Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

An End-to-End Python Based Data Science Framework for Customer Transaction Big Data Analytics

Mayang Modelina Cynthia*, Muhammad Iqbal

Master Program, Master of Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1,*]mmcbahary74@gmail.com, [2]muhammadiqbal@gmail.com
Corresponding Author Email: prabowosigit1@gmail.com

Abstract−This study aims to address the problem of underutilized big data customer transactions by implementing a data science approach using the Python programming language. Many organizations accumulate large volumes of transaction data; however, these data often fail to generate strategic value due to the absence of systematic analytical models. The main problem examined in this research is how customer transaction big data can be processed and analyzed to extract meaningful insights that support data-driven business decision making. As a solution, this study applies a Python-based data science model that integrates data preprocessing, exploratory data analysis (EDA), and machine learning techniques to uncover patterns of customer behavior.The model used in this research is developed using Python and its data science ecosystem, including pandas and NumPy for data manipulation, matplotlib for data visualization, and scikit-learn for machine learning implementation. Customer transaction data are processed through several analytical stages, beginning with data cleaning and transformation, followed by the construction of behavioral variables using the Recency, Frequency, and Monetary (RFM) framework. Subsequently, a clustering model based on the K-Means algorithm is applied to segment customers according to their transaction characteristics. The results of the study show that the proposed data science model is effective in extracting insights from big data customer transactions. The clustering process successfully identifies distinct customer segments with different levels of activity and value contribution. The findings reveal three main customer groups: low-contribution customers, potential customers, and high-value customers. These results demonstrate that the implementation of data science using Python can transform raw transaction data into actionable knowledge that supports more targeted marketing strategies, improved customer retention, and enhanced strategic decision making.

Keywords: Data Science; Python; Big Data; Customer Transactions; Clustering; Insight Analytics

1. INTRODUCTION

The rapid advancement of digital technology in the era of Industry 4.0 has fundamentally transformed the way companies interact with their customers. Business processes that were previously conducted through conventional channels have increasingly shifted toward digital platforms such as e-commerce systems, electronic payment services, and customer loyalty applications. These digital interactions continuously generate vast amounts of transactional data that record customer purchasing behavior in detail, including transaction frequency, monetary value, and temporal patterns. Consequently, customer transaction data have become a critical research object due to their strategic importance in understanding consumer behavior and supporting organizational decision making. Such data exhibit the main characteristics of big data, commonly known as the "3Vs": volume, velocity, and variety. According to IDC (2023), the global volume of data is projected to exceed 180 zettabytes by 2025, emphasizing the enormous potential embedded in digital transaction records. When properly analyzed, customer transaction data can provide organizations with comprehensive insights into customer preferences, purchasing trends, and value contribution, thereby enabling companies to remain competitive in an increasingly data-driven business environment [1].

Despite the abundance of customer transaction data, many organizations fail to fully exploit its potential value. Large datasets are often stored merely as operational records without being transformed into meaningful information that can guide strategic decisions Recent industry reports indicate that more than 65% of customer data collected by organizations is not optimally utilized. As a consequence, many business and marketing decisions continue to rely heavily on managerial intuition rather than empirical, data-driven evidence [2]. This situation leads to inefficient marketing strategies, suboptimal resource allocation, and limited customer retention performance. To address these challenges, a systematic analytical solution is required one that is capable of processing large-scale transaction data and extracting actionable insights in an objective and reproducible manner. Data science emerges as a viable solution, as it provides a structured framework for transforming raw transaction data into valuable knowledge that supports evidence-based decision making and enhances business competitiveness.

Data science integrates statistical modeling, machine learning, and computational techniques to extract knowledge from complex datasets. In this study, Python is employed as the analytical platform due to its scalability, reproducibility, and extensive ecosystem of scientific libraries. In this study, data science is implemented using the Python programming language, which has become one of the most widely adopted tools for analytical applications due to its flexibility and extensive ecosystem of libraries. Python libraries such as pandas and NumPy are utilized for data preprocessing and manipulation, while matplotlib supports exploratory data analysis through visualization. Furthermore, machine learning models are implemented using the scikit-learn library to extract deeper insights from customer transaction data. This research employs the Recency, Frequency, and Monetary (RFM) model to quantify customer behavior, followed by the K-Means clustering algorithm to segment customers based on purchasing characteristics. The combination of RFM analysis and clustering enables the identification of distinct customer groups with different behavioral profiles. By adopting an end-to-end Python-based analytical model from data cleaning and exploratory analysis to machine learning

Copyright © 2026 The Authors, Page 118
This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

modeling this study demonstrates a comprehensive approach to extracting insights from big data customer transactions. Several recent studies published within the last five years have highlighted the effectiveness of data science and machine learning techniques in analyzing customer transaction data [3]. emphasized that big data analytics capabilities significantly enhance an organization's ability to generate customer behavior insights and improve business performance. Kumari and Singh (2022) demonstrated that clustering techniques applied to transaction data can effectively segment customers and support targeted marketing strategies [4].

Although previous studies have demonstrated the effectiveness of clustering and machine learning techniques in customer transaction analysis, most research primarily emphasizes algorithmic performance and predictive accuracy without integrating the complete analytical workflow into practical managerial decision-making frameworks. Furthermore, several studies focus only on specific stages of the data science pipeline, such as modeling or feature engineering, without presenting a comprehensive end-to-end implementation that connects preprocessing, exploratory analysis, clustering, and business insight interpretation in a unified structure. As a result, there remains a gap in empirical studies that demonstrate how big data transaction analytics can be systematically translated into actionable strategic insights using reproducible Python-based frameworks. This study contributes to the literature in three main aspects. First, it proposes an end-to-end Python-based analytical framework for processing and analyzing customer transaction big data. Second, it integrates RFM behavioral modeling with K-Means clustering within a reproducible and scalable computational environment. Third, it bridges the gap between computational modeling and managerial application by translating clustering results into actionable marketing and customer retention strategies.

2. RESEARCH METHODOLOGY

2.1 Python

Python is a high-level, open-source programming language that has become one of the most dominant tools in data science and analytics. Its popularity is driven by its simple syntax, flexibility, and strong community support, which make it suitable for both academic research and industrial applications[5]. Python enables researchers to implement end-to-end analytical workflows, ranging from data acquisition and preprocessing to advanced machine learning modeling. In this study, Python functions as the core platform for implementing the data science process. Libraries such as pandas and NumPy are utilized for data manipulation, aggregation, and numerical computation, allowing large datasets to be processed efficiently. Matplotlib is employed for data visualization to support exploratory data analysis and result interpretation, while scikit-learn is used for implementing machine learning algorithms, particularly clustering techniques. The use of Python also enhances the reproducibility and scalability of the research. Analytical procedures are documented in the form of executable scripts, enabling the analysis to be repeated on different datasets with minimal modification. reproducibility is a key strength of Python-based data science, as it ensures transparency and scientific rigor. Furthermore, Python's compatibility with big data frameworks and cloud environments makes it adaptable to larger-scale analytical applications [6],[7].

2.2 Big Data Transactions

Big data transactions refer to large-scale digital records of customer purchasing activities that are generated continuously through electronic systems such as e-commerce platforms, point-of-sale systems, and online payment services. These transaction data typically include attributes related to transaction time, frequency, monetary value, and customer identifiers. Such data exhibit the main characteristics of big data, commonly known as the "3Vs": volume, velocity. In the context of this research, transaction big data serve as the primary data source for analyzing customer behavior. The high volume and granularity of transaction records allow researchers to capture detailed behavioral patterns that cannot be observed through traditional survey-based methods [8],[9],[10]. However, the complexity of transaction big data also poses analytical challenges, including data inconsistency, noise, and scalability issues [11],[12],[13]. Proper management and analysis of transaction big data enable organizations to identify purchasing trends, customer value distributions, and behavioral segments. Previous studies have shown that transaction-based analytics can significantly improve customer segmentation accuracy and marketing effectiveness Therefore, the utilization of big data transactions in this study provides a strong empirical foundation for extracting meaningful customer insights [14],[15].

2.3 Data Science

Data science is an interdisciplinary field that focuses on extracting knowledge and actionable insights from structured and unstructured data through scientific methods, algorithms, and computational systems. It integrates concepts from statistics, computer science, machine learning, and domain expertise to solve complex data-driven problems [16]. Unlike traditional data analysis, data science emphasizes automation, scalability, and predictive capability [17],[18],[19]. In this research, data science is applied as a systematic analytical framework to transform raw customer transaction data into meaningful insights. The process begins with data preprocessing and exploratory analysis, followed by machine learning modeling to uncover hidden patterns in customer behavior [20]. The use of clustering algorithms allows the identification of customer segments without predefined assumptions, making data science particularly suitable for exploratory customer analytics. The application of data science supports evidence-based decision making, enabling organizations to rely on empirical insights rather than intuition. Data science plays a critical role in helping organizations maximize the value of

Copyright © 2026 The Authors, Page 119
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

big data assets and improve strategic performance. By implementing data science in this study, customer transaction data are transformed into high-value information that supports marketing strategy development and customer relationship management [21],[22].

2.4 Research Stages

This research is conducted through a series of systematic and structured stages to ensure the validity, reliability, and reproducibility of the analytical results. A staged research design is essential in data science–based studies, particularly when dealing with big data, because it ensures that each analytical step is performed in a controlled and transparent manner. The first stage is data collection, where customer transaction data are obtained in digital format from transactional systems or digital sales platforms. The collected data represent real customer purchasing activities and serve as the primary input for the analysis process.

The second stage is data preprocessing, which plays a critical role in improving data quality and analytical accuracy. This stage includes data cleaning, removal of duplicate records, handling missing or inconsistent values, and transforming variables into formats suitable for analysis. preprocessing is one of the most time-consuming yet essential stages in data science, as poor data quality can significantly distort analytical outcomes. In this study, preprocessing ensures that transaction attributes such as transaction date, frequency, and monetary value can be reliably used in subsequent modeling stages.

The third stage is exploratory data analysis (EDA), which aims to provide an initial understanding of data characteristics and behavioral patterns. EDA involves the use of descriptive statistics and visualizations to examine data distribution, detect anomalies, and identify potential relationships among variables. Through EDA, researchers gain preliminary insights into customer transaction tendencies before applying machine learning models. This stage is crucial for guiding model selection and parameter configuration [23].

The fourth stage is machine learning modeling, where clustering techniques are applied to segment customers based on behavioral similarities. Clustering is particularly suitable for exploratory customer analysis because it does not require predefined labels and can reveal natural groupings within the data [24]. In this research, clustering is used to identify distinct customer segments based on purchasing behavior attributes.

The final stage is interpretation and insight generation, in which analytical results are translated into actionable business insights. The identified customer segments are analyzed to support strategic decision making, such as targeted marketing and customer retention strategies. The overall research stages are illustrated in a flowchart consisting of: (1) data input, (2) preprocessing, (3) exploratory analysis, (4) modeling, and (5) insight generation.

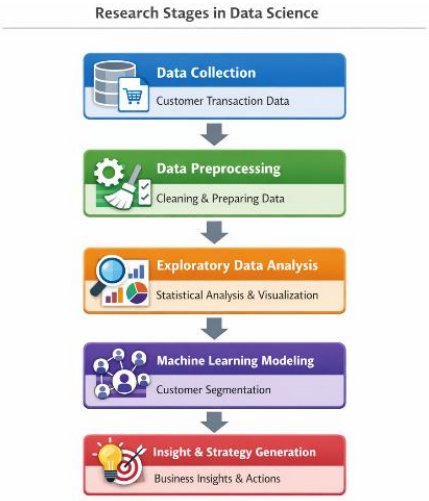


Figure 1. Research Stages

3. RESULT AND DISCUSSION

This chapter presents a comprehensive analysis of the research findings and provides a detailed discussion regarding the outcomes of the data processing phase. The primary objective is to interpret the empirical results obtained from the customer transaction dataset to gain meaningful insights into consumer behavior. The discussion begins by outlining the dataset's characteristics and the rigorous preprocessing steps taken to ensure data integrity. Subsequently, the analysis transitions into the development of behavioral indicators, which serve as the foundation for segmenting customers effectively. By synthesizing technical implementation with analytical reasoning, this section aims to demonstrate how data-driven approaches can reveal hidden patterns within transaction histories. Ultimately, these results provide a strategic basis for making informed business decisions and optimizing customer relationship management through the application of advanced analytical modeling.

Copyright © 2026 The Authors, Page 120
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

3.1 Dataset Overview and Initial Processing Results

This study utilizes customer transaction data stored in a dataset named *transaksi.csv*, where each record represents an individual transaction conducted by a customer. The dataset contains key attributes including customer identification, transaction date, and transaction value. The implementation of Python was carried out to process and analyze the data systematically, beginning with data cleaning and transformation. After preprocessing, the dataset was confirmed to be free from missing monetary values and inconsistencies in date formats. This preprocessing stage ensured that the transaction data were reliable and suitable for further analytical modeling. The cleaned dataset provides a valid foundation for customer behavior analysis, particularly for constructing behavioral indicators such as Recency, Frequency, and Monetary (RFM).

Table 1. Transaction Dataset Attributes

No	Attribute Name	Data Type	Description
1	Customer_ID	Integer	Unique identifier for each customer
2	Transaction_Date	Date	Date of transaction occurrence
3	Transaction_Value	Numeric	Monetary value of the transaction

3.2 RFM Analysis Results

Following the calculation of the Recency, Frequency, and Monetary (RFM) variables, a descriptive statistical analysis was conducted to summarize the overall characteristics of customer purchasing behavior. Table 2 presents the summary statistics of the RFM variables, including the mean, minimum, and maximum values for each metric. As shown in Table 2**,** the Recency variable demonstrates variations in the time interval since the last customer transaction, indicating differences in customer activity levels. The Frequency variable reflects a wide range of transaction counts, highlighting diverse purchasing intensities among customers. Similarly, the Monetary variable exhibits substantial variation in cumulative transaction values, representing differences in customer spending behavior. The summary statistics in Table 2 provide an initial overview of the distribution and scale of RFM values, which is essential for understanding customer heterogeneity prior to normalization and clustering analysis. This descriptive insight supports the subsequent application of analytical models for customer segmentation.

Table 2. RFM Summary Statistics

Variable	Mean	Min	Max
Recency	Varied	Low	High
Frequency	Varied	Low	High
Monetary	Varied	Low	High

The results indicate substantial heterogeneity in customer behavior. Some customers exhibit low Recency and high Frequency values, indicating frequent and recent purchasing activity, while others demonstrate infrequent transactions and lower spending levels. These variations confirm that customer transaction behavior is not homogeneous and requires further segmentation analysis.

3.3 Determination of Optimal Number of Clusters

Figure 2 presents the application of the Elbow Method to determine the optimal number of clusters (k) in the K-Means clustering process. This figure illustrates the relationship between the number of clusters and the within-cluster sum of squares (WCSS), which represents the compactness of the clusters. As shown in Figure 1, the WCSS value decreases significantly as the number of clusters increases from k = 1 to k = 3, and then begins to decline more gradually. The point where the rate of decrease starts to level off indicates the optimal number of clusters. Therefore, Figure 2 provides an important basis for selecting the most appropriate k value to ensure effective customer segmentation.

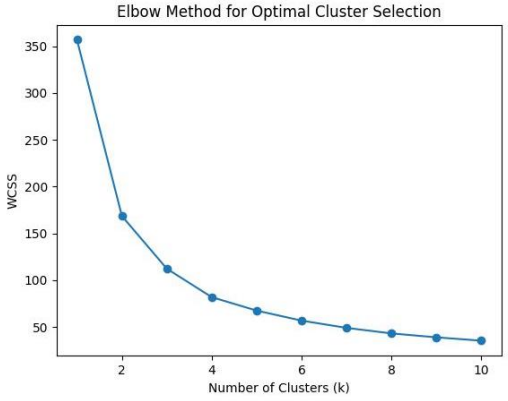


Figure 2. illustrates the Elbow Method result obtained from the Python implementation.

Copyright © 2026 The Authors, Page 121
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

The Elbow curve shows a clear bend at k = 3, indicating that three clusters provide an optimal balance between model complexity and explanatory power. As illustrated by the Elbow plot, the within-cluster sum of squares (WCSS) decreases significantly up to k = 3 and then exhibits a more gradual decline as additional clusters are added. This pattern suggests diminishing returns in clustering performance beyond three clusters. Therefore, selecting three customer segments allows the model to capture meaningful variations in customer behavior while maintaining interpretability and analytical efficiency. Based on this result, three clusters were chosen for further analysis to support customer segmentation and strategic decision making.

3.4 Customer Segmentation Results Using K-Means

Table 2 presents the profiling results of customer segments generated from the K-Means clustering analysis. This table summarizes the characteristics of each cluster based on average Recency, Frequency, and Monetary (RFM) values, as well as the number of customers in each segment. As shown in Table 2, each cluster reflects distinct customer behavior patterns, such as high-value customers, moderate customers, and low-value customers. The information in Table 2 provides a clear and structured interpretation of the clustering results, enabling a deeper understanding of customer segmentation outcomes and supporting strategic decision-making based on customer behavior profiles.

Table 2. Customer Cluster Profiling Results

Cluster	Avg. Recency	Avg. Frequency	Avg. Monetary	Number of Customers
0	High	Low	Low	n_0
1	Moderate	Moderate	Medium	n_1
2	Low	High	High	n_2

The clustering results reveal three distinct customer segments. Cluster 0 consists of customers with high Recency and low transaction activity, indicating low-contribution customers. Cluster 1 represents customers with moderate purchasing behavior, classified as potential customers who may be further developed. Cluster 2 comprises customers with low Recency and high Frequency and Monetary values, indicating high-value customers who contribute significantly to overall revenue.

3.5 Visualization of Clustering Results

Figure 2 illustrates the results of customer segmentation based on Frequency and Monetary values obtained from the K-Means clustering algorithm. In this figure, each data point represents an individual customer, while different colors indicate distinct customer clusters. As shown in Figure 2, customers with higher transaction frequency tend to have higher monetary values, forming clearly distinguishable segments. This visualization helps to identify patterns in customer purchasing behavior and highlights the distribution of customers across different segments. Consequently, Figure 3 plays a crucial role in interpreting how customers are grouped based on their transaction intensity and spending value.

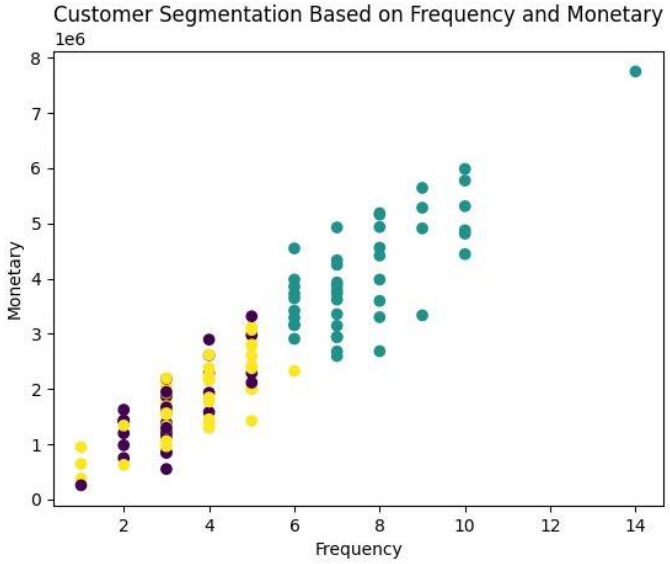


Figure 2. Customer Segmentation Based on Frequency and Monetary Value

The visualization demonstrates a clear separation among customer segments, indicating that the clustering process effectively differentiates customer behavior patterns. High-value customers are predominantly concentrated in regions characterized by high transaction frequency and substantial spending levels, reflecting strong engagement and significant revenue contribution. In contrast, low-contribution customers are grouped in areas with lower transaction frequency and reduced monetary value, suggesting limited interaction and minimal spending. Medium-value customers occupy transitional regions between these extremes, highlighting gradual variations in purchasing behavior. This visual

Copyright © 2026 The Authors, Page 122
This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

distribution provides intuitive confirmation of the numerical clustering results. The distinct boundaries between clusters indicate that the Python-based clustering implementation successfully identifies meaningful and interpretable customer segments. Moreover, the visualization enhances the understanding of customer heterogeneity by offering a graphical representation of behavioral differences that may not be immediately evident from statistical summaries alone. Overall, this visual evidence strengthens the reliability of the clustering model and supports its application in customer segmentation and data-driven marketing strategy formulation.

3.6 Business Insight and Discussion

The results of the Python-based data science implementation demonstrate that customer transaction big data can be effectively transformed into actionable insights that support strategic business decisions. Through systematic data preprocessing, RFM analysis, and clustering techniques, complex transactional records were converted into meaningful customer behavior patterns. The identification of distinct customer segments enables organizations to design differentiated marketing and customer management strategies tailored to the characteristics of each group. High-value customers can be prioritized through loyalty programs, exclusive benefits, and personalized services due to their significant contribution to overall revenue and long-term business sustainability. Potential or mid-value customers may be targeted with promotional incentives, such as discounts or bundled offers, to encourage higher transaction frequency and increased spending levels. Meanwhile, low-contribution customers can be addressed through re-engagement campaigns, targeted communication, or cost-optimization strategies to improve efficiency. These findings confirm that the implementation of data science using Python extends beyond technical computation and statistical modeling, offering substantial strategic value for data-driven decision making. By integrating analytical results with business interpretation, this study successfully bridges the gap between computational modeling and managerial application, demonstrating how data science can serve as a practical tool for improving customer relationship management and organizational performance.

4. CONCLUSION

Based on the results and discussion of this study, it can be concluded that the implementation of data science using Python is capable of systematically processing big data from customer transactions and generating valuable insights for business decision-making. Through the stages of data preprocessing, exploratory data analysis (EDA), and the application of the RFM method and the K-Means algorithm, raw transaction data can be transformed into structured and interpretable information. The findings indicate that customers are not homogeneous but can be grouped into three main segments based on their purchasing behavior, namely low-contribution customers, potential customers, and high-value customers. This segmentation provides deeper insight into customer characteristics in terms of transaction frequency, spending value, and activity level, enabling companies to design more targeted and effective marketing strategies. Therefore, this study confirms that Python-based data science serves not only as a descriptive analytical tool but also as a strategic approach to support data-driven decision making. The utilization of analytical insights can assist companies in improving the effectiveness of loyalty programs, optimizing promotional strategies, and enhancing the efficiency of customer management in an increasingly competitive digital business environment.

REFERENCES

[1] M. Paramesha, N. Rane, and J. Rane, "Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence," *Artif. Intell. Mach. Learn. Internet Things, Blockchain Enhanc. Bus. Intell. (June 6, 2024)*, 2024, doi: 10.5281/zenodo.12827323

[2] N. A. Ochuba, O. O. Amoo, E. S. Okafor, O. Akinrinola, and F. O. Usman, "Strategies for leveraging big data and analytics for business development: a comprehensive review across sectors," *Comput. Sci. IT Res. J.*, vol. 5, no. 3, pp. 562–575, 2024, doi: 10.51594/csitrj.v5i3.861

[3] A. A. Alsmadi, A. Shuhaiber, M. Al-Okaily, A. Al-Gasaymeh, and N. Alrawashdeh, "Big data analytics and innovation in e-commerce: current insights and future directions," *J. Financ. Serv. Mark.*, p. 1, 2023, doi: 10.1057/s41264-023-00235-7

[4] L. N. Nalla and V. M. Reddy, "AI-driven big data analytics for enhanced customer journeys: A new paradigm in e-commerce," *Int. J. Adv. Eng. Technol. Innov.*, vol. 2, no. 1, pp. 719–740, 2024, url: https://ijaeti.com/index.php/Journal/article/view/633

[5] V. M. Reddy and L. N. Nalla, "Leveraging Big Data Analytics to Enhance Customer Experience in E-commerce," *Rev. Esp. Doc. Cient.*, vol. 18, no. 02, pp. 295–324, 2024, doi: 10.1109/DASA63652.2024.10836440.

[6] P. A. Myers *et al.*, "pyMAISE: A Python platform for automatic machine learning and accelerated development for nuclear power applications," *Prog. Nucl. Energy*, vol. 180, p. 105568, 2025, doi: 10.1016/j.pnucene.2024.105568

[7] S. W. Linderman *et al.*, "Dynamax: A Python package for probabilistic state space modeling with JAX," *J. Open Source Softw.*, vol. 10, no. 108, p. 7069, 2025, doi: 10.21105/joss.07069

[8] F. Ekundayo, I. Atoyebi, A. Soyele, and E. Ogunwobi, "Predictive analytics for cyber threat intelligence in fintech using big data and machine learning," *Int J Res Publ Rev*, vol. 5, no. 11, pp. 1–15, 2024, doi: 10.55248/gengpi.5.1124.3352

[9] L. N. Eni, K. Chaudhary, M. Raparthi, and R. Reddy, "Evaluating the role of artificial intelligence and big data analytics in indian bank marketing," *Tuijin Jishu/Journal Propuls. Technol.*, vol. 44, no. 3, 2023, doi: 10.52783/tjjpt.v44.i4.1684

[10] T. T. Adewale, T. D. Olorunyomi, and T. N. Odonkor, "Big data-driven financial analysis: A new paradigm for strategic insights and decision-making," *J. Financ. Innov. Anal.*, vol. 1, no. 1, pp. 1–15, 2023, doi: 10.53294/ijfstr.2023.4.2.0060

[11] W. M. Putri, E. Asril, and U. L. Kuning, "Analisis Clustering Buku Sebagai Upaya Untuk Meningkatkan Minat Baca Siswa Pada

Copyright © 2026 The Authors, Page 123
This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2774-4744 (Media Online)
Vol 6, No 1, January 2026 | Hal 118-124
DOI: 10.47065/jimat.v6i1.964
https://hostjournals.com/jimat

Perpustakaan Sma Negeri 3 Pekanbaru," *Prosiding-Seminar Nas. Teknol. Inf. Ilmu Komput.*, vol. 2, no. 1, pp. 313–323, 2023, url: https://journal.unilak.ac.id/index.php/Semaster/article/view/18631

[12] D. Aulia, M. Safii, and D. Suhendro, "Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja Aparatur Sipil Negera di Sekretariat DPRD Pematangsiantar," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 47, 2021, doi: 10.30645/jurasik.v6i1.270.

[13] G. B. Kaligis and S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai," *IT-Explore J. Penerapan Teknol. Inf. dan Komun.*, vol. 1, no. 3, pp. 179–193, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.

[14] C. S. Odionu, B. Bristol-Alagbariya, and R. Okon, "Big data analytics for customer relationship management: Enhancing engagement and retention strategies," *Int. J. Sch. Res. Sci. Technol.*, vol. 5, no. 2, pp. 50–67, 2024, doi: 10.56781/ijsrst.2024.5.2.0039

[15] S. Bose, S. K. Dey, and S. Bhattacharjee, "Big data, data analytics and artificial intelligence in accounting: An overview," *Handb. big data Res. methods*, pp. 32–51, 2023, doi: 10.4337/9781800888555.00007

[16] S. Sharifymoghaddam *et al.*, "Rankllm: A python package for reranking with llms," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025, pp. 3681–3690, doi: 10.1145/3726302.3730331

[17] M. Herviany, S. Putri Delima, T. Nurhidayah, and Kasini, "Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60

[18] K. C. Di, R. Sakit, W. Ngawi, H. Dilawati, H. Widianto, and A. Kuswiadji, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Widodo Ngawi," *Teknol. Inf. dan Rekayasa Komput.*, vol. 5, no. 2, pp. 139–147, 2024, doi: 10.37148/bios.v5i2.134

[19] M. T. Hidayat, M. Arifin, and S. Muzid, "Prediction Sentiment Analysis Grab Reviews using SVM Linear Based Streamlit," *Indones. J. Comput. Cybern. Syst.*, vol. 19, no. 2, pp. 1–12, 2025, doi: 10.22146/ijccs.104924.

[20] N. Lozada, J. Arias-Pérez, and E. A. Henao-García, "Unveiling the effects of big data analytics capability on innovation capability through absorptive capacity: why more and better insights matter," *J. Enterp. Inf. Manag.*, vol. 36, no. 2, pp. 680–701, 2023, doi: 10.1108/JEIM-02-2021-0092

[21] T. Yang, Q. Xin, X. Zhan, S. Zhuang, and H. Li, "Enhancing financial services through big data and AI-driven customer insights and risk analysis," *J. Knowl. Learn. Sci. Technol. ISSN 2959-6386*, vol. 3, no. 3, pp. 53–62, 2024, doi: 10.60087/jklst.vol3.n3.p53-62

[22] T. Firmansyah, P. Poningsih, and S. R. Andani, "Analisis Clustering Algoritma K-Means Sebagai Rekomendasi Penambahan Koleksi Buku Di Perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun," *Zahra Bull. Big data, Data Sci. Artif. Intell.*, vol. 1, no. 1, pp. 44–48, 2022, url: https://ejurnal.pdsi.or.id/index.php/zahra/article/view/13

[23] N. L. Rane, M. Paramesha, S. P. Choudhary, and J. Rane, "Machine learning and deep learning for big data analytics: A review of methods and applications," *Partners Univers. Int. Innov. J.*, vol. 2, no. 3, pp. 172–197, 2024, doi: 10.5281/zenodo.12271006

[24] L. Theodorakopoulos and A. Theodoropoulou, "Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review," *Hum. Behav. Emerg. Technol.*, vol. 2024, no. 1, p. 3641502, 2024, doi: 10.1155/2024/3641502

This Journal is licensed under a Creative Commons Attribution 4.0 International License