



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2025) pp: 5666-5672

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Prediksi Pembelian E-Commerce Menggunakan XGBoost Berbasis Perilaku Sesi Pengguna

Syalaiza Nizar Ruscikasani¹, Raden Roro Najwa Oktalivia², Firman Restu Putra³, Ahmad Jurnaidi Wahidin⁴, Beni Rahmatullah⁵, Ika kurniawati⁶

^{1,2,3}Teknik Informatika, Universitas Bina Sarana Informatika

^{4,5}Teknologi Informasi, Universitas Bina Sarana Informatika

⁶Sistem Informasi, Universitas Nusa Mandiri

¹syalaizacika@gmail.com, ²najwaoktalivia@gmail.com, ³covid5548@gmail.com, ⁴ahmad.ajn@bsi.ac.id,

⁵beni.brh@bsi.ac.id, ⁶ika.iki@nusamandiri.ac.id

Abstrak

Penelitian ini membahas pemodelan prediksi pembelian pada platform e-commerce dengan memanfaatkan algoritma Extreme Gradient Boosting (XGBoost) berbasis data perilaku sesi pengguna. Dataset yang digunakan terdiri dari 12.719 sesi pengguna dengan sejumlah atribut perilaku, meliputi waktu kunjungan (Timestamp), durasi interaksi halaman (TimeOnPage_seconds), sumber rujukan (ReferralSource), tipe perangkat (DeviceType), tipe halaman (PageType), negara asal pengguna, serta jumlah item dalam keranjang belanja. Tahap prapengolahan data dilakukan untuk meningkatkan kualitas dataset sebelum pemodelan. Atribut numerik diproses menggunakan imputasi median guna menangani nilai hilang secara stabil terhadap outlier, sedangkan fitur kategorikal ditransformasikan menggunakan ordinal encoding agar sesuai dengan karakteristik algoritma berbasis pohon. Dataset selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified split untuk menjaga proporsi kelas pada variabel target. Model XGBoost dilatih menggunakan parameter terkalibrasi dan dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, dan ROC-AUC. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 0,785 dan ROC-AUC sebesar 0,804, yang menandakan kemampuan diskriminasi yang baik dalam membedakan sesi yang berpotensi menghasilkan pembelian. Analisis feature importance berbasis gain mengungkapkan bahwa fitur Timestamp dan TimeOnPage_seconds merupakan faktor paling berpengaruh dalam pembentukan prediksi. Visualisasi pohon keputusan memberikan pemahaman tambahan mengenai mekanisme pemisahan fitur pada model. Temuan ini menunjukkan bahwa data perilaku sesi pengguna dapat dimanfaatkan secara efektif untuk mendukung prediksi konversi serta menjadi dasar pengembangan sistem rekomendasi dan strategi pemasaran e-commerce yang lebih adaptif.

Kata kunci: Pembelian E-Commerce, Perilaku Sesi Pengguna, XGBoost, Prediksi Konversi, Machine Learning

1. Latar Belakang

Perkembangan teknologi digital telah mengubah secara signifikan pola konsumsi masyarakat dan meningkatkan aktivitas transaksi melalui platform e-commerce[1]. Kemudahan akses, variasi produk, serta efisiensi proses transaksi menjadikan e-commerce sebagai salah satu sektor dengan pertumbuhan tercepat dalam ekosistem ekonomi digital[2]. Peningkatan jumlah pengguna juga diikuti oleh meningkatnya volume interaksi dalam setiap sesi penggunaan, seperti aktivitas menelusuri produk, melakukan klik, menambahkan barang ke keranjang, atau melakukan pencarian[3]. Setiap aktivitas tersebut menghasilkan data perilaku yang kaya dan berpotensi memberikan gambaran mengenai niat pengguna dalam melakukan pembelian. Beberapa studi kasus menunjukkan bahwa penggunaan data perilaku belum optimal untuk konversi[4]. Rendahnya tingkat konversi tetap menjadi tantangan, sementara biaya akuisisi pelanggan dan promosi digital terus meningkat setiap tahun[5].

Berbagai pendekatan statistika klasik seperti regresi linier dan regresi logistik sebelumnya digunakan untuk memahami perilaku konsumen, tetapi metode konvensional ini sering tidak mampu menangkap pola perilaku yang kompleks dan non linear pada data sesi pengguna[6]. Model tersebut cenderung terbatas ketika diterapkan pada data berskala besar dengan variabel perilaku yang beragam, sehingga diperlukan metode yang lebih adaptif dan akurat dalam memprediksi kemungkinan pembelian. Algoritma machine learning berbasis ensemble seperti Extreme Gradient Boosting atau XGBoost mulai banyak digunakan dalam analisis perilaku konsumen e-commerce dan menunjukkan performa unggul pada berbagai permasalahan prediksi[7][8][9]. Sejumlah penelitian

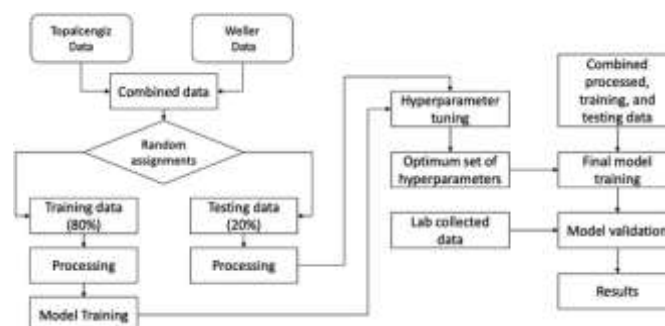
membuktikan bahwa *XGBoost* mampu menghasilkan akurasi lebih tinggi dibandingkan model dasar seperti *Random Forest* atau *regresi logistik* pada data tabular dengan pola interaksi yang kompleks[10].

Meskipun performanya menjanjikan, sebagian besar penelitian terkait *XGBoost* masih berfokus pada peramalan penjualan atau prediksi churn pelanggan, dan belum banyak yang memanfaatkan data perilaku sesi pengguna sebagai dasar prediksi pembelian. Padahal, data sesi memiliki keunggulan berupa kemampuan memberikan sinyal dini terkait niat beli karena informasi yang ditangkap bersifat *real time* dan detail. Penelitian internasional yang dilakukan oleh Wang et al. menunjukkan bahwa *XGBoost* dapat memprediksi perilaku pembelian pengguna secara akurat ketika fitur perilaku dianalisis dengan tepat[11]. Penelitian lain juga menegaskan bahwa model *boosting* mampu memetakan kecenderungan pengguna dengan performa lebih baik dibandingkan pendekatan konvensional[12]. Namun, penelitian yang secara khusus menggabungkan pemanfaatan data perilaku sesi dengan model *XGBoost* untuk prediksi pembelian masih relatif terbatas sehingga terdapat celah penelitian yang perlu diisi.

Melihat kondisi tersebut, penelitian ini penting dilakukan untuk mengembangkan model prediksi yang mampu membaca pola perilaku pengguna dalam satu sesi dan memproyeksikan kemungkinan terjadinya pembelian secara akurat. Pendekatan berbasis *XGBoost* berpotensi memberikan kontribusi signifikan baik dari sisi akademik maupun praktis. Model prediktif semacam ini dapat membantu platform *e-commerce* dalam merancang strategi personalisasi, rekomendasi produk, dan intervensi promosi secara lebih tepat sasaran. Selain itu, penelitian ini diharapkan dapat memperluas pemahaman mengenai perilaku sesi sebagai salah satu aspek penting dalam analitik *e-commerce* modern serta memberikan gambaran mengenai fitur perilaku yang paling berpengaruh terhadap keputusan pembelian. Dengan demikian, penelitian ini memiliki kebaruan baik dari segi pemanfaatan data sesi, penggunaan *XGBoost*, maupun kontribusinya dalam pengembangan metode prediksi konversi dalam lingkungan *e-commerce*.

2. Metode Penelitian

Penelitian ini dirancang sebagai pemodelan prediktif menggunakan algoritma *XGBoost* untuk memprediksi keputusan pembelian berdasarkan perilaku sesi pengguna. Seluruh tahapan dilaksanakan secara terstruktur mulai dari pemuatan dataset, prapengolahan data, pembangunan model, evaluasi performa, hingga interpretasi hasil. Diagram alir lengkap.



Gambar 1. Diagram Alir

2.1 Data Penelitian dan Prapengolahan

Dataset yang digunakan penelitian ini terdiri dari 12.719 baris data dengan sepuluh atribut perilaku pengguna yang merepresentasikan aktivitas interaksi pengguna dalam satu sesi pada platform *e-commerce*. Atribut-atribut tersebut mencakup informasi waktu, konteks sesi, serta karakteristik interaksi yang relevan untuk memodelkan keputusan pembelian.

Dua atribut identitas yaitu *SessionID* dan *UserID*, dihapus sejak tahap awal penelitian. Penghapusan ini dilakukan untuk mencegah terjadinya data *leakage* serta memastikan bahwa proses pemodelan sepenuhnya bergantung pada pola perilaku pengguna, bukan pada informasi identitas yang bersifat unik dan tidak dapat digeneralisasi. Dengan demikian, model yang dibangun diharapkan mampu menangkap pola keputusan pembelian yang bersifat umum dan dapat diterapkan pada pengguna baru.

Variabel target dalam penelitian ini ditentukan melalui proses pencocokan nama kolom yang mengandung kata “*purchase*”, sehingga diperoleh kolom *Purchased* sebagai label *biner* yang merepresentasikan apakah suatu sesi berujung pada pembelian atau tidak. Penentuan variabel target ini memastikan bahwa permasalahan penelitian diformulasikan secara jelas sebagai tugas klasifikasi biner.

Pada tahap prapengolahan data, atribut *numerik* diproses menggunakan metode imputasi median untuk menangani nilai yang hilang. Pendekatan ini dipilih karena nilai median relatif lebih stabil terhadap keberadaan *outlier* serta distribusi data yang tidak normal, yang umum dijumpai pada data perilaku pengguna. Keunggulan imputasi median dalam konteks data tabular telah dibahas secara luas dalam penelitian komparatif terbaru oleh Li et al. (2024), yang menunjukkan bahwa metode ini termasuk yang paling konsisten ketika pola *missing value* tidak seimbang[13]. Nilai median dihitung menggunakan persamaan berikut:

$$x^{\wedge} = \text{median}(x_1, x_2, \dots, x_n) \quad (1)$$

Untuk fitur kategorikal seperti *PageType*, *DeviceType*, *Country*, dan *ReferralSource*, dilakukan transformasi menggunakan *Ordinal Encoding*. Pendekatan ini dipilih karena terbukti memberikan performa yang lebih baik pada algoritma berbasis pohon dibandingkan *one-hot encoding*, yang berpotensi meningkatkan dimensi fitur secara signifikan dan menurunkan efisiensi model[14]. Sebelum proses encoding dilakukan, nilai kosong pada fitur kategorikal digantikan dengan token “*MISSING*”, sehingga seluruh kategori dapat diproses secara konsisten oleh model.

Setelah tahap prapengolahan selesai, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified split. Teknik ini dipilih untuk menjaga proporsi kelas pada variabel target tetap seimbang antara data latih dan data uji, sebagaimana lazim diterapkan dalam penelitian prediksi niat pembelian pada *platform e-commerce*[15]. Proses pembagian ini menghasilkan 10.175 data latih dan 2.544 data uji, yang selanjutnya digunakan pada tahap pemodelan dan evaluasi.

2.2 Pemodelan dan Perhitungan Gain

Tahap pemodelan dalam penelitian ini dilakukan menggunakan algoritma *XGBoost Classifier*, yang telah banyak digunakan dan terbukti unggul dalam pengolahan data tabular. Keunggulan XGBoost terletak pada kemampuannya memanfaatkan teknik *gradient boosting* berbasis pohon keputusan, sehingga mampu menangkap hubungan *non-linear* antar *fitur* serta mengelola interaksi kompleks antar variabel. Sejumlah penelitian menunjukkan bahwa pada skenario data tabular umum, pendekatan ini secara konsisten mampu mengungguli model *deep learning*, baik dari sisi performa maupun efisiensi komputasi[16]. Oleh karena itu, *XGBoost* dipilih sebagai metode utama dalam membangun model prediksi pada penelitian ini.

Tabel 1. Perhitungan Gain

Parameter	Nilai
n_estimators	200
learning_rate	0.1
max_depth	4
subsample	0.8
colsample_bytree	0.8
eval_metric	Logloss
random_stat	42

Konfigurasi parameter yang digunakan dalam proses pelatihan model ditampilkan pada Tabel 1 Perhitungan Gain. Parameter *n_estimators* ditetapkan sebesar 200 untuk menentukan jumlah pohon yang dibangun dalam proses boosting. Nilai *learning_rate* sebesar 0.1 digunakan untuk mengontrol kontribusi setiap pohon baru terhadap model secara keseluruhan. Parameter *max_depth* diset pada nilai 4 untuk membatasi kompleksitas pohon dan mengurangi risiko *overfitting*. Selain itu, nilai *subsample* dan *colsample_bytree* masing-masing ditetapkan sebesar 0.8 guna meningkatkan generalisasi model melalui pengambilan sampel data dan fitur secara acak. Metode evaluasi kesalahan model menggunakan *logloss*, sedangkan *random_state* bernilai 42 digunakan untuk menjaga reproduktibilitas hasil eksperimen.

Selain proses pelatihan model secara standar menggunakan *XGBoost*, penelitian ini juga melakukan perhitungan manual terhadap nilai gradient, hessian, dan gain pada 200 sampel awal. Langkah ini dilakukan sebagai bentuk verifikasi metodologis untuk memastikan bahwa mekanisme pembentukan pohon pada implementasi *XGBoost* selaras dengan teori *gradient boosting* yang mendasarinya. Dengan melakukan perhitungan manual, penelitian ini tidak hanya mengandalkan hasil *black-box* dari perangkat lunak, tetapi juga mengevaluasi kesesuaian proses optimasi yang terjadi di tingkat matematis.

$$Gain = \frac{2}{1} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{G^2}{H + \lambda} \right) - \gamma \quad (2)$$

Hasil dari perhitungan manual tersebut menunjukkan bahwa pemisahan optimal terjadi pada fitur *Timestamp* dengan nilai ambang 5971. Nilai pemisahan ini terbukti identik dengan pemisahan yang dihasilkan pada pohon pertama model *XGBoost* selama proses pelatihan. Kesamaan hasil ini mengindikasikan bahwa mekanisme pemilihan split yang dilakukan oleh algoritma *XGBoost* telah berjalan sesuai dengan formulasi teoritisnya.

2.3 Evaluasi dan Interpretasi Model

Tahap evaluasi model dilakukan untuk menilai sejauh mana model *XGBoost* yang dibangun mampu memprediksi keputusan pembelian pengguna secara akurat dan konsisten. Evaluasi kinerja dilakukan menggunakan beberapa metrik, yaitu *akurasi*, *precision*, *recall*, *F1-score*, dan *ROC-AUC*. Penggunaan lebih dari satu metrik dimaksudkan untuk memberikan gambaran yang komprehensif terhadap performa model, khususnya dalam konteks data perilaku pengguna yang cenderung kompleks dan berpotensi mengalami ketidakseimbangan kelas.

Sejumlah penelitian terbaru yang menerapkan model berbasis *tree-based* atau *gradient boosting* untuk prediksi perilaku maupun keputusan pembelian pengguna juga menggunakan kombinasi metrik evaluasi yang serupa, seperti AUC, precision, recall, dan F1-score, guna menangkap berbagai aspek kinerja model secara lebih menyeluruh[17]. Pendekatan ini memungkinkan evaluasi tidak hanya berfokus pada ketepatan prediksi secara umum, tetapi juga pada kemampuan model dalam mengidentifikasi kelas positif secara benar.

Dalam penelitian ini, salah satu metrik utama yang digunakan adalah *F1-score*, yang merupakan harmonisasi antara precision dan recall. *F1-score* dirumuskan sebagai berikut:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Hasil evaluasi pada data uji menunjukkan bahwa model menghasilkan nilai *akurasi* sebesar 0,785, *precision* sebesar 0,827, *recall* sebesar 0,580, *F1-score* sebesar 0,682, serta *ROC-AUC* sebesar 0,804. Nilai-nilai ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mempelajari pola perilaku pengguna dan mengidentifikasi sesi yang mengarah pada pembelian. Tingginya nilai precision mengindikasikan bahwa prediksi pembelian yang dihasilkan model relatif jarang mengalami kesalahan, sementara nilai recall yang lebih rendah menunjukkan bahwa masih terdapat sebagian sesi pembelian yang belum sepenuhnya terdeteksi oleh model.

Selain evaluasi kuantitatif, interpretasi model juga dilakukan untuk memahami kontribusi masing-masing fitur dalam proses pengambilan keputusan model. Interpretasi ini dilakukan melalui analisis *feature importance berbasis gain*, yang mencerminkan seberapa besar kontribusi suatu fitur dalam meningkatkan kualitas pemisahan pada pohon keputusan. Hasil analisis menunjukkan bahwa fitur yang paling berpengaruh terhadap prediksi pembelian meliputi *Timestamp*, *TimeOnPage_seconds*, *Country*, dan *ReferralSource*.

Visualisasi pohon keputusan pertama juga digunakan sebagai sarana interpretasi tambahan. Visualisasi ini memberikan gambaran intuitif mengenai bagaimana model melakukan pemisahan data sejak iterasi awal, serta menunjukkan urutan fitur yang digunakan dalam proses pengambilan keputusan. Dengan pendekatan ini, proses prediksi model menjadi lebih transparan dan dapat dipahami secara konseptual.

Pada tahap akhir, model yang telah dilatih dan dievaluasi disimpan dalam format *JSON* menggunakan metode *save_model()*. Penyimpanan model dalam format ini memungkinkan integrasi lebih lanjut pada tahap operasional sistem, sehingga model dapat digunakan kembali tanpa perlu dilakukan pelatihan ulang.

3. Hasil dan Pembahasan

Hasil penelitian diperoleh dari penerapan algoritma *XGBoost* terhadap dataset interaksi pengguna yang mencerminkan perilaku mereka selama menjalani sesi pada *platform e-commerce*. Dataset terdiri dari 12.719 observasi dengan variabel target *Purchased* yang menunjukkan apakah pengguna melakukan pembelian pada sesi tersebut. Variabel *input* mencakup waktu kunjungan (*Timestamp*), lama interaksi halaman (*TimeOnPage_seconds*), jumlah item di keranjang, tipe perangkat, tipe halaman, negara asal, serta sumber rujukan. Seluruh fitur ini merepresentasikan perilaku dan konteks sesi pengguna. Tahap pra-pemrosesan meliputi imputasi nilai hilang untuk variabel numerik dan encoding untuk fitur kategorikal agar seluruh variabel dapat diolah oleh model. Data kemudian dibagi menjadi 80 persen data latih dan 20 persen data uji menggunakan *stratified split* untuk menjaga proporsi kelas.

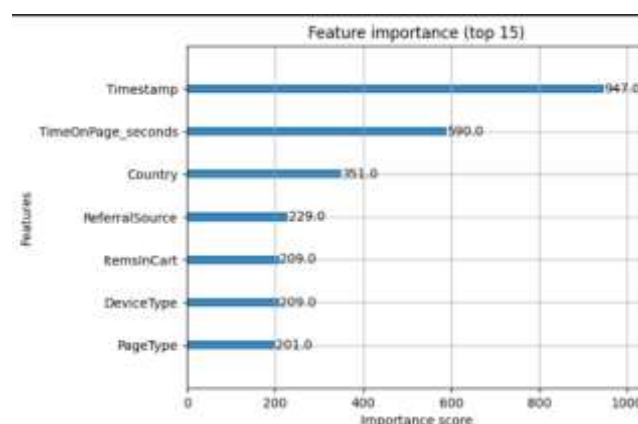
Tabel 2. Hasil dan Pembahasan

Metrik	Nilai
Accuracy	0.785
Precision	0.827
Recall	0.580
F1- Score	0.682
ROC-AUC	0.804

Model *XGBoost* dilatih menggunakan *learning_rate* 0.1, *n_estimators* 200, dan *max_depth* 4. Evaluasi pada data uji menghasilkan nilai *accuracy* sebesar 0.785 dengan *precision* 0.827 dan *recall* 0.580. Nilai *F1-score* sebesar 0.682 menunjukkan keseimbangan performa antara kemampuan model mengidentifikasi pembelian secara tepat dan menangkap variasi sesi yang benar-benar menghasilkan transaksi. Nilai *ROC-AUC* sebesar 0.804 memberikan gambaran bahwa model mampu membedakan sesi yang berpotensi menghasilkan pembelian dan sesi yang tidak dengan tingkat pemisahan yang baik.

Analisis ini memperlihatkan bahwa *XGBoost* memberikan hasil yang stabil secara keseluruhan. Tingginya nilai *precision* mengindikasikan bahwa model jarang salah ketika memprediksi adanya pembelian, sedangkan nilai *recall* yang lebih rendah menunjukkan bahwa sebagian sesi berpotensi pembelian masih sulit dikenali sepenuhnya oleh model. Hal ini wajar mengingat variasi perilaku pengguna pada *platform e-commerce* seringkali kompleks dan tidak sepenuhnya tercermin dalam fitur perilaku dasar.

3.1. Visualisasi Feature Importance

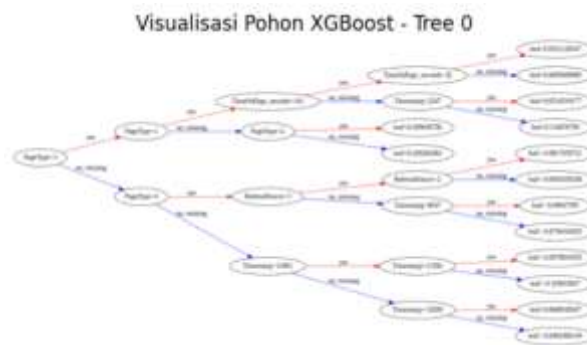


Gambar 2. Visualisasi Featire Importance

Memperlihatkan bahwa *Timestamp* menjadi fitur paling berpengaruh dalam prediksi. Ini menunjukkan bahwa waktu akses pengguna memiliki pola tertentu yang berkaitan dengan kecenderungan melakukan pembelian. Variabel *TimeOnPage_seconds* juga menempati urutan atas, menandakan bahwa durasi interaksi yang lebih lama sering berkaitan dengan minat pembelian yang lebih kuat. Fitur lainnya seperti *Country*, *ReferralSource*,

ItemsInCart, dan *DeviceType* berperan penting dalam membentuk keputusan model, yang mengindikasikan bahwa karakteristik pengguna dan titik masuk sesi turut mempengaruhi peluang konversi.

3.2. Visualisasi Pohon Keputusan XGBoos



Gambar 3. Visualisasi Pohon XGBOOS

Pohon keputusan memperlihatkan bagaimana model *XGBoost* memulai proses pemisahan data berdasarkan fitur yang dianggap paling berpengaruh. Pada bagian paling atas, model memilih *TimeOnPage_seconds* sebagai pemisahan pertama. Hal ini menunjukkan bahwa durasi interaksi pengguna pada halaman merupakan faktor yang sangat kuat dalam membedakan apakah sebuah sesi memiliki kecenderungan menghasilkan pembelian atau tidak. Pengguna yang menghabiskan waktu lebih lama pada halaman biasanya menunjukkan minat yang lebih tinggi terhadap produk. Setelah pemisahan berdasarkan durasi interaksi, model menggunakan *Timestamp* sebagai pemisahan berikutnya. Ini menandakan bahwa waktu kunjungan juga memiliki pola tertentu yang berkaitan dengan perilaku pembelian, misalnya adanya jam-jam atau periode tertentu ketika pengguna lebih cenderung melakukan transaksi.

Pada cabang-cabang berikutnya, fitur seperti *PageType*, *ReferralSource*, dan *Country* digunakan untuk mengarahkan keputusan model secara lebih spesifik. Fitur *PageType* membantu model membedakan sesi berdasarkan jenis halaman yang dikunjungi, sementara *ReferralSource* menunjukkan pengaruh sumber kedatangan pengguna terhadap peluang pembelian. Fitur *Country* menggambarkan adanya perbedaan pola perilaku antara pengguna dari wilayah yang berbeda.

Beberapa cabang juga menunjukkan arah khusus untuk nilai yang hilang, karena *XGBoost* secara otomatis menentukan jalur terbaik bagi data yang tidak memiliki nilai lengkap. Mekanisme ini membuat model tetap mampu mengambil keputusan yang akurat meskipun terdapat data yang tidak sempurna.

Secara keseluruhan, visualisasi pohon keputusan ini menunjukkan bahwa *XGBoost* memanfaatkan kombinasi sinyal perilaku, waktu kunjungan, dan karakteristik pengguna untuk memprediksi kemungkinan terjadinya pembelian. Struktur pemisahan yang terbentuk memberikan gambaran yang lebih jelas mengenai pola keputusan internal model.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *XGBoost* mampu memprediksi keputusan pembelian pada platform *e-commerce* secara akurat dengan memanfaatkan data perilaku sesi pengguna. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 0,785, *precision* 0,827, *recall* 0,580, *F1-score* 0,682, dan *ROC-AUC* 0,804, yang menandakan kemampuan model dalam membedakan sesi yang berpotensi menghasilkan pembelian dan yang tidak. Fitur seperti *Timestamp*, *TimeOnPage_seconds*, *Country*, dan *ReferralSource* terbukti memiliki pengaruh paling besar terhadap keputusan model, sehingga dapat dianggap sebagai indikator penting dalam memetakan kecenderungan pembelian pengguna. Penelitian ini memberikan bukti bahwa data perilaku sesi dapat dimanfaatkan secara efektif untuk meningkatkan prediksi konversi dan mendukung pengambilan keputusan berbasis data pada platform *e-commerce*. Model yang dihasilkan dapat menjadi dasar bagi sistem rekomendasi, personalisasi konten, atau strategi pemasaran yang lebih adaptif terhadap pola perilaku pengguna. Ke depan, penelitian dapat diperluas dengan menambahkan fitur perilaku yang lebih kaya atau menguji model pada konteks *e-commerce* yang berbeda untuk meningkatkan generalisasi prediksi.

Referensi

- [1] J. Dwi, S. Amory, M. Mudo, and J. Rhena, "Transformasi Ekonomi Digital dan Evolusi Pola Konsumsi : Tinjauan Literatur tentang Perubahan Perilaku Belanja di Era Internet," vol. 14, pp. 28–37, 2025.
- [2] "E-COMMERCE AND ECONOMIC GROWTH IN INDONESIA: ANALYSIS OF PANEL DATA REGRESSION Kandi Dwi Pratiwi 1," vol. 7, pp. 171–186, 2022.
- [3] A. V. Bharathi, J. M. Rao, and A. K. Tripathy, "Click Stream Analysis in e-Commerce Websites-a Framework," *2018 Fourth Int. Conf. Comput. Commun. Control Autom.*, pp. 1–5, 2018.
- [4] L. Bigon and C. Greco, "Prediction is very hard , especially about conversion . Predicting user purchases from clickstream data in fashion e-commerce Prediction is very hard , especially about conversion * Predicting user purchases from clickstream data in fashion e-commerce," 2025.
- [5] M. T. Nuseir, G. A. El Refae, A. Aljumah, M. Alshurideh, S. Urabi, and B. Al Kurdi, "and the Impact on Customer Experience : A Systematic Review," 2023.
- [6] T. Chen and C. Guestrin, "XGBoost : A Scalable Tree Boosting System," pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [7] T. Nyata, "APPLICATION OF XGBOOST ALGORITHM FOR PREDICTING CUSTOMER," vol. 1, pp. 50–59, 2025, doi: 10.64476/jtbc.v1i1.5.
- [8] R. Winurputra and D. E. Ratnawati, "PERAMALAN PENJUALAN PRODUK MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST) DAN KERANGKA KERJA CRISP-DM UNTUK PENGOPTIMALAN MANAJEMEN PERSEDIAAN (STUDI KASUS : UB MART) PRODUCT SALES FORECASTING USING EXTREME GRADIENT BOOSTING (XGBOOST) AND CRISP-DM FRAMEWORK FOR INVENTORY MANAGEMENT OPTIMIZATION (CASE STUDY : UB MART)," vol. 12, no. 2, pp. 417–428, 2025, doi: 10.25126/jtiik.2025129451.
- [9] I. G. Ayu, R. Astarani, and I. G. Surya, "Analisis Perbandingan XGBoost dan LightGBM dalam Prediksi Penjualan Ritel Walmart Store Sales," vol. 3, no. M1, pp. 717–728, 2025.
- [10] M. Imani and A. Beikmohammadi, "Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE , ADASYN , and GNUS Under Varying Imbalance Levels," no. M1, pp. 1–40, 2025.
- [11] W. Wang *et al.*, "A User Purchase Behavior Prediction Method Based on XGBoost," 2023.
- [12] N. Hesvindrati, A. Aminuddin, J. Mahadhni, A. Pambudi, and B. Sudaryatno, "International Journal of Current Science Research and Review Behavior-Based Purchase Intent Prediction in E-Commerce : A Machine Learning Approach Corresponding Author : Agung Pambudi Corresponding Author : Agung Pambudi," vol. 08, no. 08, pp. 3970–3980, 2025, doi: 10.47191/ijcsrr/V8-i8-03.
- [13] J. Li *et al.*, "Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," pp. 1–9, 2024, doi: 10.1186/s12874-024-02173-x.
- [14] M. Albaladejo-saura, R. Vaquero-crist, J. Alfonso, and F. Esparza-ros, "applied sciences Influence of Maturity Status on Kinanthropometric and Physical Fitness Variables in Adolescent Female Volleyball Players," 2022.
- [15] K. Valtanen *et al.*, "ScienceDirect ScienceDirect Matching Circularity Improvements and Digital Product Passport Viewpoints : Insights from Three Industrial Case Studies," *Procedia Comput. Sci.*, vol. 253, no. 2024, pp. 1720–1729, 2025, doi: 10.1016/j.procs.2025.01.234.
- [16] R. Shwartz-ziv, "T d : d l n a y n," 2021.
- [17] L. Gan, "XGBoost-Based E-Commerce Customer Loss Prediction," vol. 2022, 2022.