

DATA MINING ANALYSIS TO DETERMINE THE INVENTORY OF FOOD SALES DURING THE PANDEMIC WITH THE K-MEANS CLUSTERING METHOD

David C. Hutajulu, Yulianus Zega, Unggul H. Siregar, Jepri Banjarnahor*

Program Studi Sistem Informasi, Fakultas Teknologi Dan Ilmu Komputer, Universitas Prima Indonesia

E-mail : Jepribanjarnahor@gmail.com

Abstract

Article Info

Received, 01 Juni 2022

Revised 24 Juni 2022

Accepted 26 Juni 2022

Currently, many businesses are no longer operating because people no longer shop directly for direct sales, but now many people shop online. This is one of the problems faced by sales that do not market their goods online, so there has been a buildup of inventory for the last 2 years during this pandemic. It is difficult for the company to know which types of goods will be sold a lot and a little every day because the system is still in use only for input. So with that there needs to be a system that solves the problem. In this study, the K-Means method will be used to determine the most sales. This system has carried out the clustering process with the K-Means method by utilizing the data of past items. The results of this research in cluster 1 with a value of Cluster 1 -> 358.64 245.9775 while in Cluster 2, Cluster 2 -> 45.39 22.824285714286.

Keywords: Data mining, Sales, Pandemic Period, K-Means clustering.

1. Introduction

With the rapid development of computerized science and technology now it has become an important thing and is needed, especially in the development of organizations and management in every company or business both government and private. The amount of competition in the business world, especially in the printing industry, requires developers to find a strategy that can increase product orders at printing companies. With the printing activity every day, the longer the data will be more and more. Therefore every company must have a good data processing system so that the data generated from these transactions can be useful to be made into a monthly or annual report. The data not only functions as an archive for the company, the data can be used and processed into useful information for increasing product orders.

Printing (printing) is a technology or art that produces copies of an image very quickly, such as words or drawings on paper, cloth, and other surfaces. The development of science and technology is progressing faster, so that at this time the printing industry has become more complete and modern. Printing companies in Indonesia are growing rapidly, both on a large, medium, and small scale. This development also took place in the city of Medan, where there were various printing companies in quite a large number. The development of the number of printing companies has resulted in increasingly high business competition in the printing sector.

2. Method

Sales of a company's products are influenced by several factors, such as changes in the number of consumer requests, changes in the number of sales and the influence and market trends. One of the important parameters in sales planning is the number of products to be produced and sold for the next period. In predicting or forecasting product sales or consumer demand, you can use several forecasting

methods such as the research conducted by Karina and Mariza in 2019, where in this study the observed data variables were sales value, consumer name, year and month which then performed a forecasting comparison. sales tested the application of the Naive and Simple Moving Average methods using the MASE (Mean Absolute Square Error) accuracy value [8,9,10].

This study uses a classification technique with K-Means to group categorical data to produce a more stable cluster. K-means Clustering Algorithm is a data analysis method or data mining method that performs the modeling process without supervision (unsupervised) and is one method that performs data grouping with a partition system. The K-Means method tries to group the existing data into several groups, where the data in one group has the same characteristics as each other and has different characteristics from the data in other groups [6, 11]. The stages of doing clustering or grouping with the K-Means method are as follows:

1. Select the number of clusters k.
2. Initialization of k cluster centers can be done in many ways. But the most often done is in a random way. Cluster centers are dubbed initial values with random numbers.
3. Place all data/objects to cluster closest. The proximity of two objects is determined based on the distance between the two objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data and the center of the cluster. In this stage, it is necessary to calculate the distance of each data to each cluster center. The most distance between one data and a certain cluster will determine which data belongs to which cluster. To count the distance of all data to each cluster center point can use the Euclidean distance theory which is formulated as follows:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2} \quad (1)$$

dimana:

$D(i, j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

This research is so that the work process is more structured in the work, the author has made a block diagram for this researcher. Which can be seen in the image below:

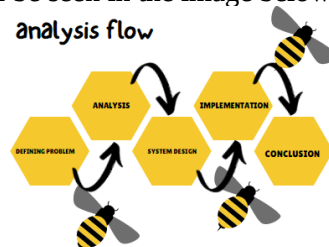


Figure 1. block diagram

3. Results and Discussion

At this stage, this research will present the results of the discussion that has been processed from the K-Kmens method. The data used in this study is data that has passed from the period 2021 – 2022.

DATA OBJEK			DATA CLUSTER		
Data Excel	Browse...	No file selected.	Import	Delete All	
Nama Objek	Data		Cluster	Centroid Awal	
Energen Strawviberry	5.08,0	Tambahkan	Input		Tambahkan
Energen Banana	8.08,0	Hapus Data	Cluster 1	250,130	Hapus Data
Energen Jahe	57.45,40.27	Hapus Data	Cluster 2	125,90	Hapus Data
Energen Kacang Hijau	308.16,166.51	Hapus Data			
Energen Coklat	290.81,193.39	Hapus Data			
Kapiko White Coffee	56.93,0	Hapus Data			
Kapiko Brown Coffee	21.60,17.31	Hapus Data			
Torabika Jahe Susu	12.10,2.56	Hapus Data			
Energen Vanilla	153.49,99.63	Hapus Data			
Torabika Cappuccino	547.78,443.33	Hapus Data			
Torabika Moka	287.81,180.68	Hapus Data			

Figure 1. Item data display

In the table above where the image on the left shows data on goods, on the right shows cluster data. The image above is taken from the system that has been created.

Table. 1 Literasi 1

ITERASI 1				
Objek	Data 1	Data 2	Cluster 1	Cluster 2
Objek 1	8.08	0	Null	OK
Objek 2	8.08	0	Null	OK
Objek 3	57.45	40.27	Null	OK
Objek 4	308.16	166.51	OK	Null
Objek 5	290.81	193.39	OK	Null
Objek 6	56.93	0	Null	OK
Objek 7	21.60	17.31	Null	OK
Objek 8	12.10	2.56	Null	OK
Objek 9	153.49	99.63	Null	OK
Objek 10	547.78	443.33	OK	Null
Objek 11	287.81	180.68	OK	Null

Table 3. Literasi 2

ITERASI 2				
Objek	Data 1	Data 2	Cluster 1	Cluster 2
Objek 1	8.08	0	Null	OK
Objek 2	8.08	0	Null	OK
Objek 3	57.45	40.27	Null	OK
Objek 4	308.16	166.51	OK	Null
Objek 5	290.81	193.39	OK	Null
Objek 6	56.93	0	Null	OK
Objek 7	21.60	17.31	Null	OK
Objek 8	12.10	2.56	Null	OK
Objek 9	153.49	99.63	Null	OK
Objek 10	547.78	443.33	OK	Null
Objek 11	287.81	180.68	OK	Null

Cluster 1 -> 358.64	245.9775	Cluster 2 -> 45.39	22.824285714286
---------------------	----------	--------------------	-----------------

In Table 1 displays the results of the cluster 1 process, Table 2 displays the second literacy of the cluster results using the k-Means method, the tables above are all taken from the system that has been created. Calculation of the center of the cluster point to the data to 1

Cluster 1

$$\sqrt{(119 - 119)^2 + (403 - 400)^2}$$

$$= 3$$

Cluster 2

$$\sqrt{(119 - 119)^2 + (403 - 200)^2}$$

$$= 203$$

Cluster 3

$$\sqrt{(119 - 119)^2 + (403 - 300)^2}$$

$$= 103$$

Calculation of the center of the cluster point to the data to 2

Cluster 1

$$\sqrt{(119 - 119)^2 + (383 - 200)^2}$$

$$= 17$$

Cluster 2

$$\sqrt{(119 - 119)^2 + (383 - 200)^2}$$

$$= 183$$

Cluster 3

$$\sqrt{(119 - 119)^2 + (383 - 200)^2}$$

$$= 83$$

And so on, the calculation is carried out until the data is 20

This system has carried out the clustering process with the K-Means method by utilizing the data of past items. The results of this research in cluster 1 with a value of Cluster 1 -> 358.64 245.9775 while in Cluster 2, Cluster 2 -> 45.39 22.824285714286.

C1

Total SKS

$$= (119+119+119+119+119+119+119+119+119+119+119+119+119) / 13 = \mathbf{119}$$

Total Value

$$= (403+383+409+428+402+416+363+404+395+394+419+433+399) / 13 = \mathbf{403.69}$$

For cluster 2, there are 3 members joined to it so that:

C3

Total SKS

$$= (119+119+97+119) / 4 = \mathbf{113.50}$$

Total Value

$$= (333+351+268+348) / 4 = \mathbf{325}$$

4. Conclusions

Based on the results of research and discussion of the implementation of the k-means clustering analysis algorithm method to determine barriers to online learning for students, the following conclusions can be drawn: This system has carried out the clustering process with the K-Means method by utilizing the data of past items. The results of this research in cluster 1 with a value of Cluster 1 -> 358.64 245.9775 while in Cluster 2, Cluster 2 -> 45.39 22.824285714286.

Reference

- [1] Suyanto, Data Mining for Classification and Clustering of Data. Bandung: Bandung Informatics, 2017.
- [2] Dalam Menentukan Pegawai Yang Layak Mengikuti Aseessment Center. In Determining Eligible Employees to Follow the Assessment Center. Utilizing the K-Means Algorithm in Determining Eligible Employees to Participate in the Assessment Center for the Sdp Clustering Program, 3(1), 87–93.
- [3] Farokhah, L., & Aditya, R. (2017). Implementation of K-Means Clustering for Recommendation on Final Project Themes on Asian Stmik Malang. Journal of Information Technology and Management, 3(2), 142–148. <https://doi.org/10.26905/jtmi.v3i2.1329>.
- [4] Sofyan, K. D., & Meutia, S. (2019). Forecasting Product Demand At PT. Lhokseumawe Healthy Joint Business Development. SNTI 2019, 1–9.
- [5] Lusiana, A., & Yuliarty, P. (2020). Application of Forecasting Method on Roof Demand at PT. X. Innovative Industry: Journal of Industrial Engineering, 10(1), 11–20.
- [6] Winda A., Jaka P., & Oktaviyani, (2017) Application of Apriori Algorithm for Transactions Drug Sales at Azka Pharmacy, National Seminar on Mathematics and Its Applications, pp. 436-442.21 October.
- [7] Djamaludin, I., & Nursikuwagus, A. (2017). Analysis of Consumer Purchase Patterns in Sales Transactions Using Apriori Algorithm. Symmetrical : Journal of Mechanical Engineering, Electrical and Computer Science, 8(2), 671. <https://doi.org/10.24176/simet.v8i2.1566>.
- [8] Afif, M., & Wulandari, D. A. N. (2019). Application of Apriori Algorithm for Procurement of Medicines and Medical Devices at Mayapada Hospital's Radiology Department. Journal of Techno Nusa Mandiri, 16(2), 99–104. <https://doi.org/10.33480/techno.v16i2.737>.
- [9] Rezkiani. (2016). Implementation of Data Mining with Apriori Algorithm to Determine the Shoes Brand Interested in Graduate Students Class 14.1a.01 Stmik Nusa Mandiri Jakarta. KNIT-2 Nusa Mandiri, 2(1), 49-INF.56.
- [10] Auliasari K., Kertaningtyas M., & Kriswantono M.(2019), Application of Forecasting Methods to Identify Potential Consumer Demand Informatics Journal Vol. 4 No. 3
- [11] Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M.R. (2018). Utilizing the K-Means Algorithm