



Url : <https://itej.uinssc.ac.id/>
 Email : itej@svekhnurjati.ac.id

K-Means Clustering As A Method For Identifying Consumer Behavior Patterns In Taqimart

Urba Yesha Hasibuan
 Information System
 Royal University
 Asahan, Indonesia
urbahasibuan2003@gmail.com

Arridha Zikra Syah
 Information System
 Royal University
 Asahan, Indonesia
azsyra@gmail.com

Sudarmin
 Information System
 Royal University
 Asahan, Indonesia
edisudarindra@gmail.com

Abstract—Understanding consumer behavior patterns is essential for organizations seeking to improve marketing strategies, enhance customer satisfaction, and optimize decision-making processes. This study explores the application of the K-Means clustering algorithm as an unsupervised machine learning method to identify distinct consumer behavior patterns from transactional and demographic data. The research process includes data preprocessing, feature selection, normalization, and the determination of the optimal number of clusters using evaluation techniques such as the Elbow Method and Silhouette Score. The K-Means algorithm is then employed to segment consumers based on similarities in purchasing behavior, including purchase frequency, transaction value, and product preferences. The resulting clusters reveal meaningful consumer groups with unique characteristics, enabling businesses to develop targeted marketing campaigns and personalized service strategies. The findings demonstrate that K-Means clustering effectively uncovers hidden structures within consumer datasets and provides valuable insights for customer segmentation and strategic planning. This study highlights the potential of data-driven approaches in supporting business intelligence and improving organizational competitiveness in increasingly dynamic markets.

Keywords— Data Mining, K-Means Clustering, Consumer Behavior Patterns.

Article info: Date Submitted: 14/03/2026 | Date Revised: 20/03/2026 | Date Accepted: 13/04/2026
This is an open access article under the CC BY-SA license



I. INTRODUCTION

The rapid development of information technology in the digital age has had a significant impact on the business world, including the retail and trade sectors [1]. Competition among businesses is becoming increasingly fierce, so companies must be able to understand the needs and behavior of their consumers in order to survive and grow [2]. The use of data has become a critical aspect of modern business decision-making [3]. Every transaction that occurs in a retail store contains valuable information that can be analyzed to identify consumer habits, preferences, and purchasing patterns [4]. Through data analysis, companies can develop more effective strategies to increase sales and consumer loyalty [5]. One of the challenges in the retail sector is the inability to optimally utilize transaction data to understand consumer behavior patterns. Many small and medium retail businesses still rely on manual analysis, resulting in less effective decision-making and suboptimal marketing strategies.

In its day-to-day operations, Taqimart serves a wide range of customers with diverse characteristics, including age, gender, income level, and purchase frequency. Every day,

the store generates a significant volume of transaction data that has the potential to be transformed into strategic insights. However, this data is often merely stored as sales records without further analysis. In fact, if managed properly, this transaction data can help understand consumer behavior and serve as the foundation for designing more targeted marketing strategies.

The challenge faced by Taqimart is the lack of an analysis method or system to segment customers based on similar shopping behavior patterns. Until now, business decisions such as product promotions and the determination of marketing strategies have still been made manually and based on subjective estimates. This has led to suboptimal inventory management, a mismatch between the products offered and consumer needs, and ineffective promotions. Taqimart lacks a strong analytical foundation to identify the most promising consumer segments. As a result, opportunities to increase sales and consumer loyalty have not been fully utilized.

One solution to address this issue is to apply data mining, specifically using the K-Means Clustering method [6] [7]. This method groups data into several clusters based on the similarity of certain characteristics [8]. In the context of consumer behavior, K-Means can be used to identify consumer groups with similar shopping habits, such as purchase frequency and total spending [9] [10]. By identifying these consumer groups, Taqimart can more easily determine marketing strategies tailored to the needs of each segment, such as offering special promotions to active consumers or introducing new products to specific groups [11] [12].

By applying the K-Means Clustering method, Taqimart is expected to be able to identify consumer behavior patterns in greater depth and with greater accuracy [13] [14]. This analysis not only helps improve operational efficiency but also supports more effective data-driven marketing strategies [15]. Thus, this research is important to help Taqimart understand its consumers' behavior, enhance its competitiveness in the market, and strengthen long-term relationships with consumers through more targeted services and strategies.

II. METHODOLOGY

This study used several stages, as shown in Figure 1 below.

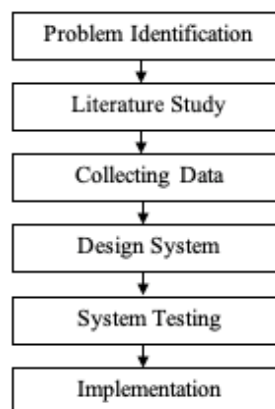


Figure 1. Research Method

A. Problem Identification

Taqimart generates a large volume of transaction data every day, but this data has not yet been analyzed to understand consumer behavior. Business decisions such as inventory procurement, promotions, and marketing strategies are still made manually and based on estimates. This leads to inaccuracies in inventory management and ineffective promotions. Therefore, the K-Means Clustering method is needed to group consumers based on their shopping patterns as a basis for more accurate decision-making.

B. Literature Study

A literature review was conducted to identify existing theories and research on data mining, consumer behavior, and the K-Means clustering method. This literature serves as the foundation for understanding how clustering works and its application in consumer segmentation within the retail sector.

C. Collecting Data

Data was collected using Taqimart transaction data, such as purchase frequency, total spending, and product types. Supporting data was also obtained through interviews or observations to understand the store's operational conditions.

D. Design System

The system design uses the Unified Modeling Language (UML), visualized using Microsoft Visio. The UML diagrams to be used include Use Case Diagrams, Activity Diagrams, and Sequence Diagrams [16].

E. System Testing

A pilot phase was conducted to assess whether the designed system met user needs. Once the design phase was complete, the system was developed using the PHP programming language and a MySQL database [17]. Testing was performed using the Blackbox Testing method to ensure that all system functions operated in accordance with the established specifications. If any discrepancies were found, the system was evaluated and refined until it functioned optimally in line with user expectations.

F. Implementation System

The resulting clusters were analyzed to identify the characteristics of each consumer group [18]. The findings were used to recommend strategies such as inventory adjustments, more targeted promotions, and the identification of potential consumer segments to improve Taqimart's performance [19].

Research Object

This research was conducted at Taqimart, a grocery retail store located in Kisaran, Indonesia. Taqimart operates in the retail sector by providing various daily necessities and household products.

The data used in this study consist of transaction data and consumer data collected from Taqimart. The dataset includes attributes such as purchase frequency, total spending, and product categories, which are used as variables in the clustering process. These data represent consumer purchasing behavior and serve as the basis for applying the K-Means Clustering method.

III. RESULT AND DISSCUSSION

A. Analysis Data

Before presenting the transaction dataset, it is important to describe the characteristics of the data used in this study. Consumer transaction records provide valuable information for understanding purchasing behavior and identifying similarities among customers. The dataset employed in this research comprises transactional information collected from 50 consumers and includes variables related to purchasing activities, such as purchase frequency, total purchase amount, and the diversity of product categories purchased. These

variables serve as the primary indicators for the K-Means clustering process to segment consumers based on their behavioral patterns. The detailed description of the transaction data used in this study is presented in Table 1.

Table 1. Transaction Data

No	Consumen initial	Buy Frequence	Buy Amount	Category Buy
1	AS	3	120000	4
2	RK	6	420000	5
3	FP	7	490000	5
4	SA	8	560000	6
5	DI	2	170000	4
6	NA	10	700000	7
7	RA	5	270000	5
8	WO	2	140000	8
9	BS	10	910000	8
10	MS	14	980000	9
....
50	NR	16	1260000	11

B. Design System

A Use Case Diagram is a visual representation of the activities performed by actors and the system to achieve a specific goal [20]. The following is the design of a use case diagram for the K-Means system that was created.

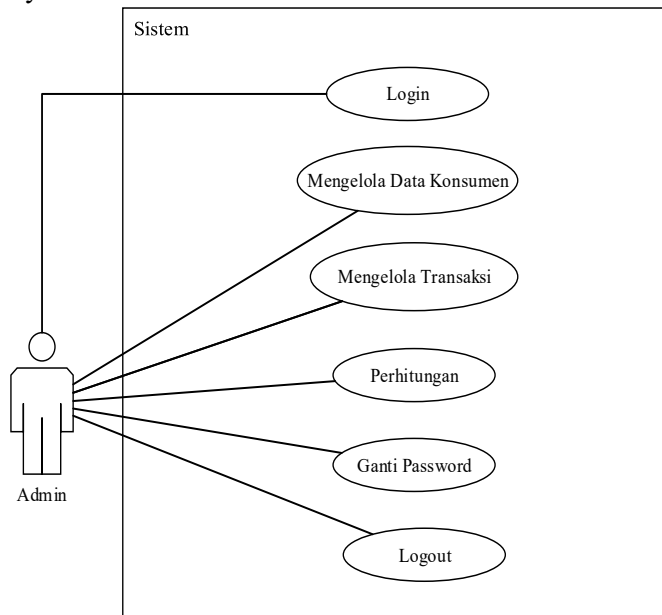


Figure 2. Use Case Diagram

C. Implementation System

1. Login Page

The login page serves as the entry point for accessing the system. Users are required to enter a valid username and password to log in to the system in accordance with their access rights.

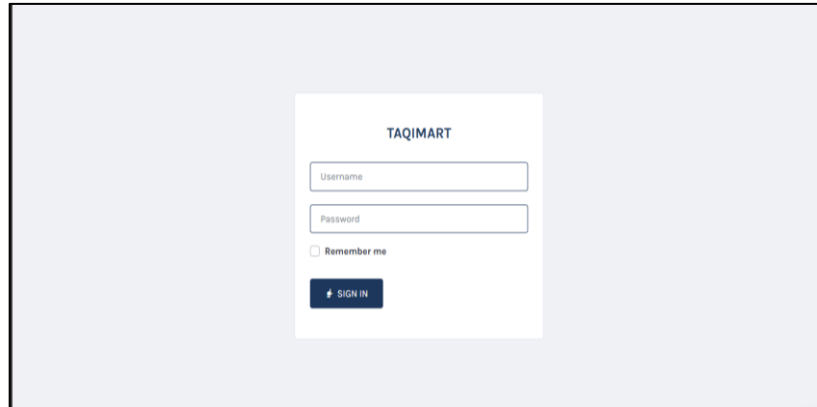


Figure 3. Login Page

2. Dashboard Page

To facilitate sales monitoring and support managerial decision-making, the developed system provides an interactive dashboard that summarizes important business information in a concise and user-friendly format. Through this dashboard, administrators can quickly access key indicators related to business performance, including monthly sales totals, monthly profit, products with low stock levels, and daily sales information. The availability of these visual summaries enables users to monitor operational conditions in real time, identify emerging trends, and respond promptly to potential issues requiring immediate attention. The implementation of the dashboard interface developed in this study is illustrated in Figure 4.

The dashboard page displays a summary of the system's key information.

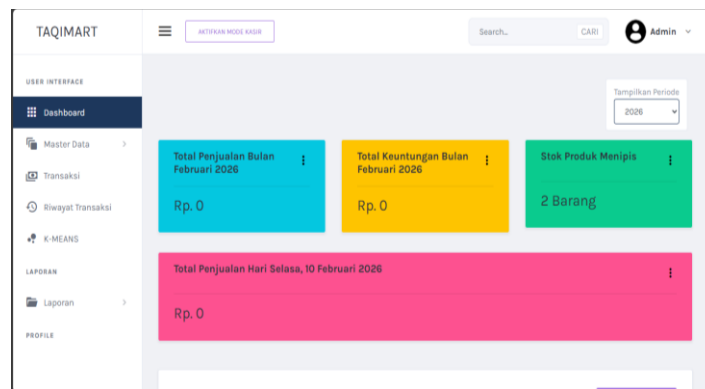


Figure 4. Dashboard Page

3. Category Page

The category page is used to manage product category data. Users can add, edit, and delete categories as the basis for grouping products. Effective product management requires accurate classification of products to ensure efficient inventory control and facilitate transaction processing. Therefore, the system provides a product category management feature that enables administrators to organize products into predefined categories according to their characteristics.

Through this interface, users can add new categories, edit existing category information, and remove categories that are no longer needed. The availability of this feature helps maintain data consistency, improves the accuracy of sales records, and supports subsequent analytical processes, including consumer behavior analysis using the K-Means clustering method. The implementation of the product category management interface is presented in Figure 5.

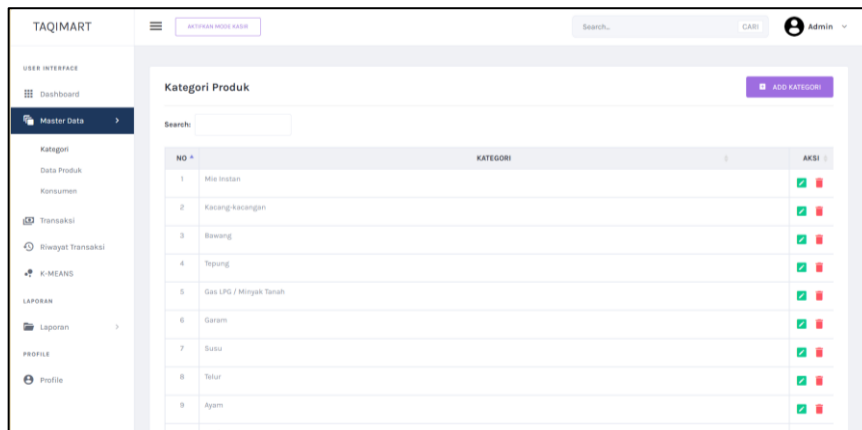


Figure 5. Category Page

4. Product Page

The product page is used to manage data on products for sale, including product name, category, price, and inventory. This data is used in the sales transaction process.

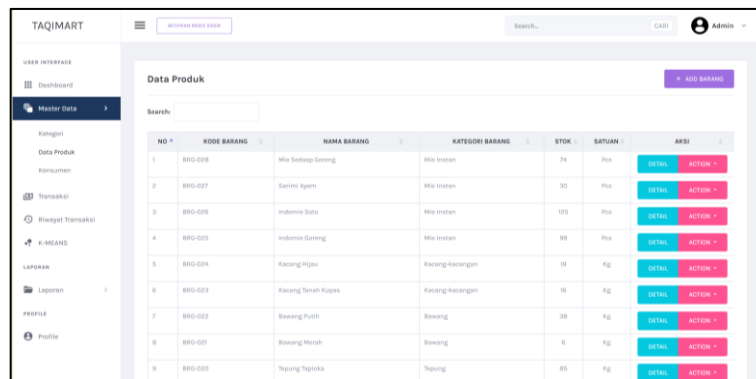


Figure 6. Product Page

5. Consumer Page

The customer page displays data on customers registered in the system. The information on this page serves as the basis for customer analysis and segmentation. Consumer data constitute an essential component of the sales information system, as they provide the foundation for understanding purchasing behavior and conducting customer segmentation analysis. To support these activities, the system includes a consumer management feature that allows administrators to record and maintain customer information in an organized manner. The interface stores various

consumer attributes, such as consumer codes, names, contact numbers, addresses, purchase frequency, total purchase value, and the number of product categories purchased. These data not only facilitate transaction management but also serve as the primary input for the K-Means clustering process to identify consumer behavior patterns and generate meaningful customer segments. The implementation of the consumer data management interface is shown in Figure 7.

NO	KODE KONSUMEN	NAMA KONSUMEN	NO HP	ALAMAT	FREKUENSI PEMBELIAN	TOTAL PEMBELIAN	JUMLAH JENIS PRODUK YANG DIBELI	AKSI
1	PLDN-07-000000041	Anah Saputra	08227300166	J. Bakri No. 1, Auk Sanggonggen	3	Rp. 100.000	4	[Green Checkmark]
2	PLDN-07-000000046	Rina Kurniawati	08570050163	J. Bakri No. 1, Auk Sanggonggen	6	Rp. 400.000	5	[Green Checkmark]
3	PLDN-07-000000047	Piper Pratama	08103700012	J. Empatman No. 05, Auk Sanggonggen	7	Rp. 400.000	5	[Green Checkmark]
4	PLDN-07-000000048	Siti Anghel	08103700066	J. Empatman No. 05, Auk Sanggonggen	8	Rp. 500.000	6	[Green Checkmark]
5	PLDN-07-000000049	Debi Irawan	08236080071	J. Empatman No. 05, Auk Sanggonggen	2	Rp. 170.000	4	[Green Checkmark]

Figure 7. Consumer Page

6. Transaction Page

The transaction page is used to record sales transactions. On this page, users can select customers, products, and the quantity purchased. The transaction module is a core component of the sales information system, as it facilitates the recording and processing of consumer purchases in a structured and efficient manner. Through this interface, administrators or cashiers can initiate transactions by selecting registered consumers or processing purchases without predefined customer information. The system automatically generates customer identification codes and transaction invoice numbers, thereby reducing manual input errors and improving data accuracy. In addition to supporting daily sales operations, the transaction data collected through this module serve as valuable input for subsequent analytical processes, particularly in identifying consumer behavior patterns using the K-Means clustering method. The implementation of the transaction processing interface is presented in Figure 8.

Figure 8. Transaction Page

7. Report Page

The report page is used to display a summary of transaction data and analysis results, including sales reports and customer segmentation results using the K-Means clustering method.

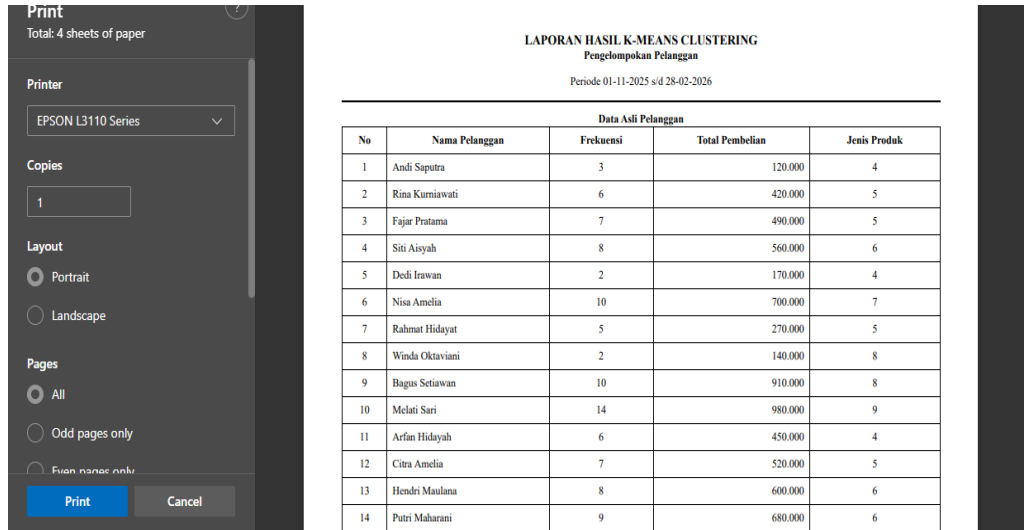


Figure 9. Report Page

Before the implementation of the K-Means clustering method, Taqimart generated a large volume of transaction data on a daily basis. However, this data was not analyzed systematically to identify consumer behavior patterns. Business decisions related to inventory procurement, promotional strategies, and marketing activities were carried out manually based on intuition and estimation. As a result, the company often faced issues such as inaccurate stock planning and ineffective promotional targeting.

After applying the K-Means clustering method, transaction data can be processed and grouped into several clusters based on customer purchasing patterns. The clustering results reveal distinct groups of consumers, such as customers with high purchase frequency, moderate purchasing behavior, and low engagement levels. This classification enables the business to better understand customer characteristics and preferences.

The comparison between the two conditions shows that the use of K-Means clustering transforms unstructured transaction data into meaningful information that supports data-driven decision-making. Inventory procurement can be adjusted based on the dominant purchasing patterns in each cluster, thereby reducing the risk of overstock or stock shortages. In addition, promotional strategies can be targeted more effectively by focusing on specific customer groups, which increases the potential success of marketing campaigns.

Thus, the implementation of the K-Means method not only improves data analysis capabilities but also provides a practical solution to the problems of manual

decision-making, inefficient promotions, and inaccurate inventory management. This demonstrates that K-Means clustering plays a significant role in enhancing business performance through better utilization of transaction data.

IV. CONCLUSION

Based on the research objectives and the results obtained, it can be concluded that the transaction data at Taqimart has been successfully utilized to analyze consumer behavior patterns, transforming previously unprocessed sales records into meaningful information. The application of the K-Means Clustering method has proven effective in grouping consumers based on similarities in purchasing behavior, including purchase frequency, total spending, and product categories. The resulting clusters reveal distinct consumer segments with different characteristics, which can be used as a basis for developing more targeted marketing strategies. Furthermore, the implementation of the clustering-based system supports more effective and data-driven decision-making, thereby improving operational efficiency and enhancing Taqimart's ability to compete in the retail market.

REFERENCES

- [1] A. Alawiyah, N. Aghnia, and F. F. Abdalah, "Implementasi Clustering Algoritma K-Means Pada Penjualan Beras Di CV Tangguh Bumi Perkasa," *J. Komisi (Jurnal Komput. dan Sist. Informasi)*, vol. 2, no. 2, pp. 17–23, 2025.
- [2] A. Yahya and R. Kurniawan, "Implementasi Algoritma K-Means untuk Pengelompokan Data Penjualan Berdasarkan Pola Penjualan," *MALCOM Indones. J. Mach. Learn. Comput. Sci. J.*, vol. 5, no. January, pp. 350–358, 2025.
- [3] R. Rahmawati, W. Prihartono, and K. Cirebon, "Optimasi Stok Dengan Clustering Data Transaksi Penjualan Menggunakan Algoritma K-Means di Konter Agung Cell," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 2, 2025.
- [4] M. Fadhil, W. Fuadi, and Maryana, "Clustering Tingkat Kecanduan Game Mobile Legends Terhadap Keharmonisan Keluarga Menggunakan Metode K-Means," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 981–990, 2025.
- [5] J. N. Iin, "Perbandingan Kmeans Dan Hierarchical Clustering Untuk Pemetaan Kawasan Rawan Stunting Di Kabupaten Kolaka," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 701–716, 2025.
- [6] I. Fazira, Z. Fitri, and Risawandi, "Optimasi Jumlah Cluster Pada K-Means Clustering Menggunakan Particle Swarm Optimization Untuk Pengelompokan Ukt Mahasiswa," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 874–886, 2025.
- [7] E. A. Wibowo and R. Aryanti, "Penerapan Metode Clustering K-Means Menggunakan RapidMiner untuk Klasifikasi Prestasi Siswa di Sekolah Swasta," *J. Inf. Technol. Informatics Eng.*, vol. 1, no. 1, pp. 20–24, 2025, [Online]. Available: <https://journal.jci.co.id/jitie/article/view/35>
- [8] Sulistiani *et al.*, "Data Analysis of Student Monitoring Using the K-Means Clustering Method," *Indones. J. Mod. Sci. Technol.*, vol. 1, no. 2, pp. 50–57, 2025, doi: 10.64021/ijmst.1.2.50-57.2025.
- [9] M. Rival, M. Misriani, and L. O. Bakrim, "Penerapan Metode Cluster Dalam Data Mining Mengelompokkan Kenakalan Remaja (Studi Kasus Polda Sultra)," *Simkom*, vol. 9, no. 1, pp. 79–89, 2024, doi: 10.51717/simkom.v9i1.375.
- [10] A. E. Febriyanti, S. Z. Harahap, and M. Masrial, "Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering dan Algoritma Hirarki Divisive Studi Kasus Toko Sembako Pujo," *INFORMATIKA*, vol. 15, no. 1, pp. 72–86, 2024, doi: 10.25130/sc.24.1.6.
- [11] N. Bili, R. T. Abineno, and A. Aha Pekuwali, "Penerapan Algoritma K-Means

- Clustering Untuk Pengelompokan Peforma Siswa Pada Pembelajaran Bahasa Indonesia (Studi Kasus : SD Inpress Waingapu 3),” *SATI Sustain. Agric. Technol. Innov.*, pp. 523–537, 2024.
- [12] M. Adelina Bui and A. Bahtiar, “Implementasi Metode Algoritma K-Means Clustering Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1451–1456, 2024, doi: 10.36040/jati.v8i2.8975.
- [13] W. P. Priyadi, J. D. Irawan, and A. Faisol, “Penerapan Data Mining Untuk Clustering Wilayah Produksi Pada Menggunakan Metode K-Means (Studi Kasus : Wilayah Jawa Timur),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 8381–8388, 2024.
- [14] A. Ghozy, F. S. Wahyuni, and S. Achmadi, “Implementasi Metode K-Means Clustering Untuk Pengelompokan Kelas Berdasarkan Pemahaman Siswa Pada Bimbingan Belajar Matematika Saschio Banyuwangi,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1072–1077, 2023, doi: 10.36040/jati.v6i2.5450.
- [15] R. Farismana, “Penerapan K-Means Clustering Untuk Pemetaan Produktivitas Padi Dan Prediksi Panen Di Kabupaten Indramayu,” *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 8, no. 3, p. 589, 2024, doi: 10.52362/jisamar.v8i3.1572.
- [16] D. D. Susilo, S. S. Hilabi, B. Priyatna, and E. Novalia, “Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT.Otomotif 1,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 476, 2024, doi: 10.35889/jutisi.v13i1.1836.
- [17] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [18] I. Ibrahim and W. Usino, “Klasterisasi Tingkat Kelayakan Provinsi Dalam Pembangunan Kawasan Industri Menggunakan Algoritma K-Means,” *SENAFTI (Seminar Nas. Mhs. Fak. Teknol. Informasi)*, vol. 3, no. September, pp. 324–333, 2024.
- [19] S. Anwar, T. Suprapti, G. Dwilestari, and I. Ali, “Pengelompokan Hasil Belajar Siswa dengan Metode Clustering K-Means,” *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.
- [20] D. Kurniadi, Y. H. Agustin, H. I. N. Akbar, and I. Farida, “Penerapan Algoritma k-Means Clustering untuk Pengelompokan Pembangunan Jalan pada Dinas Pekerjaan Umum dan Penataan Ruang,” *Aiti (Jurnal Teknol. Informasi)*, vol. 20, no. 1, pp. 64–77, 2023, doi: 10.24246/aiti.v20i1.64-77.