JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Product Demand Forecasting in E-Commerce with Big Data Analytics: Personalization, Decision Making and Optimization

Cahyasari Kartika Murni[1], Achmad Firman Choiri[1], Febriane Devi Rahmawati[1]

Informatics Department, Institut Teknologi dan Bisnis Widya Gama Lumajang, Indonesia[1]

Corresponding Author : Cahyasari Kartika Murni (cahyasarikartikamurni@gmail.com)

ARTICLE INFO

Date of entry:
12 March 2025
Revision Date:
18 April 2025
Date Received:
24 April 2025

ABSTRACT

This study explores the role of Big Data in forecasting product demand in the e-commerce sector through the application of machine learning and time series methods. A quantitative descriptive approach is used, involving data collection, preprocessing, analysis, and model evaluation. Forecasting techniques applied include ARIMA for time series prediction and XGBoost for supervised learning to identify key demand factors. Model performance is evaluated using accuracy metrics such as RMSE, MAE, and MAPE. The results indicate that the XGBoost model provides the highest forecasting accuracy at 89%, while the ARIMA model achieves 78%. These findings demonstrate that Big Data significantly supports strategic decision-making in e-commerce by enhancing personalization, optimizing inventory, and enabling data-driven marketing strategies.

Keywords: ARIMA, Big data, Data Analysis, E-commerce, XGBoost

Cite this as: Murni, C. K., Choiri, A. F., & Rahmawati, F. D. (2025). Product Demand Forecasting in E-Commerce with Big Data Analytics: Personalization, Decision Making and Optimization. *Journal of Informatics Development*, *3*(2), 1–6. https://doi.org/10.30741/jid.v3i2.1548

INTRODUCTION

The rapid advancement of technology has significantly transformed economic activities, particularly in the way products are marketed and sold. One of the most notable changes is the emergence of online shopping through e-commerce platforms, which has become a part of modern socio-cultural dynamics (Sari, 2015). E-commerce facilitates the buying and selling of products via the internet without requiring face-to-face interaction between sellers and buyers. In this context, Big Data has emerged as a critical driver in improving decision-making, enhancing customer experience, and optimizing various operational aspects of e-commerce.

Big Data refers to data that cannot be processed using traditional tools due to its massive size and complexity. It is commonly defined by three main characteristics: volume, velocity, and variety (Gartner, 2013; Widy, 2017). Volume denotes the enormous scale of data generated every second, such as transaction records and social media activity. Velocity refers to the rapid growth rate of data, necessitating real-time processing. Variety represents the different types of data, ranging from structured numerical inputs to unstructured formats such as reviews, images, and videos. The implementation of Big Data in e-commerce enables businesses to analyze customer behavior,

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

monitor market trends, and optimize strategic decision-making processes (Santoso, 2022; Syira, 2023).

One of the key applications of Big Data in e-commerce is deep personalization, which enhances customer experience by delivering tailored product recommendations, dynamic promotions, and customized user interfaces (Sehgal, 2020). This approach not only increases customer satisfaction but also drives higher sales and improves conversion rates. Retailers can utilize browsing and purchase history to offer personalized assortments and dynamically adjust content based on real-time behavior, leading to a measurable increase in average order value (El Housni & Topaloglu, 2022). Moreover, advanced algorithms that integrate segmentation and geolocation data enable hyper-personalized recommendations, particularly effective in sectors like fashion retail (Yıldız et al., 2023; Zhang et al., 2023). Through Big Data analysis, companies can also deliver incentives such as cashback, free shipping, and special discounts that align with consumer preferences (Saibabu et al., 2024).

Beyond personalization, Big Data allows for in-depth customer journey analysis by mapping user interactions across touchpoints to identify friction points in the shopping process (Painuly et al., 2021). Understanding these behavioral patterns helps e-commerce platforms optimize navigation and interface design, resulting in smoother transaction flows and reduced cart abandonment (Razina et al., 2024). Furthermore, Big Data facilitates predictive inventory management by leveraging historical and real-time data to anticipate product demand, avoid stock imbalances, and enhance supply chain efficiency.

Although many studies have addressed the application of Big Data in business, there is still a lack of in-depth research focusing on the integration of Big Data with machine learning models for forecasting product demand in e-commerce using large-scale platforms such as Hadoop and Spark. This gap is critical, especially for e-commerce companies seeking to strengthen their market competitiveness through predictive analytics. Therefore, this study aims to forecast product demand in e-commerce by applying Big Data analytics and machine learning techniques, including regression analysis and decision tree algorithms. The goal is to build accurate and scalable forecasting models that help businesses improve efficiency, optimize inventory management, and make data-driven strategic decisions.

METHOD

This research applies a quantitative descriptive approach to analyze product demand forecasting in the e-commerce sector using historical and behavioral transaction data. The study is based on secondary data obtained from publicly accessible sources, including Google Dataset Search and relevant prior studies that provide structured data on product sales, user behavior, and temporal variables.

1. Data Collection and Preprocessing
   The data used includes structured transaction records, which are cleaned and preprocessed through the following stages:
   - Data Cleaning: Removal of duplicates, handling of missing values, and outlier correction.
   - Data Transformation: Normalization of numerical data and encoding of categorical variables.
   - Feature Engineering: Creation of new features such as average sales per category, seasonal patterns, and trend indicators to enhance model performance.
2. Forecasting and Modeling Techniques
   To forecast product demand and analyze related patterns, several machine learning and statistical techniques are applied:
   - Time Series Forecasting with ARIMA

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

ARIMA (AutoRegressive Integrated Moving Average) is used to predict future product demand trends based on past observations. It consists of three components: autoregressive (AR), differencing (I), and moving average (MA), and is widely known for its accuracy in economic and sales forecasting tasks (Djawoto, 2017; Yadav & Goswami, 2024; Le, 2024).

- Supervised Learning with XGBoost
  The XGBoost (Extreme Gradient Boosting) algorithm is implemented to identify important factors affecting demand. XGBoost is known for its robustness, efficiency, and accuracy in handling complex datasets with nonlinear relationships (Saroji & Prakoso, 2023; Abhinaya et al., 2024; Murdiansyah, 2024). It has been successfully applied across various domains including healthcare, energy, and finance (Shaeri et al., 2023).
- Clustering for Customer Segmentation
  To better understand user behavior and purchasing patterns, clustering is conducted using both K-Means or C-Means algorithms:
  K-Means clusters customer data into segments based on similarities in transaction behavior, which can inform targeted marketing. C-Means, clustering technique, allows degrees of membership for each data point across multiple clusters. This method has proven effective in mapping regional dynamics in both tourism and human development contexts (Qori'atunnadyah, 2023; 2024).

3. Model Evaluation
   Model performance is evaluated using standard predictive accuracy metrics:
   - RMSE (Root Mean Squared Error): Measures the average squared difference between predicted and actual values.
   - MAE (Mean Absolute Error): Computes the average absolute difference between predictions and actual values.
   - MAPE (Mean Absolute Percentage Error): Represents the mean percentage error, useful for evaluating relative accuracy.

   While the methodology outlines the analytical steps clearly, technical details such as model parameter tuning, validation techniques, and algorithm implementation environment (e.g., Python, R, Spark) are not elaborated in this stage. Additionally, a process diagram or flowchart representing the full method pipeline would enhance clarity and replicability.

RESULTS AND DISCUSSION

1. Deep Personalization
   Big data really allows companies to understand customer needs in more detail and depth. By processing browsing history, transaction data and customer behavior. Then the company can provide relevant solutions for customers. The analyzed data allows companies to offer very appropriate product recommendations. In addition, companies can also arrange promotions and discounts so that they can attract customer attention.
   Advertising or marketing can be used as a customer puller, for example if someone routinely searches for an item, then a system supported by big data can automatically recommend the latest products or even provide special offers. This approach not only increases customer satisfaction, but also creates opportunities to increase sales. Thus, big data is very helpful in increasing product sales, product recommendations, special offers, and content displayed to customers in e-commerce.
2. Data-Based Decision Making
   By analyzing historical data, seasonal patterns, and market trends, companies can gain deep insights into customer needs. This analytical process allows companies to accurately predict product demand, helping to manage inventory more efficiently and avoiding risks such as stockouts or overstocking that can be detrimental to operations. For example, if data shows a spike in demand during the holiday season, companies can stock up early to meet customer

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

needs. In addition to inventory management, analyzed data also forms the basis for designing more effective marketing. By understanding customer preferences, shopping habits, and needs, companies can create more relevant and engaging promotional messages. These well-targeted campaigns not only increase customer engagement but also strengthen their brand loyalty. Thus, data-driven decision-making not only helps companies improve operational efficiency but also drives overall business growth through a more strategic and informed approach.

3. Optimizing User Experience
   Big data helps companies gain a deeper understanding of how customers interact with e-commerce websites. By analyzing user behavior patterns, companies can identify obstacles that may be distracting customers. One example is detecting problems with site navigation. So if customers frequently abandon their shopping carts, this could be an indication of a problem, such as an overly complicated checkout process or unclear shipping information. With these insights, companies can take steps to further simplify the process and improve user experience. Big data allows for the customization of website displays based on individual customer preferences. For example, a site could display favorite product categories on the front page or offer relevant promotions during a customer's visit. Real-time analytics plays a key role in providing this more personalized experience. As customers navigate the site, the system can automatically tailor product recommendations and promotions to suit their specific needs, creating a more engaging and efficient shopping experience. With this approach, companies not only increase customer satisfaction but also strengthen loyalty through relevant and high-value experiences.
   Several studies have shown that Big Data analysis can significantly improve the accuracy of product demand prediction in e-commerce. Social media sentiment analysis can provide valuable insights into customer preferences and help predict demand for new products. Big Data refers to the huge volume of data generated every second. This data can be structured, semi-structured, or unstructured and is characterized by 3Vs: volume, velocity, and variety.

Data Analysis Results After processing and analyzing e-commerce transaction data, here are some of the main results found:

a. Product Demand Pattern: From the time series analysis, it was found that product demand tends to increase on certain days, such as weekends and approaching big days (for example, National Online Shopping Day). Seasonality and market trends: Certain products experience spikes in demand based on seasons or special events, such as fashion during the holiday season and gadgets when there is a new product release.
b. Factors Affecting Demand, namely the results of analysis with forecasting models. Price and discounts: Products with high discounts experience an increase in demand of up to 35% higher than without discounts. Reviews and ratings. Then Promotion time: Promotions in the form of advertisements and flash sales have a direct impact on increasing the number of transactions, with peak sales in the first 2 hours after the promotion starts.
c. The accuracy of the Prediction Model is tested to predict product demand. The evaluation results using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score metrics are as follows:

Table 1. Forecast results

Model	RMSE	MAE	R^2Score
ARIMA	19.6	13.2	0.78
XGBoost	14.3	9.7	0.89

Source : Reseacher, 2025

From these results, the XGBoost model has the best performance with a forecasting accuracy of 89%, and using the ARIMA model has a performance with a forecasting accuracy of 78%. Comparison of Models and Prediction Accuracy in this study, the XGBoost model is much better than the ARIMA model because the XGBoost model has many advantages in

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

complexity. If the ARIMA model is very suitable for forecasting market trends in product demand contests.

CONCLUSION

Based on the analysis results, the implementation of Big Data analytics has proven effective in forecasting product demand in e-commerce, enabling companies to improve operational efficiency and deliver a more personalized and seamless shopping experience. Big Data, characterized by its extreme volume, velocity, and variety, supports fast and innovative data processing essential for informed decision-making and process automation. With the application of appropriate forecasting models such as ARIMA and XGBoost, businesses can enhance their ability to predict market trends, personalize product recommendations, optimize promotional strategies, and streamline inventory management. However, this study has several limitations. It relies solely on secondary data and does not include validation using real-time datasets or comparisons across different e-commerce platforms. The analysis also does not account for external influencing factors such as seasonality effects, consumer sentiment, or macroeconomic conditions. Furthermore, the technical details of model training, hyperparameter optimization, and platform-specific implementation were not explored in depth. Future research should consider the integration of real-time Big Data streams with more advanced machine learning frameworks, including deep learning architectures. Additionally, developing a web-based decision support system that incorporates these predictive models would allow for real-time monitoring and dynamic forecasting, enhancing the practical value of Big Data in e-commerce applications.

REFERENCES

Abhinaya, P. M., Reddy, C. K., Ranjan, A., & Ozer, O. (2024). Explicit monitoring and prediction of hailstorms with XGBoost classifier for sustainability (pp. 107–132). IGI Global. https://doi.org/10.4018/979-8-3693-3896-4.ch006

Djawoto, D. (2017). Peramalan laju inflasi dengan metode AutoRegressive Integrated Moving Average (ARIMA). Jurnal Ekonomi Pembangunan, 14(4), 524–538. https://doi.org/10.24034/J25485024.Y2010.V14.I4.2190

El Housni, O., & Topaloglu, H. (2022). Joint assortment optimization and customization under a mixture of multinomial logit models: On the value of personalized assortments. Operations Research, 71(4). https://doi.org/10.1287/opre.2022.2384

Gartner. (2013). What is Big Data? Gartner IT Glossary. https://www.gartner.com/en/information-technology/glossary/big-data

Le, L.-H. (2024). Time series analysis and applications in data analysis, forecasting and prediction. https://doi.org/10.56764/hpu2.jos.2024.3.1.20-29

Murdiansyah, D. T. (2024). Prediksi stroke menggunakan Extreme Gradient Boosting. JIKO (Jurnal Informatika dan Komputer), 8(2), 419. https://doi.org/10.26798/jiko.v8i2.1295

Painuly, S., Sharma, S., & Matta, P. (2021). Big data driven e-commerce application management system. 2021 International Conference on Communication and Electronics Systems (ICCES), 1–5. https://doi.org/10.1109/ICCES51350.2021.9489108

Qori'atunnadyah, M. (2023). Metode C-Means untuk pengelompokkan kabupaten/kota Provinsi Jawa Timur berdasarkan indikator Indeks Pembangunan Manusia (IPM). Journal of Informatics Development, 1(2), 51–58. https://doi.org/10.30741/jid.v2i2.1013

Qori'atunnadyah, M. (2024). Mapping domestic and foreign tourists in East Java using C-Means clustering. Jurnal Statistika dan Aplikasinya, 8(1), 54–62. https://doi.org/10.21009/JSA.08105

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Razina, R., Yunita, A., & Sabilah, K. (2024). Mengoptimalkan pengalaman belanja dengan aplikasi e-commerce. Jurnal Manajemen dan Bisnis Ekonomi, 3(1), 339–350. https://doi.org/10.54066/jmbe-itb.v3i1.2798

Saibabu, N., Chappa, M., Chaitanya, N., Das, S., Rao, A. L., & Mallam, M. (2024). Big data: An essential route for creating new business prospects. https://doi.org/10.1109/amathe61652.2024.10582054

Saroji, S., & Prakoso, S. (2023). An implementation of XGBoost algorithm to estimate effective porosity on well log data. Journal of Physics: Conference Series, 2498(1), 012011. https://doi.org/10.1088/1742-6596/2498/1/012011

Sehgal, R. (2020). Deep personalization for better human connect and optimization by using non-conventional mechanisms in the modern digital systems. International Journal of Scientific and Research Publications, 11(1), 433–456. https://doi.org/10.29322/IJSRP.11.01.2021.P10952

Shaeri, M. R., Ellis, M. C., & Randriambololona, A. M. (2023). XGBoost-based model for prediction of heat transfer coefficients in liquid cold plates. https://doi.org/10.1615/tfec2023.cmd.045483

Syira, S. D., et al. (2023). Pemanfaatan Big Data dalam peningkatan efektivitas strategi komunikasi marketing terpadu pada perusahaan e-commerce. Jurnal Ekonomi Manajemen Sistem Informasi, 4(5). https://doi.org/10.31933/jemsi.v4i5

Widy, S. (2017). Berkenalan dengan Big Data. Medium.com. https://medium.com/skyshidigital/berkenalan-dengan-big-data-15fd941122f8

Yadav, D. K., & Goswami, L. (2024). Autoregressive Integrated Moving Average model for time series analysis. https://doi.org/10.1109/icocwc60930.2024.10470488

Yıldız, E., Güngör Şen, C., & Işık, E. (2023). A hyper-personalized product recommendation system focused on customer segmentation: An application in the fashion retail industry. Journal of Theoretical and Applied Electronic Commerce Research, 18(1), 571–596. https://doi.org/10.3390/jtaer18010029

Zhang, C., Liu, B., Mohammed, B. S., & Jumani, A. K. (2023). Big data assisted empirical study for business value identification using smart technologies: An empirical study for business value identification of big data adaption in e-commerce. International Journal of E-Collaboration, 19(7), 1–19. https://doi.org/10.4018/ijec.316882