Jurnal Teknik Informatika C.I.T Medicom, 15 (4) (2023) 199-205

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Clustering of goods in the lustrorezea online shop

Erlinda[1], and Febri Haswan[2]
[1,2] Departemen Informatika, Universitas Islam Kuantan Singingi, Riau, Indonesia

Article Info

Article history

Received : Jul 25, 2023
Revised : Aug 12, 2023
Accepted : Sep 16, 2023

Keywords:

Clustering;
Dataset;
K-Means;
Online Shop;
Rapid Miner.

Abstract

Lustrorezea is an online shop that was founded in 2018 and has 150 regular customers. This Lustrorezea online shop sells goods in the form of women's, men's, children's clothing and other items, not all items sell well, there are also items that very popular, and not very popular, therefore we need technology in the form of data mining, data mining can help business people make decisions quickly and precisely. In this research, data grouping uses the Rapid Miner application. By using the Rapid Miner application you can group data more quickly and accurately. The results of this research are the formation of 3 clusters by determining the types of goods that are very popular, best-selling and less-selling. With this data, the owner of Lustrorezea can analyze stock needs so that sales can increase further and minimize losses.

Corresponding Author:

Febri Haswan,
Department Informatika
Universitas Islam Kuantan Singingi
Jl. Gatot Subroto KM. 7 Kebun Nenas, Teluk Kuantan, Riau, Indonesia, 29562
Email: febri.haswan88@gmail.com

1. Introduction

Online shopping or what is often called online shopping itself is a process of purchasing goods or services from those who sell goods or services via the internet where the seller and buyer never meet or make physical contact, where the goods being bought and sold are offered through displays with pictures. which is on a website or virtual shop. After that, the buyer can choose the desired item and then make payment to the seller via the relevant bank account[1]. After the payment process is received, the seller's obligation is to send the goods ordered by the buyer to the destination address[2][3].

In the current era of information technology, online shops are increasingly in demand by the public, where all activities are carried out online, from working, studying, to shopping[2]. Lustrorezea is an online shop that was founded in 2018 and has 150 regular customers. This Lustrorezea online shop sells goods in the form of women's, men's, children's clothing and other items, not all items sell well, there are also items that very popular, and less popular, therefore a technology in the form of data mining is needed. Data mining is a process of extracting data from very important information[4]. Data Mining is also a process for exploring patterns from data[5][6]. These patterns are obtained from various types of databases such as relational databases, data warehouses, transaction data, and object-oriented data[7][8][9]. The use of data mining can help business people make decisions quickly and precisely[10][11].

Grouping (Clustering) is a data mining method that functions to group several data to produce similar data to form new data[12][13]. The grouping of goods in the Lustrorezea online shop will later apply the K-Means algorithm. The K-Means algorithm is an algorithm in data mining that can be used

to cluster data[14][15][16]. This K-Means algorithm will make it easier to group items that are very popular, best selling, and less popular[17][18]. This will of course make it easier for the owner to make decisions to increase sales each month.

2. Research Methods

2.1. Research Framework

Primary data sources were obtained directly from first hand using observation methods and interview methods with Lustrorezea owners, and secondary data were obtained through pre-existing information from existing sources [19][20]. The following is an overview of the framework for the research carried out, for more details, see the image below.

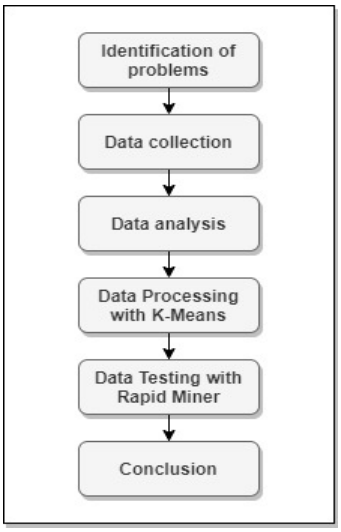

Figure 1. Research Framework

The following is an explanation of the Research Framework:

a. Identify the Problem
   Identification of the problem in this research is to find out items that are very popular, best selling and less popular as a reference for stock inventory at the Lustrorezea Online Shop.
b. Data Collection
   Data collection was carried out to determine the amount of initial stock of goods, goods sold and remaining stock. The data used as a sample was data from May to July 2023.
c. Data Analysis
   Data analysis is carried out by analyzing stock data.
d. Data processing with K-Means
   Data processing using the K-means clustering method is carried out manually to group items that are very popular, best selling and less selling.
e. Data Testing with Rapid Miner
   Test the data with Rapid Miner to find out how many clusters there are in Lustrorezea Online Shop sales.
f. Conclusion
   Describe the results of data processing using the K-means clustering method and testing using Rapid Miner.

2.2. Metode K-Means

The steps for clustering using the K-Means method are as follows[21][22][23]:

a. Determine the k value as the number of clusters to be formed.
b. Determine the initial center point of each cluster.

c. Calculate the distance of each input data to each centroid using the Euclidean distance formula until you find the closest distance of each data to the centroid.
The following is the Euclidean Distance equation[24][25]:

$$D(x,y) = \sqrt{(Y_1 - X_1)^2 + (Y_2 - X_2)^2} \quad (1)$$

Information;
D = Distance
x = Data
y = Centroid

d. Classify data based on its proximity to the centroid.
e. Calculate the cluster center again with the current cluster members. The cluster center is the average value of all data objects in a particular cluster.
f. Calculate the cluster center again with the current cluster members. The cluster center is the average value of all data objects in a particular cluster.
g. Calculate each object again using the new cluster center. If the cluster center does not change again then the clustering process is complete. Or, return to step number 3 until the cluster center does not change anymore.

3. Results and Discussion

3.1. Data source

The data source or sample used is goods sales data from May to July 2023, where the data taken consists of the name of the goods, goods sold, and the total stock of goods available. This data will later be used as a dataset for forming sales clusters. on Lustrorezea, for more details, see the table below.

Table 1.
Preliminary data

No.	Name of goods	First stock	Items Sold	Total Stock
1	Gamis	45	42	3
2	Dress	15	12	3
3	Blouse	24	20	4
4	Sweater	18	12	6
5	Outer	13	4	9
6	Tunik	26	21	5
7	Kemeja	24	19	5
8	One Set	31	29	2
9	Jilbab	27	21	6
10	Kemeja Pria	9	6	3
11	Koko	18	15	3
12	Kaos Dewasa	13	11	2
13	Celana Jeans	7	3	4
14	Setelan Kaos	20	19	1
15	Kaos Anak-Anak	9	7	2
16	Celana Joger	7	6	1
17	Celana Pendek	5	2	3
18	Celana Chinos	3	1	2
19	Sweater Anak	3	0	3
20	Baju Tidur Anak	8	6	2

3.2. Clustering Process

For data clustering in this research, the RapidMiner application was used. RapidMiner is used for modeling, evaluation, and implementation[15][26]. RapidMiner is designed to be easy to use and allows users to easily build and test various models, even without programming experience[27]. RapidMiner can group data more quickly and accurately[15], for more details, see the image below.

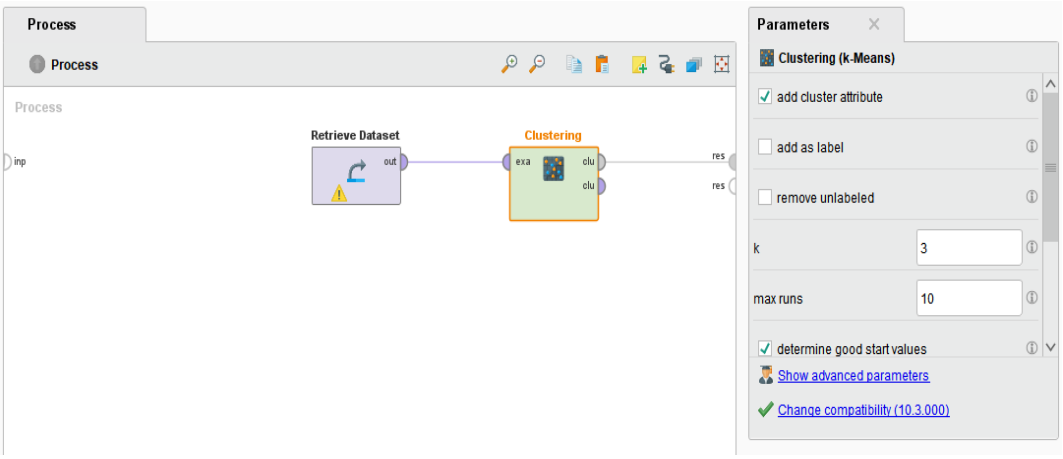

Figure 2. Clustering Process

In the picture above there is an initial dataset of sales of goods from May to July 2023. This dataset is processed using the RapidMiner application, where the data is then grouped using the K-Means algorithm with a k value of 3 and max runs 10, then the data will automatically form 3 cluster models. For more details, see the image below.

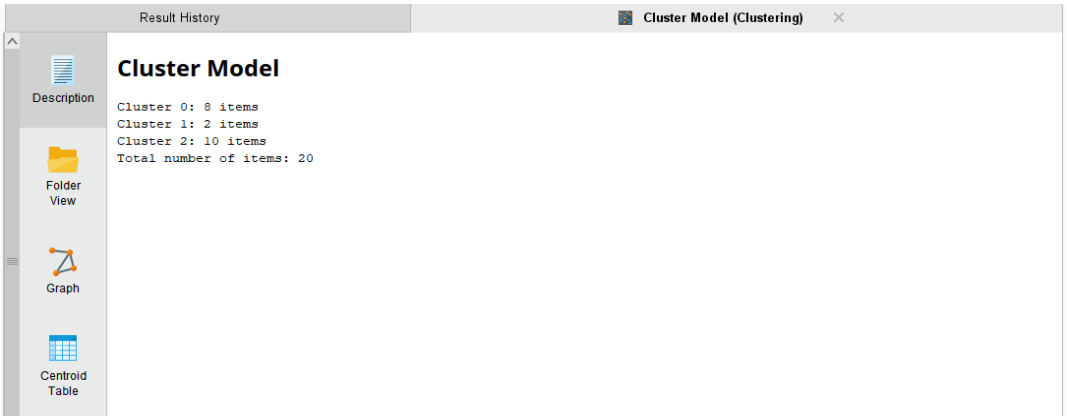

Figure 3. Cluster Model

From the 3 models above, you can describe what items are selling very well, selling well and not selling well. If you look at the picture below, you can clearly see which items are selling very well, selling well and not selling well, in cluster 1 there are 2 items which are said to be very well sold, in cluster 2 there are 8 items which are said to be selling well, and in cluster 3 there are 10 items which are said to be selling less well. , for more details can be seen in the image below.

Row No.	id	Nama Barang	cluster	Stok Awal	Stok Akhir	Total
1	1	Gamis	cluster_1	45	42	3
2	2	Dress	cluster_0	15	12	3
3	3	Blouse	cluster_0	24	20	4
4	4	Sweater	cluster_0	18	12	6
5	5	Outer	cluster_2	13	4	9
6	6	Tunik	cluster_0	26	21	5
7	7	Kemeja Wanita	cluster_0	24	19	5
8	8	One Set	cluster_1	31	29	2
9	9	Jilbab	cluster_0	27	21	6
10	10	Kemeja Pria	cluster_2	9	6	3
11	11	Koko	cluster_0	18	15	3
12	12	Kaos Dewasa	cluster_2	13	11	2
13	13	Celana Jeans	cluster_2	7	3	4
14	14	Setelan Kaos	cluster_0	20	19	1
15	15	Kaos Anak-A...	cluster_2	9	7	2
16	16	Celana Joger	cluster_2	7	6	1
17	17	Celana Pend...	cluster_2	5	2	3
18	18	Celana Chinos	cluster_2	3	1	2
19	19	Sweater Anak	cluster_2	3	0	3
20	20	Baju Tidur An...	cluster_2	8	6	2

Figure 4. Data Cluster

From the table above, the results obtained are 2 clusters for the best-selling items, namely gamis and One Set, 8 clusters for the best-selling types of items, namely Dresses, Blouses, Sweaters, Tunics, Shirts, Hijabs, Kokos, and T-shirt Suits, and 10 types goods declared to be less in demand are Outerwear, Men's Shirts, Adult T-shirts, Jeans, Children's T-shirts, Joggers, Shorts, Chinos, Children's Sweaters and Children's Sleepwear. With this data, the owner of Lustrorezea can carry out an analysis of stock needs. so that it can increase sales and minimize losses. For more details, see the image below.

Table 2.
Cluster Results

No.	Name of goods	First stock	Items Sold	Total Stock	Cluster	Information
1	Gamis	45	42	3	1	Very popular
2	Dress	15	12	3	0	Bestseller
3	Blouse	24	20	4	0	Bestseller
4	Sweater	18	12	6	0	Bestseller
5	Outer	13	4	9	2	Less popular
6	Tunik	26	21	5	0	Bestseller
7	Kemeja	24	19	5	0	Bestseller
8	One Set	31	29	2	1	Very popular
9	Jilbab	27	21	6	0	Bestseller
10	Kemeja Pria	9	6	3	2	Less popular
11	Koko	18	15	3	0	Bestseller
12	Kaos Dewasa	13	11	2	2	Less popular
13	Celana Jeans	7	3	4	2	Less popular
14	Setelan Kaos	20	19	1	0	Bestseller
15	Kaos Anak-Anak	9	7	2	2	Less popular
16	Celana Joger	7	6	1	2	Less popular
17	Celana Pendek	5	2	3	2	Less popular
18	Celana Chinos	3	1	2	2	Less popular

No.	Name of goods	First stock	Items Sold	Total Stock	Cluster	Information
19	Sweater Anak	3	0	3	2	Less popular
20	Baju Tidur Anak	8	6	2	2	Less popular

4. Conclusions

From the results of research that has been carried out to cluster goods in the Lustrorezea online shop, it can be concluded that by using the Rapid Miner application you can group data more quickly and accurately. In this research, 3 clusters were formed by determining the types of goods that are very popular, best selling and less popular. sold, with this data the Lustrorezea owner can analyze stock needs so that sales can increase further and minimize losses.

References

[1] J. Cha, "Exploring the internet as a unique shopping channel to sell both real and virtual items: A comparison of factors affecting purchase intention and consumer characteristics," *J. Electron. Commer. Res.*, vol. 12, no. 2, p. 115, 2011.

[2] F. Naufal, Y. H. Chrisnanto, and A. K. Ningsih, "Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 10–17, 2022.

[3] G. Antoniou and L. Batten, "E-commerce: protecting purchaser privacy to enforce trust," *Electron. Commer. Res.*, vol. 11, pp. 421–456, 2011.

[4] M. L. Gargano and B. G. Raggad, "Data Mining-a powerful information creating tool," *OCLC Syst. Serv. Int. Digit. Libr. Perspect.*, vol. 15, no. 2, pp. 81–90, 1999.

[5] M. Van Leeuwen, "Interactive data exploration using pattern mining," *Interact. Knowl. Discov. data Min. Biomed. informatics State-of-the-art Futur. challenges*, pp. 169–182, 2014.

[6] W. M. Hochachka *et al.*, "Data-mining discovery of pattern and process in ecological systems," *J. Wildl. Manage.*, vol. 71, no. 7, pp. 2427–2437, 2007.

[7] S. Dietrich and S. Urban, *Fundamentals of object databases: Object-oriented and object-relational design*. Morgan & Claypool Publishers, 2010.

[8] R. J. Petrasch and R. R. Petrasch, "Data integration and interoperability: Towards a model-driven and pattern-oriented approach," *Modelling*, vol. 3, no. 1, pp. 105–126, 2022.

[9] R. M. Al-Khatib, N. K. T. El-Omari, and M. A. Al-Betar, "Innovative cloud computing object-oriented model to unify heterogeneous data," *Int. J. Oper. Res.*, vol. 46, no. 3, pp. 289–322, 2023.

[10] I. H. Sarker, "Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective," *SN Comput. Sci.*, vol. 2, no. 5, p. 377, 2021.

[11] N. Normah, S. Nurajizah, and A. Salbinda, "Penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab Banten," *J. Khatulistiwa Inform.*, vol. 7, no. 2, pp. 158–163, 2021.

[12] M. J. H. Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, 2018.

[13] M. V. Sitorus, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Konsumsi Produk Kosmetik milik PT Cedefindo," *J. Ilm. MIKA AMIK Al Muslim*, vol. 5, no. 2, pp. 63–68, 2021.

[14] A. L. Hananto, P. Assiroj, B. Priyatna, A. Fauzi, A. Y. Rahman, and S. S. Hilabi, "Analysis of drug data mining with clustering technique using K-Means algorithm," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 12024.

[15] Sudirman, A. P. Windarto, and A. Wanto, "Data mining tools| rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2018, p. 12089.

[16] T. M. Ghazal, "Performances of K-means clustering algorithm with different distance metrics," *Intell. Autom. Soft Comput.*, vol. 30, no. 2, pp. 735–742, 2021.

[17] U. Hasanah and B. Humaidi, "Analysis of data mining using k-means clustering algorithm for product grouping," *Int. J. Informatics Inf. Syst.*, vol. 3, no. 1, pp. 12–22, 2020.

[18] R.-Q. Liu, Y.-C. Lee, and H.-L. Mu, "Customer classification and market basket analysis using K-means clustering and association rules: evidence from distribution big data of korean retailing company," *Knowl. Manag. Res.*, vol. 19, no. 4, pp. 59–76, 2018.

[19] M. DeJonckheere and L. M. Vaughn, "Semistructured interviewing in primary care research: a balance of relationship and rigour," *Fam. Med. community Heal.*, vol. 7, no. 2, 2019.

[20] M. Ciesielska, K. W. Boström, and M. Öhlander, "Observation methods," *Qual. Methodol. Organ. Stud. Vol. II Methods Possibilities*, pp. 33–52, 2018.

[21] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019.
[22] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access*, vol. 8, pp. 80716–80727, 2020.
[23] H. Ismkhan, "Ik-means–+: An iterative clustering algorithm based on an enhanced version of the k-means," *Pattern Recognit.*, vol. 79, pp. 402–413, 2018.
[24] M. Hossain and S. Abufardeh, "A New Method of Calculating Squared Euclidean Distance (SED) Using pTree Technology and Its Performance Analysis.," in *CATA*, 2019, pp. 45–54.
[25] T. Behrens, K. Schmidt, R. A. Viscarra Rossel, P. Gries, T. Scholten, and R. A. MacMillan, "Spatial modelling with Euclidean distance fields and machine learning," *Eur. J. Soil Sci.*, vol. 69, no. 5, pp. 757–770, 2018.
[26] G. F. Nama, H. Lukmanul, and J. Junaidi, "Implementation of K-Means Technique in Data Mining to Cluster Researchers Google Scholar Profile," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, 2019.
[27] V. Pynam, R. R. Spanadna, and K. Srikanth, "An extensive study of data analysis tools (rapid miner, weka, r tool, knime, orange)," *Int. J. Comput. Sci. Eng*, vol. 5, no. 9, pp. 4–11, 2018.