Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

CLUSTERING OF CLOTHING SALES DATA USING K-MEANS METHOD

Virandha Salbinda[1)], Rani Irma Handayani [2*)] , Normah[3)]
[1) 2) 3)]Universitas Nusa Mandiri, Jakarta, Indonesia
[1)]virandhasalbinda60@gmail.com, [2)]rani.rih@nusamandiri.ac.id,
[3)]normah.nor@nusamandiri.ac.id

Submitted : Jan 3, 2022 | **Accepted** : Jan 18, 2022 | **Published** : Apr 1, 2022

Abstract: In the era of globalization, the development of technological sophistication is growing rapidly which is an aspect that can be utilized to achieve convenience, especially in the flow of information. This technological sophistication by all accounts is increasingly spreading with the use of computers which are currently very popular in various areas of life. For example in the fields of education, entertainment, health, especially in the business sector. Top Store is a store that is engaged in selling clothes, however, of the various types of clothes that are sold, of course, not all of them are selling very well, and some are not selling well. Sales data, purchase of goods and unexpected expenses at Top Shop is not structured, so the data are only serves as an archive for the store and not be utilized for the development strategy of marketing. Therefore, it is necessary to apply Clustering of Clothing Sales Data in Top Stores with the K- Means Method . The K-means method can be applied to Top Stores to determine which clothes are selling very well, selling well and not selling well. The application of the K-Means method in Top Stores, namely by grouping clothing stock data. Then choose 3 clusters randomly as the initial centroid. After the data in each cluster does not change, it can be seen that the final result is that there are 21 best-selling articles, 17 articles that are selling well and 12 articles that are not selling well. Then applying the *K-means* method to Rapidminer is done by entering product stock data, namely initial stock, sold stock and final stock which will become a *database* on Ms. Excel, the data is then connected to the Rapidminer *Tools* , and will be processed and formed *K-means.*

Keywords: Best Selling, Data Mining, K-means, Rapidminer, Store

INTRODUCTION

In the business world, to stay ahead of the the competition especially in sales, expect business people to set an example that can set the pattem and promote within the company. There are several ways that do that is by comproving the quality of product, enhancing product and reduction in operating costs of companies with a way to use the analysis of the data of the company (Indriyani & Irfiani, 2019). Cluster is a collection of data objects that have similarities with one another in other groups (NOVIANTO, 2019).

The development of data grouping, or clustering, has a simple nature and is close to the human way of thinking whenever large amounts of data are presented to us, we usually tend to summarize the amount of data (Hasugian, 2018). The purpose of designing this program is not to replace the role of humans, but to substitute human knowledge into a system form so that it can be used by many people (Bastian et al., n.d.).

The data mining technique that will be used in this research is the K-Means method and the software used to support data processing is RapidMiner. This study aims to group the products sold at the online shop Ragam Jogja into several clusters to find out which products are most in demand by consumers so that the number of items in stock must be large, products are in demand by consumers for moderate stock quantities and products that are less desirable by consumers for stock quantities. a little (Handoko et al., 2020).

LITERATURE REVIEW

Data mining is a term used to describe the discovery of knowledge in databases (Aji, 2017). K-Means is one of the data grouping methods nonhierarchical (blockage) that can partition data into data form two or more groups. The purpose of grouping is to minimize theobjective function which is set in the clustering process, will generally seeks to minimize variationwithin a group and maximize

*Name of the corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

variation between groups (Gustientiedina et al., 2019). The purpose of grouping is to minimize the objective function which is set in the clustering process, will generally seeks to minimize variation within a group and maximize variation between groups (MURTI, 2017). Data mining is used to process big data in databases so as to produce new information that is useful for business strategy (Indriyani & Irfiani, 2019). The K-Means algorithm performs point-based clustering (centroid) by determining three parameters, namely the number of clusters, cluster initialization and system distance (Indriyani & Irfiani, 2019).

K-Means partitions data into groups so that data with the same characteristics are included in the same group and data with different characteristics are grouped into other groups (Mardalius, 2018). Data mining is an analysis of observing large amounts of data to find previously unknown relationships and new methods to summarize data so that it is easy to understand and useful for data owners (Rianti, 2017). The use of data mining techniques is expected to help speed up the decision-making process, enabling companies to manage the information contained in transaction data into new knowledge. The purpose of this study was to determine customer interest in hijab products which were determined based on transaction and sales data using the k-means method (Yulianti et al., 2019).

RapidMiner extracts patterns from large data sets by combining statistical, artificial intelligence and database methods (Rahmat C.T.I. et al., 2017). Objects will be grouped into one or more clusters so that objects in one cluster will have a high similarity between one another (Aulia, 2021). RapidMiner is one of the software of choice for extracting data with data mining methods. The results shown by RapidMiner can also be displayed visually with graphs (Issn, 2018).

METHODS

The stages in this research include research steps. The framework in this research is described as follows:

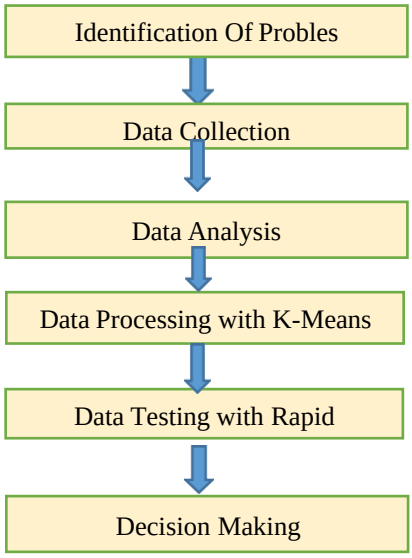

Figure 1: Research Framework

Research and Methods

1. Observation
   To make this observation, the author makes direct observations in the Sales section of the Top store, which is located at Jl. Ciledug Raya PD. Cipulir Jaya Market.
2. Interview
   The author obtains data and examines the truth of these data by conducting direct interviews by the owner of the Top shop, Mrs. Oktavianti.
3. Literature review
   Conduting library research by colleting data information obtained from libraries, journals, articles and the internet related to this research.

*Name of the corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT AND DISCUSSION

The data used in this study is clothing stock data from October 2020 to December 2020 at Top Stores. The data below is the data to be processed, in the form of initial stock, sold stock and ending stock:

Table 1 Top Store Stock Data for October 2020 – December 2020

No	Description	First stock	Stock Sold	Last stock
1	Outer Kimono	86	26	60
2	Knitted long tunic	47	40	7
13	Blue striped shirt	63	46	17
14	Gray striped shirt	44	19	25
25	Flower kimono	83	66	17
26	spotted kimono	54	34	20
37	Gray flannel shirt	53	33	20
38	Black flannel shirt	41	25	15
49	white jacket	48	27	21
50	jacket blue	42	21	22

The steps in clustering using the K-Means method are as follows:

1. Determine the value of k as the number of clusters to be formed.
2. Determine the starting point of each cluster.
3. Calculate the distance in each input data of each centroid using the Euclidean distance formula (Euclidean Distance) until the closest distance is found in each data with the centroid. Here is the
4. Euclidian Distance equation:

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

Description;
D = Distance
x = Data
y = Centroid

4. Classify data based on their proximity to the centroid.
5. Recalculate the cluster center with the current cluster members. Cluster center is the average value of all data objects in a particular cluster.
6. Count each object again using the new cluster center. If the cluster center does not change again then the clustering process is complete. Or, go back to step number 3 until the cluster center doesn't change anymore.

Table 2 Initial Centroid

Description	First stock	Stock Sold	Last stock

*Name of the corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

C1	34	21	13
C2	64	37	27
C3	87	69	18

Table 2 is Determining the Initial Center of the Cluster randomly taken from the data in Table IV.1. The data selected for the initial cluster are the 48th, 22nd, and 44th products.

The distance to the 1st data centroid in cluster 1 (C1) is:

$$(X1, C1) = \sqrt{(SA_1 - C1_1)^2 + (ST_1 - C1_1)^2 + (SAK_1 - C1_1)^2}$$
$$= \sqrt{(86 - 34)^2 + (26 - 21)^2 + (60 - 13)^2}$$
$$= \sqrt{2.704 + 25 + 2.209}$$
$$= 70{,}2709$$

The distance to the 1st data centroid in cluster 1 (C1) is:

$$(X1, C2) = \sqrt{(SA_1 - C1_2)^2 + (ST_1 - C1_2)^2 + (SAK_1 - C1_2)^2}$$
$$= \sqrt{(86 - 64)^2 + (26 - 37)^2 + (60 - 27)^2}$$
$$= \sqrt{484 + 121 + 1.089}$$
$$= 41{,}1582$$

The distance to the 1st data centroid in cluster 1 (C1) is:

$$(X1, C3) = \sqrt{(SA_1 - C1_3)^2 + (ST_1 - C1_3)^2 + (SAK_1 - C1_3)^2}$$
$$= \sqrt{(86 - 87)^2 + (26 - 69)^2 + (60 - 18)^2}$$
$$= \sqrt{1 + 1.849 + 764}$$
$$= 60{,}1165$$

Then for the next centroid distance is below:

Table 3 Data Cluster 1 Iteration 1

Description	First stock	Stock Sold	Last stock
Knitted long tunic	47	40	7
Long tunic purple	37	17	20
Long tunic green	28	13	15
Long tunic pink	48	37	11
Sabrina flower	49	23	26
Ruffle blue	34	16	18
Ruffle red	26	9	17
Gray striped shirt	44	19	25
White striped shirt	42	13	29
Black striped shirt	44	35	9
Green striped shirt	38	16	22
army pants	35	11	24
Tropical dongker	43	27	16
Tropical white	42	23	19
red flannel shirt	32	12	20
Black flannel shirt	41	25	15
Blue flannel shirt	49	22	27

*Name of the corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

denim jacket	38	16	22
jacket red	34	21	13
white jacket	48	27	21
jacket blue	42	21	22
Knitted long tunic	47	40	7
Amount	**40,0476**	**21,0952**	**18,9523**

Table 3 shows that cluster 1 in the first iteration produces 21 data that are incorporated into it.

Table 4 Data Cluster 2 Iteration 1

Description	First stock	Stock Sold	Last stock
Outer Kimono	86	26	60
Long tunic flower	57	27	30
Ruffle white	54	22	32
denim pants	74	39	35
cargo pants	67	24	43
Hot pants	55	29	26
Peach pants	64	37	27
Milo pants	76	34	42
spotted kimono	54	34	20
Kimono blue	66	38	28
Kimono green	78	32	46
Kimono red	67	42	25
Tropical brown	56	38	18
Gray flannel shirt	53	33	20
lilac sweater	56	33	23
Crop sweater	67	35	32
T-shirt crop	55	33	22
Amount	**1085**	**556**	**529**
Average	**63,8235**	**32,7058**	**31,1176**

Table 4 shows that for Cluster 2 there are 17 data that are included in it.

Table 5 Data Cluster 3 Iteration 1

Description	First stock	Stock Sold	Last stock
Sabrina knitting	66	57	9
Sabrina cotton	71	55	16
Blue striped shirt	78	53	25
red striped shirt	67	56	11
Flower kimono	83	51	32
pink kimono	67	56	11
cartoon pajamas	67	56	11
polka dot pajamas	65	57	8
jeans jacket	68	56	12
Oversized sweater	63	58	5
oversized t-shirt	63	58	5

*Name of the corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2. 11244
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Jacket varsity	78	53	25
Amount	**836**	**666**	**170**
Average	**69,6666**	**55,5**	**14,1666**

Table 5 shows that for Cluster 3 there are 12 data included in it

Tabel 6 New Centroids

Description	First stock	Stock Sold	Last stock
C1	40,04761905	21,0952381	18,9523809
C2	63,82352941	32,7058823	31,1176470
C3	69,66666667	55,5	14,16666667

Table 6 is the new centroid generated from the average of the 3 clusters.

This new centroid is calculated using a formula such as the iteration per-1 formula and because there is no data moving clusters and the 1st and 2nd clusters are the same, the new centroid calculation process is stopped and ends in the 2nd iteration.

K-Means On Rapidminer

1. Clustering Results Display

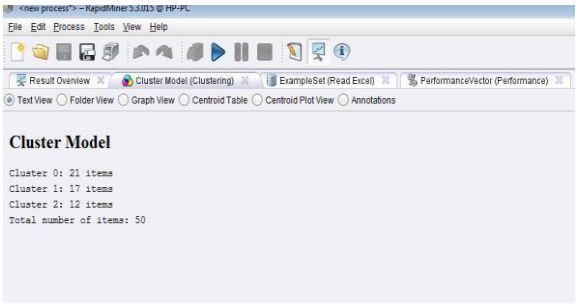

Figure 1 Display the results of the Cluster Model

Figure 1 is a display of the results of the cluster model from the processed data.

CONCLUSION

The K-means method can be applied to Top Stores to determine the sales of clothes that are selling well in stores, selling well in stores and not selling well in stores. The application of the K-Means method in Top Stores is by grouping clothing stock data. Then choose 3 clusters randomly as the initial centroid. After the data in each cluster does not change, it can be seen the final results of the best-selling, best-selling and less-selling.

REFERENCES

Aji, A. M. B. (2017). *Pengelompokan Peminatan Outline Tugas Akhir Dengan Menggunakan Algoritma K-Means Pada AMIK MI BSI Jakarta*. *5*(2), 1–9.

Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, *1*(1), 1–5. https://doi.org/10.46576/djtechno.v1i1.964

Bahar, A., Pramono, B., & Sagala, L. H. S. (2016). Penentuan strategi penjualan alat-alat tattoo di studio sonyxtattoo menggunakan metode. *SemanTIK*, *2*(2), 75–86.

Bastian, A., Sujadi, H., & Febrianto, G. (n.d.). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)*. *1*, 26–32.

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24.

*Name of the corresponding

https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, *25*(1), 76–88. https://doi.org/10.35760/tr.2020.v25i1.2677

Hasugian, P. S. (2018). Penerapan Data Mining untuk Klasifikasi Produk Menggunakan Algoritma K-Means (Studi Kasus : Toko Usaha Maju Barabai). *Jurnal Mantik Penusa*, *2*(2), 191–198.

Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA : Jurnal Informatika*, *7*(2), 109. https://doi.org/10.30595/juita.v7i2.5529

Issn, I. P. E.-. (2018). Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Padapt Batamas Niaga Jaya. *Computer Based Information System Journal*, *02*, 20–35.

Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *Jurteksi*, *4*(2), 123–132. https://doi.org/10.33330/jurteksi.v4i2.36

MURTI, M. A. W. K. (2017). Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah – Buahan Di Provinsi Daerah Istimewa Yogyakarta. *Skripsi*.

NOVIANTO, R. (2019). Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *6*(1), 85–95. https://doi.org/10.35957/jatisi.v6i1.150

Rahmat C.T.I., B., Agidtama Gafar, A., Fajriani, N., Ramdani, U., Rihin Uyun, F., Purnamasari P., Y., & Ransi, N. (2017). Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan. *Seminar Nasional Riset Kuantitatif Terapan 2017*, *April*, 58–60.

Rianti, E. (2017). Data Mining Dalam Menentukan Penjualan Laris Menggunakan Metode Clustering. *KomTekInfo*, *4*(2), 267–283.

Yulianti, Y., Utami, D. Y., Hikmah, N., & Hasan, F. N. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Mengetahui Minat Customer Di Toko Hijab. *Jurnal Pilar Nusa Mandiri*, *15*(2), 241–246. https://doi.org/10.33480/pilar.v15i2.650

*Name of the corresponding