

Implementation of the K-Means Algorithm for Clustering Hot Selling and Less Selling Goods at XYZ Wholesalers

Silva Chaerani^{1*}, Riki Andri Yusda², Rohminatin¹

^{1,3}Faculty of Computer Science, Information Systems, Royal Asahan University, North Sumatra

²Faculty of Computer Science, Computer Systems, Royal Asahan University, North Sumatra

Email: ^{1*}silvachaerani.royal@gmail.com, ²rikiandriyusda@gmail.com, ³rohminatin2019@gmail.com

Email Penulis Korespondensi: silvachaerani.royal@gmail.com

Submitted 15-07-2025; Accepted 11-08-2025; Published 30-08-2025

Abstract

Ratna Wholesale is a store that sells various household supplies, clothing, and other daily necessities. In managing the stock of goods, the owner still has difficulty in distinguishing products that are in demand and less in demand objectively, so that there is often an imbalance in the inventory of goods. This has an impact on the vacancy of products that are needed and the accumulation of products that are less in demand. To overcome these problems, this research aims to build a system that can group products based on sales levels using the K-Means clustering method. This research uses a quantitative approach with data collection techniques through observation, interviews, and sales documentation for the last three years. The system was developed using Visual Basic programming language and MySQL database. The results of the system implementation show that the K-Means method is effective in grouping products into two categories, namely hot selling items and less selling items, thus helping owners in making more efficient stock management decisions and increasing customer satisfaction. From the calculation of the system, it is obtained that the products that are in demand are 24% and the products that are less in demand are 76%.

Keywords: Data Mining, K-Means; Clustering; Product, Wholesalers

1. INTRODUCTION

In this era of rapid technological development, it helps humans facilitate every activity in various fields so that it becomes more effective and efficient [1]. One of the technologies that can be used in the industrial field is in inventorying goods. The availability of goods and completeness in a company is a very important element, so as good management in the process of managing the availability of stock items is needed to avoid the accumulation of the same goods and goods that are less in demand [2].

Ratna Wholesale is a store that sells various household items, such as children's clothes, bed sheets, curtains, bed covers, and other products. Initially, the store was located in Dusun III Tanjung Alam. With many types of goods sold, the owner often has difficulty managing products that sell quickly and those that are rarely purchased. Currently, the owner of Ratna Wholesale still has difficulty in objectively distinguishing the best-selling and less-selling items. Thus, the most in-demand items often run out, leaving customers disappointed and able to switch to other stores. Conversely, there are also items that are rarely bought, but are still available in large quantities, thus consuming storage space without providing comparable benefits and causing the stock of goods to be unbalanced, because less in-demand items are in excess and in-demand items are in short supply.

The problem faced is an imbalance in product supply based on the level of demand. Highly in-demand items are often not available when customers are looking for them, increasing the risk of losing customers [3]. Conversely, less in-demand goods accumulate in the warehouse, thereby increasing operational costs that make the turnover of goods inefficient and hamper store profits [4].

To solve this problem, a system is needed that can cluster products based on the level of demand using the K-Means clustering method [5]. The system will categorize products into two main categories: best-selling products and less-selling products. With this system, the owner can more easily determine which products should always be available in large quantities and which ones are sufficiently provided in small quantities so as not to fill the storage space [6].

This K-Means clustering-based system will help owners organize products more efficiently. By knowing which products are most in demand, the owner can make more informed decisions in organizing inventory [7]. In addition, this system can also increase customer satisfaction and speed up the turnover of goods in the store. Clustering is one of the methods of data mining that has the ability to group data into groups based on similar characteristics, while data that has different characteristics will be grouped into other groups [8]. K-Means is a simple unsupervised clustering algorithm. It is popular in clustering due to its simplicity and efficiency, and is recognized as one of the top ten data mining algorithms by IEEE [9].

Previous research conducted by Bili et al [10] they used the k-means method for clustering student performance in Indonesian language learning, the results of clustering using the K-Means method on Indonesian language scores of SD Inpres Waingapu 3 students have successfully grouped students into four clusters based on their academic performance. Based on the clustering results using the K-Means method in RapidMiner application, the clustering performance is evaluated with several important metrics. The average within-cluster distance to the centroid is 8.370, which indicates the average distance of each data point to the centroid of its respective cluster. In more detail, the average within-cluster distance for cluster 0 is 8,650 and for cluster 1 is 8,281. Evaluation using the Davies-Bouldin Index (DBI) found the best value of 0.078 in cluster_2. This low DBI value indicates that the clustering performed is

quite good, with the clusters formed having a clear distance between clusters and the data within the clusters are relatively the same. From these results, it can be concluded that the K-Means method is effective in clustering students based on their academic performance, which can help in monitoring groups of students who require special attention or different learning strategies. While this research applies k-means clustering to web-based applications with the aim that the application can be accessed anytime and anywhere to make it easier for users to use the application.

K-means is a non-hierarchical data clustering method that attempts to partition existing data into two or more groups [11]. This method partitions data into groups so that data with the same characteristics are put into the same group and data with different characteristics are grouped into other groups [10]. The purpose of clustering this data is to minimize the objective function set in the clustering process, which generally seeks to minimize variation within a group and maximize variation between groups [12]. Therefore, the purpose of this research is to design a system to categorize best-selling and less-selling items using the K-Means method at Ratna Wholesale.

While there have been many studies applying the K-Means clustering algorithm in various fields, ranging from student performance analysis to agricultural sector productivity mapping, previous studies have focused on education data, public service optimization, or supply chain segmentation at a macro scale. To date, there is still a lack of applied research specifically targeting micro-scale inventory management in small and medium-sized retail environments, specifically utilizing real historical transaction data to cluster products based on demand levels. Most previous studies have not highlighted the practical application of the K-Means algorithm in daily decision-making systems in the retail sector, particularly in traditional wholesale operations that still rely heavily on manual stock management. Therefore, there is a gap in the implementation of K-Means clustering adapted to the operational context of local retail stores, where the clustering results can be used directly to support stock procurement decisions, minimize the risk of excess or shortage of goods, and improve overall sales efficiency.

2. RESEARCH METHOD

This research consists of five stages, namely Problem Identification, Data Collection, Data Analysis, System Design and Implementation which are described in Figure 1 below.

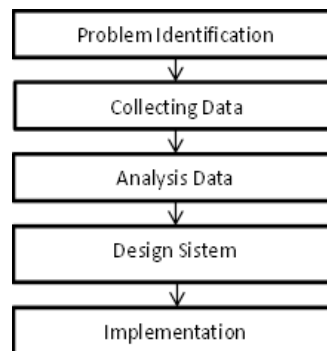


Figure 1. Research Method

2.1 Problem Identification

At this stage, the author observes that Ratna Wholesale faces problems in stock management, especially in objectively distinguishing between in-demand and out-of-demand items. This results in a stock imbalance, where the in-demand items run out quickly while the less-sold items accumulate in the warehouse.

2.2 Collecting Data

The system requires initial data to run the clustering process. In this case, the data used comes from the sales history of products in Ratna Grosir for the last three years, which includes the product name as well as the number of sales. This information plays an important role in identifying the best-selling and least popular products, which will then be grouped in the clustering process. The sample data used in this process is presented in Table 1.

Table 1. Sales Data 3 Year Before

No	Product Name	Category	Sales Data (Year)		
			2022	2023	2024
1	Kavolux Standing Fan	Electronic	294	398	326
2	Cosmetic Rack	Home Furniture	218	160	221
3	Rice Dispenser	Home Furniture	106	130	124
4	Bonita Bed Sheet	Home Furniture	172	222	183
5	Cliks Wardrobe	Home Furniture	322	263	296
6	Double-Leg Towel Rack	Home Furniture	86	122	111

No	Product Name	Category	Sales Data (Year)		
			2022	2023	2024
7	Complete Curtain Rod Set	Home Furniture	284	285	260
8	Super Mixer	Electronic	177	208	245
9	Embossed Curtain	Home Furniture	398	358	351
10	Wall Clock	Electronic	51	52	16
...					
50	Kavolux Wall Fan	Electronic	75	61	58

Table 1 presents the 50 initial data used as samples in the clustering process to group products based on their sell-through rate at Ratna Grosir. This data consists of product names and sales amount for the last three years. This information is the main basis for the K-Means algorithm in determining the two main groups, namely products that are classified as in-demand and products that are less desirable. By utilizing this historical sales data, the system can identify sales patterns and generate more accurate product groupings to support decision-making in stock management and sales strategy.

2.3 Analysis Data

The data that has been collected is then analyzed to determine the sales pattern of each product. The results of this analysis are the basis for clustering using the K-Means method. The K-Means algorithm is used in this system to cluster products based on best-selling and less-selling items in Ratna Wholesale. The stages in this process start with selecting the number of clusters, initializing the initial centroid, calculating the distance using the Euclidean method, grouping the data to the nearest cluster, and updating the centroid until the cluster result is stable [13]. The system will automatically process the sales data and generate two main clusters, namely products that are classified as in-demand and products that are less in-demand based on their sales intensity [14]. To calculate the k-means algorithm for clustering best-selling and less-selling items in Ratna Wholesale, we use the following equation [15]:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Where:

D (i,j) : Distance of data i to cluster center j

X_{ki} : Data to i on the kth data attribute

X_{kj} : The jth center point at the kth attribute

2.4 Design System

After the analysis process is complete, the next step is to design and build a system that is able to group goods into two clusters, namely in-demand and out-of-demand goods. System design is done using Microsoft Visio as a tool to create UML diagrams, such as Use Case Diagram, to visualize the structure and workflow of the system. Furthermore, system development is carried out using the Visual Basic programming language in Microsoft Visual Studio 2010, and using a MySQL database integrated through XAMPP [16].

2.5 Implementation

After the system is declared feasible, it is implemented directly at Ratna Wholesale. The system will automatically categorize products based on sales data, so that owners can more easily manage stock items and increase sales efficiency.

3. RESULT AND DISCUSSION

3.1 Design System

Prior to system implementation, a design stage is carried out to describe the flow of use and interaction between users and the system. Figure 2 below presents a Use Case Diagram that shows how the roles of admin and shop owners interact with the system to carry out the process of grouping goods based on sales levels [17].

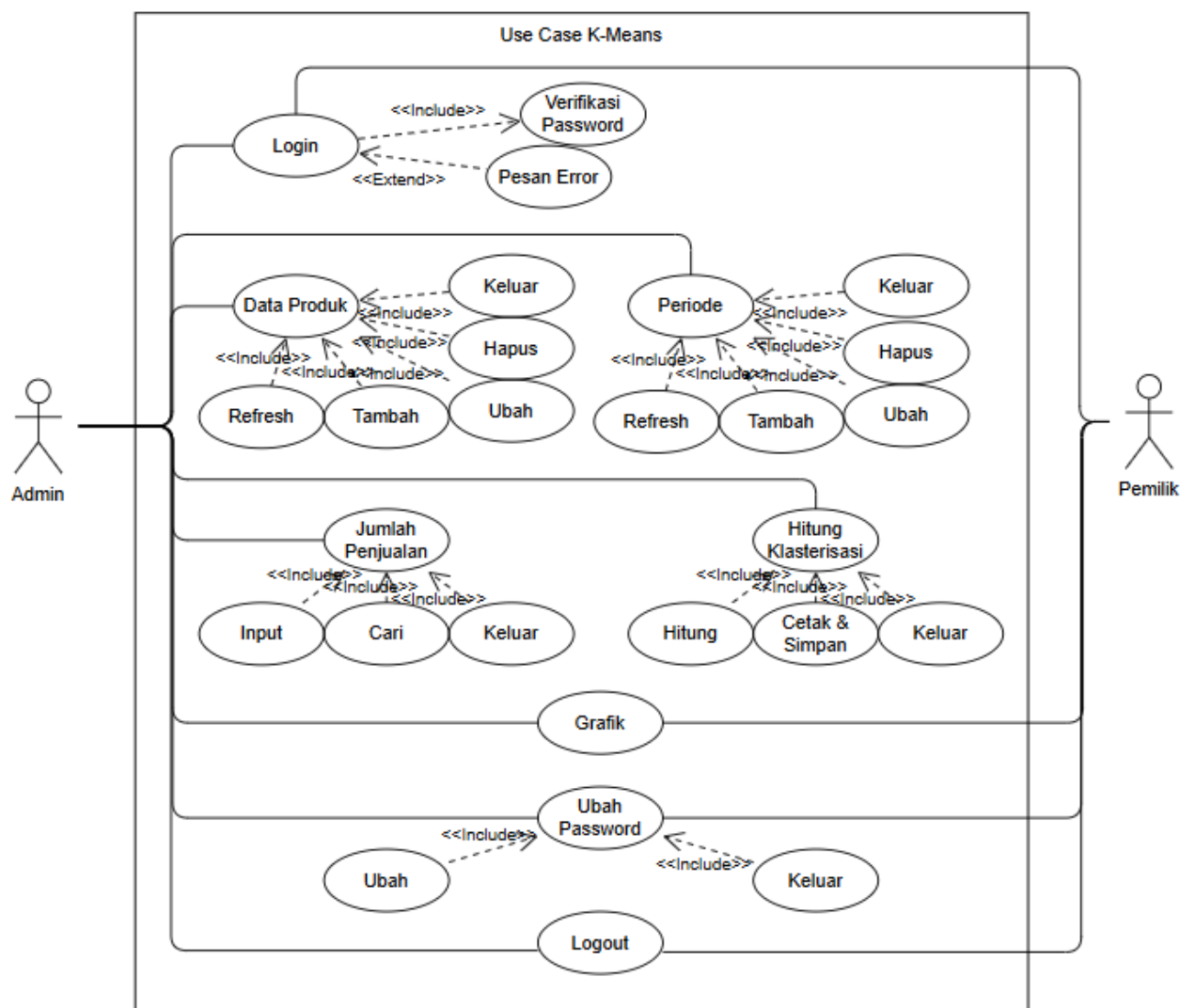


Figure 2. Design System

In Figure 2, there are two actors, namely the admin who has a role in managing the initial data, including inputting sales data and managing system users and owners to access clustering results and reports to support decision making in stock management. Through this system design, user interaction becomes more structured and systematic, supporting the main objective of the research in clustering goods efficiently and data driven.

3.2 Determining the Number of Cluster

The number of clusters has been determined to be 2 clusters, namely C1 Less Sold and C2 Sold because the purpose of this research is to determine the best-selling and unsold items from 50 products that have been presented in table 1.

3.3 Determining the Initial Cluster Centers

To determine the initial centroid, the data is taken from the least accumulated sales data, namely the 10th data and the most accumulated sales data, namely the 39th data. Before the clustering process using the K-Means algorithm is carried out, an important first step is determining the initial center point or initial centroid. The selection of the initial centroid greatly affects the final result of clustering, because it becomes the initial reference in determining the closeness of data to a cluster. Table 2 below shows the selection of two initial centroids based on the highest and lowest sales data of the 50 products analyzed.

Table 1. Initial Centroid

Initial Centroid	Year		
	2022	2023	2024
Cluster 1	51	52	16
Cluster 2	360	375	388

Therefore, Table 2 is a critical basis for starting the clustering process that will divide products into strategic categories, so that shopkeepers can determine stock priorities more precisely.

3.4 Cluster Dintance Calculation

To measure the distance between the data and the cluster center, the Euclidian distance is used, for example on the following Kavolux Standing Fan products:

$$d(1,1) = \sqrt{(294 - 360)^2 + (398 - 375)^2 + (326 - 388)^2} = 93,43$$

$$d(1,2) = \sqrt{(294 - 51)^2 + (398 - 52)^2 + (326 - 16)^2} = 524,28$$

And so on, the distance calculation is carried out for all product sales data.

3.5 Finding the Shortest Value

To find the shortest distance from the iteration calculation results, it is done by finding the minimum value of the two clusters for each product:

Kavolux Standing Fan = ('524,28', '93,43')

Kavolux Standing Fan = ('C1', 'C2')

The shortest value distance cluster result is 93.43 in cluster 2.

Cosmetic Rack = ('285,62', '307,05')

Cosmetic Rack = ('C1', 'C2')

The shortest value distance cluster result is 285.62 in cluster 1.

Rice Dispenser = ('144,13', '440,72')

Rice Dispenser = ('C1', 'C2')

The shortest value distance cluster result is 144.13 in cluster 1.

Bonita Bed Sheet = ('267,26', '317,46')

Bonita Bed Sheet = ('C1', 'C2')

The shortest value distance cluster result is 267.26 in cluster 1.

And further until the 3rd cluster results on all products, then the shortest distance results are obtained until the 3rd iteration. Then obtained the results of clustering goods that are in demand and goods that are less in demand with the same results [18]. As for the products that are in demand called C1 there are 38 products and products that are less in demand or C2 there are 12 types of products. After iterating with the K-Means algorithm, the system produces two main clusters that represent product categories based on their sales levels. The final results of this clustering process are presented in Table 3 below. This table shows the clustering results for each product in the 1st and 2nd iterations, as well as the final cluster in which the product is classified.

Table 3. Product Cluster Result

No.	Product Name	Cluster	
		Iteration 1	Iteration 2
1.	Kavolux Standing Fan	2	2
2.	Cosmetic Rack	1	1
3.	Rice Dispenser	1	1
4.	Bonita Bed Sheet	1	1
5.	Cliks Wardrobe	2	2
6.	Double-Leg Towel Rack	1	1
7.	Complete Curtain Rod Set	2	2
8.	Super Mixer	1	1
9.	Embossed Curtain	2	2
10.	Wall Clock	1	1
...			
50.	Kavolux Wall Fan	1	1

Through the results presented in table 3 above, the system has successfully grouped products objectively based on historical sales data, thus providing a more accurate and efficient basis for decision making in inventory product.

3.6 Implementation

This system is made using visual basic programming language with visual studio as text editor and mysql as database. This program has a login menu to enter the system [19]. Only registered users can use this clustering system. After the user enters the system, the user is directed to the main menu. From the main menu users can access other menus [20]. The system developed in this research is designed to be used easily by registered users. After the system development and testing process is complete, the next step is the implementation stage. This stage aims to implement the system directly in the operational environment of Ratna Grosir. Figure 3 below illustrates the main display main menu of the system after the login process is successfully carried out by the user.



Figure 3. Main Menu

Figure 3 displays the main menu that is systematic and focused on user needs, this system is expected to improve the accuracy of stock management, reduce the accumulation of unsold goods, and prevent the vacancy of goods that are in great demand by customers.

3.5.1 Clustering Graphics

After the system runs the clustering process with the K-Means algorithm, the clustering results are not only presented in tabular form, but also visualized in graphical form to make it easier for users to understand. This visualization makes it easier for shop owners to quickly see the comparison of the number of products between the most and least popular categories. Figure 4 below shows a graph of the clustering results generated by the system.

In this system, you can see a graph of which groupings of goods are in demand and not in demand. Orange-colored graphs indicate for groups that are in demand and blue-colored graphs indicate for goods that are less in demand as shown in Figure 4.

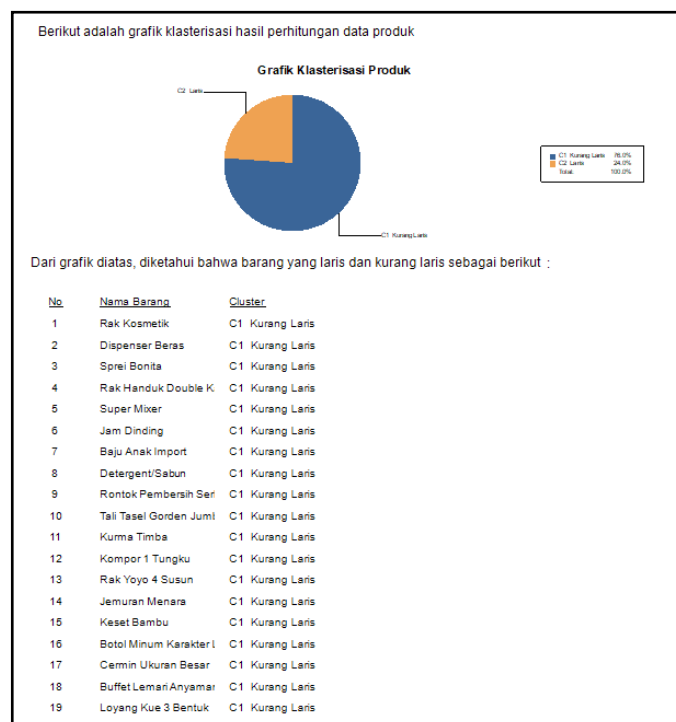


Figure 4. Clustering Graphics

Figure 4 adds value to the presentation of the analysis results because it facilitates non-technical understanding and speeds up the process of identifying problems in product stock and distribution.

3.7 Discussion

The application of the K-Means Clustering algorithm in this research successfully provides a solution to the problems faced by Ratna Grosir in managing stock items based on sales levels. Through the clustering process of sales data for the last three years, the system is able to divide products into two main categories: high-selling products and low-selling products. The results show that most of the products (76%) belong to the low-selling cluster (C1), while only 24% are in the high-selling cluster (C2). This finding provides strategic insights for store owners to make adjustments to stock volumes and purchasing patterns, as well as avoid wasting storage space for immobile goods.

From a technical perspective, the use of initial centroids based on the highest and lowest sales data proved effective in accelerating the convergence of the iteration process. The system achieved cluster stability in only two iterations, which indicates that the sales patterns of the products have a fairly clear consistency in terms of demand intensity. Furthermore, the visualization of cluster results in the form of graphs makes it easier for end users to understand the clustering results without the need for in-depth technical background. This feature improves the usability of the system and supports data-driven decision making.

The application of the K-Means method in the context of small to medium-sized retail management such as this provides a new contribution to the applied data mining literature. The system built not only functions as an analytical tool, but also as a decision support system that can be integrated directly into the store's operational activities.

4. CONCLUSION

In the research that has been done, it can be concluded that the system built successfully implements the K-Means Clustering method to group products at Ratna Grosir into two categories, namely goods that are in demand and goods that are less in demand, based on sales data for the last three years. From the calculation of the system, it is obtained that the products that are in demand are 24% and the products that are less in demand are 76%. The use of clustering applications using the K-Means algorithm can help shop owners in making stock procurement decisions, so as to avoid shortages of goods in demand by customers and reduce the accumulation of goods that are less in demand and the system made is easy to use, because it provides data processing features, clustering processes, and printing reports on cluster results.

REFERENCES

- [1] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020, doi: 10.36294/jurti.v4i2.1444.
- [2] S. Wijayanto and M. Yoka Fathoni, "Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jupiter*, vol. 13, no. 2, pp. 212–219, 2021.
- [3] V. Afifah and D. Setyantoro, "Rancangan Sistem Pemilihan dan Penetapan Harga dalam Proses Pengadaan Barang dan Jasa Logistik Berbasis Web," *J. IKRA-ITH Inform.*, vol. 5, no. 2, pp. 108–117, 2021.
- [4] A. Y. Simanjuntak, I. S. S. Simatupang, and Anita, "Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 85–91, 2022.
- [5] J. R. S. Penda Sudarto Hasugian, "Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 3, pp. 262–268, 2022, [Online]. Available: <https://djournals.com/klik>
- [6] W. W. Kristianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, vol. 5, no. 2, pp. 90–98, 2022, doi: 10.37792/jukanti.v5i2.547.
- [7] A. Y. Sari and E. Supriatna, "Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Classifier untuk Mendukung Strategi Promosi," *J. Dimamu*, vol. 3, no. 1, pp. 18–28, 2023, doi: 10.32627/dimamu.v3i1.837.
- [8] A. Ghazy, 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.
- [9] 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.
- [10] 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.
- [11] Hasim Azari, Dwi Hartanti, and Aprilisa Arum Sari, "Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 379–389, 2024, doi: 10.29408/jit.v7i2.26016.
- [12] I. Ibrahim and W. Usino, "Klasterisasi Tingkat Kelayakan Provinsi Dalam Pembangunan Kawasan Industri Menggunakan Algoritma K-Means," *SENAFTI (Semiinar Nas. Mhs. Fak. Teknol. Informasi)*, vol. 3, no. September, pp. 324–333, 2024.
- [13] 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.

- [14] 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.
- [15] 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.
- [16] 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>
- [17] S. Agustini, "Perancangan Sistem Informasi Data Stok Barang Berbasis Web Pada Hellomee," *J. Eng. Technol. Innov. (JETI)*, vol. 1, no. 1, pp. 19–35, 2022.
- [18] A. H. La Dimuru, "Perhatian Pemerintah Dalam Peningkatan Kualitas Sumber Daya Manusia Petani Rumpuk Laut di Desa Gomar Sungai Kec. Aru Selatan Kab. Kepulauan Aru," *Jaurnal Cakrawala Ilm.*, vol. 3, no. 9, pp. 5–24, 2024, [Online]. Available: [http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf](http://repo.iain-tulungagung.ac.id/5510/5/BAB%202.pdf)
- [19] 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.
- [20] J. Multidisiplin Saintek, Y. Candra Pratama, and Z. Reno Saputra, "Sistem Informasi Desa Delta Upang Berbasis Web," *J. Sains dan Teknol.*, vol. 2, no. 12, pp. 86–96, 2024, [Online]. Available: <https://ejournal.warunayama.org/index.php/kohesi/article/download/2788/2634>