Journal of Intelligent Decision Support System (IDSS) 8 (1) (2025), pp. 9-21

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Product Sales Grouping Application Design Using K-Means Clustering Algorithm

Sondang

Politeknik Unggul LP3M, Medan, Indonesia

Article Info

Article history:

Received Dec 16, 2024
Revised Dec 22, 2024
Accepted Jan 30, 2025

Keywords:

Application
K-Means Clustering
Minimarket
Product Grouping
Sales

ABSTRACT

This research aims to design and develop a product sales grouping application at Minimarket Diky using the K-Means Clustering algorithm. Product grouping based on sales patterns is one of the effective methods to improve marketing strategies, stock management, and more efficient business decision making. By using the K-Means algorithm, product sales data is processed to group products based on the initial number of items, the number sold, and the amount of stock. The designed application is able to identify sales patterns that are difficult to find manually, so as to provide deeper insights to minimarket management. This grouping process helps minimarkets in developing a more targeted product procurement strategy, managing stock more efficiently, and identifying products that have very good sales, good sales, and not good sales. The application development method used in this research is the web-based RAD (Rapid Application Development) method using the PHP programming language and MySQL database.

Corresponding Author:

Sondang,
Manajemen Informatika,
Politeknik Unggul LP3M,
Jl. Iskandar Muda No. 3 CDEF, Medan, 20153
Email: dedematondang@gmail.com

Introduction

In the business world, especially in the supermarket, minimarket, and self-service sectors, developers must find patterns that can increase sales and marketing of products in their self-service by utilizing transaction data (Syafrinal & Febrianti, 2023). Data mining technology arises because of the large amount of data available, the need for information to help the business decision-making process, and the need for information technology infrastructure. Minimarket Diky has a problem with leaders who do not check products sold, what products customers need, and ineffective data storage. Data mining can help Minimarket Diky know which items are selling well and which are not, and to compare sales from year to year. This becomes an effective way to increase sales at Minimarket Diky.

Various types of software greatly assist computer users in completing tasks that must be done due to the rapid advancement of information technology. Quality information will be produced by quality software. The use of information systems can help work, and can even replace human employees. The application of information systems in various aspects of modern life is changing, including the way people see work and its quality. Processed data can produce knowledge or information to help make decisions. Data mining is one of the many applications used in making intelligent information systems to produce high-quality data(Syafrinal & Febrianti, 2023).

Data mining is useful and efficient for analyzing and implementing data, cluster analysis is the classification of uncensored data is an important topic in data mining (Berkhin, 2006). The goal of the study is to divide the collection of patterns into clusters, or groups, of related data points. Compared to patterns from other clusters, patterns within a cluster are more similar to one another. Applications for clustering techniques include document extraction, image segmentation, pattern classification, information retrieval, and more (Xu & Wunch, 2005). Proper analysis and processing of sales trends can assist in identifying the best-selling and least-sold products. This can assist businesses in developing marketing strategies and managing inventory stock (Normah, Rifai, et al., 2020). Decision-making by business executives can be aided by the outcomes of data mining analysis (Normah, Yulianti, et al., 2020).

In terms of data clustering, the K-means algorithm is one of the most popular clustering techniques for finding structure in data sets (Bagirov, 2008; Hamerly & Elkan, 2002; Jain, 2010). The K-Means algorithm, which uses a straightforward iterative scheme to find a local minimum solution, can divide a number of data points into a number of clusters. It is commonly known that the K-Means approach is effective for clustering big data sets. However, identifying the number of clusters is the primary bottleneck in the K-means operation, and the final clustering heavily relies on the initial identification of cluster elements, making K-means challenging to use (Bhatia, 2004). Various groups and individuals are encouraged to use software due to the rapid development of software. In addition, communicating and transacting in the internet era without time or distance limitations. In the industrial era 4.0, information technology has influenced the progress of businesses and companies (Sulistiani & Darwis, 2020;Aldino et al., 2021; Dinasari et al., 2020).

Businesses that use the internet have the ability to reach people around the world who have the opportunity to cooperate and compete. Modern technologies make it easier for companies to work together, create innovations, and become more recognized by society. They can also help customers communicate and interact with companies to purchase or order goods and services (Sulistiani & Darwis, 2020;Rahmanto & Hotijah, 2020; Triyanto et al., 2020). To support other business operations, the retail industry is a very important business field for other industries.

The purpose of this research is to help Minimarket Diky in determining sales products that fall into the categories of very in demand, in demand, and less in demand through the concept of data mining using the K-Means Algorithm. The challenges of applying the K-Means method in Diky's minimarket are complex data dimensions, data scale and normalization, noise and outlier data, changes in consumer behavior, interpretation limitations, and determining the right number of clusters.

The application of the Rapid Application Development (RAD) method accelerates the stages of software development, from planning, design, and implementation, so that it is completed more quickly(Dwi Wijaya, 2020). Software development that is not large-scale and complex is appropriate with the RAD software development method (Hariyanto et al., 2021). In web design, the application of the RAD method can facilitate data searches(Nuku et al., 2020).

The application of the RAD method can produce software that greatly meets user needs, provides added value in achieving revenue, and allows software implementation to run smoothly and quickly (Wahid, 2019). The use of the Rapid Application Development (RAD) method helps in the design of information systems, so that the distribution of information systems is more efficient and effective (Aryanti et al., 2021). The Rapid Application Development (RAD) method allows the creation of online store software, which can increase sales (Mansur & Azzahra, 2023).

The K-Means algorithm can group medium to lowest data. This method is very useful to help agencies find very important information from their data warehouse, which has not been controlled by stock (Ferdy Pangestu et al., 2023). K-Means algorithm can determine products that are easy to sell (Syafrinal & Febrianti, 2023). The K-Means method was chosen for stock management in the retail sector because it is capable of clustering data, efficient for large-scale data, easy to interpret, flexible, cost and time efficient, and adaptive to retail dynamics.

The Rapid Application Development (RAD) method requires a relatively short time, ranging from 60 to 180 days (Mishra & Dubey, 2013). The Rapid Application Development (RAD) method is used for system development that requires a high level of dynamism, limited development time and

cost, current information needs, and the need for close user relationship interaction(Daud et al., 2010). The Rapid Application Development (RAD) method has evolved into an incremental software development method with a faster processing time (Sukamto, 2016) and a development process model that focuses on a short development cycle (Putri & Effendi, 2018).

One of the highest levels of DFD is the context diagram to describe all system inputs or outputs from a process, as well as provide an overview of the system. The system is limited by being represented with a dotted line. The context diagram shows only one process, so there should be no store (Ladjamudin, 2013). Data Flow Diagram (DFD) is a graphical way to show the flow of information and changes used to move data from input to output. The basic elements of a data flow diagram are: 1) External Entity, a notation box that provides data into the system or provides data from the system but is outside the system. If the information system is designed for one part, the external entity is still part of the system; 2) Data Flow, data flow flows between processes, stores data, and shows the data flow of data in the form of inputs or results of system processes. Data flows are indicated by arrows and named lines connecting system components; 3) Process, what the system does is called a process. Each process has one or more inputs and produces one or more output data. Processes can also process data or incoming data streams into outgoing data streams according to the desired specifications. This process is also referred to as a bubble; 4) Data Store, the data store is where the data in the system is stored. Two parallel lines or two lines with one side of the side open can be used to describe a data warehouse. Processes can retrieve or provide data to the database (Fathansyah, 2018).

Once all the important memory has been put into the data store object, relationships can be created between one data store object and another. Entity Relationship (E-R) diagrams are created to show relationships. E-R diagrams display objects (or entities) and their attributes, as well as the relationships between them. Peter Chen introduced the E-R diagram, which is independent of the database technology used. Cardinality and modality are terms for the maximum degree of relation. The cardinality that occurs between two sets of entities can be (1) One to One (1-1), a relation that occurs if an entry in a data store object is associated with only one entry in another data store object. (2) One to Many (1-M), a relation that occurs when an entry in a data store object is related to one or more entries in another data store object. (3) Many to Many (M-M), a relation that occurs when one or more entries in a data store object are related to one or more entries in another data store object.

K-Means is a non-hierarchical data clustering technique that aims to divide the current data into one or more clusters or groups. This method combines data into groups with similar characteristics and groups with different characteristics (Magister Ilmu Komputer et al., 2015). The purpose of data clustering is to reduce the goal features set during the clustering process, usually trying to minimize variation within a cluster and maximize variation between clusters. The K-Means algorithm is known for its simplicity and ability to quickly classify large amounts of data and outliers (Dahria et al., 2019).

Method

This research uses the Rapid Application Development (RAD) method and system design tools that use Context Diagram, Data Flow Diagram (DFD), and Entity Relation Diagram (ERD). The RAD method was chosen because it allows application designs that are easy to develop, has limitations that ensure that the system does not undergo significant changes, and produces a high quality system (Suhartono et al., 2018). In this research, the Efficient Development model is used, which prioritizes creating applications that are balanced in quality, speed, and lower cost. The validation method for cluster results generated from the K-Means method using Silhouette Score is useful for measuring how well objects in a particular cluster are grouped. The RAD method consists of structured stages that allow software development to be carried out in a short time with a focus on short cycles. This method is suitable for small-scale software development (Sagala, 2018). RAD has five stages, namely: 1) Requirement Planning, the purpose of this stage is to collect data from the authorities to determine the services, weaknesses, and objectives of the software system. At the needs analysis stage, researchers conducted observations and interviews to collect data. Observations were made at Minimarket Diky to

collect sales data, and interviews were conducted with employees to ask what features were needed for the application to be designed. Data was collected from Minimarket Diky in the Langkat area. A sample of 2000 data was collected for two months from June to July 2023; 2) Modeling Analysis, the purpose of the modeling analysis stage is to evaluate overall system architecture activities through the identification and explanation of fundamental software system abstractions and their relationship to each other; 3) Modeling Design, after conducting the previous analysis, the modeling design stage aims to create a system design that suits user needs. The stages of analysis and design are carried out repeatedly. The goal at the design stage, or design, is to show the application process to be designed. Calculation of the K-means algorithm, database, Context Diagram design, Data Flow Diagram, and Entity Relationship Diagram are components of web-based application design; 4) Development, to ensure that the prototype meets the previously identified analysis and design specifications, the development stage aims to show the platform, software, and hardware used, as well as implementation constraints; 5) Implementation, before the system is used, the system modeling design designed in the previous stage is implemented through the testing process.

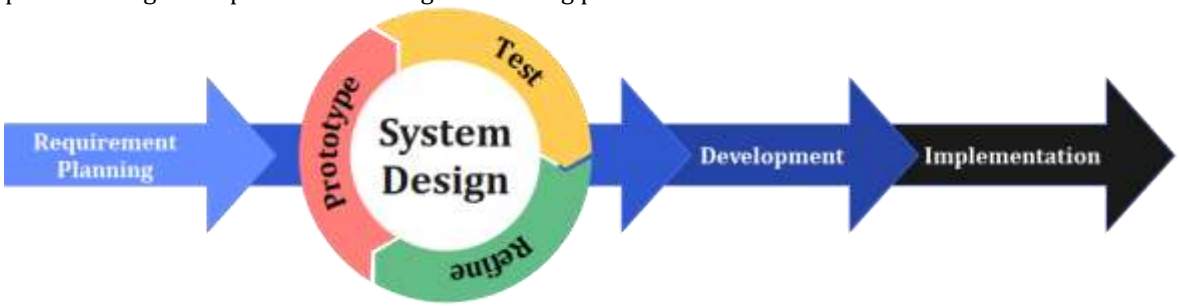

Figure 1. Stages of the Rapid Application Development Method

Results and Discussions

Data Grouping

The following is a sample of goods data that will be grouped based on the K-Means algorithm:

Table 1. Item Data

Item	Item Code	Item Name	Begin Stock	Sold Amount	Stock Amount
1	BRG-001	BATRAI X2000	60	56	4
2	BRG-002	BAWANG GORENG SANGKAYA 100GR	10	10	0
3	BRG-003	BAY FRESH ROSE	16	12	4
4	BRG-004	BAYCLIN 100ML	48	47	1
5	BRG-005	BAYCLIN 200ML	42	30	12
6	BRG-006	BAYCLIN 500ML	54	54	0
7	BRG-007	BAYGON CHERRY BLOSSOM 600ML	6	6	0
8	BRG-008	BAYGON CITERUS FRESH 200ML	24	12	12
9	BRG-009	BAYGON CITRUS FRESH 600ML	12	12	0
10	BRG-010	BAYGON FLOWER GARDEN 600ML	18	18	0
11	BRG-011	BAYGON GREEN TEA & WHITE ORCHID	18	11	7
12	BRG-012	BAYGON LAVENDER	12	11	1
13	BRG-013	BAYGON MAT ELEKTRIK	9	9	0
14	BRG-014	BAYGON MAX HIJAU	534	262	272
15	BRG-015	BAYGON MAX MERAH	280	275	5
16	BRG-016	BAYGON NATURAL EUCALYPTUS 600ML	12	11	1
17	BRG-017	BAYGON TEA BLOSSOM 600ML	12	12	0
18	BRG-018	BB CREAM PIXY 01	3	2	1
19	BRG-019	BB CREAM PIXY 01	9	6	3
20	BRG-020	BB CREAM PIXY 02	3	1	2
21	BRG-021	BB CREAM WARDAH LIGHT	18	16	2
22	BRG-022	BB CREAM WARDAH LIGHT	12	12	0
23	BRG-023	BB CREAM WARDAH NATURAL	17	17	0
24	BRG-024	BD CAKRA	12	5	7

25	BRG-025	BD KARET CB	14	14	0
26	BRG-026	BD LIDI	48	36	12
27	BRG-027	BDL PAPAYA	24	14	10
28	BRG-028	BEAR BRAND GOLD	24	7	17
29	BRG-029	BEAR BRAND GOLD 140ML	96	85	11
30	BRG-030	BEAR BREND	3183	2943	240

Determination of the Number of Clusters (K)

The data is grouped into 3 clusters (K = 3), namely Cluster 1 represents Very Hot Selling Goods, Cluster 2 represents Hot Selling Goods, and Cluster 3 represents Non-Selling Goods.

Centroid Determination

Determining the initial Centroid point in this study was randomly selected from the data in Table 1 so that the centroid point was obtained as follows:

Table 2. Initial Centroid Data

Centroid	Initialization	Item Name	Item Amount	Quantity Sold	Total Stock
C1	BRG-001	BATRAI X2000	60	56	4
C2	BRG-008	BAYGON CITERUS FRESH 200ML	24	12	12
C3	BRG-017	BAYGON TEA BLOSSOM 600ML	12	12	0

Euclidean Distance Calculation

After preparing the item data to be processed with the K-Means Clustering method, the data processing process can begin. Calculating the closest distance to the Centroid using the Euclidean formula is:

$$D(X_2 - X_1) = \sqrt{(X_2a - X_1a)^2 + (X_2b - X_1b)^2 + (X_2c - X_1c)^2}$$

Table 3 shows the calculation results of the data in Table 2 using the Euclidean formula.

Table 3. Iteration-1 Clustering Calculation Results

Item Number	Item Code	Starting	Sold	Stock	C1	C2	C3	Nearest Distance	Cluster
1	BRG-001	60	56	4	0	57,4108	65,23803	0	C1
2	BRG-004	48	47	1	15,29706	43,84062	50,21952	15,29705854	C1
3	BRG-006	54	54	0	7,483315	52,99057	59,39697	7,483314774	C1
4	BRG-014	534	262	272	582,1821	624,6599	639,5061	582,1821021	C1
5	BRG-015	280	275	5	310,4223	367,0885	375,5236	310,422293	C1
6	BRG-026	48	36	12	24,65766	33,94113	44,89989	24,65765601	C1
7	BRG-029	96	85	11	46,75468	102,5378	111,8302	46,75467891	C1
8	BRG-030	3183	2943	240	4259,53	4315,325	4324,766	4259,529786	C1
9	BRG-005	42	30	12	32,61901	25,45584	36,98648	25,45584412	C2
10	BRG-008	24	12	12	57,4108	0	16,97056	0	C2
11	BRG-011	18	11	7	61,62792	7,874008	9,273618	7,874007874	C2
12	BRG-027	24	14	10	55,64171	2,828427	15,74802	2,828427125	C2
13	BRG-028	24	7	17	62,17717	7,071068	21,40093	7,071067812	C2
14	BRG-003	16	12	4	62,2254	11,31371	5,656854	5,656854249	C3
15	BRG-010	18	18	0	56,78028	14,69694	8,485281	8,485281374	C3
16	BRG-021	18	16	2	58,03447	12,32883	7,483315	7,483314774	C3
17	BRG-023	17	17	0	58,18935	14,76482	7,071068	7,071067812	C3
18	BRG-025	14	14	0	62,41795	15,74802	2,828427	2,828427125	C3
19	BRG-002	10	10	0	68,0588	18,54724	2,828427	2,828427125	C3
20	BRG-007	6	6	0	73,7021	22,44994	8,485281	8,485281374	C3
21	BRG-009	12	12	0	65,23803	16,97056	0	0	C3
22	BRG-012	12	11	1	65,8635	16,30951	1,414214	1,414213562	C3
23	BRG-013	9	9	0	69,46942	19,44222	4,242641	4,242640687	C3
24	BRG-016	12	11	1	65,8635	16,30951	1,414214	1,414213562	C3
25	BRG-017	12	12	0	65,23803	16,97056	0	0	C3
26	BRG-018	3	2	1	78,57481	25,72936	13,49074	13,49073756	C3
27	BRG-019	9	6	3	71,42829	18,49324	7,348469	7,348469228	C3

28	BRG-020	3	1	2	79,23383	25,72936	14,3527	14,35270009	C3
29	BRG-022	12	12	0	65,23803	16,97056	0	0	C3
30	BRG-024	12	5	7	70,09993	14,76482	9,899495	9,899494937	C3

In order to perform the second iteration calculation for the new Centroid, the average value of each Cluster that has been grouped is calculated. The Centroid recalculation is as follows:

Table 4 Initial centroid Iteration 2

Centroid	Starting	Sold	Stock
C1	537,875	469,75	68,125
C2	26,4	14,8	11,6
C3	11,47058824	10,23529	1,235294

Next, calculate the cluster distance using the centroid in Table 4 as in Table 3, then it can be grouped based on iteration 2, namely:

Table 5. Iteration-2 Clustering Calculation Results

Item Number	Item Code	Starting	Sold	Stock	C1	C2	C3	Nearest Distance	Cluster
1	BRG-014	534	262	272	291,1015	621,7506	640,1062	291,1015179	C1
2	BRG-015	280	275	5	329,2595	363,4014	377,1241	329,2595386	C1
3	BRG-030	3183	2943	240	3625,354	4311,687	4326,282	3625,354168	C1
4	BRG-001	60	56	4	635,3468	53,70438	66,76193	53,70437599	C2
5	BRG-004	48	47	1	650,5389	40,19652	51,82757	40,19651726	C2
6	BRG-006	54	54	0	641,5794	49,32504	61,03791	49,32504435	C2
7	BRG-026	48	36	12	656,7089	30,26814	45,97931	30,26813506	C2
8	BRG-029	96	85	11	588,6844	98,85626	113,2711	98,85625929	C2
9	BRG-005	42	30	12	665,1482	21,7844	37,92845	21,78439809	C2
10	BRG-008	24	12	12	690,4727	3,709447	16,61262	3,709447398	C2
11	BRG-027	24	14	10	689,3142	2,993326	15,74736	2,993325909	C2
12	BRG-028	24	7	17	693,411	9,785704	20,39557	9,785703858	C2
13	BRG-011	18	11	7	696,0301	10,3034	8,743559	8,743559172	C3
14	BRG-003	16	12	4	697,1374	13,18181	5,592259	5,592258614	C3
15	BRG-010	18	18	0	692,0904	14,67515	10,22007	10,22006981	C3
16	BRG-021	18	16	2	693,2039	12,81249	8,743559	8,743559172	C3
17	BRG-023	17	17	0	693,4943	15,09172	8,823922	8,82392156	C3
18	BRG-025	14	14	0	697,7063	16,99882	4,700732	4,700732476	C3
19	BRG-002	10	10	0	703,3229	20,65333	1,934928	1,934927553	C3
20	BRG-007	6	6	0	708,9401	25,06312	7,027873	7,027873324	C3
21	BRG-009	12	12	0	700,5145	18,70187	2,218201	2,218200898	C3
22	BRG-012	12	11	1	701,0719	18,28004	0,959383	0,959382731	C3
23	BRG-013	9	9	0	704,7271	21,70161	3,02584	3,02584027	C3
24	BRG-016	12	11	1	701,0719	18,28004	0,959383	0,959382731	C3
25	BRG-017	12	12	0	700,5145	18,70187	2,218201	2,218200898	C3
26	BRG-018	3	2	1	713,7136	28,70122	11,81636	11,8163572	C3
27	BRG-019	9	6	3	706,4093	21,31103	5,211114	5,211114021	C3
28	BRG-020	3	1	2	714,2761	28,8125	12,55493	12,55493121	C3
29	BRG-022	12	12	0	700,5145	18,70187	2,218201	2,218200898	C3
30	BRG-024	12	5	7	704,4667	18,01555	7,805153	7,80515312	C3

After performing calculations in finding a ratio that does not change anymore, the iteration will stop with the calculation of iteration 2 and the next result is the same, so that the centroid value does not change anymore, the results of the clustering of goods are as follows:

Table 6. K-Means Clustering Process Results

Item Number	Item Code	Item Name	C1	C2	C3
1	BRG-001	BATRAI X2000		Bestsellers	
2	BRG-002	BAWANG GORENG SANGKAYA 100GR			Not in-demand
3	BRG-003	BAY FRESH ROSE			Not in-demand
4	BRG-004	BAYCLIN 100ML		Bestsellers	

5	BRG-005	BAYCLIN 200ML		Bestsellers	
6	BRG-006	BAYCLIN 500ML		Bestsellers	
7	BRG-007	BAYGON CHERRY BLOSSOM 600ML			Not in-demand
8	BRG-008	BAYGON CITERUS FRESH 200ML		Bestsellers	
9	BRG-009	BAYGON CITRUS FRESH 600ML			Not in-demand
10	BRG-010	BAYGON FLOWER GARDEN 600ML			Not in-demand
11	BRG-011	BAYGON GREEN TEA & WHITE ORCHID			Not in-demand
12	BRG-012	BAYGON LAVENDER			Not in-demand
13	BRG-013	BAYGON MAT ELEKTRIK			Not in-demand
14	BRG-014	BAYGON MAX HIJAU	Highly In-Demand		
15	BRG-015	BAYGON MAX MERAH	Highly In-Demand		
16	BRG-016	BAYGON NATURAL EUCALYPTUS 600ML			Not in-demand
17	BRG-017	BAYGON TEA BLOSSOM 600ML			Not in-demand
18	BRG-018	BB CREAM PIXY 01			Not in-demand
19	BRG-019	BB CREAM PIXY 01			Not in-demand
20	BRG-020	BB CREAM PIXY 02			Not in-demand
21	BRG-021	BB CREAM WARDAH LIGHT			Not in-demand
22	BRG-022	BB CREAM WARDAH LIGHT			Not in-demand
23	BRG-023	BB CREAM WARDAH NATURAL			Not in-demand
24	BRG-024	BD CAKRA			Not in-demand
25	BRG-025	BD KARET CB			Not in-demand
26	BRG-026	BD LIDI		Bestsellers	
27	BRG-027	BDL PAPAYA		Bestsellers	
28	BRG-028	BEAR BRAND GOLD		Bestsellers	
29	BRG-029	BEAR BRAND GOLD 140ML		Bestsellers	
30	BRG-030	BEAR BREND	Highly In-Demand		

Based on Table 6, the grouping of Highly In-Demand (C1) is 3 data, Bestsellers (C2) is 9 data, and Not in-demand (C3) is 18 data.

Evaluation Results

Based on the results of the Silhouette Score calculation from the clustering results using K-Means, Cluster C1 shows the best cluster quality of C2 and C3, which is 0.54, C2 shows the best quality compared to C3 with a value of 0.50, and C3 shows good cluster quality with a value of 0.39.

System Design

Context Diagram, Data Flow Diagram (DFD), Entity Relationship Diagram (ERD), and User Interface Design are system design components that aim to design input, output, and interface components of applications that are built based on the plan. Context Diagram is useful for explaining the activities that occur in the system. The following is the Context Diagram design:

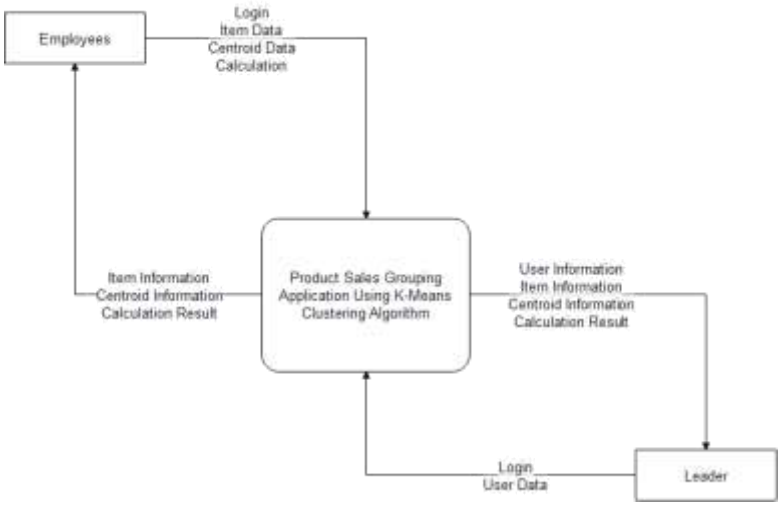

Figure 2. Context Diagram

DFD is useful for explaining in general what processes can be carried out on the system.

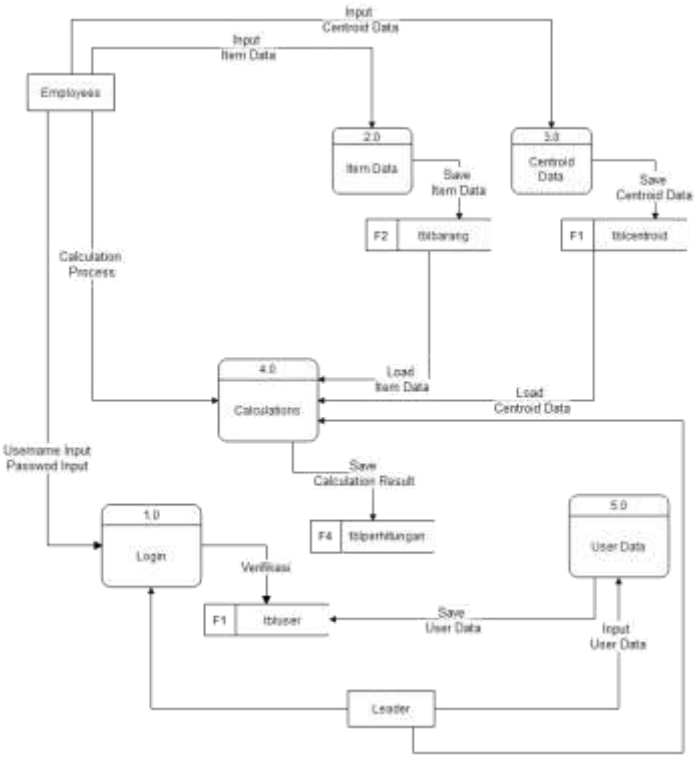

Figure 3. Data Flow Diagram

Model data and describe the relationship between data in the system using an E-R diagram or Entity Relationship Diagram (ERD).

Figure 4. Entity Relationship Diagram

Program Results

The designed application produces a data mining application to classify product sales using the web-based K-Means Clustering algorithm. The application has two levels of access rights, namely admin and employee.

Login Page

The login page shown in Figure 5 is used by admin and employees to login to the application.

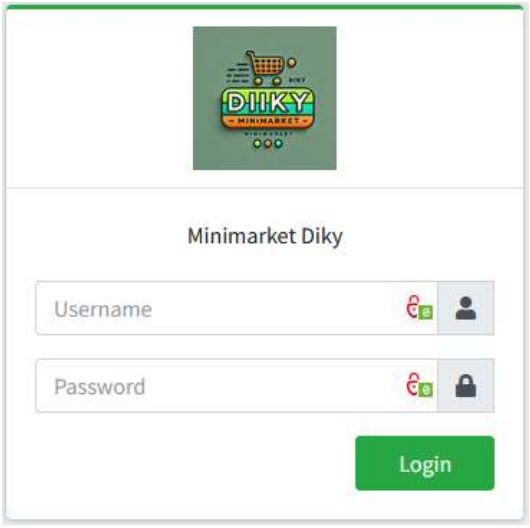

Figure 5. Application Login Page

Main Menu Page

The main menu page is a list of menus that can be selected based on the needs that can display data according to the type of menu selected.

Figure 6. Application Main Menu Page

Item Data Page

The Item Data page is a page for displaying goods data. Item data can be input through the Add Item button and the Import Item Data menu which functions to import item data from Excel files.

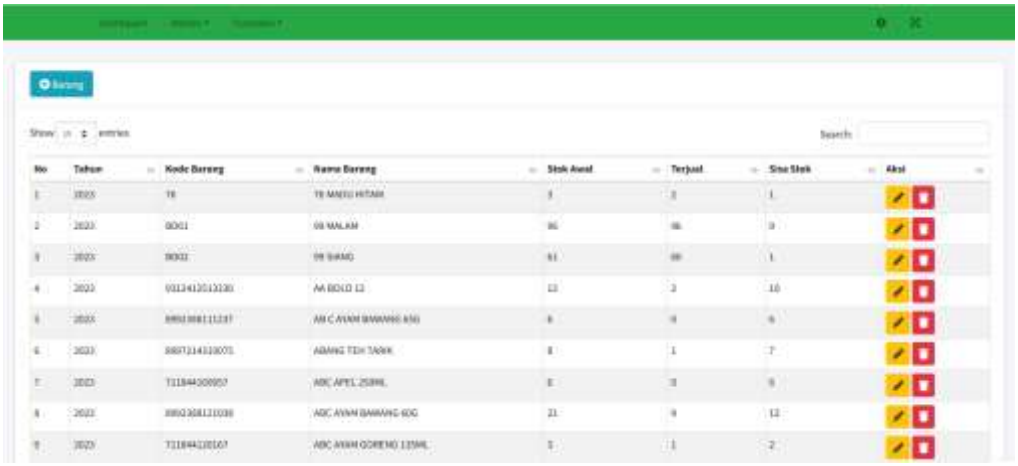

Figure 7. Item Data Page

Centroid Data Page

The Centroid Data page functions to display Centroid data and can change Centroid values.

Figure 8. Centroid Data Page

Calculation Result

The calculation results page is the result of the calculation process using the K-Means Clustering method through the designed application.

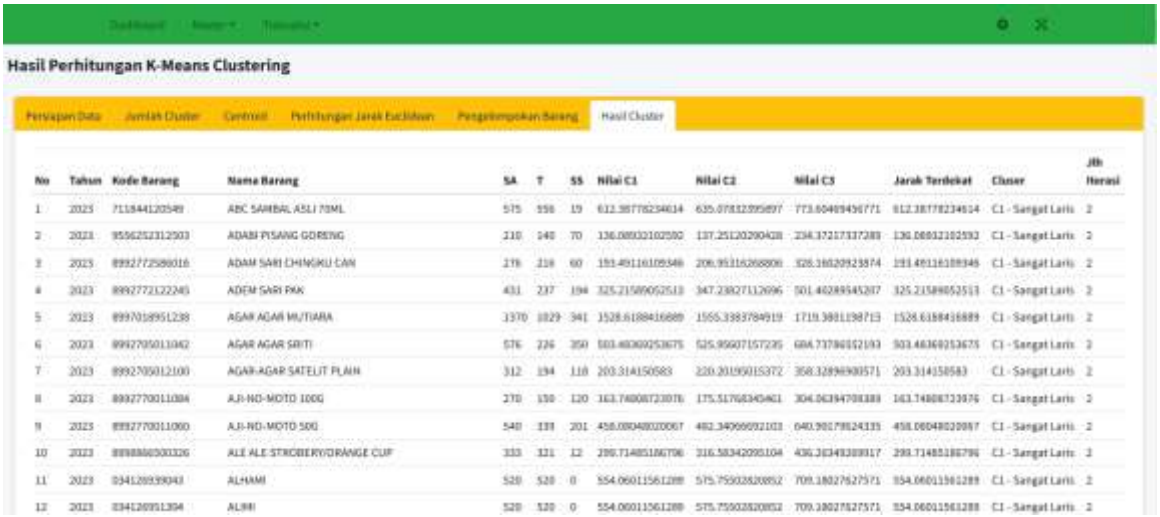

Figure 9. Calculation Result Page

Product Clustering Results

A page that displays product clusters that are very in demand, in demand and not in demand.

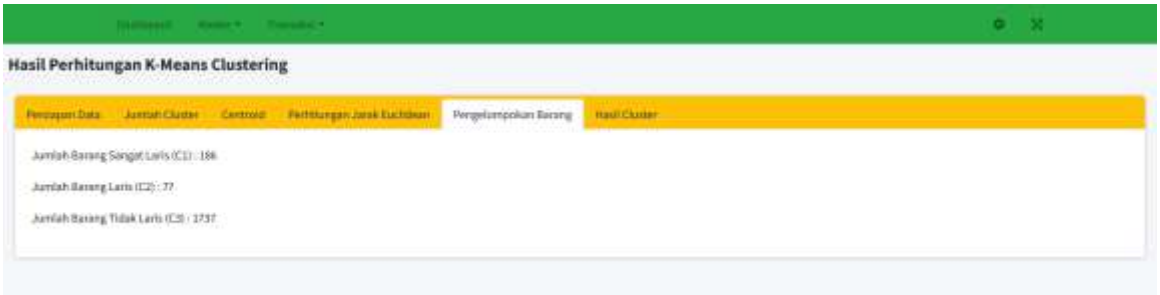

Figure 10. Product Grouping Result Page

Conclusions

Based on the results of research to design Product Sales Grouping Applications Using the K-Means Clustering Algorithm, it can be concluded that the use of the K-Means Clustering Algorithm can group product sales based on clusters of Highly In-Demand, In-Demand, and Not In-Demand. The designed system can be replicated in other minimarkets even though the operational structure is different. The use of the RAD method in designing applications can be done quickly at a relatively low cost. The combination of RAD method and K-Means Clustering Algorithm can be used as a reference as an application development method and product sales grouping that is superior in terms of speed, accuracy and low cost. The implementation of the application can affect the efficiency of stock management by segmenting products based on sales patterns, forecasting seasonal stock needs, and optimal product placement in the warehouse, while to increase sales, it can segment customers for the right offer, optimize product placement in stores, data-based promotions, and introduce new products. Evaluation results based on Silhouette Score calculations show that Cluster C1 has the best cluster quality of C2 and C3 at 0.54, C2 shows the best quality compared to C3 with a value of 0.50, and C3 shows good cluster quality with a value of 0.39.

References

Ahmed, A. H., & Ashour, W. (2011). An Initialization Method for the K-means Algorithm using RNN and Coupling Degree. In *International Journal of Computer Applications* (Vol. 25, Issue 1).

Aldino, A. A., Saputra, A., Nurkholis, A., & Setiawansyah, S. (2021). Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur. *Building of Informatics, Technology and Science (BITS)*, *3*(3), 325–330. https://doi.org/10.47065/bits.v3i3.1041

Aryanti, R., Fitriani, E., Ardiansyah, D., & Saepudin, A. (2021). Penerapan Metode Rapid Application Development Dalam Pengembangan Sistem Informasi Akademik Berbasis Web. *Paradigma - Jurnal Komputer Dan Informatika*, *23*(2). https://doi.org/10.31294/p.v23i2.11170

Bagirov, A. M. (2008). *Modified global k-means algorithm for minimum sum-of-squares clustering problems*.

Berkhin, P. (2006). *A Survey of Clustering Data Mining Techniques*.

Bhatia, S. K. (2004). *Adaptive K-Means Clustering*. www.aaai.org

Dahria, M., Gunawan, R., & Lubis, Z. (2019). *Seminar Nasional Sains & Teknologi Informasi (SENSASI) Implementasi K-Means Untuk Pengelompokan Produk Terbaik PT. Koko Pelli*. http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage|495

Daud, N. M. N., Bakar, N. A. A. A., & Rusli, H. M. (2010). Implementing rapid application development (RAD) methodology in developing practical training application system. *2010 International Symposium on Information Technology*, *3*, 1664–1667. https://doi.org/10.1109/ITSIM.2010.5561634

Dinasari, W., Budiman, A., & Ayu Megawaty, D. (2020). SISTEM INFORMASI MANAJEMEN ABSENSI GURU BERBASIS MOBILE (STUDI KASUS : SD NEGERI 3 TANGKIT SERDANG). *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *1*(2), 50–57. http://jim.teknokrat.ac.id/index.php/JTSI

Dwi Wijaya, Y. (2020). *PENERAPAN METODE RAPID APPLICATION DEVELOPMENT (RAD) DALAM PENGEMBANGAN SISTEM INFORMASI DATA TOKO Penulis Korespondensi*. http://www.jurnal.umk.ac.id/sitech

Fathansyah. (2018). *Basis data*. Informatika.

Ferdy Pangestu, F. P., Nur Yasin, N. Y., Ronald Chistover Hasugian, R. C., & Yunita, Y. (2023). Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *12*(1), 53–62. https://doi.org/10.32736/sisfokom.v12i1.1461

Hamerly, G., & Elkan, C. (2002). *Alternatives to the k-means algorithm that find better clusterings*.

Hariyanto, D., Sastra, R., Putri, F. E., Informasi, S., Kota Bogor, K., & Komputer, T. (2021). Implementasi Metode Rapid Application Development Pada Sistem Informasi Perpustakaan. In *Jurnal JUPITER* (Vol. 13, Issue 1).

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/https://doi.org/10.1016/j.patrec.2009.09.011

Ladjamudin, A.-B. Bin. (2013). *Analisis dan Desain Sistem Informasi*. Graha Ilmu.

Magister Ilmu Komputer, W., Nusa Mandiri, S., & Satria Wahono, R. (2015). Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems*, *1*(1). http://journal.ilmukomputer.org

Mansur, M., & Azzahra, S. liyah. (2023). *PENERAPAN METODE RAPID APPLICATION DEVELOPMENT (RAD) DALAM PERANCANGAN SISTEM PENJUALAN ONLINE DI TOKO ELEKTRONIK 85 INFORMASI ARTIKEL ABSTRAK* (Vol. 5, Issue 1). https://e-journal.unper.ac.id/index.php/informatics

Mishra, A., & Dubey, D. (2013). A Comparative Study of Different Software Development Life Cycle Models in Different Scenarios. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(5). www.ijarcsms.com

Normah, Rifai, B., & Sari, P. (2020). Algoritma Apriori Sebagai Solusi Kontrol Persediaan Suku Cadang Mobil PT. Buanasakti Aneka Motor Jakarta. *Paradigma*, *22*(2).

Normah, Yulianti, I., Novianti, D., Winnarto, M. N., Zumarniansyah, A., & Linawati, S. (2020). Comparison of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012006

Nuku, R., Masihor, E. J. A., & Pasaribu, R. L. (2020). Penerapan Metode RAD dalam Perancangan Sistem Informasi Aplikasi Penelusuran Putusan (SIAPP). In *JOINTER : JOURNAL OF INFORMATICS ENGINEERING* (Vol. 01, Issue 02).

Putri, M. P., & Effendi, H. (2018). Implementasi Metode Rapid Application Development Pada Website Service Guide "Waterfall Tour South Sumatera." In *Jurnal SISFOKOM* (Vol. 07).

Rahmanto, Y., & Hotijah, S. (2020). PERANCANGAN SISTEM INFORMASI GEOGRAFIS KEBUDAYAAN LAMPUNG BERBASIS MOBILE. In *JDMSI* (Vol. 1, Issue 3).

Sagala, J. R. (2018). *garuda855743*.

Suhartono, R. D. A., Sfenrianto, S., Mustofa, M., Andriyani, D., & Kaburuan, E. R. (2018). Development of 3D Solar System Application Using RAD Model for Elementary Schools. *Int. Conf. Orange Technol* , 1–4.

Sulistiani, H., & Darwis, D. (2020). Penerapan Metode Agile untuk Pengembangan Online Analytical Processing (OLAP) pada Data Penjualan (Studi Kasus: CV Adilia Lestari). *Jurnal CoreIT*, *6*(1).

Syafrinal, I., & Febrianti, E. L. (2023). *PENERAPAN ALGORITMA K-MEANS PADA APLIKASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN (STUDI KASUS: ZAHRA MART)* (Vol. 13, Issue 1).

Triyanto, R., Stekpi, J., Trilogi, /, Kalibata, T., & Selatan, J. (2020). Rancang Bangun Aplikasi Penjualan Berbasis Website (Studi Kasus : Toko Waroeng Bola). *Jurnal Sistem Informasi Dan Sains Teknologi*, *2*(1).

Wahid, B. A. (2019). Penerapan Metode Rapid Application Development Terhadap Penjualan Fashion Distro Secara Online. In *Jurnal Esensi Infokom* (Vol. 3, Issue 1).

Xu, R., & Wunch. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645–678.