan pada data *marketplace.* Sementara itu, Netayawijit [12] mengintegrasikan SMOTE dan SHAP untuk menghasilkan model yang lebih mudah diinterpretasikan, namun objek penelitian yang digunakan berada pada bidang kesehatan.

Berdasarkan kajian penelitian terdahulu, masih ditemukan beberapa celah penelitian yang memiliki peluang untuk dikembangkan lebih lanjut. Pertama, studi yang secara khusus membandingkan performa Random Forest dan XGBoost pada kasus prediksi pembatalan pesanan *marketplace* masih terbatas. Kedua, belum banyak penelitian yang mengkombinasikan penanganan *class imbalance*, pengendalian *data leakage*, dan analisis interpretabilitas model dalam satu kerangka penelitian. Ketiga, sebagian penelitian masih berfokus pada tingkat akurasi tanpa mempertimbangkan metrik evaluasi lainnya seperti *precision, recall, dan F1-score* secara relevan untuk data tidak seimbang.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengavaluasi perbandingan kinerja algoritma Random Forest dan XGBoost untuk melakukan prediksi pembatalan pesanan *marketplace* menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data. Penelitian ini juga menerapkan pengendalian *data leakage* pada tahap prapemrosesan data serta menggunakan metode *SHapley Additive exPlanations* (SHAP) sebagai pendekatan untuk menganalisis kontribusi masing-masing fitur terhadap keputusan prediksi model [12]. Hasil penelitian diharapkan dapat memberikan informasi mengenai model yang memiliki performa terbaik berdasarkan proses evaluasi dilakukan berdasarkan *Accuracy, Precision, Recall, F1-Score,* dan ROC-AUC sehingga dapat membantu menghasilkan keputusan yang lebih efektif terhadap *platform marketplace*.

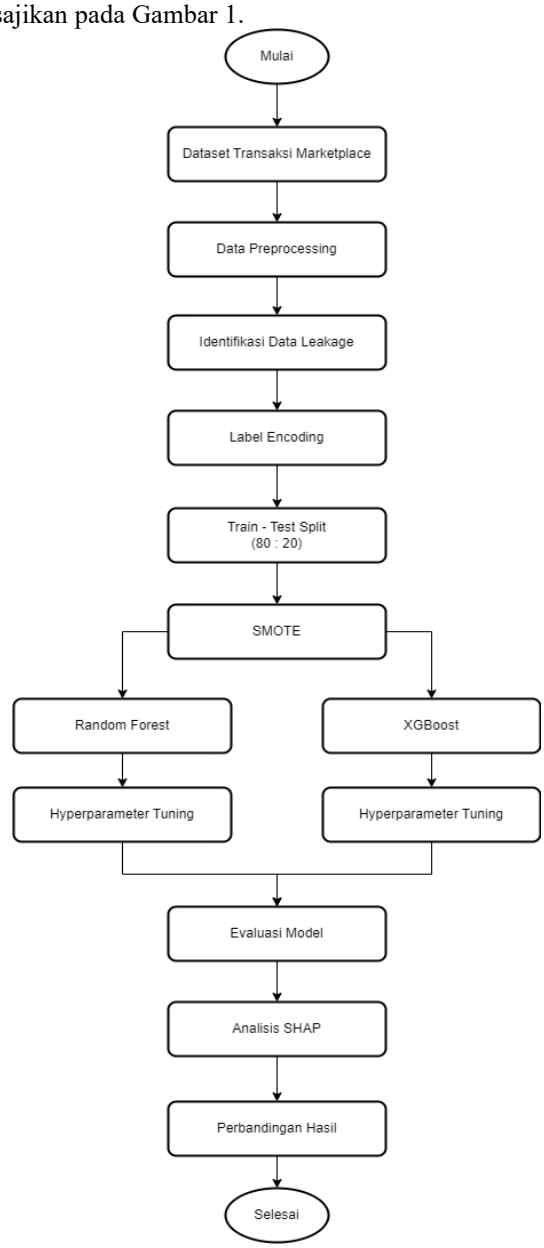
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini mengimplementasikan metode machine learning untuk memprediksi potensi pembatalan pesanan pada platform marketplace dengan memanfaatkan data historis transaksi. Pendekatan terstruktur diterapkan sepanjang alur penelitian untuk memastikan keandalan model yang dihasilkan. Proses diawali dengan pengumpulan data transaksi, kemudian dilanjutkan ke tahap prapemrosesan data termasuk pembersihan data (data cleaning), penanganan nilai yang hilang (missing values), serta pemrosesan fitur (feature engineering). Selanjutnya, data yang telah siap digunakan

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

untuk melatih dan membangun model machine learning. Pada tahap akhir, dilakukan evaluasi performa serta interpretasi terhadap hasil prediksi guna memahami faktor-faktor utama yang memengaruhi pembatalan pesanan. Alur penelitian secara menyeluruh disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, proses penelitian dimulai dengan tahap pengumpulan dataset transaksi *marketplace* yang berisi informasi pelanggan, metode pembayaran, layanan pengiriman, lokasi pembeli, dan atribut transaksi lainnya. Dataset kemudian melalui tahap *preprocessing* untuk membersihkan data, membentuk variabel target, serta menyiapkan fitur yang akan digunakan dalam proses pelatihan model. Selanjutnya dilakukan identifikasi atribut yang berpotensi menyebabkan *data leakage*. Setelah atribut tersebut dihapus, fitur yang bersifat kategorikal dikonversi ke dalam bentuk numerik menggunakan teknik *Label Encoding*. Selanjutnya data dibagi menjadi data latih dan data uji dengan perbandingan 80:20 dengan metode *stratified split*. Untuk mengatasi ketidakseimbangan kelas pada data latih digunakan metode SMOTE. Data hasil oversampling menggunakan SMOTE kemudian dimanfaatkan dalam proses pembangun model Random Forest dan XGBoost. Kedua model kemudian dioptimasi menggunakan *hyperparameter tuning* dan dievaluasi menggunakan beberapa metrik klasifikasi. Pada tahap akhir dilakukan interpretasi terhadap model menggunakan metode SHAP untuk mengidentifikasi fitur-fitur yang memberikan pengaruh terhadap hasil prediksi pembatalan pesanan.

2.2 Dataset dan Prapemrosesan Data

Dataset yang dilakukan merupakan data transaksi *marketplace* yang diperoleh dari Kaggle dan berisi informasi pesanan, pelanggan, pembayaran, serta pengiriman barang. Variabel target yang digunakan adalah status pembatalan pesanan yang dibentuk menjadi dua kelas, yaitu pesanan tidak dibatalkan (0) dan pesanan dibatalkan (1).

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Prapemrosesan data dilakukan untuk meningkatkan kualitas data sebelum memasuki tahap pelatihan model. Proses tersebut meliputi pemeriksaan *missing value*, penghapusan atribut yang tidak relevan, ekstraksi fitur waktu dari atribut tanggal transaksi, serta transformasi data kategorikal menggunakan *Label Encoding*. *Label encoding* dipilih karena mampu mengubah setiap kategori kedalam bentuk numerik berbentuk integer, sehingga data tersebut dapat digunakan sebagai masukan bagi algoritma klasifikasi yang hanya menerima data numerik [13].

Selain itu dilakukan identifikasi terhadap atribut yang berpotensi menyebabkan *data leakage*. Beberapa atribut seperti total pembayaran, estimasi potongan biaya pengiriman, dan ongkos kirim yang dibayar pembeli dihapus karena mengandung informasi langsung maupun tidak langsung terhadap variabel target, sehingga dapat menghasilkan performa model yang bias dan tidak mampu melakukan generalisasi dengan baik [14].

2.3 Pembagian Data

Setelah proses prapemrosesan selesai, dataset selanjutnya dibagi menjadi data latih *(training set)* dan data uji *(testing set)* dengan perbandingan 80:20. Pembagian dilakukan menggunakan metode *stratified split* untuk menjaga proporsi masing-masing kelas pada kedua kelompok data tersebut sehingga distribusi kelas tetap konsisten [15]. Pembagian data latih dan data uji bertujuan agar kemampuan generalisasi model dapat dievaluasi menggunakan data yang tidak terlibat selama proses pelatihan sehingga hasil evaluasi menjadi lebih objektif [16].
Pembagian dataset dapat dinyatakan sebagai:

$$D = D_{train} + D_{test} \tag{1}$$

dimana D menyatakan dataset keseluruhan, D_{train} menyatakan data latih *(training set),* sedangkan D_{test} menyatakan data uji *(testing set).*

2.4 Penanganan Ketidakseimbangan Data Menggunakan SMOTE

Distribusi masing-masing kelas pada dataset menunjukkan jumlah transaksi tidak dibatalkan lebih banyak dibandingkan transaksi yang dibatalkan. Kondisi tersebut dapat mengakibatkan model lebih berpeluang memprediksi kelas mayoritas sehingga kemampuan mendeteksi transaksi yang dibatalkan menjadi rendah [11].

Sebagai upaya untuk mengatasi permasalahan tersebut digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini membentuk data sintetis pada kelas minoritas dengan mempertimbangkan kedekatan antar sampel melalui pendekatan *nearest neighbor* [6],[7].

Penerapan SMOTE dilakukan setelah proses pembagian data sehingga data sintetis hanya dibentuk pada data latih. Pendekatan ini direkomendasikan untuk menghindari bias evaluasi akibat *oversampling* pada data uji [17].
Persamaan pembentukan data sintetis SMOTE ditunjukkan pada Persamaan (2).

$$x_{baru} = x_i + \lambda(x_{nn} - x_i) \tag{2}$$

dimana X_1 menyatakan sampel pada kelas minoritas, X_{nn} menyatakan sampel tetangga terdekat *(nearest neighbor),* sedangkan λ bilangan acak antara 0 dan 1.

2.5 Pembangunan Model Random Forest

Random Forest adalah metode *ensemble learning* yang menghasilkan sejumlah *decision tree* dan mengintegrasikan prediksi dari setiap pohon untuk menentukan kelas akhir [4],[10].
Prediksi Random Forest diperoleh melalui mekanisme *majority voting* yang dirumuskan sebagai:

$$\hat{y} = mode(T_1, T_2, \dots, T_n) \tag{3}$$

dimana $T_1, T_2, \dots T_n$ merupakan hasil prediksi masing-masing *decision tree* yang selanjutnya ditentukan nilai mayoritas *(majority voting)* sebagai hasil prediksi akhir.

2.6 Pembangunan Model XGBoost

XGBoost merupakan algoritma yang menggunakan pendekatan *gradient boosting* dengan menggabungkan model secara bertahap dengan memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya [5].
Fungsi objektif XGBoost dapat dituliskan sebagai:

$$Obj^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{4}$$

dimana l adalah fungsi kerugian *(loss function),* Ω merupakan fungsi regularisasi untuk mengendalikan kompleksitas model, dan f_k menyatakan pohon keputusan ke-k pada setiap iterasi pembentukan model XGBoost.

2.7 Evaluasi Model

Kinerja model dievaluasi menggunakan metrik *Accuracy, Precision, Recall, F1-Score*, dan ROC-AUC untuk memberikan gambaran yang lebih komprehensif pada data yang tidak seimbang. ROC-AUC digunakan untuk mengukur kemampuan model dalam membedakan dua kelas pada berbagai nilai *threshold* [18], [19].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

dimana TP *(True Positive)* merupakan jumlah data kelas positif yang di prediksi dengan benar sebagai kelas positif, TN *(True Negative)* adalah jumlah data kelas negatif yang diprediksi dengan benar sebagai kelas negatif, FP *(False Positive)* menunjukkan jumlah data kelas negatif yang diprediksi sebagai kelas positif, sedangkan FN *(False Negative)* merupakan jumlah data kelas positif yang diprediksi sebagai kelas negatif.

2.8 Interpretasi Model Menggunakan SHAP

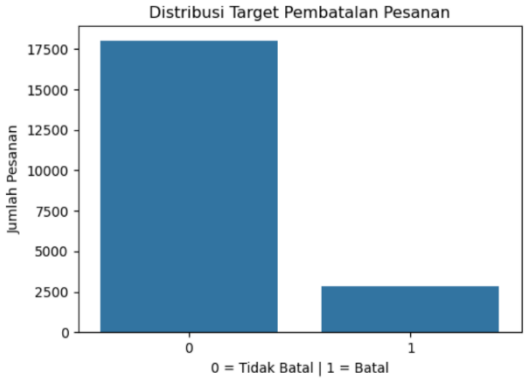
Sebagai upaya meningkatkan transparansi model, penelitian ini menggunakan metode *SHapley Additive exPlanations* (SHAP). Metode tersebut digunakan untuk menjelaskan tingkat kontribusi setiap fitur dalam memengaruhi hasil prediksi model sehingga faktor-faktor yang memengaruhi pembatalan pesanan dapat diidentifikasi dengan lebih jelas [20]. Hasil analisis SHAP disajikan dalam bentuk *summary plot* dan *feature importance* untuk menunjukkan fitur-fitur yang memberikan pengaruh paling besar terhadap proses pengambilan keputusan oleh model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Prapemrosesan Data

Tahap prapemrosesan diimplementasikan untuk menghasilkan dataset sebelum pembangunan model *machine learning*. Proses ini meliputi pemeriksaan *missing value*, pembentukan variabel target, ekstraksi fitur waktu, *Label Encoding*, serta penghapusan atribut yang berpotensi menyebabkan *data leakage*. Tahapan ini penting karena kualitas data mempengaruhi kemampuan model dalam melakukan prediksi.

Dataset yang digunakan berasal dari Kaggle *(all_months_clean.csv)*. Hasil pemeriksaan menunjukkan bahwa atribut Alasan Pembatalan memiliki *missing value* tinggi sehingga dihapus agar tidak menimbulkan bias. Selanjutnya, variabel target dibentuk dari atribut Status Pesanan, yaitu label 1 untuk pesanan batal dan 0 untuk selain batal sehingga permasalahan diformulasikan sebagai klasifikasi biner. Distribusi kelas target ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Kelas Target Pembatalan Pesanan

Berdasarkan Gambar 2, jumlah transaksi tidak batal (kelas 0) jauh lebih banyak daripada transaksi batal (kelas 1). Kondisi ini menunjukkan adanya ketidakseimbangan kelas yang dapat mengakibatkan model lebih cenderung memprediksi kelas mayoritas, sehingga kemampuan mendeteksi transaksi batal menurun. Oleh karena itu, penanganan ketidakseimbangan kelas dilakukan setelah *train-test split* dan sebelum pelatihan model.

Selanjutnya, atribut Waktu Pesanan Dibuat diekstraksi menjadi Tahun, Bulan, Hari, dan Jam, sedangkan atribut kategorikal ditransformasikan menggunakan *Label Encoding* sehingga dapat diolah oleh algoritma Random Forest dan XGBoost. Atribut yang berpotensi menyebabkan *data leakage*, yaitu Total Pembayaran, Estimasi Potongan Biaya Pengiriman , dan Ongkos Kirim Dibayar oleh Pembeli, dihapus karena hanya tersedia setelah transaksi berlangsung. Langkah ini bertujuan agar model hanya menggunakan informasi yang tersedia pada saat prediksi sehingga memiliki kemampuan generalisasi yang lebih baik. Hasil prapemrosesan menghasilkan dataset yang siap digunakan untuk pembagian data dan pembangunan model klasifikasi.

3.2 Hasil Pembagian Data

Setelah proses prapemrosesan, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan *Stratified Split* Untuk mempertahankan proporsi setiap kelas sehingga distribusinya tetap mendekati dataset asli.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Pendekatan ini penting karena dataset memiliki karakteristik *class imbalance*, sehingga proses pelatihan dan evaluasi dapat memberikan gambaran performa model yang lebih representatif.

Data latih digunakan dalam proses pembangunan model Random Forest dan XGBoost, sementara itu data uji digunakan untuk menilai kinerja model terhadap data yang belum pernah digunakan selama pelatihan yang pada akhirnya kemampuan generalisasi model dapat dinilai secara lebih objektif. Pembagian data dilakukan sebelum penerapan SMOTE agar data sintetis hanya dibentuk pada data latih, sedangkan evaluasi tetap menggunakan data uji asli. Pendekatan ini direkomendasikan untuk menghindari bias akibat *oversampling* dan menghasilkan evaluasi yang lebih valid pada data tidak seimbang.

3.3 Hasil Penanganan Ketidakseimbang Data Menggunakan SMOTE

Distribusi kelas pada data latih menunjukkan kondisi *class imbalance.* dengan 13.036 transaksi tidak batal dan 2.058 transaksi batal sebelum penerapan SMOTE. Situasi ini beresiko menyebabkan model memiliki kecenderungan mempelajari pola pada kelas mayoritas. Distribusi kelas sebelum dan sesudah SMOTE ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Kelas Sebelum dan Sesudah SMOTE

Kondisi	Tidak Batal	Batal
Sebelum SMOTE	13.036	2.058
Sesudah SMOTE	13.036	13.036

Berdasarkan Tabel 1, setelah metode SMOTE diterapkan jumlah data pada kedua kelas menjadi seimbang, dengan masing-masing kelas terdiri atas 13.036 data. SMOTE membentuk data sintetis pada kelas minoritas dengan mempertimbangkan kedekatan antar sampel *(nearest neighbor),* sehingga proporsi data antar kelas menjadi lebih seimbang tanpa sekadar menduplikasi data.

3.4 Hasil Pembangunan Model Random Forest

3.4.1 Hasil Evaluasi Model Random Forest

Model Random Forest dikembangkan dengan memanfaatkan data latih hasil penerapan SMOTE, kemudian dievaluasi sebelum dan sesudah *hyperparameter tuning* menggunakan metrik *Accuracy, Precision, Recall, F1-Score.* Penerapan sejumlah metrik diperlukan karena *accuracy* saja belum mampu menggambarkan performa model pada dataset tidak seimbang. Hasil evaluasi model Random Forest ditunjukkan pada Tabel 2.

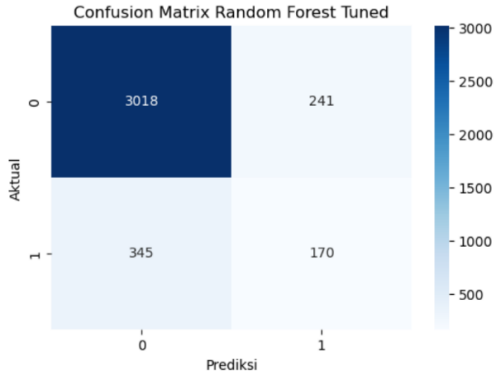
Tabel 2. Hasil Evaluasi Model Random Forest

Model	Accuracy	Precision	Recall	F1-Score
Random Forest + SMOTE	0,8455	0,4154	0,3243	0,3642
Random Forest Tuned + Smote	0,8447	0,4136	0,3301	0,3672

Berdasarkan Tabel 2, Random Forest + SMOTE memperoleh *accuracy* 84,55%, sedangkan Random Forest *Tuned* + SMOTE mencapai 84,47%. Meskipun *accuracy* relatif tidak berubah, *tuning* meningkatkan nilai *Recall* serta *F1-Score*, sehingga menunjukkan kemampuan model dalam mengenali transaksi batal menjadi lebih baik. Hal ini menunjukkan bahwa *tuning* lebih berpengaruh terhadap peningkatan deteksi kelas minoritas dibandingkan *accuracy* secara keseluruhan.

3.4.2 Analisis Confusion Metrix Random Forest

Untuk memperoleh gambaran kinerja model secara lebih detail, evaluasi dilakukan melalui analisis *confusion matrix*. Berbeda dengan *accuracy* yang hanya menunjukkan persentase prediksi benar, *confusion matrix* menyajikan jumlah prediksi benar dan salah pada setiap kelas sehingga lebih sesuai digunakan pada dataset tidak seimbang. Hasil confusion matrix Random Forest Tuned ditunjukkan pada Gambar 3.



Gambar 3. Confusion Matrix Random Forest Tuned

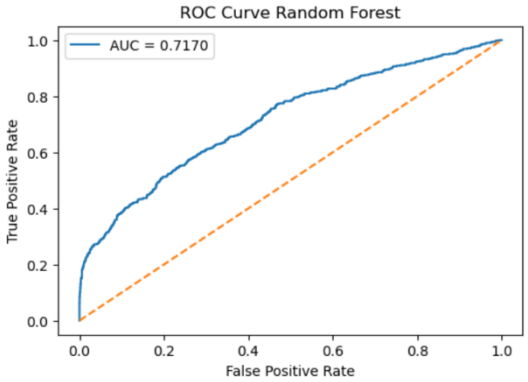
Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Berdasarkan Gambar 3, model menghasilkan 3.018 *True Negative*, 170 *True Positive,* 241 *False Positive,* dan 345 *False Negative*. Hasil tersebut menunjukkan bahwa model mampu mengenali transaksi tidak batal dengan baik, tetapi masih terdapat cukup banyak transaksi batal yang diklasifikasikan sebagai tidak batal.

Temuan ini selaras dengan nilai *Recall* mencapai 33,01%, yang mengindikasikan bahwa model baru mampu mendeteksi sekitar sepertiga transaksi batal. Secara keseluruhan, Random Forest telah mampu mengklasifikasikan kedua kelas, tetapi kemampuan mendeteksi kelas minoritas masih perlu ditingkatkan. Oleh karena itu, penelitian selanjutnya mengevaluasi algoritma XGBoost yang diharapkan memberikan performa lebih baik dalam mengenali kelas minoritas.

3.4.3 Analisis ROC Curve Random Forest

Selain melalui *confusion matrix*, performa model juga dievaluasi menggunakan kurva *Receiver Operating Characteristic* (ROC) yang menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai tingkat nilai *threshold*. Semakin mendekati kurva ke sudut kiri atas , semakin baik kemampuan model dalam membedakan transaksi batal dan tidak batal. Hasil ROC Curve Random Forest ditunjukkan pada gambar 4.



Gambar 4. ROC Curve Random Forest Tuned

Berdasarkan Gambar 4, kurva ROC berada di atas garis diagonal yang merepresentasikan prediksi acak, sehingga menunjukkan bahwa model Random Forest memiliki kemampuan diskriminasi yang lebih baik dibandingkan random classifier. Model Random Forest *Tuned* memperoleh nilai ROC-AUC sebesar 0,717, yang mengidentifikasi kemampuan cukup baik dalam membedakan transaksi batal dan tidak batal. Semakin tinggi nilai ROC-AUC, semakin baik kemampuan model dalam memisahkan kedua kelas. Metrik ini memberikan evaluasi yang lebih komprehensif karena mempertimbangkan berbagai nilai *threshold*, sehingga sesuai digunakan pada dataset dengan distribusi kelas yang tidak seimbang.

3.4.4 Analisis Feature Importance Random Forest

Salah satu kelebihan algoritma Random Forest adalah kemampuannya untuk mengukur tingkat kepentingan *(feature importance)* masing-masing atribut yang digunakan selama proses pembentukan pohon keputusan. Nilai *feature importance* menunjukkan besarnya kontribusi relatif setiap fitur terhadap proses klasifikasi. Hasil analisis *feature importance* ditunjukkan pada Tabel 3.

Tabel 3. Fitur Importance Random Forest

Feature	Importance
Opsi Pengiriman	0.221077
Metode Pembayaran	0.109196
product_categories	0.090460
Perkiraan Ongkos Kirim	0.082697
Kota/Kabupaten	0.077982
Bulan	0.073988
total_weight_gr	0.070739
Hari	0.066061
Provinsi	0.063947
Jam	0.057235

Berdasarkan Tabel 3, Opsi Pengiriman merupakan fitur dengan tingkat kepentingan tertinggi, diikuti oleh Metode Pembayaran, *product_categories*, Perkiraan Ongkos Kirim, dan Kota/Kabupaten. Hasil tersebut menunjukkan bahwa karakteristik pengiriman, metode pembayaran, kategori produk, serta lokasi pelanggan memberikan pengaruh yang lebih besar terhadap prediksi pembatalan pesanan dibandingkan atribut lainnya.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Selain itu, fitur temporal seperti Bulan, Hari dan Jam juga memiliki kontribusi dalam proses klasifikasi. Temuan ini mengindikasikan bahwa waktu terjadinya transaksi turut memengaruhi pola pembatalan pesanan, meskipun pengaruhnya lebih rendah dibandingkan atribut transaksi dan pengiriman.

Berdasarkan keseluruhan, hasil *feature importance* menandakan bahwa model Random Forest tidak hanya dapat melakukan proses klasifikasi, tetapi juga mengidentifikasi atribut yang paling berpengaruh terhadap prediksi pembatalan pesanan. Informasi tersebut dapat menjadi dasar bagi pengelola *marketplace* untuk memahami faktor-faktor yang berkontribusi terhadap risiko pembatalan transaksi.

3.5 Hasil Pembangunan Model XGBoost

3.5.1 Hasil Evaluasi Model XGBoost

Selain Random Forest, penelitian ini juga membangun model menggunakan algoritma XGBoost yang menerapkan pendekatan *gradient boosting* untuk memperbaiki kesalahan prediksi secara bertahap sehingga menghasilkan kinerja klasifikasi yang baik. Model dibangun menggunakan data latih hasil penerapan SMOTE, selanjutnya melakukan *hyperparameter tuning* guna mendapatkan kombinasi parameter terbaik. Setelah itu, dievaluasi menggunakan metrik *Accuracy, Precision, Recall, F1-Score*, dan ROC-AUC. Hasil evaluasi Model XGBoost ditunjukkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model XGBoost

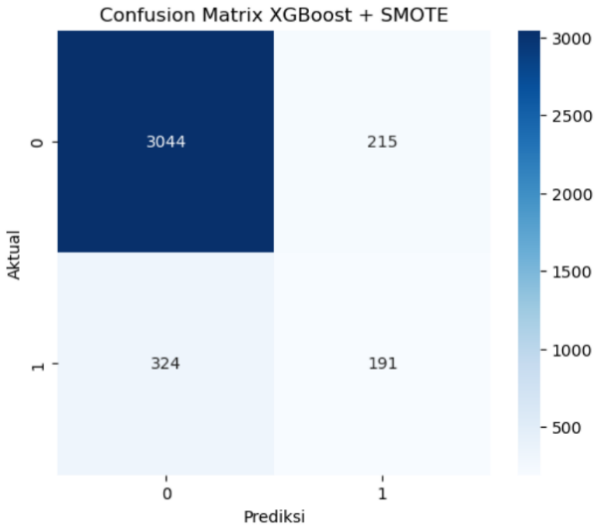
Model	Accuracy	Precision	Recall	F1-Score
XGBoost + SMOTE	0,8572	0,4704	0,3709	0,4148
XGBoost Tuned + SMOTE	0,8662	0,5143	0,3495	0,4162

Berdasarkan Tabel 4, proses *hyperparameter tuning* meningkatkan *Accuracy* dari 85,72% menjadi 86,62% *Precision* dari 47,04% menjadi 51,43%, serta *F1-Score* dari 41,48% menjadi 41,62%. Sebaliknya, *Recall* menurun dari 37,09% menjadi 34,95%, yang menunjukkan berkurangnya kemampuan model dalam mendeteksi transaksi batal.

Meskipun XGBoost *Tuned* + SMOTE menghasilkan *Accuracy* dan *F1-Score* yang sedikit lebih tinggi, penelitian ini menetapkan XGBoost +SMOTE sebagai model terbaik karena menghasilkan *Recall* sebesar 37,09%, sesuai dengan tujuan penelitian untuk memaksimalkan deteksi transaksi yang berpotensi dibatalkan pada data tidak seimbang. Berdasarkan keseluruhan hasil evaluasi, XGBoost menunjukkan kinerja yang lebih baik dibandingkan Random Forest pada sebagian besar metrik evaluasi.

3.5.2 Analisis Confusion Matrix XGBoost

Untuk menganalisis kinerja model secara lebih terperinci, digunakan *confusion matrix* yang menampilan jumlah hasil prediksi yang benar maupun salah pada masing-masing kelas sehingga kemampuan model dalam membedakan kategori transaksi dapat dianalisis secara lebih jelas. Hasil confusion matrix XGBoost + SMOTE ditunjukkan pada Gambar 5.



Gambar 5. Confusion Matrix XGBoost + SMOTE

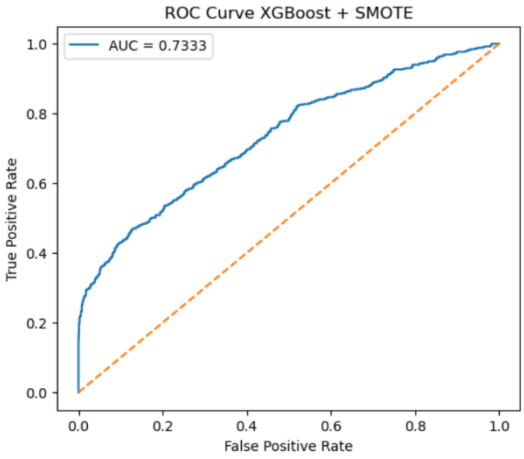
Berdasarkan Gambar 5, model menghasilkan 3.044 *True Negative* (TN) dan 191 *True Positive* (TP), 215 *False Positive* (FP), dan 324 *False negative* (FN). Dibandingkan Random Forest, jumlah *True Positive* meningkat dari 170 menjadi 191, sedangkan *False Positive* menurun dari 241 menjadi 215, sejalan dengan nilai *precision* dan *recall* XGBoost yang lebih tinggi dibandingkan Random Forest. Meskipun masih terdapat 324 *False Negative*, hasil ini menunjukkan bahwa XGBoost memiliki kemampuan klasifikasi yang lebih baik, khususnya dalam mengenali transaksi yang berpotensi dibatalkan

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

3.5.3 Analisis ROC Curve XGBoost

Evaluasi model juga dilakukan menggunakan ROC Curve untuk mengukur kemampuan XGBoost dalam mengidentifikasi perbedaan antara transaksi batal dan tidak batal untuk berbagai nilai ambang batas klasifikasi *(threshold)*. Hasil ROC Curve XGBoost ditunjukkan pada Gambar 6.

Berdasarkan Gambar 6, posisi kurva ROC yang berada di atas garis diagonal yang menggambarkan bahwa kemampuan klasifikasi lebih baik daripada prediksi acak. Model XGBoost + SMOTE menghasilkan nilai ROC-AUC sebesar 0,7333, lebih unggul dibandingkan Random Forest dengan nilai ROC-AUC sebesar 0,7170. Perolehan nilai tersebut mengindikasikan bahwa model memiliki kemampuan diskriminasi yang lebih baik. *ROC Curve* dan nilai AUC digunakan sebagai metrik evaluasi untuk mengukur kemampuan model dalam membedakan dua kelas.



Gambar 6. ROC Curve XGBoost + SMOTE

3.5.4 Analisis Feature Importance XGBoost

Selain menghasilkan prediksi, XGBoost juga menghitung *feature importance* untuk mengetahui besarnya pengaruh setiap fitur terhadap pembentukan model. Hasil analisis *feature importance* XGBoost ditunjukkan pada Tabel 5.

Tabel 5. Fitur Importance XGBoost

Feature	Importance
Opsi Pengiriman	0.203322
Metode Pembayaran	0.182088
Tahun	0.109181
product_categories	0.067004
Bulan	0.064642
num_product_categories	0.061953
Provinsi	0.047597
total_qty	0.044210
total_returned_qty	0.042034
Perkiraan Ongkos Kirim	0.037218
total_weight_gr	0.032580
Kota/Kabupaten	0.031256
Hari	0.028098
Total Diskon	0.027747
Jam	0.021069

Berdasarkan Tabel 5, atribut Opsi Pengiriman merupakan atribut paling berpengaruh, diikuti Metode Pembayaran, Tahun, *product_categories*, dan Bulan. Hasil tersebut menunjukkan bahwa layanan pengiriman, metode pembayaran, waktu transaksi, dan jenis produk menjadi faktor utama dalam memprediksi pembatalan pesanan. Atribut lain seperti *num_product_categories*, Provinsi, *total qyt, total_returned_qyt,* dan Perkiraan Ongkos Kirim juga berkontribusi, meskipun nilainya lebih rendah. Secara keseluruhan, distribusi *feature importance* pada XGBoost lebih merata sehingga model tidak bergantung pada satu atribut dan mendukung performanya yang lebih baik dibandingkan Random Forest.

3.6 Perbandingan Kinerja Model

Setelah seluruh proses pelatihan dan evaluasi dilakukan, tahap berikutnya adalah membandingkan performa setiap model untuk menentukan model terbaik dalam memprediksi pembatalan pesanan pada *marketplace*. Hasil perbandingan kinerja model ditunjukkan pada Tabel 6.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Tabel 6. Perbandingan Kinerja Model

Model	Acccuracy	Precision	Recall	F1-Score
Random Forest + SMOTE	0,8455	0,4154	0,3243	0,3642
Random Forest Tuned + SMOTE	0,8447	0,4136	0,3301	0,3672
XGBoost + SMOTE	0,8572	0,4704	0,3709	0,4148
XGBoost Tuned + SMOTE	0,8662	0,5143	0,3495	0,4162

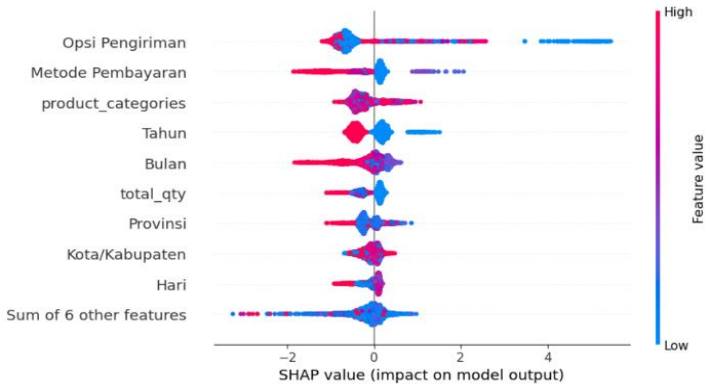
Berdasarkan Tabel 6, algoritma XGBoost secara umum memberikan kinerja yang lebih unggul daripada Random Forest pada sebagian besar metrik evaluasi. Model XGBoost + SMOTE memperoleh *accuracy* 85,72%, *precision* 47,04%, dan *recall* 37,09%, sedangkan XGBoost *Tuned* + SMOTE memperoleh *accuracy, precision, dan F1-Score* tertinggi, tetapii *recall* menurun menjadi 34,95%.

Selain itu, XGBoost menunjukkan nilai ROC-AUC lebih unggul dibandingkan Random Forest, nilai ROC-AUC tersebut mendukung hasil evaluasi metrik lainnya dan memberikan ukuran kemampuan diskriminasi model yang komprehensif pada berbagai *threshold* klasifikasi. Hasil yang diperoleh dalam penelitian ini menjelaskan bahwa kombinasi XGBoost dan teknik penanganan data tidak seimbang dapat meningkatkan kemampuan model dalam mengidentifikasi kelas minoritas tanpa mengurangi performa klasifikasi secara keseluruhan.

Berdasarkan tujuan penelitian ini, XGBoost + SMOTE dipilih sebagai model terbaik karena menghasilkan *recall* tertinggi 37,09%, sehingga lebih mampu mendeteksi transaksi yang berpotensi dibatalkan pada dataset tidak seimbang. Model ini selanjutnya digunakan pada tahap interpretasi menggunakan SHAP.

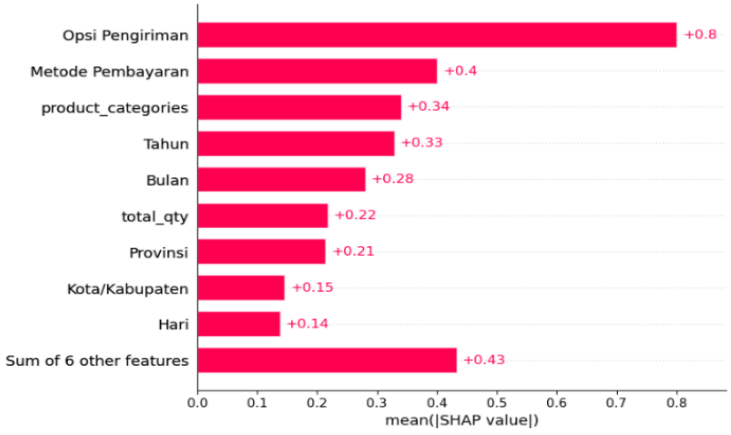
3.7 Interpretasi Model Menggunakan SHAP

Setelah mendapatkan model terbaik, proses interpretasi dilakukan menggunakan metode *SHapley Additive exPlanations* (SHAP) guna mengetahui kontribusi masing-masing fitur terhadap hasil prediksi. SHAP meningkatkan transparansi model dan mendukung pendekatan *Explainable Artificial Intelligence* (XAI) dalam meningkatkan interpretabilitas, akuntabilitas, serta kepercayaan terhadap hasil prediksi. SHAP Summary Plot ditunjukkan pada Gambar 7.



Gambar 7. SHAP Summary Plot

Berdasarkan gambar 7, Opsi pengiriman, Metode Pembayaran, *product_categories*, Tahun, dan Bulan merupakan fitur yang paling berpengaruh terhadap prediksi pembatalan pesanan. Warna dan sebaran titik menunjukkan bahwa besar maupun arah pengaruh setiap fitur dapat berbeda pada setiap transaksi.
SHAP Bar Plot ditunjukkan pada gambar 8.



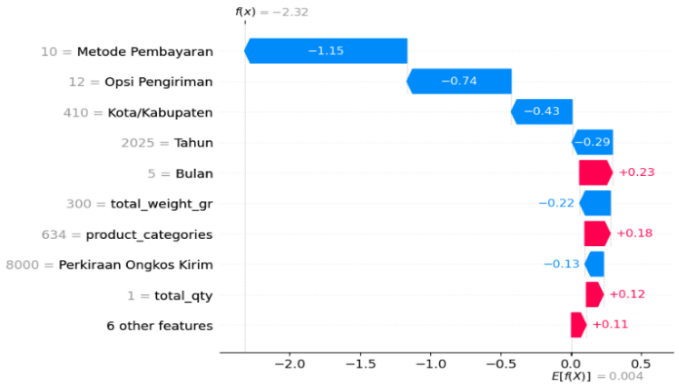
Gambar 8. SHAP Bar Plot

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

Berdasarkan Gambar 8, *SHAP Bar Plot* menunjukkan rata-rata kontribusi global setiap fitur *(mean | SHAP value|).* Hasilnya konsisten dengan *summary plot*, yaitu Opsi Pengiriman, Metode Pembayaran, *product_categories*, Tahun, dan Bulan sebagai fitur dengan kontribusi terbesar.
SHAP Waterfall Plot ditunjukkan pada Gambar 9.

Berdasarkan Gambar 9, *SHAP Waterfall Plot* menjelaskan kontribusi fitur pada satu transaksi. Metode Pembayaran, Opsi Pengiriman, Kota/Kabupaten, dan Tahun memberikan kontribusi positif sehingga menghasilkan nilai prediksi akhir sebesar -2,32 pada transaksi yang dianalisis.

Penggunaan SHAP pada penelitian ini memberikan interpretasi yang lebih komprehensif dibandingkan *feature importance* bawaan model karena mampu menjelaskan pengaruh fitur secara global maupun pada setiap prediksi. Dengan demikian, mekanisme pengambilan keputusan model menjadi lebih transparan dan dapat dipahami sehingga mendukung evaluasi serta pengambilan keputusan pada *platform marketplace.*



Gambar 9. SHAP Waterfall Plot

3.8 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma XGBoost memberikan performa lebih baik dibandingkan Random Forest. Model XGBoost *Tuned* memperoleh nilai *Accuracy, Precision, dan F1-Score* tertinggi, sedangkan XGBoost + SMOTE memperoleh *Recall* tertinggi sebesar 37,09%. Karena penelitian ini berfokus pada deteksi transaksi yang berpotensi dibatalkan pada data tidak seimbang, XGBoost + SMOTE dipilih sebagai model terbaik karena lebih mampu mengenali kelas minoritas. Prioritas terhadap *Recall* diberikan karena kesalahan *False Negative* beresiko menyebabkan *marketplace* kehilangan peluang melakukan tindakan pencegahan terhadap transaksi yang berpotensi batal. Penerapan SMOTE membantu menyeimbangkan distribusi data latih sehingga kemampuan model dalam mempelajari karakteristik kelas minoritas menjadi lebih baik[8],[14]. Temuan ini juga sejalan dengan Salehi dan Khedmati [21] yang menunjukkan bahwa penerapan SMOTE pada metode ensemble efektif meningkatkan klasifikasi pada dataset tidak seimbang.

Selain meningkatkan performa prediksi penelitian ini menghapus atribut yang berpotensi menyebabkan *data leakage,* seperti total pembayaran, estimasi potongan biaya pengiriman, dan ongkos kirim yang dibayar pembeli, sehingga model hanya memanfaatkan informasi yang tersedia saat prediksi dan memiliki kemampuan generalisasi yang lebih baik. Transparansi model juga ditingkatkan melalui interpretasi menggunakan SHAP yang mampu menjelaskan kontribusi setiap fitur pada tingkat global maupun individual [12], [20].

Dibandingkan penelitian Koten et al.[9], yang menerapkan algoritma Random Forest dan SMOTE pada kasus yang sama, penelitian ini tidak hanya mengevaluasi Random Forest, tetapi juga membandingkannya dengan XGBoost, menerapkan pengendalian *data leakage,* melakukan *hyperparameter tuning,* serta menambahkan interpretasi menggunakan SHAP. Berdasarkan hasil yang diperoleh, penelitian ini tidak sekedar menghasilkan model prediksi, tetapi turut memberikan penjelasan terhadap faktor-faktor yang memengaruhi hasil prediksi.

Temuan penelitian ini juga melengkapi penelitian Nuryana et al. [3] melalui penerapan identifikasi *data leakage* dan perbandingan dua algoritma *ensemble,* berbeda dengan Dewi et al. [10] yang berfokus pada prediksi pembelian uang. Hasil penelitian sejalan dengan Mohammadpour et al. [11] mengenai pentingnya penanganan *class imbalance* menggunakan SMOTE serta mendukung temuan Netayawijit [12] bahwa SHAP meningkatkan transparansi model.

Secara keseluruhan, novelty penelitian ini terletak pada integrasi perbandingan Random Forest dan XGBoost, pengendalian *data leakage*, penanganan class imbalance menggunakan SMOTE, evaluasi dengan berbagai metrik (*Accuracy, Precision, Recall, F1-Score,* dan ROC-AUC), serta interpretasi SHAP dalam satu kerangka penelitian. Pendekatan yang diterapkan menghasilkan model yang tidak hanya mampu menghasilkan kinerja prediksi yang baik, namun juga lebih transparan dan mudah diinterpretasikan sebagai pendukung pengambilan keputusan pada *marketplace*. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan satu sumber dataset dan dua algoritma *ensemble*. Oleh karena itu, penelitian selanjutnya dapat memanfaatkan dataset yang berasal dari berbagai *marketplace* serta mengevaluasi algoritma lain, seperti LightGBM, CatBoost, atau *deep learning*, serta menerapkan metode optimasi *hyperparameter* yang lebih komprehensif.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 714−726
DOI 10.47065/bits.v8i2.10680
https://ejurnal.seminar-id.com/index.php/bits

4. KESIMPULAN

Studi ini berhasil membandingkan kinerja algoritma Random Forest dan XGBoost dalam memprediksi pembatalan pesanan pada *marketplace* dengan menerapkan *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas dan *Shapley Additive exPlanations* (SHAP) untuk meningkatkan interpretabilitas model. Hasil penelitian menunjukkan bahwa XGBoost memberikan performa yang lebih baik dibandingkan Random Forest pada hampir seluruh metrik evaluasi. Model XGBoost *Tuned* + SMOTE memperoleh nilai *Accuracy, Precision, dan F1-Score* tertinggi, sedangkan XGBoost + SMOTE ditetapkan sebagai model terbaik karena menghasilkan nilai *Recall* tertinggi sebesar 37,09%, sehingga lebih efektif dalam mendeteksi transaksi yang berpotensi dibatalkan pada data tidak seimbang. Penerapan SMOTE berhasil meningkatkan representasi kelas minoritas, sedangkan pengendalian *data leakage* membuat model lebih sesuai diterapkan pada kondisi nyata. Interpretasi menggunakan SHAP menunjukkan bahwa Opsi Pengiriman, Metode Pembayaran, *product_categories,* Tahun, dan Bulan merupakan faktor yang paling berpengaruh terhadap prediksi pembatalan pesanan. Penelitian ini memberikan kontribusi melalui perbandingan Random Forest dan XGBoost, pengendalian *data leakage,* serta integrasi SHAP untuk menghasilkan model yang lebih transparan. Meskipun demikian, penelitian ini masih terbatas pada satu sumber dataset dan dua algoritma *ensemble.* Oleh karena itu, penelitian selanjutnya disarankan menggunkan dataset yang lebih beragam, membandingkan algoritma lain, serta menerapkan metode optimasi *hyperparameter* yang lebih komprehensif agar kemampuan model dalam mendeteksi pembatalan pesanan dapat terus ditingkatkan.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: D.P., dan U.S.;
Metodologi: D.P., dan U.S.;
Perangkat Lunak: D.P.;
Validasi: D.P. dan U.S.;
Analisis Formal: D.P.;
Investigasi: D.P.;
Sumber Daya: D.P.;
Kurasi Data: D.P.;
Penulisan Draf Awal: D.P.;
Penulisan, Peninjauan, dan Penyuntingan: D.P., dan U.S.;
Visualisasi: D.P.;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Penelitian dan penerbitan artikel ini didanai secara mandiri oleh penulis tanpa dukungan pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku *(Not applicable).*

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku *(Not applicable).*

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan ChatGPT untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

[1] S. Wang, Y. Ye, B. Ning, J. H. Cheah, and X. J. Lim, “Why Do Some Consumers Still Prefer In-Store Shopping? An Exploration of Online Shopping Cart Abandonment Behavior,” *Front. Psychol.*, vol. 12, no. January, pp. 1–14, 2022, doi: 10.3389/fpsyg.2021.829696.
[2] T. Le Tan, K. Nguyen Chau Ngoc, H. L. T. Thanh, H. N. T. Thu, and U. V. T. Hoang, “Enhancing Repurchase Intention on

Digital Platforms Based on Shopping Well-Being Through Shopping Value, Trust and Impulsive Buying," *SAGE Open*, vol. 14, no. 3, pp. 1–25, 2024, doi: 10.1177/21582440241278454.

[3] M. R. Nuryana, Muhammad Nur Aulia Rahman, and R. S. Hadikusuma, "Predicting Successful Timely Order Completion in E-Commerce Using XGboost and Catboost Models," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 11, no. 1, pp. 69–77, 2026, doi: 10.24252/instek.v11i1.63315.

[4] R. N. S. Hakim and A. Asmunin, "Optimasi Algoritma Random Forest dengan Teknik Boosting dalam Prediksi Churn Pelanggan di Industri Telekomunikasi," *J. Manag. Inform.*, vol. 13, no. 2, pp. 1–9, 2024, [Online]. Available: https://ejournal.unesa.ac.id/index.php/jurnal-manajemen-informatika/article/view/59841

[5] M. R. Atsauri, H. Mawengkang, and S. Efendi, "Enhancing Unbalanced Data Classification with Cross-Validation and Extreme Gradient Boosting: A Comprehensive Analysis," *J. Informatics Telecommun. Eng.*, vol. 7, no. 1, pp. 30–42, 2023, doi: 10.31289/jite.v7i1.8690.

[6] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, 2024, doi: 10.1007/s10994-022-06296-4.

[7] Y. Han, Z. Wei, and G. Huang, "An imbalance data quality monitoring based on SMOTE-XGBOOST supported by edge computing," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-60600-x.

[8] J. Starcke, J. Spadafora, J. Spadafora, P. Spadafora, and M. Toma, "The Effect of Data Leakage and Feature Selection on Machine Learning Performance for Early Parkinson's Disease Detection," *Bioengineering*, vol. 12, no. 8, 2025, doi: 10.3390/bioengineering12080845.

[9] F. Koten, I. W. Sudiarsa, N. K. Trisnawati, M. M. Palo Pera, and M. A. Nidin, "Prediksi Pembatalan Pesanan E-Commerce Indonesia Mengunakan Random Forest dan SMOTE," *J. Manaj. Akunt. danIlmu Ekon.*, vol. 3, no. 1, pp. 8–14, 2026, doi: 10.70585/jumali.v3i1.192.

[10] A. C. Dewi, A. Hermawan, and D. Avianto, "Classification of Customers' Repeat Order Probability Using Decision Tree, Naïve Bayes and Random Forest," *J. Pilar Nusa Mandiri*, vol. 20, no. 1, pp. 52–59, 2024, doi: 10.33480/pilar.v20i1.5243.

[11] S. I. Mohammadpour, M. Khedmati, and M. J. H. Zada, "Classification of truck-involved crash severity: Dealing with missing, imbalanced, and high dimensional safety data," *PLoS One*, vol. 18, no. 3, pp. 1–22, 2023, doi: 10.1371/journal.pone.0281901.

[12] P. Netayawijit, W. Chansanam, and K. Sorn-In, "Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support," *Healthc.*, vol. 13, no. 20, pp. 1–26, 2025, doi: 10.3390/healthcare13202588.

[13] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.

[14] Q. Dong, "Leakage Prediction in Machine Learning Models When Using Data from Sports Wearable Sensors," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/5314671.

[15] A. Ichwani, R. I. Kesuma, A. Setiawan, I. E. Wicaksono, and R. Hanifah, "Preventing Data Leakage in Classification via Integrated Machine Learning Pipelines: Preprocessing, Feature Transformation, and Hyperparameter Tuning," *J. Tek. Inform.*, vol. 7, no. 1, pp. 391–410, 2026, doi: 10.52436/1.jutif.2026.7.1.5490.

[16] L. Chmielowski, M. Kucharzak, and R. Burduk, "Novel method of building train and test sets for evaluation of machine learning models related to software bugs assignment," *Sci. Rep.*, vol. 13, no. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-48617-0.

[17] A. Demircioğlu, "Applying oversampling before cross-validation will lead to high bias in radiomics," *Sci. Rep.*, vol. 14, no. 1, pp. 1–11, 2024, doi: 10.1038/s41598-024-62585-z.

[18] J. T. Hancock, T. M. Khoshgoftaar, and J. M. Johnson, "Evaluating classifier performance with highly imbalanced Big Data," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00724-5.

[19] S. Roumeliotis *et al.*, "ROC curve analysis: a useful statistic multi-tool in the research of nephrology," *Int. Urol. Nephrol.*, vol. 56, no. 8, pp. 2651–2658, 2024, doi: 10.1007/s11255-024-04022-8.

[20] A. Nambiar, S. Harikrishnaa, and S. Sharanprasath, "Model-agnostic explainable artificial intelligence tools for severity prediction and symptom analysis on Indian COVID-19 data," *Front. Artif. Intell.*, vol. 6, 2023, doi: 10.3389/frai.2023.1272506.

[21] A. R. Salehi and M. Khedmati, "A cluster-based SMOTE both-sampling (CSBBoost) ensemble algorithm for classifying imbalanced data," *Sci. Rep.*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-55598-1.

