Jurnal Ilmiah Teknik Informatika dan Komunikasi
Volume. 6, Nomor. 3, November 2026
E-ISSN : 2827-7945 ; P-ISSN : 2827-8127; Hal. 344-358
DOI: https://doi.org/10.55606/juitik.v6i3.2748
Tersedia: https://journal.sinov.id/index.php/juitik

Model Prediksi Permintaan Produk pada *E- Commerce* Menggunakan Pendekatan *Big* Data dan Analisis Metode Pembayaran

Maksidion Lende[1*], Herman Huki Ratu[2], Lidia Lali Momo[3]
[1-3] Teknik Informatika, Universitas Stella Maris Sumba, Indonesia
**Penulis Korespondensi: maksidionlende7@gmail.com*

Abstract*. The rapid development of e-commerce has generated large amounts of transaction and customer behavior data that can be utilized to support business decision-making. One of the important challenges faced by e-commerce businesses is predicting product demand accurately because demand fluctuations can cause overstocking or stock shortages. This study aims to develop a product demand prediction model by utilizing big data analytics and incorporating payment method variables. The study uses a quantitative approach with descriptive and predictive analysis. The data consist of historical transaction records, product information, payment methods, and sales volume. The payment methods considered include credit or debit cards, bank transfers, electronic wallets, and cash on delivery. Data processing is designed to include data collection, cleaning, transformation, exploratory analysis, feature engineering, model development, and evaluation. Several prediction approaches are considered, including ARIMA or SARIMA as a traditional time-series model, XGBoost and Random Forest as machine learning models, and Long Short-Term Memory as a deep learning model. Model performance will be evaluated using Root Mean Square Error, Mean Absolute Error, and Mean Absolute Percentage Error. The proposed approach is expected to provide a more comprehensive representation of product demand by considering payment behavior as an additional predictive variable. The resulting model can support inventory planning and product availability decisions in e-commerce.*

Keywords*: Big Data; Demand Prediction; E-Commerce; Payment Method; XGBoost.*

Abstrak. Perkembangan e-commerce yang semakin pesat menghasilkan data dalam jumlah besar yang berasal dari transaksi, perilaku pelanggan, produk, dan metode pembayaran. Data tersebut dapat dimanfaatkan untuk mendukung pengambilan keputusan, khususnya dalam memprediksi permintaan produk. Permintaan produk yang mengalami perubahan dari waktu ke waktu dapat menyebabkan terjadinya kelebihan persediaan maupun kekurangan stok apabila tidak diperkirakan dengan baik. Penelitian ini bertujuan mengembangkan model prediksi permintaan produk pada e-commerce dengan memanfaatkan pendekatan big data serta mengintegrasikan variabel metode pembayaran. Penelitian menggunakan pendekatan kuantitatif dengan rancangan analisis deskriptif dan prediktif. Data yang digunakan meliputi data transaksi historis, informasi produk, metode pembayaran, dan jumlah produk terjual. Metode pembayaran yang dianalisis meliputi kartu kredit atau debit, transfer bank, dompet elektronik, dan cash on delivery (COD). Tahapan penelitian meliputi pengumpulan data, pembersihan data, transformasi data, analisis eksploratif, rekayasa fitur, pemodelan, dan evaluasi. Model yang dirancang meliputi ARIMA atau SARIMA, XGBoost, Random Forest, dan Long Short-Term Memory (LSTM). Evaluasi dilakukan menggunakan Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE). Penelitian ini diharapkan menghasilkan model prediksi yang mampu memanfaatkan pola transaksi dan metode pembayaran sebagai informasi tambahan untuk memperkirakan permintaan produk serta mendukung perencanaan persediaan pada e-commerce.

Kata kunci: *Big Data; E-Commerce;* Metode Pembayaran; Prediksi Permintaan; XGBoost.

1. LATAR BELAKANG

Perkembangan internet dan teknologi informasi telah mendorong perubahan dalam aktivitas perdagangan. Salah satu bentuk perubahan tersebut adalah meningkatnya pemanfaatan *electronic commerce* atau e-commerce sebagai media transaksi antara penjual dan konsumen. E-commerce memungkinkan proses pencarian produk, pemesanan, pembayaran, dan pertukaran informasi dilakukan melalui jaringan internet. Aktivitas tersebut menghasilkan data transaksi yang semakin besar dan beragam sehingga dapat dimanfaatkan sebagai sumber informasi untuk mendukung pengambilan keputusan bisnis (Akbar & Alam, 2020).

Naskah Masuk: 27 Agustus 2026; Revisi: 23 September 2026; Diterima: 13 Oktober 2026; Terbit: 01 November 2026

Pertumbuhan transaksi e-commerce juga menimbulkan tantangan dalam pengelolaan persediaan. Permintaan terhadap produk dapat berubah dari waktu ke waktu sehingga jumlah stok yang tersedia belum tentu sesuai dengan kebutuhan konsumen. Kondisi ketika persediaan lebih besar daripada permintaan dapat menyebabkan *overstock*, sedangkan persediaan yang lebih rendah daripada permintaan dapat menyebabkan *understock* atau *stockout*. Kedua kondisi tersebut dapat memengaruhi efisiensi pengelolaan persediaan dan peluang penjualan.

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah melakukan prediksi permintaan berdasarkan data transaksi historis. Data transaksi dapat memberikan informasi mengenai pola penjualan berdasarkan waktu, produk, harga, dan karakteristik transaksi. Pemanfaatan data secara tepat memungkinkan pelaku e-commerce memperoleh informasi yang lebih baik dalam memperkirakan kebutuhan produk pada periode berikutnya.

Data e-commerce memiliki karakteristik yang sesuai dengan konsep *big data*. Data dapat memiliki volume yang besar, dihasilkan dengan kecepatan tinggi, memiliki format yang beragam, serta membutuhkan proses pengolahan untuk menghasilkan informasi yang bernilai. Karakteristik tersebut dikenal melalui konsep *volume, velocity, variety, veracity,* dan *value*. Dalam proposal penelitian ini, data e-commerce direncanakan mencapai sekitar 250 baris dengan transaksi yang berlangsung secara terus-menerus.

Selain data penjualan, metode pembayaran merupakan salah satu informasi transaksi yang dapat dianalisis. E-commerce menyediakan berbagai metode pembayaran, seperti kartu kredit atau debit, transfer bank, e-wallet, dan cash on delivery (COD). Perbedaan metode pembayaran dapat berkaitan dengan karakteristik transaksi konsumen. Oleh karena itu, metode pembayaran dapat dipertimbangkan sebagai salah satu fitur tambahan dalam model prediksi.

Penelitian terdahulu yang terdapat dalam proposal menunjukkan penggunaan XGBoost dan Random Forest dalam prediksi penjualan. Penelitian Sabrina dan Dayini (2025), misalnya, menggunakan pendekatan *ensemble learning* untuk melakukan prediksi penjualan dan evaluasi menggunakan MSE, MAE, dan MAPE. Sementara itu, penelitian mengenai pemilihan metode pembayaran menunjukkan bahwa faktor kemudahan, pengalaman, dan keamanan berkaitan dengan preferensi konsumen dalam memilih metode pembayaran (Surya, 2025).

Penelitian lain yang dibahas dalam proposal menggunakan XGBoost, LSTM, dan ARIMA untuk prediksi permintaan produk pada e-commerce. Pendekatan tersebut menunjukkan bahwa machine learning dapat digunakan untuk menangkap pola permintaan berdasarkan data transaksi dan faktor waktu.

Berdasarkan kondisi tersebut, penelitian ini berfokus pada pengembangan model prediksi permintaan produk dengan memanfaatkan data transaksi dan memasukkan metode pembayaran sebagai salah satu variabel prediktif. Model XGBoost digunakan sebagai model utama karena sesuai untuk pengolahan data tabular dan dapat menangkap hubungan kompleks antarvariabel.

Tujuan penelitian adalah mengintegrasikan data transaksi e-commerce untuk membangun model prediksi permintaan produk, menganalisis penggunaan metode pembayaran sebagai variabel dalam prediksi, serta membandingkan kinerja model dengan dan tanpa variabel metode pembayaran.

2. KAJIAN TEORITIS

E-Commerce

E-commerce merupakan aktivitas perdagangan yang dilakukan melalui media elektronik dan jaringan internet. Aktivitas tersebut mencakup proses pencarian informasi, pemilihan produk, pemesanan, pembayaran, dan transaksi antara penjual dengan konsumen. Perkembangan *e-commerce* menyebabkan semakin banyak aktivitas konsumen yang dapat direkam menjadi data digital.Data yang dihasilkan dari *e-commerce* dapat berupa data transaksi, data produk, data pelanggan, data pembayaran, serta data aktivitas pengguna. Data tersebut dapat digunakan untuk mengetahui pola transaksi dan membantu proses pengambilan keputusan bisnis (Akbar & Alam, 2020).

***Big* Data**

Big data merupakan kumpulan data dengan karakteristik volume besar, kecepatan pertumbuhan tinggi, dan keberagaman bentuk data. Dalam pengelolaan e-commerce, big data dapat berasal dari transaksi pembelian, pencarian produk, interaksi pelanggan, pembayaran, dan aktivitas lainnya. Lima karakteristik utama big data terdiri atas *volume, velocity, variety, veracity,* dan *value*. Volume berkaitan dengan jumlah data, velocity berkaitan dengan kecepatan data dihasilkan, variety berkaitan dengan bentuk data, veracity berkaitan dengan kualitas data, sedangkan value berkaitan dengan manfaat yang dapat diperoleh dari pengolahan data. (Gandomi & Haider, 2015; Kitchin, 2014).

Metode Pembayaran

Metode pembayaran merupakan mekanisme yang digunakan konsumen untuk menyelesaikan transaksi. Dalam e-commerce, metode pembayaran dapat terdiri atas kartu kredit atau debit, transfer bank, e-wallet, dan COD. Metode pembayaran dapat digunakan sebagai variabel kategorikal dalam dataset. Agar dapat diproses oleh algoritma machine

learning, variabel tersebut perlu diubah menjadi bentuk numerik melalui proses encoding (Kuhn & Johnson, 2013; James et al., 2021). Dalam penelitian ini, analisis metode pembayaran difokuskan pada hubungannya dengan volume pembelian atau permintaan produk, bukan pada aspek keamanan transaksi. Hal tersebut sesuai dengan batasan penelitian pada proposal.

Prediksi Permintaan Produk

Prediksi permintaan produk merupakan proses memperkirakan jumlah kebutuhan atau jumlah produk yang akan terjual pada periode mendatang berdasarkan data historis. Prediksi permintaan dapat digunakan untuk mendukung perencanaan persediaan. Pada penelitian ini, jumlah produk terjual digunakan sebagai variabel target. Variabel prediktor mencakup harga, waktu transaksi, produk atau kategori, metode pembayaran, serta data historis penjualan. Pemodelan prediktif pada prinsipnya memerlukan proses penyiapan data, pembagian data, pemodelan, dan pengukuran kinerja model (Kuhn & Johnson, 2013; James et al., 2021). Literatur predictive analytics menekankan bahwa prediksi perlu dibedakan dari analisis deskriptif dan dievaluasi berdasarkan tujuan prediksi yang jelas (Shmueli & Koppius, 2011).

XGBoost

XGBoost atau Extreme Gradient Boosting merupakan algoritma machine learning berbasis gradient boosting. Algoritma ini membangun model secara bertahap menggunakan sejumlah decision tree untuk menghasilkan prediksi. XGBoost dirancang sebagai sistem tree boosting yang skalabel dan dapat digunakan pada berbagai permasalahan machine learning, termasuk data tabular (Massaro et al., 2021). Dalam penelitian ini, XGBoost digunakan untuk mempelajari hubungan antara karakteristik transaksi dan jumlah produk terjual. Penggunaan XGBoost juga relevan dengan penelitian terdahulu dalam proposal yang menggunakan algoritma tersebut untuk prediksi penjualan dan prediksi permintaan produk.

Random Forest

Random Forest merupakan algoritma ensemble learning yang menggabungkan sejumlah *decision tree*. Setiap pohon dibangun berdasarkan variasi acak dan hasil model diperoleh dari penggabungan prediksi pohon-pohon tersebut. *Random Forest* dapat digunakan untuk klasifikasi maupun regresi (Yang, 2023). Random Forest digunakan dalam penelitian sebagai model pembanding terhadap XGBoost. Penggunaan Random Forest juga didukung oleh penelitian terdahulu yang menggunakannya bersama XGBoost dalam prediksi penjualan.

Long Short-Term Memory

Long Short-Term Memory (LSTM) merupakan pengembangan dari Recurrent Neural Network yang dirancang untuk menangani dependensi pada data sekuensial dan

mempertahankan informasi dalam interval waktu yang panjang (Guo, 2022). Dalam proposal, LSTM direncanakan digunakan untuk menangkap pola permintaan berdasarkan urutan waktu.

Evaluasi Model

Evaluasi digunakan untuk mengetahui tingkat kesalahan antara nilai aktual dan nilai hasil prediksi. Penelitian ini menggunakan RMSE, MAE, dan MAPE. RMSE digunakan untuk mengukur akar dari rata-rata kuadrat kesalahan, MAE mengukur rata-rata nilai absolut selisih antara nilai aktual dan prediksi, sedangkan MAPE mengukur rata-rata persentase kesalahan prediksi (Hyndman & Athanasopoulos, 2021; Petropoulos et al., 2022). Ketiga metrik dapat memberikan informasi yang berbeda tentang kesalahan prediksi, sehingga interpretasi kinerja sebaiknya tidak hanya bergantung pada satu ukuran (Hyndman & Athanasopoulos, 2021; Petropoulos et al., 2022).

3. METODE PENELITIAN

Jenis Penelitian

Penelitian menggunakan pendekatan kuantitatif dengan desain deskriptif dan prediktif. Pendekatan tersebut digunakan untuk menganalisis data historis transaksi dan metode pembayaran serta membangun model prediksi permintaan produk. Proses penelitian mengikuti tahapan umum pemodelan prediktif, yaitu penyiapan data, pembagian data, pemodelan, dan evaluasi kinerja (Kuhn & Johnson, 2013; James et al., 2021). Proposal secara khusus menyebutkan bahwa penelitian berfokus pada hubungan antara COD, e-wallet, kartu kredit/debit, dan transfer bank dengan volume permintaan produk.

Sumber Data

Data penelitian berupa data transaksi e-commerce yang terdiri atas informasi tanggal transaksi, produk, kategori, harga, jumlah produk terjual, dan metode pembayaran. Proposal merencanakan penggunaan data historis dalam periode tertentu dan menyebutkan sumber data dapat berasal dari log transaksi, perilaku penelusuran atau clickstream, serta detail pembayaran. Data transaksi e-commerce dapat merepresentasikan informasi mengenai proses pemesanan dan transaksi digital, sedangkan bentuk data yang digunakan dalam penelitian perlu disesuaikan dengan tujuan analisis (OECD, 2025). Untuk tahap pengujian awal, dataset dapat dibuat dalam jumlah terbatas. Namun, untuk penelitian akhir, data harus berasal dari dataset penelitian yang dapat dipertanggungjawabkan.

Variabel Penelitian

Variabel penelitian terdiri atas variabel target dan variabel prediktor.

Tabel 1. Variabel Penelitian.

Variabel	Jenis	Keterangan
Tanggal transaksi	Prediktor	Waktu transaksi
Produk	Prediktor	Identitas produk
Kategori	Prediktor	Kelompok produk
Harga	Prediktor	Harga produk
Metode pembayaran	Prediktor utama	E-wallet, COD, transfer, kartu
Jumlah terjual	Target	Jumlah produk yang terjual

Tahapan Penelitian

Tahapan penelitian terdiri atas identifikasi dataset dan studi literatur, pengumpulan data, preprocessing, feature engineering, pemodelan, evaluasi, dan interpretasi hasil. Tahapan tersebut sejalan dengan proses umum predictive modeling yang menempatkan preprocessing, data splitting, pemodelan, dan evaluasi sebagai bagian penting dalam pengembangan model (Kuhn & Johnson, 2013; James et al., 2021).

1) Identifikasi Dataset dan Studi Literatur

   Tahap awal dilakukan dengan menentukan data yang diperlukan serta mengkaji penelitian terdahulu yang relevan dengan prediksi permintaan produk, big data, machine learning, dan metode pembayaran.

2) Pengumpulan Dataset

   Data transaksi dikumpulkan dan disusun dalam satu dataset. Data dapat berasal dari database atau dataset yang tersedia untuk penelitian.

3) Preprocessing

   Preprocessing meliputi pemeriksaan data kosong, data duplikat, kesalahan format, dan ketidaksesuaian data. Data transaksi yang tidak relevan dapat dihapus agar tidak memengaruhi model.

4) Feature Engineering

   Feature engineering dilakukan untuk membentuk fitur yang dapat digunakan dalam model. Fitur waktu dapat berupa tahun, bulan, minggu, dan hari. Metode pembayaran diubah menjadi fitur numerik menggunakan *encoding*.

5) Pemodelan

   XGBoost digunakan sebagai model utama. Random Forest dan LSTM dapat digunakan sebagai model pembanding sesuai kebutuhan penelitian.

6) Evaluasi

   Model dievaluasi dengan membandingkan hasil prediksi terhadap nilai aktual menggunakan RMSE, MAE, dan MAPE.

7) Interpretasi

   Hasil evaluasi digunakan untuk mengetahui kinerja model serta menganalisis apakah metode pembayaran memberikan tambahan informasi terhadap prediksi permintaan.

Metode Analisis Data

Analisis dilakukan melalui dua tahap utama, yaitu analisis deskriptif dan analisis prediktif. Analisis deskriptif digunakan untuk mengetahui karakteristik data transaksi dan distribusi metode pembayaran. Analisis prediktif dilakukan menggunakan XGBoost. Pemisahan tahap deskriptif dan prediktif membantu memastikan bahwa karakteristik awal data dipahami sebelum kinerja model dinilai (Kuhn & Johnson, 2013; James et al., 2021). Prinsip analitik data juga menekankan pentingnya memahami karakteristik data sebelum memilih dan mengevaluasi model prediktif (Miah & Islam, 2024).

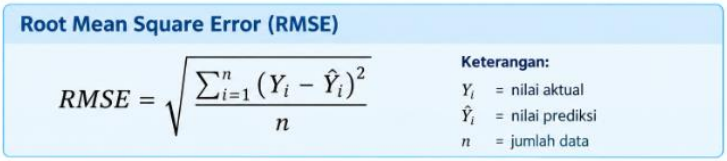
Dua skenario pemodelan digunakan:

1) Model 1: XGBoost tanpa variabel metode pembayaran.
2) Model 2: XGBoost dengan variabel metode pembayaran.

Perbandingan kedua model digunakan untuk mengetahui perubahan kinerja prediksi setelah metode pembayaran ditambahkan sebagai fitur. Perbandingan model berdasarkan metrik kesalahan memungkinkan perubahan kinerja dievaluasi secara kuantitatif; interpretasinya perlu mempertimbangkan bahwa setiap metrik memiliki karakteristik dan keterbatasan yang berbeda (Hyndman & Athanasopoulos, 2021; Petropoulos et al., 2022).

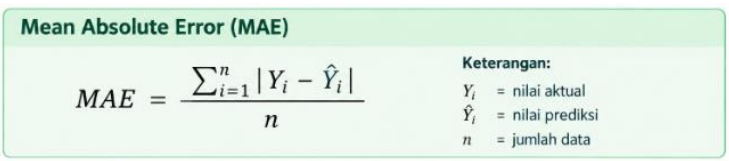
Teknik Evaluasi

1) RMSE dihitung dengan rumus:



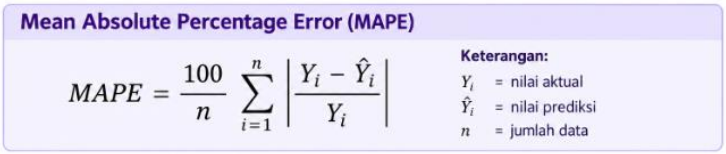
Gambar 1. Rumus RMSE.

2) MAE dihitung dengan rumus:



Gambar 2. Rumus MAE.

3) MAPE dihitung dengan rumus:



Gambar 3. Rumus MAPE.

4. HASIL DAN PEMBAHASAN

Deskripsi Dataset

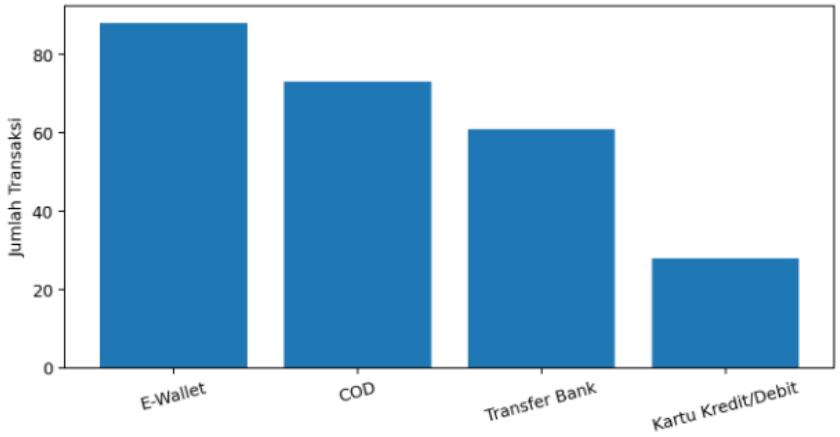
Dataset yang digunakan dalam penelitian ini terdiri atas 250 transaksi. Setiap transaksi memiliki empat variabel utama, yaitu tanggal transaksi, metode pembayaran, harga, dan jumlah produk terjual. Variabel jumlah produk terjual digunakan sebagai variabel target yang akan diprediksi, sedangkan harga, waktu transaksi, dan metode pembayaran digunakan sebagai informasi yang dapat membantu proses prediksi. Metode pembayaran dalam dataset dikelompokkan menjadi empat kategori, yaitu E-Wallet, COD, Transfer Bank, dan Kartu Kredit/Debit. Pengelompokan ini diperlukan agar metode pembayaran dapat dianalisis sebagai karakteristik transaksi dan, pada Model 2, digunakan sebagai fitur tambahan dalam pemodelan XGBoost. Dari 250 transaksi tersebut, sebanyak 88 transaksi menggunakan E-Wallet, 73 transaksi menggunakan COD, 61 transaksi menggunakan Transfer Bank, dan 28 transaksi menggunakan Kartu Kredit/Debit. Distribusi tersebut menunjukkan bahwa setiap kategori metode pembayaran memiliki jumlah transaksi yang berbeda.

Tabel 2. Variabel yang digunakan dalam dataset.

Variabel	Keterangan
Tanggal	Tanggal terjadinya transaksi
Metode Pembayaran	E-Wallet, COD, Transfer Bank, Kartu Kredit/Debit
Harga	Nilai harga transaksi/produk
Jumlah Produk Terjual	Jumlah produk yang terjual pada setiap transaksi dan menjadi variabel target

Berdasarkan Tabel 2, dataset yang digunakan terdiri atas empat variabel utama, yaitu tanggal, metode pembayaran, harga, dan jumlah produk terjual. Variabel tanggal digunakan untuk menunjukkan waktu terjadinya transaksi, sedangkan metode pembayaran menunjukkan jenis pembayaran yang digunakan pelanggan. Variabel harga menunjukkan nilai harga pada transaksi. Sementara itu, jumlah produk terjual merupakan variabel target yang akan diprediksi menggunakan algoritma XGBoost. Dengan demikian, keempat variabel tersebut digunakan

untuk memberikan gambaran karakteristik transaksi sekaligus mendukung proses analisis dan pemodelan.



Gambar 4. Distribusi metode pembayaran pada 250 transaksi

Berdasarkan Gambar 4, dapat dilihat distribusi penggunaan metode pembayaran pada 250 transaksi yang dianalisis. Metode pembayaran yang digunakan terdiri dari E-Wallet, COD, Transfer Bank, dan Kartu Kredit/Debit. Tinggi rendahnya batang pada grafik menunjukkan jumlah transaksi pada masing-masing metode pembayaran. Perbedaan jumlah transaksi tersebut menunjukkan bahwa tingkat penggunaan setiap metode pembayaran oleh pelanggan tidak sama. Informasi mengenai distribusi ini penting karena metode pembayaran selanjutnya digunakan sebagai salah satu informasi yang diuji dalam proses pemodelan untuk mengetahui apakah keberadaannya dapat membantu memprediksi jumlah produk terjual.

Analisis Metode Pembayaran

Analisis awal dilakukan untuk melihat apakah terdapat perbedaan jumlah produk terjual berdasarkan metode pembayaran. Analisis ini bersifat deskriptif, sehingga tujuannya adalah menggambarkan pola yang terdapat pada data sebelum dilakukan pemodelan. Dengan demikian, perbedaan rata-rata yang terlihat belum dapat digunakan untuk menyatakan bahwa metode pembayaran menyebabkan kenaikan atau penurunan penjualan.

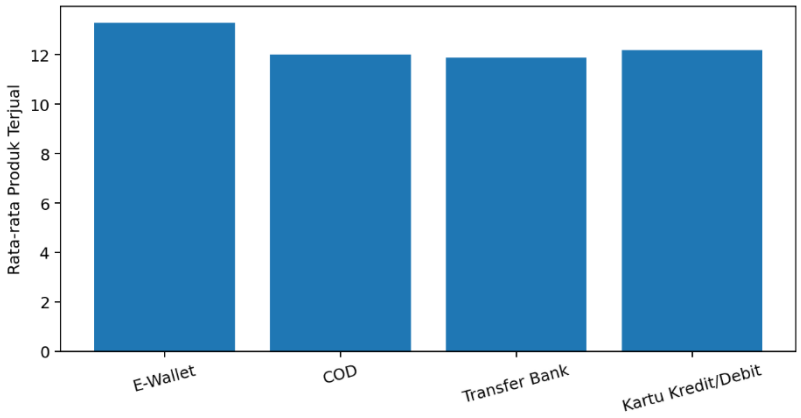
Tabel 3. Distribusi transaksi dan rata-rata produk terjual.

Metode Pembayaran	Jumlah Transaksi	Rata-rata Produk Terjual
E-Wallet	88	13.28
COD	73	12.00
Transfer Bank	61	11.89
Kartu Kredit/Debit	28	12.18

Berdasarkan Tabel 3, terlihat bahwa jumlah transaksi pada masing-masing metode pembayaran tidak sama. Perbedaan jumlah transaksi tersebut menunjukkan bahwa terdapat

variasi penggunaan metode pembayaran oleh pelanggan. Selain itu, terdapat perbedaan rata-rata jumlah produk yang terjual pada setiap metode pembayaran. Nilai rata-rata tersebut digunakan untuk melihat pola awal hubungan antara metode pembayaran dan jumlah produk terjual. Namun, perbedaan rata-rata belum dapat digunakan untuk menyimpulkan adanya hubungan sebab-akibat karena jumlah produk terjual juga dapat dipengaruhi oleh faktor lain seperti harga, waktu transaksi, promosi, dan kondisi pasar

.



Gambar 5. Rata-rata jumlah produk terjual berdasarkan metode pembayaran.

Berdasarkan Gambar 5. terlihat adanya perbedaan rata-rata jumlah produk terjual berdasarkan metode pembayaran yang digunakan. Setiap batang pada grafik menunjukkan rata-rata produk yang terjual pada masing-masing metode pembayaran. Batang yang lebih tinggi menunjukkan rata-rata jumlah produk terjual yang lebih besar, sedangkan batang yang lebih rendah menunjukkan rata-rata yang lebih kecil. Perbedaan tersebut memberikan gambaran awal bahwa karakteristik transaksi dapat berbeda berdasarkan metode pembayaran. Namun, hasil ini masih bersifat deskriptif sehingga belum dapat digunakan untuk menyatakan bahwa metode pembayaran secara langsung menyebabkan perbedaan jumlah produk terjual. Oleh karena itu, diperlukan proses pemodelan untuk menguji apakah informasi metode pembayaran dapat memberikan tambahan informasi dalam melakukan prediksi.

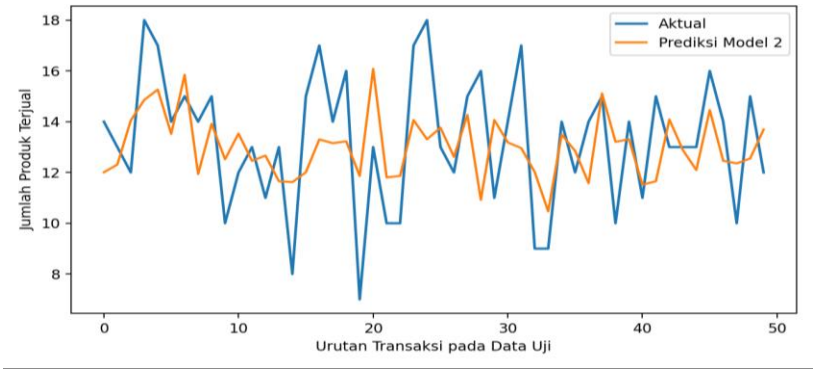
Pemodelan dengan XGBoost

Pemodelan dilakukan menggunakan algoritma XGBoost untuk memprediksi jumlah produk terjual. Untuk mengetahui apakah metode pembayaran memberikan informasi tambahan bagi model, digunakan dua skenario. Model 1 tidak menggunakan metode pembayaran sebagai fitur, sedangkan Model 2 menambahkan metode pembayaran sebagai fitur.

Tabel 4. Skenario pemodelan.

Model	Fitur yang Digunakan	Tujuan
Model 1	Harga, hari, dan bulan transaksi	Menjadi model pembanding tanpa informasi metode pembayaran
Model 2	Harga, hari, bulan, dan metode pembayaran	Mengukur perubahan kinerja setelah metode pembayaran ditambahkan

Dataset diurutkan berdasarkan tanggal transaksi. Sebanyak 80% data awal digunakan sebagai data latih dan 20% data terakhir digunakan sebagai data uji. Dengan 250 transaksi, pembagian tersebut menghasilkan 200 transaksi untuk pelatihan dan 50 transaksi untuk pengujian. Pembagian berdasarkan urutan waktu digunakan agar evaluasi lebih mendekati kondisi prediksi transaksi berikutnya, bukan mencampurkan data masa depan ke dalam data pelatihan. Pada Model 2, variabel metode pembayaran yang bersifat kategorikal diubah menjadi beberapa variabel indikator agar dapat diproses oleh algoritma. Setelah model dilatih menggunakan data latih, model menghasilkan prediksi jumlah produk terjual untuk 50 transaksi pada data uji.



Gambar 6. Perbandingan nilai aktual dan prediksi pada data uji Model 2.

Gambar 6. memperlihatkan hubungan antara jumlah produk terjual aktual dan hasil prediksi. Garis aktual menunjukkan kondisi penjualan sebenarnya pada data uji, sedangkan garis prediksi menunjukkan nilai yang diperkirakan oleh Model 2. Semakin dekat nilai prediksi dengan nilai aktual, semakin kecil kesalahan prediksi yang dihasilkan model.

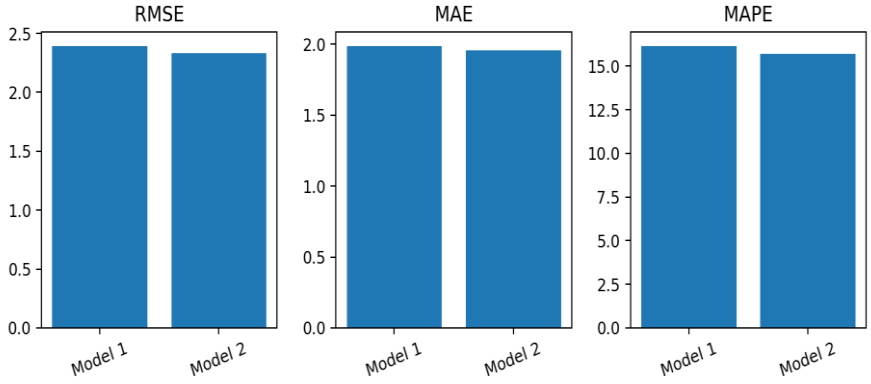
Hasil Evaluasi Model

Kinerja model dievaluasi menggunakan tiga metrik, yaitu RMSE (Root Mean Square Error), MAE (Mean Absolute Error), dan MAPE (Mean Absolute Percentage Error). RMSE memberikan penalti lebih besar terhadap kesalahan yang besar, MAE menunjukkan rata-rata besarnya kesalahan absolut, sedangkan MAPE menunjukkan rata-rata kesalahan dalam bentuk persentase.

Tabel 5. Hasil evaluasi Model 1 dan Model 2.

Model	RMSE	MAE	MAPE (%)
Model 1	2.395	1.989	16.17
Model 2	2.330	1.956	15.70

Hasil menunjukkan bahwa RMSE Model 2 sebesar 2.330, sedangkan Model 1 sebesar 2.395. Artinya, setelah metode pembayaran ditambahkan, RMSE berubah sebesar 2.74% (menurun). Nilai MAE Model 2 sebesar 1.956, dibandingkan dengan 1.989 pada Model 1. Perubahan ini menunjukkan bahwa rata-rata kesalahan absolut menurun sekitar 1.65% setelah fitur metode pembayaran ditambahkan. Untuk MAPE, Model 1 menghasilkan 16.17% dan Model 2 menghasilkan 15.70%. Pada MAPE menurun sekitar 2.90%. Karena ketiga metrik dapat memberikan gambaran yang berbeda, interpretasi kinerja model sebaiknya tidak hanya didasarkan pada satu metrik.



Gambar 7. Perbandingan metrik evaluasi Model 1 dan Model 2.

Berdasarkan Gambar 7, dapat dilihat perbandingan kinerja Model 1 dan Model 2 berdasarkan nilai RMSE, MAE, dan MAPE. Model 1 merupakan model tanpa menggunakan metode pembayaran, sedangkan Model 2 merupakan model yang menambahkan metode pembayaran sebagai fitur. Pada data Model 2 menghasilkan RMSE sebesar 2,330, MAE sebesar 1,956, dan MAPE sebesar 15,70%, sedangkan Model 1 menghasilkan RMSE sebesar 2,395, MAE sebesar 1,989, dan MAPE sebesar 16,17%. Nilai yang lebih rendah pada ketiga metrik menunjukkan bahwa tingkat kesalahan prediksi Model 2 lebih kecil pada dataset. Dengan demikian, secara visual Gambar 4.4 memperlihatkan perubahan kinerja model setelah metode pembayaran ditambahkan sebagai fitur.

Pembahasan

Hasil analisis menunjukkan bahwa metode pembayaran dapat diperlakukan sebagai salah satu informasi transaksi yang diuji untuk membantu prediksi jumlah produk terjual.

Analisis deskriptif terlebih dahulu menunjukkan adanya perbedaan rata-rata jumlah produk terjual pada empat metode pembayaran. Selanjutnya, informasi tersebut diuji dalam model XGBoost melalui perbandingan Model 1 dan Model 2.

Pada data 250 transaksi ini, penambahan metode pembayaran menghasilkan RMSE sebesar 2.330, MAE sebesar 1.956, dan MAPE sebesar 15.70%. Hasil tersebut perlu dibaca bersama hasil Model 1 karena tujuan utama perbandingan adalah mengetahui apakah penambahan metode pembayaran benar-benar memberikan perubahan pada kemampuan prediksi.

Jika pada dataset penelitian sebenarnya Model 2 menghasilkan RMSE, MAE, dan MAPE yang lebih rendah secara konsisten dibandingkan Model 1, maka metode pembayaran dapat dibahas sebagai fitur yang memberikan informasi tambahan bagi prediksi. Sebaliknya, apabila metrik Model 2 tidak lebih baik, maka penambahan metode pembayaran belum menunjukkan peningkatan kinerja model pada dataset tersebut.

Hasil prediksi dapat digunakan sebagai informasi pendukung untuk perencanaan persediaan, terutama dalam memperkirakan kebutuhan produk berdasarkan pola transaksi. Namun, hasil prediksi tidak sebaiknya digunakan sebagai satu-satunya dasar keputusan. Faktor lain seperti promosi, perubahan harga, musim, tren permintaan, ketersediaan produk, dan kondisi pasar juga perlu dipertimbangkan.

Secara keseluruhan, tahapan analisis dalam penelitian ini dapat dipahami melalui urutan: data transaksi dikumpulkan, data dideskripsikan, metode pembayaran dianalisis, dua skenario XGBoost dibangun, kinerja model dievaluasi, kemudian hasilnya dibandingkan untuk mengetahui kontribusi informasi metode pembayaran terhadap prediksi jumlah produk terjual.

5. KESIMPULAN DAN SARAN

Kesimpulan

Hasil penelitian pada dataset menunjukkan bahwa model XGBoost dapat digunakan untuk memprediksi jumlah produk terjual dengan memanfaatkan informasi harga dan waktu transaksi. Perbandingan dua skenario menunjukkan bahwa Model 2, yang menambahkan metode pembayaran sebagai fitur, menghasilkan RMSE sebesar 2,330, MAE sebesar 1,956, dan MAPE sebesar 15,70%, sedangkan Model 1 menghasilkan RMSE sebesar 2,395, MAE sebesar 1,989, dan MAPE sebesar 16,17%. Dengan demikian, pada dataset yang digunakan dalam artikel ini, penambahan metode pembayaran diikuti oleh penurunan nilai ketiga metrik kesalahan. Temuan ini mendukung penggunaan perbandingan model sebagai cara untuk menilai apakah suatu fitur memberikan informasi tambahan terhadap prediksi, tetapi belum

dapat digeneralisasikan sebagai bukti bahwa metode pembayaran selalu meningkatkan akurasi pada dataset e-commerce yang berbeda.

Saran

Penelitian selanjutnya perlu menggunakan dataset transaksi nyata dengan jumlah observasi yang lebih besar dan periode pengamatan yang memadai agar hasil evaluasi lebih representatif. Pengujian juga dapat diperluas dengan membandingkan XGBoost dengan Random Forest, ARIMA/SARIMA, atau LSTM sesuai karakteristik data, serta menggunakan validasi deret waktu dan penyetelan hyperparameter untuk mengurangi risiko evaluasi yang bias. Variabel tambahan seperti promosi, diskon, musim, kategori produk, dan ketersediaan stok dapat dipertimbangkan karena permintaan dipengaruhi oleh lebih dari satu faktor. Selain itu, penelitian berikutnya dapat menggunakan metode interpretabilitas model untuk mengetahui kontribusi masing-masing fitur terhadap hasil prediksi.

DAFTAR REFERENSI

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. SAGE Publications.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer. https://doi.org/10.1007/978-1-4614-6849-3

Laudon, K. C., & Traver, C. G. (2021). E-commerce 2021–2022: Business, technology, society. Pearson.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

OECD. (2025). The 2025 OECD definition of e-commerce and guidelines for interpretation. OECD Publishing. https://doi.org/10.1787/2254f1de-en

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., et al. (2022). Forecasting: Theory and practice. International Journal of Forecasting, 38(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. MIS Quarterly, 35(3), 553–572.

Turban, E., Outland, J., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2018). Electronic commerce 2018: A manager's guide to e-business. Springer.

Zheng, A., & Casari, A. (2018). Feature engineering for machine learning: Principles and techniques for data scientists. O'Reilly Media.