

Data Mining Sales of Skin Care Products Using the K-Means Method

Dasril Aldo^{1)*}

¹⁾Teknik Informatika, Institut Teknologi Telkom Purwokerto ¹<u>dasril@ittelkom-pwt.ac.id</u>

Submitted : Dec 16, 2022 | Accepted : Dec 27, 2022 | Published : Jan 1, 2023

Abstract: Data mining is a form of method advancement in computerization that can dig past data into very valuable information. The problem in this study is that the sale of beauty and skincare carried out is still done manually so that it can cause it to not match the stock in the storage warehouse with changing market demand. Data mining with the K-Means method is one solution to this problem by grouping similar data, in this study grouping into two, namely best-selling and unsold products. The purpose of this study is that the store can provide stock of products in the warehouse according to market demand. Using a sample of 30 data resulted in 18 data as skin care products were not selling well and 12 data as skin care products were not selling well. With the results of a 100% similarity between manual calculations and using the rapid miner application, it can be concluded that the K-Means algorithm can be used as a solution to the problems.

Keywords: Data Mining; K-Means; Clustering; Skin Care; Rapid Miner.

INTRODUCTION

The current advancement of technology is very helpful for various sectors in helping the activities that are required to move quickly. This is due to the changing market needs for a product, sometimes market demand is very high and sometimes market demand becomes very low, so business actors must be observant in managing the stock of goods in the warehouse so that they do not accumulate and can meet market demand. various technological innovations have been widely applied such as e-tandur in the agricultural sector (Amalia et al., 2020), e-monitoring for basic food prices in the consumer goods sector (Juansyah, 2018), Fintecth in the financial sector (Narastri, 2020), digitalization in the industrial sector (Adha et al., 2020), E-Government in the government sector (Rachmatullah & Purwani, 2022), virtual reality technology in the real estate sector (Saputra et al., 2021), integrated information systems (Aldo et al., 2022) and interactive multimedia (Arifah et al., 2022) in the tourism sector and many other sectors. In this study, the application of technology in the trade sector, especially the sale of cosmetic products.

Cosmetic products are ingredients that are used on the outside of the body (skin, hair, and lips) or teeth with the aim of cleansing, scenting, and improving appearance. Quoted from the Indonesian Digital Marketing Association, it shows that the use of beauty and skincare products in Indonesia is very large, with serum usage data reaching 11.18 million, Lip Cream 9.40 million, lipstick 8.88 million, mascara 5.69 million, eyebrow 5.43, toner, 5.08 million, facemask 4.50 million, eyeliner 4.09 million, fondation 3.45 million and eyeshadow 2.97 million. From the data, it is certain that users of anticipatory and skincare products will continue to increase.

Cosmetic Store which is engaged in the business of selling beauty and skincare products. This shop was founded by young people whose hobbies are doing business and are very initiative in the world of sales business. The problem in this study is that the sale of beauty and skincare carried out is still done manually so that it can cause it to not match the stock in the storage warehouse with changing market demand. If this happens continuously, it will cause a buildup of products and not meet market demand which causes expired products and consumer disappointment because the products they want to buy are not available. This can also result in stores causing losses and consumers being able to move to other stores. As an alternative solution to this problem, data mining can provide a solution to the problem of grouping goods so that there is no problem with the stock in the storage warehouse.

Data mining is the process of data mining using statistical techniques, mathematics, artificial intelligence, and machine learning to identify useful information and related knowledge from various large databases. In general, companies collect information through a database system that is useful for collecting transaction data, then later the data is processed so that it can be known the level and volume of a sale at a certain time and so on (Karsito & Sari, 2018). Research on data mining including by Pranata and Utomo in 2020 with the results of data





mining can increase the supply of spare parts in motor workshops based on data from sinar service workshops (Pranata & Utomo, 2020). Further research by Risal et al in 2021 with the results of data mining can classify data on the COVID-19 case rate with an accuracy of 91% (Risal et al., 2021). Furthermore, research conducted by Mardiah and Yulia in 2019 with the results of Data mining A priori algorithms can be used to process data on the results of transactions selling motorcycle parts in motor repair shops, so as to determine the rules of association between goods sold, making it easier for users to sell the number of goods sold. Further research by Noviyanto in 2020 with data mining results can group data on deaths caused by COVID-19 in the Asian Continent (Noviyanto, 2020) and many other studies. In data mining, there are many methods that can be used, including the naïve bayes method (Sitoto et al., 2022), the K-Means Clustering method (Gong, 2021), the FP-Growth method (Chen et al., 2021), the A priori methods. In this study, the method used was the K-Means Clustering method to group skin care products based on the category of best-selling and unsold goods, so that the store knows the products to be stocked in the warehouse.

The K-Means Clustering method is one of the most commonly used cluster analysis algorithms. The inefficiencies that K-Means use to handle larger data should be noted (Harahap, 2019). The use of the K-means method in this study is because this algorithm has a fairly high accuracy of the object size, so this algorithm is relatively more measurable. The purpose of using this method is to later be able to generate knowledge from the data stored in the form of reports on best-selling and under-selling goods in order to increase stock on the inventory of goods in the warehouse so that there is no accumulation and shortage of stock of skin care products.

LITERATURE REVIEW

Data mining is an approach used in extracting the meaning of data from a given dataset using one or more of those approaches as statistics, machine learning, mathematics or artificial intelligence techniques (Al-hashedi & Magalingam, 2021). Data mining is also known as a very useful technique to help companies find very important information contained in data warehouses (Ependi & Akbar, 2021). Data Mining is a series of processes to add and search for information that has not been known manually from a database. The resulting information is obtained by extracting and recognizing important or interesting patterns from the data contained in the database. Data mining is mainly used to find knowledge contained in large databases so it is often called Knowledge Discovery Databases (KDD) (Syahril et al., 2020). The stages in the KDD process on the database are as shown in Figure 1. below (Renhoran et al., 2018):



The pattern of information generated from the data mining process above is part of the KDD process called interpretation. This stage includes examining whether the patterns or information found contradict previously existing facts or hypotheses. The 5 (five) stages above in the Knowledge Discovery in Database (KDD) process can be explained as follows:

- 1. Selection First, select data from multiple data sources by cleaning data that has noise or missing value.
- 2. Preprocessing Second, merge all the data sources that have been collected before the next stage of the process continues.
- 3. Transformation Third, transform data into a form that is more suitable for data mining.
- 4. Data Mining Fourth, apply the data mining method to the system, so that it can generate patterns from the collected data.
- 5. Interpretation / Evaluation Fifth, interpret and evaluate the pattern obtained, so that it can be identified whether the pattern can represent the knowledge to be achieved.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



The method used in this study is the K_Means method that functions to process existing data. This method is a clustering algorithm in data mining to be able to produce groups of large amounts of data with a point-based partitioning method with fast and efficient computational time. In addition, the definition of K-Means is a distance-based clustering method that divides data into a number of clusters and this algorithm only works on numerical attributes. It can be concluded that the KMeans algorithm will make the data that the cluster wants to do will later go to several centers / points where the point is used as a reference in grouping and the data carried out clustering must be numerical. Basically, the K-Means Algorithm only takes part of the number of components obtained to then be used as the initial cluster center, in determining the center of this cluster is randomly selected from the data population (Suhartini & Yuliani, 2021).

The steps to do in clustering with the K-Means Clustering method are as follows (Handoko et al., 2020):

- 1. Specifies the value of k for the number of clusters that you want to form.
- 2. Select the value for the initial cluster center (centroid) as much as k.
- 3. Calculates the distance of each input data to each centroid using the Euclidean Distance formula until the closest distance of each data to the centroid is found. Here is the Euclidean Distance equation:

$$d(x_{i,j},\mu_j) = \sqrt{\sum (x_i - \mu_j)^2} (1)$$

Where = criteria data to, = centroid cluster j and $x_{i,\mu_j} \Sigma$ = summation

- 4. Classify/group each data based on its proximity to centroids (the smallest distance).
- 5. Renews centroid values. The new centroid value is obtained from the average of the cluster in question by using the formula:

$$\mu_j(t+1) = \frac{1}{n_{sj}} \sum j \in sjx_j \ (2)$$

Where the $\mu_i(t+1)$ = new centroid on the 1st iteration (t+1), n_{si} = a lot of data on the Sj cluster.

6. Perform a loop from steps 3 through 5, until none of the members of each cluster have changed.

If step 6 has been met, then the cluster center value (μ_j) in the last iteration will be used as a parameter to determine data classification. In this study, testing was carried out using the Web-based K-Means Clustering application. The calculation results obtained through the application are used as a comparison with the acquisition of data through manual calculations so that the accuracy of the data obtained by the application can be known.

METHOD

In order for the research to be more targeted, easy to understand, the sequence of steps will be made systematically so that it can be used as a clear and easy guideline to solve existing problems. The sequence of steps to be made in this study is represented in Figure 2.



The explanation of the research flow above is as follows:

1. Preliminary research

Preliminary research is a problem identification process aimed at the introduction of problems or an inventory of problems, problem identification is carried out by approaching the object of study. This stage has the aim of knowing the problems that occur precisely, so that the most appropriate solution is used as a problem solution to the problem at hand. At this stage, a search for information or problems is carried out, information obtained from a Cosmetics Store engaged in the business of selling beauty and skin care products. This shop was founded by young people who like doing business and are very initiative in the sales business world. The problem in this study is that the sale of beauty and skincare





carried out is still done manually so that it can cause it to not match the stock in the storage warehouse with changing market demand. If this happens continuously, it will cause a buildup of products and not meet market demand which causes expired products and consumer disappointment because the products they want to buy are not available. This can also result in stores causing losses and consumers being able to move to other stores. From these problems, a solution will be found.

- 2. Research Data Collection Data collection is carried out by collecting data in the form of skin care product types with attributes of product name, stock and number sold. The place of this research is in the Cosmetic Store.
- 3. Data Analysis With K-Means

The analysis stage is intended to provide alternatives to predetermined problems. The steps performed, as follows:

- a. Specifies the value of k for the number of clusters that you want to form.
- b. Select the value for the initial cluster center (centroid) as much as k.
- c. Calculates the distance of each input data to each centroid using the Euclidean Distance formula until the closest distance of each data to the centroid is found.
- d. Classify/group each data based on its proximity to centroids (the smallest distance).
- e. Renews centroid values.
- f. Perform a loop from steps 3 through 5, until none of the members of each cluster have changed.
- 4. System Implementation

At this stage, the results of analysis with manual calculation of the K-means method will be implemented to software rapid miners.

5. System Testing The test was carried out by comparing the results of the K-means method with the software rapid miner.

RESULT

The use of data used in this study is in the form of a skincare product dataset consisting of 3 attributes, namely product name, number of stocks, number of sold at Cosmetic Shop. The data is shown in Table 1.

No	Product Name	Stock	Sold
1	XXX Cheek and Lip Tint	1590	693
2	XXX Cheek Blossom Blush On	1582	1266
3	XXX Complete Makeup 058	1612	999
4	XXX Deep Black Mascara	1566	1334
5	XXX Eau De Perfume Woman	1590	978
6	XXX Exclusive Box Package Jelly Tint	1578	1243
7	XXX Eyeliner Pen	1540	660
8	XXX Eyeshadow Palette 7007	1562	750
9	XXX Eyeshadow Palette 7669B	1540	770
10	XXX Eyeshadow Palette 7669D	1566	696
11	XXX Eyeshadow Palette 7672	1558	677
12	XXX Face Serum Anti Acne Set	1552	854
13	XXX Face Serum Anti Aging Set	1540	550
14	XXX Face Serum Dark Spot Treatment Set	1606	888
15	XXX Face Serum Exfoliating Treatment Set	1558	1016
16	XXX Face Serum Skin Barrier Treatment Set	1544	869
17	XXX Icare Antiseptic Wipes 10 Sheets	1558	339
18	XXX Icare Hand sanitizer gel 100ml	1550	1142
19	XXX Icare Hand Sanitizer Gel 30ml	1540	660
20	XXX Icare Hand Sanitizer Gel 5 Liters	1562	750
21	XXX Icare Handwash 5 Liters	1540	770
22	XXX Instant Bright Body Lotion	1566	696
23	XXX Jelly Tint	1624	870
24	XXX Korean Dewy Ombre Lips (Honest Lip Velvet + Lip Tint)	1624	696
25	XXX Korean Matte Ombre Lips (Hazelnude Lip Cream + LipTint)	1684	783
26	XXX Lip Crayon Satin	1606	888
27	XXX Lip Velvet	1556	707

Table 1. Skin Care Product Data





No	Product Name	Stock	Sold
28	XXX Lipstick Intense Matte	1606	786
29	XXX Make Up Kit Complete Beauty care 502	1594	700
30	XXX Makeup Palette Cappadocia	1646	393

Furthermore, in analyzing the system using the k-means algorithm with the following stages:



Fig. 3 Stages of Data Analysis

Clustering is a data mining processing technique that is included in the category of unsupervised learning, namely learning without a teacher or without using class labels as in classification. Clustering in the K-Means algorithm looks for similarity values between data and distance calculation techniques, one of which uses euclidean distance, the problem that always occurs is how to estimate the best number of k, so that an evaluation of the stability of cluster repetition is needed to estimate the right number of clusters (k).

- The K-means steps can be seen in the algorithm and flowchart as below:
- 1. Add the sum of K
- 2. Calculate the centroid/average of the data present in each cluster
- 3. Allocate individual data to the nearest centroid/average
- 4. Data grouping
- 5. Back to Step 2, if there is still data moving clusters or if the centroid value changes, there is above the specified threshold value or if the value change in the objective function used is above the specified threshold value
- 6. Stage of calculation of the total value
- 7. Ranking stage.

Based on the explanation of the *k-means* method, it will then be translated into a flowchart which is then implemented and tested with a rapid miner tool. The flow chart of the *k-means* method can be seen in the form of figure 3. below:



Fig. 4 Process Flowchart on K-Man Clustering

Figure 4.3 explains how the k-means process is performed basically. It then paralates into K-means by parallelizing against the search for the smallest distance between the data to the centroid by dividing the data into several parts of the thread; and the centroid update count where each centroid is formed has its own thread. While the search for a stable (best) number of k is to compare the difference in the largest mean value between the MCA (maximum cluster assignments) index K-means and the MCA index random prototype baseline. MCA index Kmeans is the maximum value of k clusters formed from the K-means algorithm, while MCA index





random prototype baseline is the lower limit of clusters that should be formed if using random data as a cluster shaper. While the extract pattern is the process of retrieving class labels that have been formed during clustering

Determination of the center of the new cluster

To get a new cluster center, it can be calculated from the average value of the cluster members and the cluster center. The new cluster center is used to perform subsequent iterations, if the results obtained have not converged. The iteration process will stop if it has met the maximum iteration entered by the user or the results achieved are already converging (the new cluster center is the same as the old cluster center). Cluster center determination algorithm:

- 1. Search for the number of members of each cluster
- 2. Calculate new center with formula

From the rules of the association, the author designed a system by calculating the data that has been obtained. The data included in the calculation, the author uses only some transaction data as a sample of the system design. The steps that will be taken after the data is obtained are the segmentation or grouping of sales data of goods, namely "**best-selling goods**" and "**not in demand**" sold.

1st iteration process

- 1. Determination of the Initial Center of the Cluster
- In this process, the initial center of the cluster or centroid can be determined randomly, in this study it will be assumed as follows:
 - a. Center Cluster 1: (1570; 670)
 - b. Cluster Center 2: (1580 ; 1020)
- 2. Cluster Center Distance Calculation

To measure the distance between the data and the center of the cluster, Euclidean distance is used. From the 30 data sampled in the example, a random initial center of the cluster was selected, namely CI (1570; 670) and C2(1580; 1020), then calculated the distance from the data sample with M(a,b), meaning that a is the value of the Stock, while b is the amount sold. So that the form of data is obtained as follows:

M1	(1590	;	693)	M11	(1558	;	677)	M21	(1540	;	770)
M2	(1582	;	1266)	M12	(1552	;	854)	M22	(1566	;	696)
M3	(1612	;	999)	M13	(1540	;	550)	M23	(1624	;	870)
M4	(1566	;	1334)	M14	(1606	;	888)	M24	(1624	;	696)
M5	(1590	;	978)	M15	(1558	;	1016)	M25	(1684	;	783)
M6	(1578	;	1243)	M16	(1544	;	869)	M26	(1606	;	888)
M7	(1540	;	660)	M17	(1558	;	339)	M27	(1556	;	707)
M8	(1562	;	750)	M18	(1550	;	1142)	M28	(1606	;	786)
M9	(1540	;	770)	M19	(1540	;	660)	M29	(1594	;	700)
M10	(1566	;	696)	M20	(1562	;	750)	M30	(1646	;	393)

Next look for the distance of all data to the first central point (1570; 670)

$$D11 = \sqrt{(1590 - 1570)^2 + (693 - 670)^2} \qquad \dots \qquad D130 = \sqrt{(1646 - 1570)^2 + (393 - 670)^2} = 287,237$$

	Next look for the distance of all data to the second center point (1580; 1020)									
D21	$=\sqrt{(1590-1580)^2+(693-1020)^2}$		D230	$=\sqrt{(1646-1580)^2+(393-1020)^2}$						
	= 327,153			= 630,464						

From the search results of the distance value of all data to the center point, we will do a comparison between C1 and C2, where C1 is the unsold item, while C2 is the best-selling item.

	C1	C2		C1	C2	_	C1	C2
M1	30,480	327,153	M11	13,892	343,705	M21	104,403	253,180
M2	596,121	246,008	M12	184,878	168,345	M22	26,306	324,302





Sinkron : Jurnal dan Penelitian Teknik Informatika Volume 7, Issue 1, January 2023 DOI: https://doi.org/10.33395/sinkron.v8i1.12007

	C1	C2		C1	C2		C1	C2
M3	331,670	38,275	M13	123,693	471,699	M23	207,162	156,320
M4	664,012	314,312	M14	220,952	134,536	M24	59,933	326,974
M5	308,649	43,174	M15	346,208	22,361	M25	160,515	258,815
M6	573,056	223,009	M16	200,691	155,232	M26	220,952	134,536
M7	31,623	362,215	M17	331,217	681,355	M27	39,560	313,919
M8	80,399	270,599	M18	472,424	125,634	M28	121,458	235,440
M9	104,403	253,180	M19	31,623	362,215	M29	38,419	320,306
M10	26,306	324,302	M20	80,399	270,599	M30	287,237	630,464

From the table, it is found that: {M1; M7; M8; M9; M10; M11; M13; M17; M19; M20; M21; M22; M24; M25; M27; M28; M29; M 30} Member C1 {M2; M3; M4; M5; M6; M12; M14; M15; M16; M18; M23; M26} } Member C2.

Iteration Process 2

1. Determination of the New Center Point of the Cluster

In this process, the way that is done is to calculate the average value of the data in the same centroid where the cluster members are obtained in the iteration process 1:

a. New Center Cluster 1:

C1 = 1576.22

C2 = 670,889

b. New Center Cluster 2:

C1 = 1589,668

C2 = 1028,917

= 336,046

2. Cluster Center Distance Calculation

To measure the distance between the data and the center of the new cluster, Euclidean distance is used. From the 30 data sampled, a new cluster initial center was obtained from the previous step, namely CI (1576.22; 670.889) and C2(1589,668; 1028.917), then calculated the distance from the existing data. D130

D11

 $\sqrt{(1590 - 1576,220)^2 + (693 - 670,889)^2}$ = 26,052

 $=\sqrt{(1646 - 1576,220)^2 + (393 - 670,889)^2}$ = 28 6,516

Next look for the distance of all data to the second center point (1589,668; 1028,917) D230

D21

_	•••••	D_{250}	—
$\sqrt{(1590 - 1589,668)^2 + (693 - 1028,917)^2}$			$\sqrt{(1646 - 1589,668)^2 + (393 - 1028,917)^2}$
= 336,046			= 639,264

From the search results of the distance value of all data to the center point, we will do a comparison between C1 and C2, where C1 is the unsold item, while C2 is the best-selling item. Table 2 Remetian Desalt 2

	C1	C2			C1	C2			C1	C2			
M1	26,052	336,046		M11	19,220	352,646		M21	105,523	262,091			
M2	595,139	237,087		M12	184,706	177,250		M22	27,112	333,240			
M3	330,056	43,322		M13	126,199	480,640		M23	204,763	164,719			
M4	663,190	305,436		M14	219,144	143,176		M24	53,975	335,725			
M5	307,420	51,765		M15	345,592	26,089		M25	155,515	266,745			
M6	572,114	214,100		M16	200,714	164,066		M26	219,144	143,176			
M7	37,824	371,151		M17	332,389	690,289		M27	41,388	322,860			
M8	80,379	279,541		M18	471,840	117,168		M28	118,900	244,234			





M9	105,523	262,091	M19	37,824	371,151	M29	34,110	329,187
M10	27,112	333,240	M20	80,379	279,541	M30	286,516	639,264

From the table, it is found that: $\{M1; M7; M8; M9; M10; M11; M13; M17; M19; M20; M21; M22; M24; M25; M27; M28; M29; M30\}$ Member C1 $\{M2; M3; M4; M5; M6; M12; M14; M15; M16; M18; M23; M26\}$ Member C2. Because in iteration 1 and iteration 2 the cluster positions do not change or are the same, the iteration process is stopped. Thus the conclusion was reached that:

a. Selling Goods { XXX Cheek Blossom Blush On ; XXX Complete Makeup 058 ; XXX Deep Black Mascara ; XXX Eau De Perfume Woman ; XXX Exclusive Box Package Jelly Tint ; XXX Face Serum Anti Acne Set ; XXX Face Serum Dark Spot Treatment Set ; XXX Face Serum Exfoliating Treatment Set ; XXX Face Serum Skin Barrier Treatment Set ; XXX Icare Hand sanitizer gel 100ml ; XXX Jelly Tint ; XXX Lip Crayon Satin.

b. Unsold Items: XXX Cheek and Lip Tint ; XXX Eyeliner Pen ; XXX Eyeshadow Palette 7007 ; XXX Eyeshadow Palette 7669B ; XXX Eyeshadow Palette 7669D ; XXX Eyeshadow Palette 7672; XXX Face Serum Anti Aging Set ; XXX Icare Antiseptic Wipes 10 Sheets ; XXX Icare Hand Sanitizer Gel 30ml ; XXX Icare Hand Sanitizer Gel 5 Liters ; XXX Icare Handwash 5 Liters ; XXX Instant Bright Body Lotion; XXX Korean Dewy Ombre Lips (Honest Lip Velvet + Lip Tint); XXX Korean Matte Ombre Lips (Hazelnude Lip Cream + LipTint); XXX Lip Velvet ; XXX Lipstick Intense Matte ; XXX Make Up Kit Complete Beauty care 502 ; XXX Makeup Palette Cappadocia.

	🚱 Cluster Model (Clustering) 🕅	📑 Repositories 💥 🚰 📾 🔯					
	🔀 Result Overview 🛛	👰 🤞 🗸 🏐 🏐 👘 🔶					
Data	View 🔘 Meta Data View 🔵 Plot View 🔘 Advanced	Charts 🔵 An	notations		🗶 🖪 d	+ 🍕	E Samples (none)
Exampl	eSet (30 examples, 2 special attributes, 2 regular attribu	utes) View	Filter (30 / 30): all		Ŧ	
Row No	Nama Produk	cluster	Stok	Teriual			E Glassi
1	Implora Cheek and Lip Tint	cluster 0	1590	693			- 🕥 data (dasri)
2	Implora Cheek Blossom Blush On	cluster 1	1582	1266			Difference (dash)
3	Implora Complete Makeup 058 [Exp Juni 2022]	cluster_1	1612	999			bary (dasri - v1, 12/15/22 6:35 PM - 1 kB)
4	Implora Deep Black Mascara	cluster_1	1566	1334			
5	Implora Eau De Parfum Woman	cluster_1	1590	978			
6	Implora Exclusive Box Package Jelly Tint	cluster_1	1578	1243			
7	Implora Eyeliner Pen	cluster_0	1540	660			
8	Implora Eyeshadow Palette 7007	cluster_0	1562	750			
9	Implora Eyeshadow Palette 7669B	cluster_0	1540	770			
10	Implora Eyeshadow Palette 7669D	cluster_0	1566	696			
11	Implora Eyeshadow Palette 7672	cluster_0	1558	677			
12	Implora Face Serum Anti Acne Set	cluster_1	1552	854			
13	Implora Face Serum Anti Aging Set	cluster_0	1540	550			
14	Implora Face Serum Dark Spot Treatment Set	cluster_1	1606	888			
15	Implora Face Serum Exfoliating Treatment Set	cluster_1	1558	1016			
16	Implora Face Serum Skin Barrier Treatment Set	cluster_1	1544	869			
17	Implora Icare Antiseptic Wipes 10 Sheets	cluster_0	1558	339			
18	Implora Icare Hand sanitizer gel 100ml	cluster_1	1550	1142			
19	Implora Icare Hand Sanitizer Gel 30ml	cluster_0	1540	660			
20	Implora Icare Hand Sanitizer Gel 5 Liter	cluster_0	1562	750			
21	Implora Icare Handwash 5 Liter	cluster_0	1540	770			
22	Implora Instant Bright Body Lotion	cluster_0	1566	696			

Fig. 5 K-Means Results With Rapid Miner

DISCUSSIONS

From the search results using the K-Means algorithm through two iterations get the same grouping results so that the process is stopped and gets the results according to clustering. After being tested using the rapid miner application, it gets the same results. However, the thing that must be considered in the k-means algorithm is that it must be thorough and appropriate when determining the starting center point because if you incorrectly determine the starting point, it will cause improper cluster results, so you have to recalculate from the beginning.

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



The problem with this algorithm is that before the algorithm is run, K points are initialized randomly so that the grouping of data obtained can be different. However, if the values obtained randomly for the initializer are not good, the groupings obtained are not optimal. Suggestions for maximum results, further research can use other methods as comparisons and can use more criteria and data samples. Because the more data that is used as training, the higher the accuracy will be.

CONCLUSION

Using the K-Means method to solve the problems that exist in the Hayati Store Cosmetics Store, namely the sale of beauty and skincare carried out, which is still done manually so that it can cause it to not match the stock in the storage warehouse with changing market demands. This problem can be given a solution in the form of using mining data using the K-Means method, and it is evident from the calculation results that 30 sample data can be grouped appropriately. There are 18 data categorized as not selling well and 12 data categorized as best-selling products. The results of manual calculations using the K-Means algorithm method and tested with the rapid miner application show 100% the same grouping, so this method is very suitable as a solution to the problem at the Cosmetics Store.

REFERENCES

- Adha, L. H., Asyhadie, Z., & Kusuma, R. (2020). DIGITALIZATION OF INDUSTRY AND ITS EFFECT ON EMPLOYMENT AND LABOR RELATIONS IN INDONESIA. *Journal of Legal Compilation*, 5(2).
- Al-hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques : A comprehensive review from 2009 to 2019. *Computer Science Review Journal*, 40, 1–23. https://doi.org/10.1016/j.cosrev.2021.100402
- Aldo, D., Lelisa Army, W., Lestari, W. J., Saputra, A. H., & Munir, Z. (2022). Development of Integrated Information System for Batam Tourism Industry Implements Website-Based User Centered Design. JOURNAL OF INFORMATICS MEDIA BUDIDARMA, 6(2). https://doi.org/10.30865/mib.v6i2.3849
- Amalia, N., Rachman, O., & Surahman, R. (2020). Artificial Intelligence Based Agricultural Information System (E-Tandur). *JAMIKA*, 10(April), 1–10. https://doi.org/10.34010/jamika.v10i1
- Arifah, D. M. Al, Ramadlana, A. N., Oktawandira, D., Pradana, M. A. Y., Syaifudin, A. I., & Aldo, D. (2022). INTERACTIVE MULTIMEDIA-BASED TOURIST ATTRACTION INFORMATION APPLICATION IN PURWOKERTO. JURSIMA (Journal of Information Systems and Management), 10(1).
- Chen, J. B., Hu, W., & Li, K. Y. and G. N. (2021). A Log Analysis Technology Based on FP-growth Improved Algorithm. *Nternational Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*, 219–223. https://doi.org/0.1109/CAIBDA53561.2021.00053
- Cong, Y. (2022). Research on Data Association Rules Mining Method based on improved a priori algorithm. 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 373–376. https://doi.org/10.1109/ICBASE51474.2020.00085
- Ependi, S., & Akbar, M. (2021). IMPLEMENTATION OF DATA MINING IN PRODUCT SALES USING A PRIORI ALGORITHM. *Bina Darma Conference on Computer Science*, 220–225.
- Gong, J. (2021). In-depth Data Mining Method of Network Shared Resources Based on K-Means CLustering. *ICMTMA*, 13, 694–698. https://doi.org/10.1109/ICMTMA52658.2021.00160
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Data Mining implementation to determine the sales level of Telkomsel data packages using the K-means clustering method. *Scientific Journal of Technology* and Engineering, 25(1), 76–88. https://doi.org/10.35760/tr.2020.v25i1.2677
- Harahap, B. (2019). Application of the K-Means algorithm to determine building materials is in demand (case study on UD. YD Indarung Building Shop). 394–403.
- Hendri, & Oscar, D. (2021). Application of C4.5 Algorithm in Measuring Visitor Satisfaction with Facilities in Jakarta Wildlife Park. *Infortech*, *3*(1).
- Juansyah. (2018). E-MONITORING OF BASIC FOOD PRICES AT THE TRADE AND INDUSTRY OFFICE OF MUSI BANYUASIN REGENCY. *Journal of Informanika*, 4(2).
- Karsito, & Sari, W. M. (2018). PREDICTION OF POTENTIAL SALES OF DELIFRANCE PRODUCTS USING THE NAIVE BAYES METHOD AT PT. SUSTAINABLE FOOD. *SIGMA Journal of Technology*, 9(September), 67–78.
- Narastri, M. (2020). FINANCIAL TECHNOLOGY (FINTECH) IN INDONESIA IS VIEWED FROM THE ISLAMIC PERSPECTIVE OF Maulidah. *Indonesian Interdisciplinary Journal of Sharia Economics* (*IIJSE*), 2(2), 155–170.
- Noviyanto. (2020). Application of Data Mining in Grouping the Number of Deaths of COVID-19 Patients by Country on the Asian Continent. *Paradigm*, 22(2).
- Pranata, B. S., & Utomo, D. P. (2020). Application of FP-Growth Algorithm Data Mining for Spare Parts Inventory in Motor Repair Shops (Case Study of Sinar Service Workshop). *Bulletin of Information*

*<u>name of corresponding author</u>



Technology (BIT), 1(2), 1–8.

- Rachmatullah, N., & Purwani, F. (2022). Analysis of the Importance of Digitalization & Information Technology Infrastructure in Government Institutions: E-Government. *FASILKOM*, *12*(1), 14–19.
- Renhoran, B. S., Nurhandayani, N., & Septiana, L. (2018). APPLICATION OF THE C4.5 ALGORITHM TO DETERMINE STOCK DATA AND MATERIAL DEMAND TARGETS THAT ARE MOST NEEDED BY LOGISTICS WAREHOUSES AT PT PLN (Persero) KEBON JERUK AREA. 12(2), 13–20.
- Risal, A. A. N., Yusuf, N. I., Kaswar, A. B., & Adiba, F. (2021). Application of Data Mining in Classifying Covid-19 Case Rates in South Sulawesi Using the Naive Bayes Algorithm. *INDONESIAN JOURNAL OF* FUNDAMENTAL SCIENCES, 7(1), 18–28.
- Saputra, R., Yahya, S., & Malang, S. (2021). Application of Virtual Reality Technology in Residential Property. *SISFOTEK*, 5(1), 307–315.
- Sitoto, Y., Allo, T., Sofica, V., Hasan, N., & Septiani, M. (2022). The use of the Naïve Bayes method in classifying unemployment in Bojong Kulur Village. *Ferris wheel of informatics*, 10(1), 30–35.
- Suhartini, & Yuliani, R. (2021). Application of Data Mining to Cluster Data on Poor People Using K-Means Algorithm in Bagik Endep Sukamulia Timur Hamlet. *Infotech : Journal of Informatics and Technology*, 4(1), 39–50.
- Syahril, M., Erwansyah, K., & Yetri, M. (2020). Application of data mining to determine the sales pattern of school equipment on the wigglo brand using a priori algorithm. *J-SISKO TECH*, *3*(1), 118–136.
- Yusuf, B., Qalbi, M., Dwitawati, I., & Ellyadi, M. (2020). IMPLEMENTATION OF THE NAÏVE BAYES AND RANDOM FOREST ALGORITHMS IN PREDICTING THE ACADEMIC PERFORMANCE OF AR-RANIRY STATE ISLAMIC UNIVERSITY STUDENTS. *Cyberspace: Journal of Information Technology Education*, 4(1), 50–58.

