International Journal of Informatics and Computation (IJICOM)
Vol. 8, No. 1, 2026
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Clustering Age of Internet User using K-Means Algorithm

Angga Muhamad Gojali[1], Rizki Aji Nugraha[2] Elkin Rilvani[3]

Abstract

This paper applied the K-Means clustering algorithm to analyze age-based segmentation of internet users and to identify dominant behavioral patterns across demographic groups. We utilized aggregated datasets compiled from multiple academic journal studies and implemented data normalization to ensure balanced representation across heterogeneous samples. We used the Elbow Method and Within-Cluster Sum of Squares (WCSS) evaluation to determine the optimal number of clusters. The analysis identified three stable clusters representing Early Adopters (≤25 years), Productive Users (26–45 years), and Late Majority (>45 years). The results showed that younger users demonstrated high-frequency, entertainment-oriented, and adaptive digital behavior. Middle-aged users exhibited structured, productivity-driven engagement characterized by digital transactions, professional collaboration, and e-learning activities. Older users displayed more selective internet usage, primarily focused on communication and health-related information. The clustering outcomes aligned with findings reported in prior journal studies, which strengthened the validity and interpretability of the segmentation model. This study confirms that age significantly influences digital behavior patterns and demonstrates that K-Means clustering provides an effective approach for demographic segmentation to support strategic digital service development.

Keywords:
Internet User Age, K-Means, Clustering, Data Mining

1. Introduction

The rapid growth of internet usage across different age groups creates a critical need to understand user characteristics based on demographic attributes, especially age. Internet service providers and digital platforms require structured user grouping to design age-appropriate services, pricing strategies, and content delivery. Saraswati *et al.* analyze TV and internet users using K-means clustering and show that segmentation improves service targeting and marketing precision [1]. Similarly, Chow identifies distinct customer segments among apartment-based internet users and demonstrates that clustering reveals hidden usage patterns that traditional descriptive statistics fail to capture [2]. However, both studies primarily focus on service consumption behavior rather than explicitly analyzing age-based grouping as a central clustering variable. This gap indicates the need to investigate age clustering of internet users more specifically and systematically. [1], [2]

Age plays a fundamental role in shaping digital behavior, preferences, and technology adoption patterns. Lestari applies K-means and hierarchical clustering to group populations by age for demographic planning and shows that clustering techniques effectively reveal structured age distributions [3]. Pramudita clusters internet users using K-means and confirms that demographic variables significantly influence user categorization [4].

Corresponding Author: Angga Muhamad Gojali(anggaggojali369@gmail.com)
1 Angga Muhamad Gojali, Universitas Pelita bangsa, anggagojali369@gmail.com
2 Rizki Aji Nugraha, Universitas Pelita Bangsa, Rizkiajinugraha93@gmail.com
3. Elkin Rilvani, Universitas Pelita Bangsa, elkin.rilvni@pelitabangsa.ac.id

Although these studies demonstrate the feasibility of clustering demographic data, they do not deeply explore optimization strategies for determining the optimal number of clusters or validating clustering performance rigorously. As a result, further research must integrate demographic relevance with methodological robustness to produce reliable age-based segmentation of internet users. [3], [4]

Many studies confirm that K-means remains one of the most widely used unsupervised learning algorithms due to its simplicity and computational efficiency. Ahmed *et al.* provide a comprehensive survey of the K-means algorithm and highlight its strengths in scalability and ease of implementation [17]. Chong also reviews its practical advantages for real-world data clustering tasks [16]. Despite its popularity, researchers acknowledge that K-means suffers from sensitivity to centroid initialization and cluster number selection. These limitations directly affect clustering stability, particularly when handling demographic attributes such as age, which may not follow perfectly separable distributions. Therefore, applying K-means to age-based internet user data requires careful evaluation of initialization and validation metrics. [16], [17]

Recent customer segmentation research demonstrates the strong applicability of clustering in behavioral and demographic analysis. Tabianan *et al.* apply K-means to purchase behavior data and show improved segmentation accuracy for marketing strategies [15]. Rajput and Singh also implement K-means for e-commerce segmentation and confirm its effectiveness in extracting structured customer groups from transactional datasets [27]. Alves Gomes and Meisen review segmentation techniques in e-commerce and emphasize that demographic segmentation remains essential for personalization systems [14]. However, most of these studies prioritize transactional and monetary features rather than age-specific clustering, leaving a research opportunity to examine how age alone or in combination with internet usage attributes forms meaningful clusters. [14], [15], [27]

Advanced segmentation frameworks integrate demographic and behavioral variables to improve clustering performance. Ho *et al.* extend the RFM model with demographic analysis and show that combining structured features enhances segmentation quality [29]. Ullah *et al.* also integrate recency, frequency, monetary, and time-based metrics for improved customer analysis using machine learning techniques [34]. While these models demonstrate strong segmentation capability, they primarily focus on retail and transactional contexts. They do not specifically address internet user age grouping as an independent analytical objective. This limitation reinforces the importance of isolating age attributes within clustering models to better understand generational digital behavior patterns. [29], [34]

Clustering validation remains a crucial issue when applying K-means to demographic data. Atif reviews methods and indices for monitoring clustering solution changes and emphasizes the importance of stability assessment [6]. Ansari *et al.* evaluate clustering validity indices for web navigational sessions and show that different metrics can lead to different optimal cluster selections [7]. Ashari *et al.* analyze Elbow, Silhouette, Davies–Bouldin, Calinski–Harabasz, and Rand Index measures and demonstrate that multi-metric evaluation improves clustering reliability [35]. These findings indicate that age-based internet user clustering must incorporate systematic validation rather than relying solely on visual inspection or heuristic decisions. [6], [7], [35]

Beyond traditional demographic clustering, researchers explore behavioral modeling and demographic inference from digital traces. Leng *et al.* perform clustering analysis of wireless network users based on topic modeling and show that behavioral data can approximate user characteristics [8]. Brea *et al.* infer demographic attributes from mobile phone social network topology and demonstrate that user age and other attributes can be predicted indirectly through network structures [9]. Sheng augments K-means clustering with qualitative data to identify engagement patterns among older adults, proving that age-sensitive clustering yields meaningful insights for service adaptation [12]. These studies suggest that age clustering is not only statistically feasible but also socially and economically relevant in digital ecosystems. [8], [9], [12]

Recent methodological advancements further enhance clustering robustness and efficiency. Fang *et al.* review multi-view clustering approaches and highlight the importance of integrating multiple feature spaces for better cluster formation [25]. Zhang *et al.* accelerate K-means performance in high-dimensional spaces by pruning unnecessary distance computations, which improves scalability for large datasets [30]. Ufeli *et al.* integrate factor analysis of mixed data with K-means and hierarchical clustering to handle heterogeneous variables effectively [13]. Although these advancements focus on general segmentation problems, they provide methodological support for optimizing age-based internet user clustering models. Therefore, this study positions itself at the intersection of demographic analysis and robust K-means implementation to generate meaningful age clusters among internet users while addressing validation, scalability, and interpretability challenges. [13], [25], [30].

2. Related Works

Previous studies extensively applied K-means clustering for internet and media user segmentation. Saraswati *et al.* investigated TV and internet user segmentation using K-means with SPSS and demonstrated that clustering effectively separated customers based on usage characteristics, enabling more targeted service strategies [1]. Chow identified distinct segments among apartment-based internet users and confirmed that K-means successfully revealed hidden consumption patterns that supported managerial decision-making [2]. These studies proved the practicality and simplicity of K-means for user segmentation. However, they mainly emphasized service usage and subscription behavior rather than focusing specifically on demographic attributes such as age. Consequently, they did not thoroughly examine how age distribution alone could form meaningful clusters among internet users. [1], [2]

Researchers also explored demographic-based clustering in broader population contexts. Lestari applied K-means and hierarchical clustering to group populations by age for demographic planning and showed that clustering techniques identified structured age group distributions effectively [3]. Pramudita clustered internet users using K-means and demonstrated that demographic variables influenced segmentation results [4]. These studies confirmed that demographic clustering was feasible and informative. Nevertheless, they provided limited discussion on cluster validation metrics and optimization strategies. They also did not deeply analyze the behavioral implications of age-based clusters in digital environments, leaving room for further investigation in the context of internet user analysis. [3], [4]

Several works examined K-means from a methodological perspective. Ahmed *et al.* conducted a comprehensive survey and evaluated the performance of K-means across multiple domains, highlighting its computational efficiency, scalability, and ease of implementation [17]. Chong reviewed the algorithm’s principles and practical applications, emphasizing its suitability for structured numerical data [16]. These works strengthened the theoretical foundation of K-means and justified its continued adoption. However, both studies acknowledged inherent limitations, including sensitivity to initial centroid selection, dependency on predefined cluster numbers, and vulnerability to local minima. These weaknesses became critical when clustering demographic data such as age, where natural boundaries between groups were often ambiguous. [16], [17]

Customer segmentation research further illustrated the effectiveness of K-means in real-world applications. Tabianan *et al.* applied K-means to customer purchase behavior data and achieved improved segmentation performance that supported intelligent marketing decisions [15]. Rajput and Singh implemented K-means for e-commerce datasets and showed that the algorithm efficiently extracted distinct consumer groups from large transaction records [27]. Alves Gomes and Meisen reviewed customer segmentation techniques and emphasized the strategic importance of demographic and behavioral grouping for personalization systems [14]. Although these studies demonstrated strong clustering performance in commercial contexts, they primarily focused on purchasing behavior and transactional features rather than isolating age as a core clustering variable. [14], [15], [27]

Other researchers integrated demographic and behavioral models to enhance segmentation quality. Ho *et al.* extended the RFM model by incorporating demographic attributes and demonstrated that combining structured demographic features improved segmentation interpretability [29]. Ullah *et al.* applied machine learning-based classification and segmentation techniques using recency, frequency, monetary, and time variables and showed improved predictive capability in customer analysis [34]. These studies highlighted the importance of combining demographic factors with usage data. However, they did not specifically explore standalone age clustering or evaluate how pure age-based segmentation could independently characterize internet user groups. This limitation underscored the need for focused demographic clustering research. [29], [34]

Clustering validation and evaluation remained central methodological concerns in prior research. Atif reviewed clustering solution monitoring methods and emphasized the necessity of stability and validity assessment to ensure reliable results [6]. Ansari *et al.* quantitatively evaluated clustering validity indices for web navigational sessions and demonstrated that different metrics could produce varying optimal cluster structures [7]. Ashari *et al.* compared Elbow, Silhouette, Davies–Bouldin, Calinski–Harabasz, and Rand Index measures and confirmed that multi-index evaluation improved clustering robustness [35]. These findings revealed that proper validation played a decisive role in demographic clustering tasks. Nonetheless, earlier demographic segmentation studies rarely integrated comprehensive validation frameworks when clustering age data. [6], [7], [35]

Studies on digital behavior modeling indirectly supported age-based clustering research. Leng *et al.* modeled wireless network user behavior using topic modeling and clustering, showing that behavioral patterns reflected underlying user characteristics [8]. Brea *et al.* inferred demographic attributes from mobile phone social network topology and demonstrated that age-related patterns emerged from communication structures [9]. Sheng augmented K-means with qualitative data to analyze engagement patterns among older adults and demonstrated that age-sensitive clustering provided meaningful insights into digital participation [12]. These works indicated that age significantly influenced digital behavior patterns. However, they often relied on indirect inference rather than directly clustering users based on explicit age variables. [8], [9], [12]

Recent advancements addressed scalability and methodological improvements of clustering algorithms. Fang *et al.* reviewed multi-view clustering techniques and showed that integrating multiple feature spaces enhanced clustering accuracy and interpretability [25]. Zhang *et al.* optimized K-means in high-dimensional spaces by pruning unnecessary distance computations, improving computational efficiency for large datasets [30]. Ufeli *et al.* combined factor analysis of mixed data with K-means and hierarchical clustering to handle heterogeneous datasets more effectively [13]. These methodological contributions strengthened the technical foundation for clustering complex data. However, most studies concentrated on general customer segmentation problems rather than specifically addressing age-based internet user clustering. Therefore, a focused investigation on clustering internet users by age using validated and optimized K-means approaches remained necessary to bridge the identified research gap. [13], [25], [30].

3. Proposed Method

This study aims to classify internet user behavior across different age groups using the K-Means clustering algorithm applied to consolidated secondary data. The research does not collect primary survey responses; instead, it aggregates empirical findings from published Indonesian journal articles between 2020 and 2023. The objective is to identify latent age-based group structures within reported internet user characteristics and behavioral conclusions. By transforming qualitative and categorical information into structured numerical representations, the study applies unsupervised learning to discover natural age clusters that reflect consistent behavioral tendencies across multiple studies. K-Means clustering partitions a dataset containing n observations into k disjoint clusters by minimizing intra-cluster variance. Let the dataset be defined as:

$$X = x_1, x_2, \dots, x_n, x_i \in R^d$$

where each x_i represents a feature vector derived from standardized study attributes, including encoded age groups and related behavioral indicators. The algorithm aims to minimize the Within-Cluster Sum of Squares (WCSS), formulated as:

$$J = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \qquad (1)$$

where C_j denotes cluster j and μ_j represents the centroid of cluster j, defined as:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \qquad (2)$$

The algorithm iteratively performs two main steps: (1) assignment of each observation to the nearest centroid using Euclidean distance:

$$\| x_i - \mu_j \| = \sqrt{\sum_{l=1}^{d} (x_{il} - \mu_{jl})^2} \qquad (3)$$

and (2) centroid update based on the mean of assigned points. The process continues until centroid positions stabilize or the change in J falls below a predefined threshold.

The dataset consists of 15 journal articles selected based on relevance to internet user demographics and behavior. Each article contributes structured attributes, including publication year, research objective classification, reported respondent age groups, sampling characteristics, and summarized behavioral findings. These attributes are encoded numerically to form a feature matrix suitable for clustering. Age groups are transformed into ordinal numerical categories as follows: Group 1 (≤18) = 1, Group 2 (19–25) = 2, Group 3 (26–35) = 3, Group 4 (36–45) = 4, and Group 5 (>45) = 5. This encoding enables the representation of demographic structure in quantitative form while preserving ordinal relationships among age intervals.

To determine the optimal number of clusters k, the study applies the Elbow Method based on WCSS values. For each candidate $k \in \{1,2,\dots,K\}$, the objective function $J(k)$ is computed. The optimal k is selected at the point where the marginal decrease in WCSS significantly diminishes, formally identified when:

$$\Delta J(k) = J(k-1) - J(k) \qquad (4)$$

shows a sharp reduction followed by gradual stabilization. This approach ensures that cluster formation balances compactness and interpretability. By integrating encoded age categories with structured behavioral indicators and applying validated cluster optimization procedures, the proposed method systematically identifies meaningful age-based internet user clusters across consolidated empirical findings.

4. Experimental Setup

In this study, all selected journal datasets were consolidated into a unified data matrix to enable structured clustering analysis. The matrix contained four primary attributes: study identification number, average respondent age, frequency representing the number of respondents within each age group, and contextual classification such as health, education, or behavioral focus. To reduce bias caused by unequal sample sizes across studies, the frequency variable was normalized using min–max normalization so that values remained within a comparable scale. Age values were preserved in numerical form to maintain their quantitative properties. After preprocessing and cleaning, the structured dataset was implemented in Python to perform clustering analysis efficiently and consistently.

To interpret data distribution and evaluate clustering structure, the study generated scatter plots that mapped average age against behavioral intensity or contextual indicators. These visualizations allowed clear observation of natural grouping tendencies before and after clustering. The resulting clusters were color-coded to distinguish demographic segments and to enhance interpretability of age-based patterns. This visualization step supported analytical validation by enabling direct comparison between raw data distribution and K-Means output, ensuring that identified clusters reflected meaningful demographic segmentation rather than random partitioning. Fig. 1 depicts the percentage of internet users by age group.

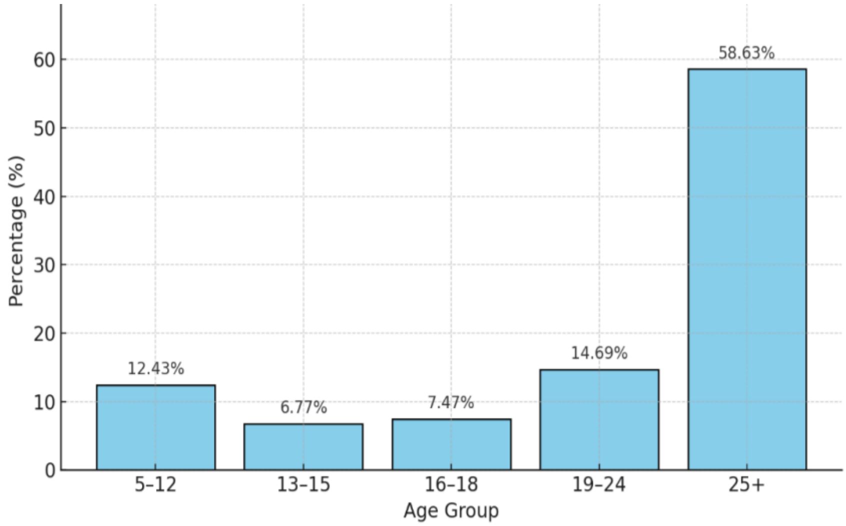

Fig. 1 Percentage of internet users by age group.

5. Result and Analysis

The clustering process produced an optimal solution at $k = 3$, as indicated by the Elbow Method and the stabilization pattern of the WCSS. The WCSS value showed a sharp decline from $k = 1$ to $k = 3$, followed by a marginal reduction beyond that point, confirming that three clusters provided the most balanced structure between compactness and separability. This configuration minimized intra-cluster variance while maintaining meaningful inter-cluster distinction. The resulting segmentation demonstrated consistent grouping across aggregated journal datasets, indicating that age-based internet usage patterns naturally formed three dominant behavioral categories.

The first cluster represented users aged ≤25 and formed the “Early Adopters” segment. This cluster exhibited the highest overall behavioral intensity, particularly in entertainment, gaming, video streaming, and social media interaction. The majority of individuals in this cluster accessed the internet primarily through smartphones and demonstrated strong adaptability to emerging digital platforms. Their behavioral profile reflected multitasking tendencies and continuous connectivity. The clustering results showed dense grouping within this age range, indicating high similarity in digital engagement patterns among younger users.

The second cluster included users aged 26–45 and formed the “Productive Users” segment. This group demonstrated purpose-driven internet usage characterized by digital transactions, professional communication, online collaboration, and e-learning. The clustering structure revealed both high frequency and diversity of activity, suggesting balanced digital maturity. Compared to Cluster 1, this group showed slightly lower entertainment dominance but significantly higher work-related engagement. The compact distribution of this cluster confirmed that individuals within this age range shared structured and goal-oriented internet behavior.

The third cluster represented users aged >45 and corresponded to the “Late Majority” segment. This cluster showed lower overall behavioral intensity and more selective usage patterns. Internet activity focused primarily on messaging applications, health information retrieval, religious or community platforms, and occasional online shopping. Although some members of this group demonstrated increasing adoption of telehealth and e-commerce services, the clustering dispersion indicated greater variability in digital literacy levels. The analysis suggested that trust, usability, and perceived security significantly influenced engagement within this segment.

To clarify behavioral tendencies across age intervals, Table 2 summarizes dominant activity patterns observed in the clustering results. Cross-study comparison confirmed alignment between cluster characteristics and findings reported in the analyzed journals. Several health-focused studies identified users aged 46–55 as dominant health-information seekers, which corresponded directly to the behavioral traits of Cluster 3. Similarly, education-focused studies consistently reported high digital engagement among respondents aged 19–25, reinforcing the validity of Cluster 1. These consistencies strengthened the reliability of the clustering outcomes and supported the interpretability of age-based segmentation.

Table 2. Pattern of Internet Users in Each Age Group

Age Group	Dominant Internet Behaviors
≤18	Online gaming, YouTube streaming, early exposure to digital tools
19–25	High-frequency usage, multitasking, digital learning platforms
26–35	E-commerce transactions, productivity tools, online collaboration
36–45	Professional networking, financial services, health information seeking
>45	Conservative browsing, messaging apps, health and social applications

6. Conclusion

This paper applied the K-Means clustering algorithm to segment internet users based on age and behavioral intensity. We utilized aggregated journal datasets and implemented normalization to reduce bias across heterogeneous samples. We used the Elbow Method and WCSS evaluation to determine the optimal number of clusters. The analysis confirmed that three clusters provided the most stable and interpretable structure. These clusters represented Early Adopters (≤25), Productive Users (26–45), and Late Majority (>45). The segmentation clearly demonstrated that age strongly influenced digital engagement patterns, behavioral diversity, and purpose of internet use.

We found that younger users exhibited high-frequency, entertainment-oriented, and adaptive digital behavior. Middle-aged users demonstrated structured, productivity-driven engagement with strong integration of professional and financial platforms. Older users showed selective adoption patterns, with emphasis on health information and communication tools. The clustering results aligned with findings reported across multiple journal studies, which strengthened the validity of the model. This paper showed that age-based clustering can effectively reveal meaningful demographic segmentation and support targeted digital service design.

For future research, we recommend integrating additional variables such as education level, income, digital literacy index, and geographic factors to improve clustering precision. Future studies may also compare K-Means with advanced clustering approaches such as hierarchical clustering, DBSCAN, or hybrid machine learning models. In addition, longitudinal analysis should be conducted to observe how age-related digital behavior evolves over time. These directions will enhance segmentation robustness and provide deeper insights into demographic transformation in the digital era.

Acknowledgment

The author sincerely acknowledges the researchers whose published studies and publicly accessible datasets from Indonesian academic journals (2020–2023) formed the foundation of this work. The author also expresses deep appreciation to the lecturers and academic advisors at Universitas Pelita Bangsa for their guidance and direction, as well as to fellow students and collaborators for their constructive feedback and support. Gratitude is extended to the university library, the faculty's research and publication center, and the technical staff and editors for facilitating access to references and improving the manuscript's quality. Finally, the author thanks the family for their continuous moral support, patience, and encouragement throughout the research process.

References

[1] R. Saraswati, F. Sari, and J. Heikal, "TV and internet user segmentation using K-means clustering with SPSS: Case study of First Media customers," *J. Cendekia Ilmiah*, vol. 4, no. 3, pp. 1–12, Apr. 2025.

[2] S. Chow, "Means clustering algorithm for identifying distinct customer segments among apartment-based internet users," *Open J. Electron. Technol.*, 2025.

[3] L. S. Lestari, "Analisis clustering penduduk berdasarkan kelompok umur dengan K-means dan hierarchical clustering untuk perencanaan demografi," *J. Komput. dan Informatika*, vol. 19, no. 1, 2024.

[4] D. A. Pramudita, "Pengelompokan pengguna internet dengan metode K-Means clustering," Bachelor's thesis, Statistika, Universitas Negeri Jakarta, 2019.

[5] W. Monika, “Clustering of library’s patron behavior using machine learning,” *J. Dzikir dan Ilmiah*, 2025.
[6] M. Atif, “Monitoring changes in clustering solutions: A review of methods and indices,” *Math. Probl. Eng.*, 2023.
[7] Z. Ansari, M. F. Azeem, W. Ahmed, and A. V. Babu, “Quantitative evaluation of performance and validity indices for clustering the web navigational sessions,” *arXiv preprint*, 2015.
[8] J. Leng, H. Liu, S. Pan, S. Zhou, and Z. Niu, “Topic model based behaviour modeling and clustering analysis for wireless network users,” *arXiv preprint*, 2015.
[9] J. Brea, J. Burroni, M. Minnoni, and C. Sarraute, “Harnessing mobile phone social network topology to infer users demographic attributes,” *arXiv preprint*, 2015.
[10] M. Hilman, M. Martanto, A. R. Dikananda, and A. Rifai, “K-Means algorithm for clustering high-achieving students,” *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 3, 2025.
[11] R. Ishak, “Clustering prestasi akademik lulusan menggunakan K-Means,” *J. Electr. Electron. Eng.*, 2024.
[12] Y. Sheng, “Augmenting K-Means clustering with qualitative data to discover engagement patterns of older adults,” *J. Med. Internet Res.*, 2024.
[13] C. P. Ufeli, M. U. Sattar, R. Hasan, and S. Mahmood, “Enhancing customer segmentation through factor analysis of mixed data (FAMD)-based approach using K-Means and hierarchical clustering algorithms,” *Information*, vol. 16, no. 6, p. 441, 2025.
[14] M. Alves Gomes and T. Meisen, “A review on customer segmentation methods for personalized customer targeting in e-commerce use cases,” *Inf. Syst. E-Bus. Manag.*, vol. 21, pp. 527–570, 2023.
[15] K. Tabianan, S. Velu, and V. Ravi, “K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data,” *Sustainability*, vol. 14, p. 7243, 2022.
[16] B. Chong, “K-means clustering algorithm: A brief review,” *Acad. J. Comput. Inf. Sci.*, vol. 4, pp. 37–40, 2021.
[17] M. Ahmed, R. Seraj, and S. M. S. Islam, “The K-means algorithm: A comprehensive survey and performance evaluation,” *Electronics*, vol. 9, p. 1295, 2020.
[18] N. Hicham and S. Karim, “Analysis of unsupervised machine learning techniques for an efficient customer segmentation using clustering ensemble and spectral clustering,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, pp. 122–131, 2022.
[19] F. Abdi and S. Abolmakarem, “Customer behavior mining framework (CBMF) using clustering and classification techniques,” *J. Ind. Eng. Int.*, vol. 15, pp. 1–18, 2019.
[20] J. M. John, O. Shobayo, and B. Ogunleye, “An exploration of clustering algorithms for customer segmentation in the UK retail market,” *Analytics*, vol. 2, pp. 809–823, 2023.
[21] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, “RFM model customer segmentation based on hierarchical approach using FCA,” *Expert Syst. Appl.*, vol. 237, p. 121449, 2024.
[22] Saxena, A. Agarwal, B. K. Pandey, and D. Pandey, “Examination of the criticality of customer segmentation using unsupervised learning methods,” *Circ. Econ. Sustain.*, vol. 4, pp. 1447–1460, 2024.
[23] J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, “RFM ranking—An effective approach to customer segmentation,” *J. King Saud Univ. Comput. Inf. Sci.*, vol. 33, pp. 1251–1257, 2021.
[24] F. Barrera, M. Segura, and C. Maroto, “Multiple criteria decision support system for customer segmentation using a sorting outranking method,” *Expert Syst. Appl.*, vol. 238, p. 122310, 2024.
[25] U. Fang, M. Li, J. Li, L. Gao, T. Jia, and Y. Zhang, “A comprehensive survey on multi-view clustering,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, pp. 1–20, 2023.
[26] D. Das, P. Kayal, and M. Maiti, “A K-means clustering model for analyzing the Bitcoin extreme value returns,” *Decis. Anal. J.*, vol. 6, pp. 1–11, 2023.
[27] L. Rajput and S. N. Singh, “Customer segmentation of e-commerce data using K-means clustering algorithm,” in *Proc. 2023 13th Int. Conf. Cloud Computing, Data Science & Engineering (Confluence)*, 2023, pp. 658–664.
[28] D. Kadarsah and J. Heikal, “Customer segmentation with K-Means clustering: Suzuki Mobil Bandung customer case study,” *J. Indones. Sos. Teknol.*, vol. 5, pp. 768–774, 2024.
[29] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, D. Man, and T. Le, “An extended RFM model for customer behaviour and demographic analysis in retail industry,” *Bus. Syst. Res.*, vol. 14, pp. 26–53, 2023.

[30] H. Zhang, J. Li, J. Zhang, and Y. Dong, "Speeding up K-means clustering in high dimensions by pruning unnecessary distance computations," *Knowl. Based Syst.*, vol. 284, p. 111262, 2024.
[31] Afzal *et al.*, "Customer segmentation using hierarchical clustering," in *Proc. 2024 IEEE 9th Int. Conf. Convergence in Technology (I2CT)*, 2024, pp. 1–6.
[32] D. A. Awaliyah, B. Prasetiyo, R. Muzayanah, and A. D. Lestari, "Optimizing customer segmentation in online retail transactions through the implementation of the K-Means clustering algorithm," *Sci. J. Inform.*, vol. 11, p. 539, 2024.
[33] V. L. Narayana *et al.*, "Mall customer segmentation using machine learning," in *Proc. 2022 Int. Conf. Electronics and Renewable Systems (ICEARS)*, 2022, pp. 1280–1288.
[34] Ullah *et al.*, "Customer analysis using machine learning-based classification algorithms for effective segmentation using recency, frequency, monetary, and time," *Sensors*, vol. 23, p. 3180, 2023.
[35] F. Ashari, E. D. Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of elbow, silhouette, Davies–Bouldin, Calinski–Harabasz, and Rand-index evaluation on K-Means algorithm for classifying flood-affected areas in Jakarta," *J. Appl. Inform. Comput.*, vol. 7, pp. 89–97, 2023.