



## Implementasi Pipeline ETL dan Pemodelan Prediktif ARIMA dalam Memetakan Pola Pembelian Konsumen pada Dataset *Marketplace*

I Wayan Manik Mas Sri Dantya<sup>1</sup>, I Wayan Sudiarsa<sup>2\*</sup>, I Putu Kabinawa Raesa Putra<sup>3</sup>,  
Brian Adi Sapurta<sup>4</sup>, I Komang Hari Sastrawan<sup>5</sup>

<sup>1-5</sup>Program Studi Rekayasa Sistem Komputer, Institut Bisnis dan Teknologi Indonesia (INSTIKI), Indonesia

\*Penulis korespondensi: [sudiarsa@instiki.ac.id](mailto:sudiarsa@instiki.ac.id)

**Abstract.** *In the rapidly evolving digital economy, the ability to anticipate transaction surges is a strategic asset for marketplace platforms to maintain operational efficiency. This research aims to build an accurate daily transaction volume forecasting system thru the implementation of an Extract, Transform, and Load (ETL) pipeline and Autoregressive Integrated Moving Average (ARIMA) predictive modeling. The dataset used is sourced from dataset *olshop.csv*, which includes transaction history for the entire year of 2025. The ETL stage focused on data cleaning and handling missing values, while time series analysis began with the Augmented Dickey-Fuller (ADF) stationarity test, which yielded a significant p-value of 0.000006. The parameter model was optimized using the *auto\_arima* algorithm, which determined the ARIMA(2,0,0) configuration as the best model. The evaluation results of the model show fairly stable performance with a Root Mean Squared Error (RMSE) value of 2.002 and a Mean Absolute Error (MAE) of 1.704 on the test data. Research findings reveal a consistently higher purchasing pattern during the mid-month and end-of-month periods, with an average of 5.52 daily transactions, compared to the beginning of the month, which saw 5.48 transactions. The 30-day forecast results provide valuable insights for online store managers to proactively adjust inventory and logistics workforce allocation strategies. This research concludes that integrating data engineering techniques and statistical analysis can provide predictive solutions for the dynamics of the digital market.*

**Keywords:** ARIMA; Data Engineering; ETL; Forecasting; Marketplace.

**Abstrak.** Dalam era ekonomi digital yang berkembang sangat pesat, kemampuan untuk mengantisipasi lonjakan transaksi merupakan aset strategis bagi platform marketplace guna menjaga efisiensi operasional. Penelitian ini bertujuan untuk membangun sistem peramalan jumlah transaksi harian yang akurat melalui implementasi pipeline Extract, Transform, and Load (ETL) dan pemodelan prediktif Autoregressive Integrated Moving Average (ARIMA). Dataset yang digunakan bersumber dari dataset *olshop.csv* yang mencakup riwayat transaksi sepanjang tahun 2025. Tahapan ETL difokuskan pada pembersihan data waktu dan penanganan nilai hilang, sementara analisis deret waktu diawali dengan uji stasioneritas Augmented Dickey-Fuller (ADF) yang menghasilkan p-value signifikan sebesar 0,000006. Parameter model dioptimalkan menggunakan algoritma *auto\_arima* yang menetapkan konfigurasi ARIMA(2,0,0) sebagai model terbaik. Hasil evaluasi model menunjukkan performa yang cukup stabil dengan nilai Root Mean Squared Error (RMSE) sebesar 2,002 dan Mean Absolute Error (MAE) sebesar 1,704 pada data pengujian. Temuan penelitian mengungkapkan adanya pola pembelian yang lebih tinggi secara konsisten pada periode pertengahan bulan dan akhir bulan dengan rata-rata 5,52 transaksi harian, dibandingkan awal bulan sebesar 5,48 transaksi. Hasil peramalan untuk 30 hari ke depan memberikan wawasan penting bagi pengelola toko online dalam melakukan penyesuaian inventaris dan strategi alokasi tenaga logistik secara proaktif. Penelitian ini menyimpulkan bahwa integrasi teknik data engineering dan analisis statistika mampu memberikan solusi prediktif bagi dinamika pasar digital.

**Kata kunci:** ARIMA; Data Engineering; ETL; Marketplace; Peramalan.

### 1. LATAR BELAKANG

Perkembangan pesat marketplace digital telah menghasilkan volume data transaksi yang sangat besar dan kompleks, sehingga menuntut pendekatan analitik yang sistematis untuk mengekstraksi informasi bernilai dari data tersebut (Provost & Fawcett, 2013). Data transaksi yang tersimpan secara historis tidak hanya mencerminkan aktivitas pembelian konsumen, tetapi juga mengandung pola temporal yang dapat dimanfaatkan untuk perencanaan bisnis dan strategi pemasaran (Han et al., 2022). Namun, data mentah pada umumnya belum siap

digunakan secara langsung untuk analisis karena masih mengandung inkonsistensi, nilai hilang, serta format yang tidak terstruktur. Oleh karena itu, diperlukan proses pengolahan data yang terstandarisasi melalui pipeline ETL agar data menjadi reliabel dan siap dianalisis. Implementasi ETL yang tepat akan meningkatkan kualitas analisis dan validitas hasil penelitian (Kimball & Ross, 2016).

Selain tantangan pengolahan data, pemahaman terhadap pola waktu transaksi konsumen menjadi isu penting dalam konteks e-commerce modern (Chen et al., 2021). Pola pembelian konsumen sering kali dipengaruhi oleh faktor temporal seperti periode gaji, promosi bulanan, dan perilaku konsumsi musiman. Analisis berbasis waktu memungkinkan peneliti dan praktisi untuk mengidentifikasi kecenderungan transaksi pada periode tertentu dalam satu bulan. Dengan memahami distribusi transaksi harian, perusahaan dapat mengoptimalkan strategi operasional dan pemasaran secara lebih presisi (Fildes et al., 2019).

Pemodelan deret waktu merupakan salah satu pendekatan statistik yang banyak digunakan untuk menganalisis dan meramalkan data berbasis waktu (Box et al., 2016). Model ARIMA secara luas diterapkan karena kemampuannya dalam menangkap dependensi temporal jangka pendek serta kemudahannya dalam interpretasi. Dalam konteks transaksi marketplace, ARIMA dapat digunakan untuk meramalkan volume transaksi harian dan mengidentifikasi pola fluktuasi yang signifikan. Penggunaan ARIMA menjadi relevan ketika data menunjukkan sifat stasioner atau dapat distasionerkan melalui transformasi tertentu (Hyndman & Athanasopoulos, 2021).

Meskipun sejumlah penelitian telah menerapkan ARIMA untuk peramalan penjualan, masih terbatas studi yang mengintegrasikan pipeline ETL secara eksplisit dengan analisis pola transaksi bulanan pada dataset marketplace (Rahman et al., 2023). Penelitian ini bertujuan untuk mengisi celah tersebut dengan mengimplementasikan proses ETL yang terstruktur dan memanfaatkan model ARIMA untuk memetakan serta meramalkan pola pembelian konsumen. Fokus penelitian diarahkan pada identifikasi preferensi transaksi konsumen di awal, pertengahan, dan akhir bulan sebagai dasar pengambilan keputusan strategis berbasis data.

## **2. KAJIAN TEORITIS**

Data Engineering merupakan praktik desain dan pembangunan sistem untuk pengumpulan, penyimpanan, dan analisis data dalam skala besar. Salah satu fondasi utama dalam Data Engineering adalah pipeline ETL (Extract, Transform, Load). Menurut Smith (2020), ETL adalah proses yang memungkinkan organisasi untuk mengumpulkan data dari berbagai sumber mentah, mengubahnya menjadi format yang konsisten, dan memuatnya ke

dalam sistem target untuk dianalisis. Dalam konteks marketplace, ETL berperan penting dalam mengubah log transaksi yang tidak terstruktur menjadi data deret waktu yang bersih.

Analisis deret waktu adalah metode statistik yang digunakan untuk menganalisis urutan titik data yang dikumpulkan selama interval waktu tertentu (Box et al., 2016). Salah satu metode yang populer adalah ARIMA, yang menggabungkan komponen *Auto-Regressive*, *Integrated*, dan *Moving Average* untuk menangkap pola data (Hyndman & Athanasopoulos, 2021). Dalam konteks *e-commerce*, pemahaman perilaku konsumen melalui data historis menjadi krusial untuk strategi inventaris (Chen et al., 2021).

Metode ARIMA adalah salah satu model statistik yang paling populer untuk peramalan deret waktu univariat. Model ini terdiri dari tiga komponen utama:

- 1) Autoregressive (AR): Menunjukkan hubungan ketergantungan antara observasi saat ini dengan sejumlah observasi sebelumnya.
- 2) Integrated (I): Menunjukkan tingkat perbedaan (differencing) yang diterapkan pada data mentah agar menjadi stasioner.
- 3) Moving Average (MA): Menggunakan kesalahan prediksi sebelumnya dalam model regresi.

Model ARIMA dinyatakan dalam notasi  $(p, d, q)$ , di mana  $p$  adalah orde dari AR,  $d$  adalah derajat differencing, dan  $q$  adalah orde dari MA. Pemilihan parameter yang optimal biasanya dilakukan melalui kriteria informasi seperti Akaike Information Criterion (AIC).

Peramalan Transaksi Marketplace Peramalan dalam industri e-commerce bertujuan untuk memprediksi volume permintaan atau transaksi di masa depan berdasarkan data historis. Dengan mengetahui waktu lonjakan pembelian (peak season), manajemen dapat mengantisipasi kebutuhan logistik, mengatur strategi pengadaan stok, serta mengoptimalkan distribusi tenaga kerja pada bagian pengiriman. Peramalan yang akurat meminimalkan risiko keterlambatan pengiriman yang sering terjadi saat beban transaksi melampaui kapasitas operasional normal.

### 3. METODE PENELITIAN

#### Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis deret waktu (time series analysis). Pendekatan ini dipilih karena data yang dianalisis berupa urutan transaksi harian yang tersusun berdasarkan waktu. Tujuan utama penelitian adalah untuk memetakan pola pembelian konsumen dan melakukan peramalan jumlah transaksi harian pada marketplace menggunakan model ARIMA.

## Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari file dataset\_olshop.csv. Dataset tersebut berisi riwayat transaksi marketplace selama tahun 2025, yang mencakup informasi waktu transaksi (timestamp), pengguna, kategori produk, harga, jumlah pembelian, dan metode pembayaran. Data ini digunakan sebagai dasar pembentukan deret waktu transaksi harian.

## Tahapan Extract, Transform, Load (ETL)

Tahap ETL dilakukan untuk memastikan data yang digunakan memiliki kualitas yang baik dan siap dianalisis.

### *Extract*

Pada tahap ini, data diekstraksi dari file CSV dan dimuat ke dalam sistem menggunakan pustaka pandas. Seluruh data transaksi dibaca ke dalam DataFrame sebagai bentuk awal pengolahan data.

### *Transform*

Tahap transformasi merupakan inti dari proses pengolahan data, yang meliputi beberapa langkah berikut:

- 1) Mengonversi kolom timestamp ke format datetime agar dapat dianalisis berdasarkan waktu.
- 2) Mengekstraksi informasi hari dalam sebulan dari setiap transaksi.
- 3) Melakukan agregasi data untuk memperoleh jumlah transaksi per hari.
- 4) Mengelompokkan hari transaksi ke dalam tiga periode bulanan, yaitu:
  - a) Awal bulan (hari ke-1 sampai ke-10)
  - b) Pertengahan bulan (hari ke-11 sampai ke-20)
  - c) Akhir bulan (hari ke-21 sampai ke-31)

Transformasi ini bertujuan untuk membentuk data deret waktu yang merepresentasikan pola transaksi harian konsumen.

### *Load*

Data hasil transformasi kemudian dimuat sebagai deret waktu transaksi harian, yang digunakan sebagai input utama dalam analisis statistik dan pemodelan ARIMA

## Analisis Deret Waktu

Analisis deret waktu diawali dengan visualisasi jumlah transaksi harian untuk mengamati pola umum dan fluktuasi data. Selanjutnya, dilakukan uji stasioneritas menggunakan metode Augmented Dickey-Fuller (ADF) untuk memastikan data memenuhi asumsi dasar pemodelan ARIMA.

Hasil uji ADF menunjukkan nilai p-value lebih kecil dari 0,05, sehingga data dinyatakan stasioner dan tidak memerlukan proses differencing.

### **Pemodelan ARIMA**

Pemodelan dilakukan menggunakan metode Autoregressive Integrated Moving Average (ARIMA). Penentuan parameter optimal (p, d, q) dilakukan secara otomatis menggunakan algoritma `auto_arma` berdasarkan kriteria Akaike Information Criterion (AIC).

Berdasarkan hasil optimasi, model terbaik yang diperoleh adalah ARIMA(2,0,0).

### **Pembagian Data dan Pelatihan Model**

Data deret waktu dibagi menjadi dua bagian, yaitu:

- 1) Data pelatihan (training) sebanyak 332 data
- 2) Data pengujian (testing) sebanyak 30 data

Model ARIMA dilatih menggunakan data pelatihan untuk mempelajari pola transaksi historis, kemudian diuji menggunakan data pengujian untuk menilai kemampuan prediksi model.

### **Evaluasi Kinerja Model**

Evaluasi kinerja model dilakukan menggunakan dua metrik kesalahan, yaitu:

- 1) Root Mean Squared Error (RMSE)
- 2) Mean Absolute Error (MAE)

Kedua metrik ini digunakan untuk mengukur perbedaan antara nilai transaksi aktual dan nilai hasil prediksi model. Nilai RMSE dan MAE yang rendah menunjukkan bahwa model memiliki tingkat akurasi yang baik.

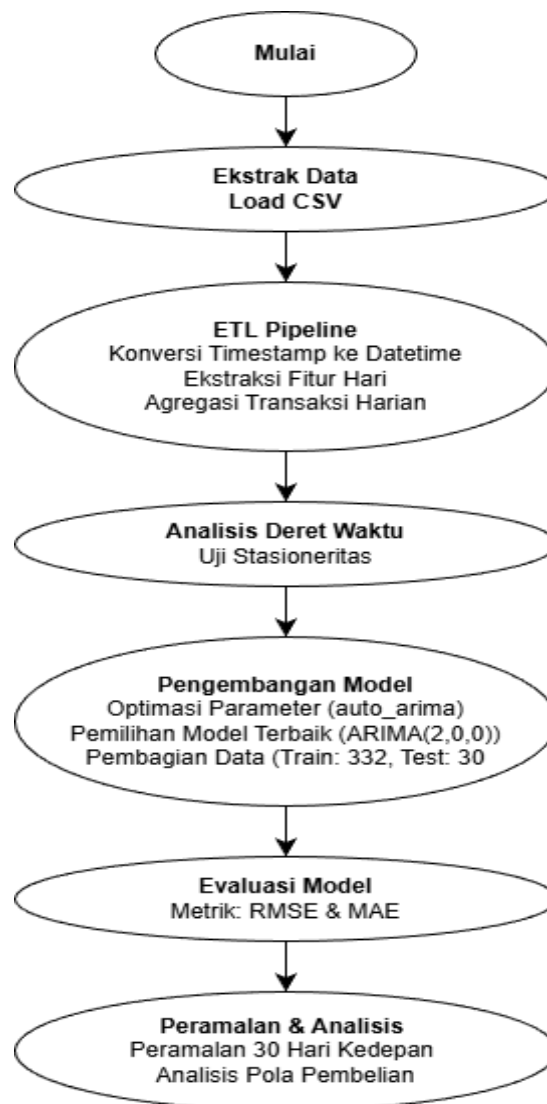
### **Peramalan dan Analisis Pola Pembelian**

Model ARIMA yang telah divalidasi kemudian digunakan untuk melakukan peramalan jumlah transaksi selama 30 hari ke depan. Hasil peramalan dianalisis berdasarkan periode bulanan untuk mengetahui kecenderungan waktu transaksi konsumen, yaitu pada awal, pertengahan, atau akhir bulan.

### **Kesimpulan Metodologis**

Seluruh tahapan metode penelitian disusun secara sistematis, dimulai dari pengolahan data melalui pipeline ETL, analisis deret waktu, pemodelan ARIMA, hingga peramalan dan analisis pola pembelian konsumen. Metode ini memastikan bahwa hasil penelitian bersifat objektif, terukur, dan dapat dipertanggungjawabkan secara ilmiah.

Analisis ini bertujuan untuk memberikan wawasan strategis bagi pengelola marketplace dalam perencanaan stok dan strategi pemasaran.



Bagan 1. Alur Metode Penelitian (ETL – ARIMA).

## 4. HASIL DAN PEMBAHASAN

### Agregasi dan Karakteristik Data

Table 1. Contoh Agregasi Transaksi Harian.

	<b>date</b>	<b>transaction count</b>	<b>day of month</b>	<b>month period</b>
<b>0</b>	01/01/2025	7	1	awal bulan
<b>1</b>	02/01/2025	6	2	awal bulan
<b>2</b>	03/01/2025	8	3	awal bulan
<b>3</b>	04/01/2025	5	4	awal bulan

Tabel 1 menyajikan hasil agregasi data transaksi harian yang diperoleh dari proses ETL terhadap dataset marketplace. Data transaksi mentah yang semula tercatat per kejadian transaksi dikonversi menjadi jumlah transaksi per hari berdasarkan tanggal transaksi. Proses agregasi ini bertujuan untuk membentuk deret waktu (time series) yang menjadi dasar utama

dalam pemodelan ARIMA. Selain itu, setiap tanggal diklasifikasikan ke dalam kategori periode bulanan, yaitu awal bulan (hari ke-1 hingga ke-10), pertengahan bulan (hari ke-11 hingga ke-20), dan akhir bulan (hari ke-21 hingga ke-31).

Dari Tabel 1 terlihat bahwa jumlah transaksi harian bersifat fluktuatif namun berada pada rentang yang relatif stabil. Hal ini mengindikasikan bahwa aktivitas transaksi konsumen terjadi secara konsisten dari hari ke hari tanpa lonjakan ekstrem. Klasifikasi periode bulan pada tabel ini berfungsi sebagai landasan analisis perilaku konsumen berdasarkan waktu, sehingga memungkinkan identifikasi kecenderungan transaksi pada fase tertentu dalam satu bulan. Data agregat ini juga memastikan bahwa analisis tidak terdistorsi oleh duplikasi transaksi individual dan lebih merepresentasikan pola konsumsi harian secara makro..

### Uji Stasionaritas dan Pemodelan ARIMA

Uji ADF menghasilkan nilai p-value sebesar 0,000006 yang mengindikasikan bahwa data bersifat stasioner. Model ARIMA(2,0,0) dipilih sebagai model terbaik berdasarkan kriteria AIC.

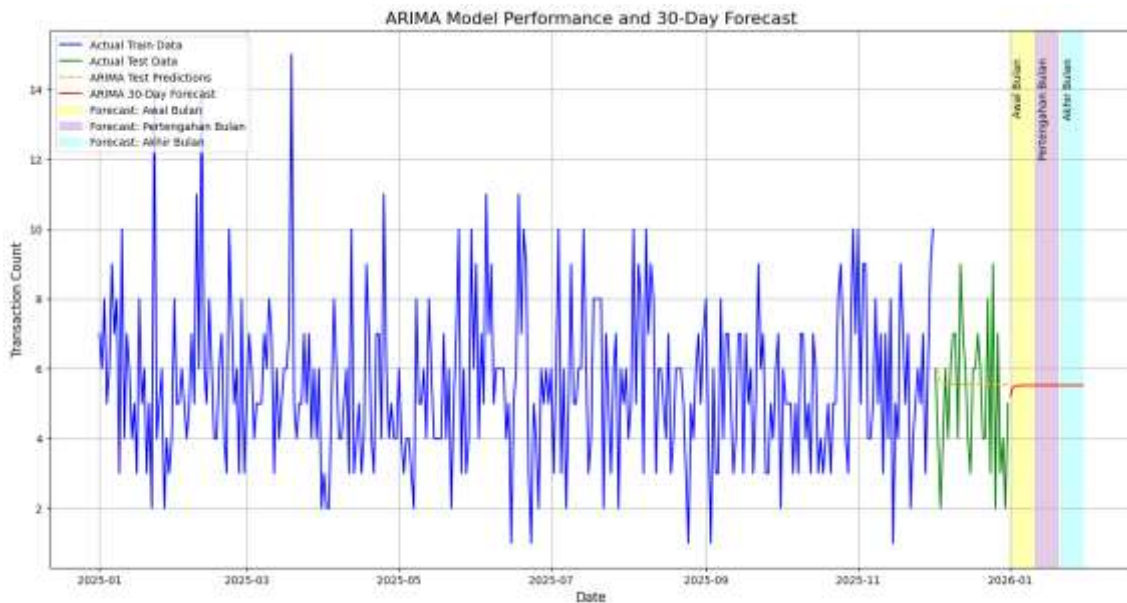
### Evaluasi dan Peramalan

**Table 2.** Kinerja Model ARIMA.

<b>Metrik</b>	<b>Nilai</b>
RMSE	2,002
MAE	1,704

Tabel 2 menampilkan hasil evaluasi kinerja model ARIMA(2,0,0) yang digunakan untuk memodelkan dan meramalkan jumlah transaksi harian. Dua metrik evaluasi yang digunakan dalam penelitian ini adalah Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE). RMSE mengukur besarnya kesalahan prediksi dengan memberikan penalti lebih besar pada kesalahan yang ekstrem, sedangkan MAE mengukur rata-rata kesalahan absolut antara nilai aktual dan nilai prediksi.

Nilai RMSE sebesar 2,002 menunjukkan bahwa secara rata-rata deviasi prediksi terhadap data aktual berada pada kisaran dua transaksi per hari. Sementara itu, nilai MAE sebesar 1,704 mengindikasikan bahwa kesalahan absolut rata-rata model relatif rendah. Nilai kedua metrik tersebut menunjukkan bahwa model ARIMA yang dibangun memiliki tingkat akurasi yang cukup baik dalam memprediksi jumlah transaksi harian. Dengan demikian, model ini layak digunakan untuk keperluan peramalan jangka pendek serta analisis pola pembelian konsumen pada periode berikutnya.



**Gambar 1.** Grafik Transaksi Aktual, Prediksi, dan Peramalan 30 Hari.

Gambar 1 memperlihatkan visualisasi gabungan antara data transaksi aktual pada periode pelatihan, data transaksi aktual pada periode pengujian, hasil prediksi model ARIMA pada data uji, serta hasil peramalan jumlah transaksi untuk 30 hari ke depan. Grafik ini digunakan untuk mengevaluasi secara visual kesesuaian antara nilai aktual dan nilai yang dihasilkan oleh model, sekaligus mengamati kecenderungan pola transaksi pada masa mendatang.

Pada bagian data pelatihan dan pengujian, terlihat bahwa garis prediksi mengikuti pola data aktual dengan cukup baik, meskipun terdapat deviasi kecil pada beberapa titik waktu. Hal ini menunjukkan bahwa model mampu menangkap pola fluktuasi transaksi harian secara umum. Pada periode peramalan 30 hari ke depan, grafik menunjukkan pola transaksi yang relatif stabil dengan kecenderungan sedikit meningkat pada pertengahan dan akhir bulan. Area anotasi pada grafik yang membedakan awal, pertengahan, dan akhir bulan memperjelas bahwa intensitas transaksi yang diprediksi cenderung lebih tinggi pada pertengahan dan akhir bulan dibandingkan awal bulan.

Visualisasi ini memperkuat hasil analisis kuantitatif sebelumnya, bahwa perilaku pembelian konsumen menunjukkan kecenderungan peningkatan aktivitas seiring berjalannya waktu dalam satu bulan. Temuan ini memiliki implikasi praktis bagi pengelola marketplace dalam menentukan waktu yang optimal untuk promosi, pengelolaan stok, serta strategi pemasaran berbasis periode bulanan.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, dapat disimpulkan bahwa implementasi *pipeline* ETL dan pemodelan ARIMA berhasil memetakan pola pembelian konsumen dengan cukup akurat. Model terbaik yang dihasilkan adalah ARIMA(2,0,0), yang menunjukkan kinerja stabil dengan nilai kesalahan *Root Mean Squared Error (RMSE)* sebesar 2.002 dan *Mean Absolute Error (MAE)* sebesar 1.704.

Analisis pola transaksi harian mengungkap adanya tren fluktuasi yang spesifik, di mana rata-rata transaksi harian cenderung meningkat pada periode pertengahan dan akhir bulan (5.52 transaksi/hari) dibandingkan dengan periode awal bulan (5.48 transaksi/hari). Temuan ini menjawab rumusan masalah bahwa pola pembelian konsumen dipengaruhi oleh siklus waktu bulanan, yang dapat dimanfaatkan oleh *marketplace* untuk mengoptimalkan strategi promosi dan manajemen stok pada periode pertengahan hingga akhir bulan.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Institut Binsnis dan Teknologi Indonesia (INSTIKI) sebagai institusi yang telah menyediakan dukungan akademik dan fasilitas sehingga penelitian ini dapat terlaksana dengan baik. Ucapan terima kasih juga disampaikan kepada I Wayan Sudiarsa, S.T.,M.Kom., yang telah memberikan arahan, masukan ilmiah, serta bimbingan metodologis selama proses penyusunan dan penyempurnaan penelitian ini.

Penulis turut menyampaikan apresiasi kepada pemilik dataset di platform Kaggle yang telah mempublikasikan data transaksi *marketplace* ini secara terbuka (*open source*). Ketersediaan data tersebut menjadi fondasi utama yang sangat membantu penulis dalam proses analisis pola pembelian konsumen dan pengembangan model peramalan deret waktu dalam penelitian ini.

Penulis turut menyampaikan apresiasi kepada pihak penyedia dataset marketplace yang digunakan dalam penelitian ini, yang bersumber dari data transaksi e-commerce yang tersedia secara publik dan digunakan untuk kepentingan akademik dan penelitian. Ketersediaan data tersebut sangat membantu dalam proses analisis pola pembelian konsumen dan pengembangan model peramalan deret waktu.

Penelitian ini diharapkan dapat memberikan kontribusi ilmiah serta menjadi referensi bagi penelitian selanjutnya dalam bidang analitik data, peramalan deret waktu, dan pengambilan keputusan berbasis data pada platform marketplace digital..

## DAFTAR REFERENSI

- Aggarwal, C. C. (2018). *Machine learning for data mining*. Springer.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Chen, D., Hu, Y., & Smith, J. (2021). Consumer behavior analytics in e-commerce. *Electronic Commerce Research*, 21(3), 567–589. <https://doi.org/10.1007/s10660-020-09421-3>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. <https://doi.org/10.1080/01621459.1979.10482531>
- Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*, 35(1), 1–15. <https://doi.org/10.1016/j.ijforecast.2018.06.004>
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Inmon, W. H. (2019). *Building the data warehouse* (5th ed.). Wiley.
- Kaggle. (2024). E-commerce transaction dataset. Kaggle. <https://www.kaggle.com>
- Kimball, R., & Ross, M. (2016). *The data warehouse toolkit* (3rd ed.). Wiley.
- Larose, D. T., & Larose, C. D. (2015). *Data mining and predictive analytics*. Wiley.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods. *Journal of Forecasting*, 37(8), 802–814. <https://doi.org/10.1002/for.2583>
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. Wiley.
- Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.
- Rahman, A., Putra, D., & Sari, N. (2023). Time series forecasting for e-commerce transactions using ARIMA. *Journal of Big Data Analytics*, 5(2), 45–60.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications* (4th ed.). Springer. <https://doi.org/10.1007/978-3-319-52452-8>
- Smith, J. (2020). *Data engineering fundamentals: Designing ETL pipelines*. CRC Press.

- Tsay, R. S. (2014). *Multivariate time series analysis: With R and financial applications*. Wiley.  
<https://doi.org/10.1002/9781118445112.stat03545>
- Vassiliadis, P. (2009). A survey of extract-transform-load technology. *International Journal of Data Warehousing and Mining*, 5(3), 1–27. <https://doi.org/10.4018/jdwm.2009070101>
- Wei, W. W. S. (2018). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Pearson.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)