J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology
Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

Segmentation and Prediction of Store Performance on the Shopee Marketplace Using a Hybrid Clustering Approach, Spatial Analysis, and Feature Importance

Eka Yuniar[1*], Sherin Ramadhania[2], Pascawati Savitri[3], Mas'ud Hermansyah[4], Akas Bagus Setiawan[5]

[1,4]*Politeknik Negeri Jember, Business, Digital Business, Jl. Mastrip, Sumbersari, Jember, Indonesia,*
[2]*Institut Teknologi Sumatera, Industrial Technology, Industrial Technology,Jl Terusan Ryacudu Lampung, Indonesia,*
[3]*Politeknik Negeri Jember, Business, International Marketing Management, Jl. Mastrip, Sumbersari, Jember, Indonesia,*
[5]*Politeknik Negeri Jember, Technology Information, Informatics Engineering, Jl. Mastrip, Sumbersari, Jember, Indonesia*

Keywords
Data mining, e-commerce, K-Means clustering, Random Forest, marketplace analytics.

****Corresponding Author:***
eka_yuniar@polije.ac.id

Abstract
Marketplace platforms have become a central component of digital commerce, particularly in Southeast Asia where Shopee has emerged as one of the dominant e-commerce ecosystems. This study aims to analyze and predict the performance of Shopee stores using a hybrid data mining approach integrating clustering, spatial analysis, and classification. The dataset consists of 655 Shopee stores collected on February 18, 2026. K-Means clustering is applied to segment store performance, while spatial analysis examines geographic distribution patterns. Furthermore, a Random Forest classifier is used to predict performance categories and identify influential features. The clustering results reveal three distinct store performance groups with a Silhouette Score of 0.6704, indicating a good cluster structure. Although K = 2 produced a higher score (0.7891), K = 3 was selected to provide more meaningful segmentation (low, medium, and high performance). The Random Forest model achieved an accuracy of 79%, with precision, recall, and F1-score demonstrating reliable predictive performance across all classes. Feature importance analysis shows that promotional activity, chat responsiveness, and follower count significantly influence store performance classification. Spatial analysis indicates that provinces such as West Java and Jakarta dominate high-performance clusters. The findings contribute to hybrid data mining frameworks and provide practical insights for improving seller competitiveness in digital commerce ecosystems.

1. Introduction

Marketplace platforms have become central to digital transactions in the modern economic era, particularly in Southeast Asia. Shopee has established a dominant position in the region, with a continuously increasing number of users and sellers each year, making it an important subject of research in the fields of information systems and data mining. The rapid growth of digital sellers on Shopee intensifies competition and highlights the necessity of data-driven approaches to evaluate and segment store performance effectively [1], [2].

Recent literature emphasizes that Artificial Intelligence (AI) has become a strategic enabler in e-commerce platforms. AI-driven personalization systems significantly enhance customer experience and operational efficiency [3]. Similarly, AI-powered recommendation and automation systems have been shown to improve firm performance and customer engagement in digital marketplaces [4].

Within the Shopee context, empirical studies confirm that AI implementation positively influences customer loyalty and digital trust [5]. AI-based chatbot services also significantly improve customer satisfaction and responsiveness [6]. Furthermore, seller attributes such as ratings, reviews, pricing strategies, and promotional intensity are critical determinants of purchase decisions [7].

From a broader digital commerce perspective, marketplace success is shaped by the interplay between technological capability, consumer trust, and data-driven marketing strategies [8]. These studies collectively highlight the transformative role of AI and analytics in shaping competitive advantage in e-commerce ecosystems.

Despite the increasing number of sellers on Shopee, there remains limited comprehensive understanding of store performance patterns. Seller performance is not solely determined by the number and type of products offered, but also by factors such as chat response rate, follower count, customer ratings, promotional activities, and store tenure. These multidimensional attributes require advanced analytical techniques capable of uncovering hidden structures and predictive relationships within the data.

Previous studies on Shopee have largely focused on customer sentiment analysis and product review classification using text mining techniques such as Naïve Bayes and TF-IDF. Other research has applied clustering methods, particularly K-Means, to analyze product segmentation within Shopee. However, studies integrating seller performance segmentation with spatial analysis and predictive feature importance remain limited. Most existing works treat clustering and classification as separate analytical processes without incorporating geographic distribution patterns or interpretability of predictive models [9].

To address this gap, this study proposes a hybrid data mining framework that integrates:

1. K-Means clustering to segment store performance based on activity indicators (number of products, chat response rate, promotional activity) and reputation metrics (rating, follower count, and store tenure);
2. Spatial analysis to examine the geographic distribution of performance clusters based on seller address data; and
3. Random Forest classification with feature importance to predict performance categories and identify the most influential determinants.

This research utilizes a dataset of 655 Shopee stores collected up to February 18, 2026. By combining unsupervised learning, spatial evaluation, and supervised classification, this study aims to provide a comprehensive understanding of seller performance patterns and their key determinants.

The findings are expected to contribute theoretically to the development of hybrid data mining frameworks in e-commerce research and practically to marketplace stakeholders by offering strategic insights for improving seller competitiveness through data-driven evaluation.

2. Research Method

The overall research workflow consists of four main stages:
(1) data collection and preprocessing,

(2) K-Means clustering for segmentation,
(3) spatial analysis for geographic distribution, and
(4) Random Forest classification for prediction and feature importance analysis.

Each stage is interconnected, where clustering results are used as labels for the classification model.

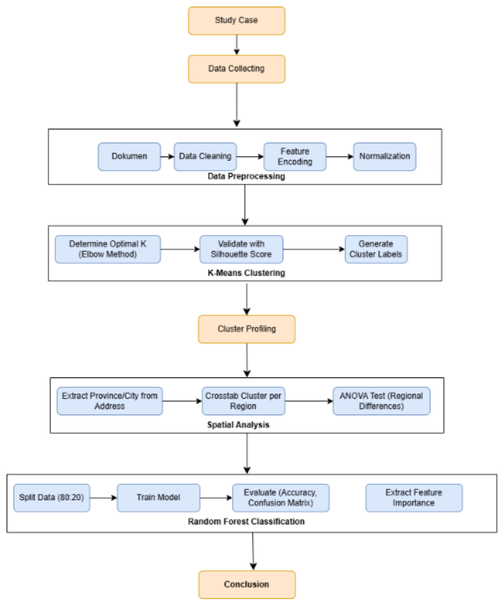

Figure 1 Research Method

2.1 Research Design

This study employs a quantitative, data-driven approach by integrating unsupervised learning (*clustering*), spatial analysis, and supervised learning (*classification*). The methodology follows three main stages: (1) data pre-processing and normalization, (2) segmentation using K-Means clustering, and (3) spatial and predictive analysis utilizing Random Forest classification with feature importance interpretation [10].

2.2 Data Collection and Pre-processing

The dataset consists of 655 sellers on the Shopee marketplace as of February 18, 2026, including attributes such as the number of products, chat response rate, follower count, store rating, length of tenure, promotional activity, and store address. Data pre-processing involves:

1. Data cleaning to handle missing values and outliers;
2. Feature encoding for categorical variables such as promotion indicators;
3. Normalization using Min-Max scaling to standardize the range of numerical attributes.

These preparation steps are consistent with best practices in data mining research for e-commerce data analysis and clustering applications, as discussed in recent studies on K-Means clustering in online retail contexts [11].

No	nama_toko	jumlah_produk	performa_chat	follower	rating_toko	penilai	alamat_toko	provinsi	promo_produk	produk_terjual	produk_favorit	gratis_ongkir
1	17Seven Original Official Shop	172	0.98	15900	4.8	4400	KAB. BANDUNG	Jawa Barat	0.32	83	116	Gratis
2	2nd Red Official Shop	401	0.98	18800	4.9	7900	KOTA JAKARTA UTARA	Jakarta	0.72	47	55	Gratis
3	308ABSLTUNSCRD Official Shop	638	0.72	283500	4.8	216300	KOTA JAKARTA UTARA	Jakarta	0.67	10000	2300	Gratis
4	3'NThree Official Shop	328	0.92	110900	4.8	16500	KAB. BANDUNG	Jawa Barat	0.40	155	545	Gratis
5	3Second Official Shop	1500	1.00	57500	4.8	21400	KOTA JAKARTA BARAT	Jakarta	0.20	114	47	Gratis

Figure 2 Initial Research Dataset

The dataset was collected using web scraping techniques from publicly available Shopee store information. Only non-sensitive and publicly accessible data were used. No personal or private data were collected, ensuring compliance with ethical standards. Data validity was ensured through manual verification and preprocessing steps.

2.3 K-Means Clustering for Store Segmentation

The primary segmentation method employed is K-Means clustering, a widely accepted partition-based algorithm that groups entities into homogenous clusters based on Euclidean distance in a multi-dimensional feature space. The optimal number of clusters (K) is determined using [12], [13]:

- Elbow Method, which provides visual information comparing the number of clusters that form a right angle at one point on the graph or the value that experiences the greatest decrease, then the cluster value is the best and by comparing the calculation of the Sum Square Error (SSE) value with the following equation [14] examines the rate of change in within-cluster sum of squares (WCSS);

$$SSE = \sum_{K=1}^{K} \sum |x_i - c_k|^2 \quad \text{......(1)}$$

- Silhouette Score An evaluation metric used to measure how well data points in a cluster match the cluster they belong to and how close or far they are from other clusters [15], [16].

$$S(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)} \quad \text{.....(2)}$$

The use of such quantitative validation metrics aligns with methodologies in recent e-commerce clustering studies [11], [17].

2.4 Spatial Analysis of Seller Distribution

Spatial analysis is conducted based on extracted location attributes (province and city) from the seller address field. Cluster distribution is evaluated for each region to identify geographic patterns in store performance. Statistical tests such as Analysis of Variance (ANOVA) are used to determine if performance metrics differ significantly among regions.

Recent research demonstrates the value of combining clustering with spatial analysis and predictive validation techniques to understand regional segmentation patterns, especially in digital adoption and performance profiling [18].

2.5 Random Forest Classification and Feature Importance

To assess the predictive capacity of identified clusters, a Random Forest classifier is trained using cluster labels as targets. Random Forest is selected due to its robustness against overfitting, capability to handle highdimensional data, and provision of feature importance scores that indicate the relative influence of each predictor on classification outcomes [19].

The process consists of:

- Splitting the dataset into training (80 %) and testing (20 %) subsets;
- Training the Random Forest model;
- Evaluating performance using accuracy, confusion matrix, and classification metrics;
- Extracting feature importance measures to interpret the influence of each variable.

The integration of classification and feature importance for predictive evaluation follows recent data mining research practices where hybrid methods improve interpretability and predictive insights [20].

2.6 Evaluation Metrics

The effectiveness of the segmentation and classification models is evaluated using:

- Silhouette coefficient for clustering quality;
- Classification accuracy and F1-score for predictive performance;
- Feature importance ranking to highlight dominant variables influencing performance categories;
- ANOVA results for spatial performance differences [21].

3. Result and Discussions

3.1 Clustering Performance Evaluation

The K-Means algorithm was applied to segment the performance of 655 Shopee stores based on key activity and reputation metrics, including number of products, chat response rate, follower count, store rating, tenure, and promotion usage. The optimal number of clusters was determined using the Elbow Method and validated with the Silhouette Coefficient. The Elbow plot indicated a clear bend at K = 3, and the Silhouette Score at K = 3 showed a satisfactory separation among clusters (Silhouette Score = 0.6704), indicating reliable cluster structure consistent with prior studies utilizing K-Means for customer segmentation in e-commerce settings.

The optimal number of clusters was determined using the Silhouette Coefficient method, testing clusters ranging from 2 to 9. The test results showed that the highest silhouette value was obtained at k = 2, with a score of 0.7891, followed by k = 3, with a score of 0.6704.

Although the highest silhouette score was achieved at K = 2 (0.7891), this study selected K = 3 to ensure better interpretability and practical relevance. A three-cluster model allows classification into low, medium, and high performance segments, which provides more actionable insights compared to binary segmentation.

Conceptually, dividing into three clusters provides a more meaningful business interpretation than two clusters, as it allows for the identification of emerging store segments in addition to flagship and low-performing stores.

3.2 Cluster Distribution and Characteristics

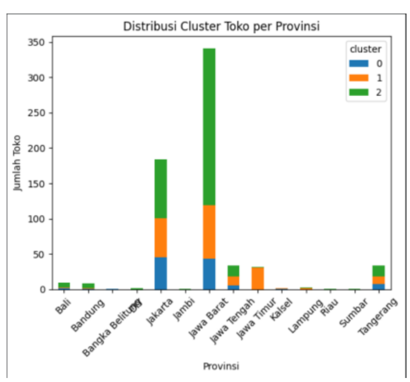

Figure 3 Distribution Cluster by Provinsi

Cluster profiling revealed three distinct performance groups among Shopee stores across several Indonesian provinces. The clustering results highlight variations in store activity levels, promotional engagement, and customer interaction across different geographical regions.

As illustrated in Figure X, the distribution of stores across clusters shows that Cluster 2 dominates the overall population, particularly in provinces with strong digital commerce ecosystems such as West Java and Jakarta. This cluster represents stores with relatively higher operational activity, including a larger number of products, stronger promotional utilization, and more consistent customer engagement.

Cluster 1 represents stores with moderate performance characteristics, where stores demonstrate an intermediate level of activity. These stores generally maintain acceptable levels of product offerings and engagement metrics but do not exhibit the same level of intensity observed in the high-performing group. In the spatial distribution, this cluster appears in several provinces including Central Java and Tangerang, suggesting a transitional segment of sellers that may have growth potential within the marketplace ecosystem.

Meanwhile, Cluster 0 corresponds to the **lower activity segment**, where stores tend to have fewer products, lower promotional intensity, and relatively limited customer interaction. This cluster is distributed in smaller numbers across most provinces, indicating that low-performance stores exist but do not dominate the marketplace landscape.

From a spatial perspective, West Java shows the highest concentration of stores across all clusters, reflecting the province's strong participation in Indonesia's digital commerce ecosystem. Jakarta also exhibits a high density of stores, particularly in Cluster 2, indicating the presence of more competitive and mature online sellers. Other regions such as Bali, Lampung, and Kalimantan Selatan contribute a smaller share of stores, which may reflect differences in digital adoption, logistics infrastructure, or market maturity.

Overall, the spatial cluster distribution suggests that regions with stronger digital infrastructure and higher economic activity tend to host a larger proportion of active and high-performing online stores. However, the presence of all three clusters across multiple provinces indicates that marketplace participation is geographically widespread, even though performance levels vary.

These findings demonstrate that store performance segmentation is influenced not only by internal operational factors but also by regional digital ecosystem development, which plays a role in shaping the competitive dynamics of online marketplaces such as Shopee.

The selection of K = 3 is not solely based on the silhouette score but also considers interpretability and practical relevance. While K = 2 provides higher compactness, it oversimplifies store performance into binary categories. In contrast, K = 3 enables the identification of emerging stores (medium segment), which is crucial for strategic decision-making in marketplace development. Therefore, this study prioritizes a balance between statistical validity and business interpretability.

3.3 Classification Performance Evaluation

The Random Forest classifier was evaluated using accuracy, precision, recall, F1-score, and confusion matrix.
The model achieved an accuracy of 79%, indicating strong predictive performance.
The classification results show:

- Class 0: Precision = 0.82, Recall = 1.00, F1-score = 0.90
- Class 1: Precision = 0.60, Recall = 0.47, F1-score = 0.53
- Class 2: Precision = 0.85, Recall = 0.88, F1-score = 0.86

The confusion matrix demonstrates that most predictions are correctly classified, with minor misclassification occurring between adjacent clusters.
These results confirm that the Random Forest model is effective in predicting store performance categories.

The hybrid framework used in this study, including K-Means for unsupervised segmentation, spatial mapping, and Random Forest for supervised prediction and feature importance analysis, provides significant advancements to the existing literature. This approach fills a specific gap in research focused on store or seller performance on the Shopee platform, which has rarely been explored in an integrated manner.

Most previous research has focused solely on customer or product segmentation [22], [23], or applying clustering methods without prediction validation and spatial context. Several hybrid clustering studies with supervised learning have also been applied more to predicting customer churn than to salesperson performance [24], [25]. By integrating these three elements (clustering, spatial analysis, and predictive classification), this study's results show that store performance is influenced by both internal factors (such as promotional intensity and chat responsiveness) and external geographic factors (such as concentration in West Java and DKI Jakarta). This approach provides a more comprehensive understanding than previously commonly used single-technique methods.

The 79% prediction accuracy and resulting feature importance rankings provide actionable guidance for Shopee sellers. These findings align with the practical implications of hybrid churn prediction research using Random Forest, where the model consistently demonstrated high performance in classification and interpretability [26]. This research extends this application to marketplace seller competitiveness, not just customer retention. Limitations of previous research, such as the lack of spatial analysis of Shopee-specific clustering or the lack of seller-level data in spatial e-commerce typologies, have been directly addressed in this study.

4. Conclusions and Future Works

The clustering results using the K-Means algorithm reveal the existence of three distinct store performance segments. These clusters represent different levels of operational performance within the marketplace ecosystem, namely low-performing stores, moderately performing stores, and high-performing stores. The segmentation results demonstrate that marketplace sellers exhibit heterogeneous operational characteristics in terms of product offerings, customer engagement, promotional activities, and store reputation.

The spatial distribution analysis further indicates that store performance is not evenly distributed across regions. Provinces such as West Java and Jakarta show a higher concentration of active and competitive stores compared to other regions. This finding suggests that regional digital ecosystem maturity, infrastructure availability, and economic activity play an important role in shaping the development of online marketplace businesses.

In addition, the feature importance analysis highlights several key variables that significantly influence store performance classification. Variables such as number of products, store rating, chat responsiveness, promotional activity, and follower count emerge as dominant factors in determining store competitiveness. These findings indicate that both operational capacity and customer engagement strategies contribute to the success of marketplace sellers.

For future research, further studies may incorporate larger datasets, temporal transaction data, and additional machine learning models such as deep learning or graph-based analysis to better capture the dynamics of marketplace performance over time.

5. References

[1] Reuters, “Google, Shopee-owner Sea to develop AI tools for e-commerce, gaming,” *www.reuters.com*, 2026.

[2] B. Yáñez-Araque, J. P. S.-I. Hernández, S. Gutiérrez-Broncano, and P. Jiménez-Estévez, “Corporate social responsibility in micro-, small- and medium-sized enterprises: Multigroup analysis of family vs. nonfamily firms,” *J. Bus. Res.*, vol. 124, pp. 581–592, 2021, doi: https://doi.org/10.1016/j.jbusres.2020.10.023.

[3] H. Li *et al.*, “Flash Flood Risk Classification Using GIS-Based Fractional Order k -Means Clustering Method,” *MDPI Fractal Fract. J.*, vol. 9, pp. 1–18, 2025, doi: https://doi.org/10.3390/fractalfract9090586.

[4] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. Slade, and A. Jeyaraj, “Opinion Paper: ‘So what if ChatGPT wrote it?’ Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy,” *Int. J. Inf. Manage.*, vol. 71, pp. 1–63, 2023, doi: https://doi.org/10.1016/j.ijinfomgt.2023.102642.

[5] S. R. Rabani, D. Amalia, M. Erika, N. Kusuma, and F. Ilayana, "Pengaruh Penggunaan AI , Literasi Digital , dan Pengalaman Pengguna Terhadap Loyalitas Pelanggan Pada E-Commerce Shopee," *J. Econ. Bus. Res.*, vol. 3, no. 2, pp. 147–159, 2024, doi: https://doi.org/10.22515/juebir.v3i2.10813.

[6] A. M. Alghaniy, "The Impact of Artificial Intelligence Technology in Shopee's Chatbot Service on Customer Satisfaction in Greater Bandung Area, Indonesia," *Int. J. Adm. Bus. Organ.*, vol. 5, no. 1, pp. 48–55, 2024, doi: https://doi.org/10.61242/ijabo.24.337.

[7] A. Muslikhun and S. Sutopo, "Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Online di Marketplace Shopee," *J. Transform. Bisnis Digit.*, vol. 1, no. 4, pp. 11–24, 2024, doi: https://doi.org/10.61132/jutrabidi.v1i4.202.

[8] P. Bicen, S. Hunt, and S. Madhavaram, "Coopetitive innovation alliance performance: Alliance competence, alliance's market orientation, and relational governance," *J. Bus. Res.*, vol. 123, pp. 23–31, 2021, doi: https://doi.org/10.1016/j.jbusres.2020.09.040.

[9] Y. A. Wijaya and D. Sudrajat, "Analisis Bibliometrik: Pemetaan Penelitian Machine Learning dalam E-commerce Berdasarkan Data dari Scopus (2019-2024)," in *Prosiding Seminar Nasional Sisfotek (Sistem Informasi dan Teknologi Informasi)*, 2024, pp. 451–461.

[10] A. Shojaei, "Data Mining Systematic Literature Review," 2024. doi: https://doi.org/10.13140/RG.2.2.14684.40324.

[11] L. A. Putri, M. Tsaqofah, D. S. Hasibuan, H. Fadillah, M. Ulfa, and M. Furqan, "Application of K-Means Clustering Algorithm for E- Commerce Data Analysis," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 3, pp. 5–8, 2025, doi: https://doi.org/10.59934/jaiea.v4i3.1170.

[12] A. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: https://doi.org/10.1016/j.patrec.2009.09.011.

[13] B. Neupane *et al.*, "Machine learning algorithms for supporting life cycle assessment studies: An analytical review," *Sustain. Prod. Consum.*, vol. 56, pp. 37–53, 2025, doi: https://doi.org/10.1016/j.spc.2025.03.015.

[14] R. Siagian, P. Sirait, and A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm and Length, Recency, Frequency, Monetary Model," *JITE (Journal Informatics Telecommun. Eng. Available*, vol. 5, no. 1, pp. 21–30, 2021, doi: 10.31289/jite.v5i1.5182 Received:

[15] F. Muttaqien, N. Fitria, V. L. Rizki, and I. Abrori, "Pengaruh Kompetensi, Program Diklat, Dan Motivasi Kerja Terhadap Peningkatan Kinerja Karyawan Pt. Bpr Nur Semesta Indah Kabupaten Jember," *J. Istiqro*, vol. 11, no. 2, pp. 107–123, 2025, doi: 10.30739/istiqro.v11i2.4119.

[16] S. Wahyuni, T. T. Wulansari, and F. Fahrullah, "Segmentasi Pelanggan Berdasarkan Analisis Recency, Frequency, Monetary Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group," *J. Rekayasa Teknol. Inf.*, vol. 7, no. 2, pp. 180–187, 2023, doi: http://dx.doi.org/10.30872/jurti.v7i2.8748.

[17] R. Setyawan and B. Murtiyasa, "A Systematic Literature Review of Clustering Algorithms in Stock Market Analysis," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 08, no. 1, pp. 36–52, 2026, doi: https://doi.org/10.47709/cnahpc.v8i1.7333.

[18] T. A. N. Azzikra, "Segmentasi Wilayah Digitalisasi di Indonesia dengan DBSCAN dan Validasi menggunakan Random Forest," *Digit. Transform. Technol.*, vol. 5, no. 2, pp. 85–91, 2025, doi: https://doi.org/10.47709/digitech.v5i2.6532.

[19] A. Khairunnisa, K. A. Notodiputro, and B. Sartono, "A Comparative Study of Random Forest and Double Random Forest Models from View Points of Their Interpretability," *Sci. J. Informatics*, vol. 11, no. 1, pp. 207–218, 2024, doi: 10.15294/sji.v11i1.48721.

[20] J. Ipmawati and K. Kusnawi, "Integration of K-Means Clustering, Random Forest, and RFM Analysis for Optimizing Consumer Segmentation in Digital Advertising Strategies," *J. SISFOKOM (Sistem Inf. dan*

Komputer), vol. 15, no. 1, pp. 112–118, 2026, doi: 10.32736/sisfokom.v15i1.2548.

[21] B. N. Yulisasih, H. Herman, S. Sunardi, and H. Yuliansyah, "Predictive Analytics on Shopee for Optimizing Product Demand Prediction through K-Means Clustering and KNN Algorithm Fusion," Journal of Information Systems and Informatics," *Ilk. J. Ilm.*, vol. 16, no. 3, pp. 330–342, 2024, doi: https://doi.org/10.33096/ilkom.v16i3.2325.330-342.

[22] M. Febima and L. Magdalena, "Predictive Analytics on Shopee for Optimizing Product Demand Prediction through K-Means Clustering and KNN Algorithm Fusion," *J. Inf. Syst. Informatics*, vol. 6, no. 2, pp. 751–765, 2024, doi: 10.51519/journalisi.v6i2.720.

[23] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *MDPI Sustain.*, vol. 14, pp. 1–15, 2022, doi: https://doi.org/10.3390/su14127243.

[24] Z. R. Li, "Customer Segmentationand Churn Prediction Based On K-Means And Random Forest: Acase Study Of E-Commerce Data," *Eurasia J. Sci. Technol.*, vol. 7, no. 2, pp. 14–19, 2025.

[25] U. I. Hartanto, I. G. P. A. Buditjahjanto, and W. Yustanti, "Hybrid Clustering and Classification of At-Risk Customer Segments in Network Marketing," *J. Inf. Eng. Educ. Technol.*, vol. 9, no. 1, pp. 42–50, 2025.

[26] M. Ali and M. Hussain, "Machine Learning-Based Customer Churn Prediction for E-Commerce Businesses," *Preprint*, pp. 1–8, 2025, doi: 10.20944/preprints202511.0735.v1.