JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1414

MAPPING OF POTENTIAL CUSTOMERS AS A CLOTHING PROMOTION STRATEGY USING K-MEANS CLUSTERING ALGORITHM

Mardalius[1]; Tika Christy[2*]

Information System
STMIK Royal
www.stmikroyal.ac.id
[1]mardalius18@gmail.com, [2*]tikachristy.royal@gmail.com
(*) Corresponding Author

Abstract—The high demand for clothes causes the development of the clothing industry in Indonesia continues to increase. Increasing the number of competitors among apparel traders is also unavoidable. This is also experienced by clothing traders with an online concept. Therefore the right sales strategy is needed to be able to survive or even win the competition. One thing that can be done is to apply technology to promote to obtain and maintain potential customers. However, promotions that are carried out without a clear and measurable concept can cause harm if carried out on target. The same thing happened in the Mustika Gerai online clothing store which was the location of observation, so far the concept of promotion was carried out by lowering prices and discounts for all customers. As a result, what happens is that sales turnover decreases dramatically while new customers who expect it may not necessarily be achieved. The purpose of this study is to research by applying data mining technology in the Gerai Mustika customer data warehouse to map potential customers as targeted promotional strategies. The data mining technique used is the k-Means Clustering method. The process of extracting information in the form of pattern discovery/mapping is then integrated using the Rapidminer software. From the results of the analysis that has been done, it can be concluded that the application of the k-means method can map potential customers based on regions or sub-districts, namely cluster 1 has 3 districts, cluster 2 has 7 districts and cluster 3 has 13 districts. These results are strengthened by RapidMiner software testing with data accuracy following the results of calculations from 23 data.

Keywords: Potential Customers, Clothing, Data Mining, K-Means, Promotion Strategy.

Abstrak—*Tingginya kebutuhan pakaian menyebabkan pekembangan industri pakaian di Indonesia terus mengalami peningkatan. Peningkatan jumlah pesaing antar pedagang pakaian juga tidak dapat dihindarkan.Hal ini juga dialami pedagang pakaian dengan konsep online.Maka dari itu dibutuhkan strategi penjualan yang tepat untuk mampu bertahan atau bahkan memenagkan persaingan.Salah satu yang dapat dilakukan yaitu menerapkan teknologi untuk melakukan promosi untuk mendapatkan dan mempertahankan pelangan potensial.Akan tetapi promosi yang dijalankan tanpa konsep yang jelas dan terukur justru dapat menyebabkan kerugian jika dilakukan tidak tepat sasaran.Sama halnya yang terjadi di toko pakaian online Gerai Mustika yang menjadi lokasi pengamatan, selama ini konsep promosi yang dilakukan yaitu dengan menurunkan harga dan diskon bagi seluruh pelanggan.Akibatnya yang terjadi adalah omset penjualan yang menurun drastis sedangkan pelanggan baru yang harapkan justru belum tentu tercapai. Adapun tujuan dari penelitian ini adalah untuk melakukan penelitian dengan menerapkan teknologi data mining pada gudang data pelanggan Gerai Mustika untuk memetakan pelanggan potensial sebagai strategi promosi yang tepat sasaran. Adapun teknik data mining yang digunakan yaitu metode k-Means Clustering. Proses penggalian informasi berupa penemuan pola/pemetaan ini selanjutnya diintegrasikan menggunakan software Rapidminer. Dari hasil analisis yang telah dilakukan dapat disimpulkan bahwa penerapan metode k-means dapat memetakan pelanggan potensial berdasarkan daerah atau kecamatan yaitu cluster 1 memiliki 3 kecamatan, cluster 2 memiliki 7 kecamatan dan cluster 3 memiliki 13 kecamatan. Hasil tersebut diperkuat dengan pengujian software Rapidminer dengan keakuratan data yang sesuai dengan hasil perhitungan dari 23 data.*

Kata Kunci*: Pelanggan Potensial, Pakaian, Data Mining, K-Means, Strategi Promosi*

INTRODUCTION

Clothing is a basic human need besides food and shelter [1].). Humans need clothes to protect and cover themselves [2]. But along with the development of human life, clothing is also used as a symbol of status, position, or position of someone who wears it. Based on data released by the Indonesian Textile Association (API) Indonesian clothing consumption is 7.5 kg/capita/year, with details of the pants being 2 kg, the remaining 5.5 kg of clothes, and batik products.

The high demand for clothing has caused the development of the clothing industry in Indonesia to continue to increase [3]. Increasing the number of competitors among clothing traders is also unavoidable. This is also experienced by clothing traders with an online concept. Therefore, it requires the right sales strategy to be able to survive or even win the competition. One thing that can be done is to apply technology to conduct promotions to get and maintain potential customers [4].

However, promotions that are run without a clear and measurable concept can cause harm if carried out on target. This occurs because of the large costs required for promotion. During this time most of the clothing merchants do promotions by lowering prices and large-scale discounts. Just as happened in the online clothing store Mustika Outlet which is the location of observation, so far the concept of promotion is carried out by lowering prices and discounts for all customers. What happens is that sales turnover decreases even though sales volume has increased slightly while new customers who expect it may not necessarily be achieved. Based on the above problems, researchers are interested in researching by applying data mining technology in the Gerai Mustika customer data warehouse to map potential customers as targeted promotional strategies. The data mining method used is the k-Means Clustering method [5]. The process of extracting information in the form of pattern discovery/mapping is then integrated using the Rapidminer software. The use of Rapidminer software aims to facilitate the analysis process of large data warehouses [6]. The K-Means method seeks to group existing data into several groups, where data in one group has the same characteristics with each other and has different characteristics with data in other groups [7]. The K-Means method is one technique that can be used to assist in designing an effective inventory strategy by utilizing sales transaction data that is already available in the company [8].

MATERIALS AND METHODS

The framework of this research is as follows:

1. Preliminary Study
   Study the literature aimed at gaining a deep understanding of the problem to be studied and the methods used as solutions. Literature is a very important publication of scientific papers both journals and processing, books related to data mining, tutorials on the use of Rapidminer software, and other references.
2. Instrument preparation
   The preparation of the instrument is intended as a tool that will be used in data collection.
3. Data Collection
   The data collected was obtained through filling instruments in the form of observation sheets, interviews, and direct observations at the Mustika Outlet. Data collected was only related to customer data with the number of purchase transactions.
4. Data Analysis
   The data collected will be analyzed using data mining techniques with the K-Means method. Before that, the data is first processed by data selection, data preprocessing, data transformation and ending with data mining to implement algorithms and mining techniques to extract knowledge in the form of pattern discovery, trends, predictions in others- At this stage members undertake data cleaning, data preprocessing activities and data transformation.
5. System Testing
   The results of data analysis performed manually using the K-Means method are then integrated using the Rapidminer software to find out whether the results obtained are compatible and have a good degree of accuracy.

RESULTS AND DISCUSSION

Based on the data collected, the power to be processed is a sample taken from the Mustika Outlet customer data in each region based on the District from 2015-2019. This customer dataset consists of attributes that have been simplified to Name Districts, Number of Customers in 2015-2019 which will be applied through Algortma K-Means. The sample data to be tested amounted to 23 districts which later will be grouped into 3 clusters. The following is Table 1 of the Mustika Gerai customer data set.

Table 1. Customer Data Set

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1414

No	Kecamatan	2015	2016	2017	2018	2019
1	Air Putih	1067	1170	1068	1172	1274
2	Lima Puluh	1272	1290	1321	1365	1378
3	Sei Balai	1277	1378	1388	1399	1378
4	Talawi	533	444	533	421	625
5	Sei Suka	543	421	566	531	628
6	Tanjung Tiram	423	321	412	541	476
7	Teluk Nibung	1175	1175	1175	1181	1190
8	T. Balai Utara	333	342	545	350	434
9	T.BalaiSelatan	222	216	612	412	555
10	SeiTualangRaso	1176	1188	1177	1185	2190
11	Datuk Bandar,	1173	1190	1188	1178	1182
12	Kisaran Timur	2399	2444	2457	2543	2690
13	Kisaran Barat	2289	2332	2277	2312	2390
14	Meranti	333	444	454	455	339
15	Air Joman	321	222	334	212	224
16	Pulo Bandring	425	338	221	229	224
17	Tinggi Raja	563	555	449	434	378
18	Air Batu	220	341	441	429	520
19	Setia Janji	555	755	725	888	1024
20	Buntu Pane	543	522	336	390	342
21	Teluk Dalam	2281	2388	2284	2392	2494
22	Sei Kepayang	1034	930	988	734	544
23	Sei Dadap	1144	1050	955	843	956

The following are the steps for completion using the k-means algorithm.

1. Starting by determining the number of clusters, namely 3 clusters.

2. Determine the initial centroid at random which is the center point of the first cluster. The initial centroids of this study are as follows:
   C10 : 1176, 1188, 1177, 1185, 2190
   C12 : 2399, 2444, 2457, 2543, 2690
   C15 : 321, 222, 334, 212, 224

3. Calculate the distance of the data with its centroid point using the following Euclidean distances:

   $$D = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad \text{.....(1)}$$

   Where:
   D = 1st data distance with centroid point
   n = number of changes / parameters
   xi = data
   yi = centroid data

The results of calculations from all the data sets with each beginning are presented in table 2 below.

Table 2 Results of Distance Between Data and Initial Centroids

No	C1	C2	C3
1	3035,02	929,145	2004,51
2	2646,01	855,628	2383,34
3	2562,83	892,167	2470,2
4	4466,64	2101,1	581,645
5	4406,03	2056,66	638,503
6	4637,19	2292,7	445,01
7	2976,09	1000,09	2053,51
8	4716,07	2367,82	349,584
9	4712,4	2333,43	486,5
10	2608,8	0	2680,8
11	2970,39	1008,09	2059,89
12	0	2608,8	5027,64
13	447,666	2251,14	4604,56
14	4708,22	2397,08	368,92
15	5027,64	2680,8	0
16	4976,68	2641,88	193,21
17	4556,81	2270,2	505,646
18	4734,22	2352,33	412,936
19	3842,85	1491,71	1260,44
20	4667,54	2366,15	429,995
21	328,947	2331,47	4716,03

No	C1	C2	C3
22	3767,12	1742,17	1346,19
23	3416,29	1307,31	1637,85

4. Conduct data cluster clustering. After calculating the data spacing in the centroid above, the next step is to group the data. Calculations can be done manually using Microsoft Excel. The purpose of number 2 in the region with the least number of customers, number 1 indicates the current area, and the number 0 indicates the area with the most potential customers.

Table 3. Results of Grouping Data

No	C1	C2	C3	C1	C2	C3
1	3035,02	929,145	2004,51	0	2	1
2	2646,01	855,628	2383,34	0	2	1
3	2562,83	892,167	2470,2	0	2	1
4	4466,64	2101,1	581,645	0	1	2
5	4406,03	2056,66	638,503	0	1	2
6	4637,19	2292,7	445,01	0	1	2
7	2976,09	1000,09	2053,51	0	2	1
8	4716,07	2367,82	349,584	0	1	2
9	4712,4	2333,43	486,5	0	1	2
10	2608,8	0	2680,8	1	2	0
11	2970,39	1008,09	2059,89	0	2	1
12	0	2608,8	5027,64	2	1	0
13	447,666	2251,14	4604,56	2	1	0
14	4708,22	2397,08	368,92	0	1	2
15	5027,64	2680,8	0	0	1	2
16	4976,68	2641,88	193,21	0	1	2
17	4556,81	2270,2	505,646	0	1	2
18	4734,22	2352,33	412,936	0	1	2
19	3842,85	1491,71	1260,44	0	1	2
20	4667,54	2366,15	429,995	0	1	2
21	328,947	2331,47	4716,03	2	1	0
22	3767,12	1742,17	1346,19	0	1	2
23	3416,29	1307,31	1637,85	0	2	1

5. After all, data has been successfully grouped into the closest cluster, then recalculate the new cluster center based on the average of each member in the cluster.

Table 4 New Centroids

	2015	2016	2017	2018	2019
C1	2323	2388	2339,33	2415,67	2524,67
C2	1183,43	1205,86	1181,71	1189	1364
C3	465,231	450,077	508,923	463,538	485,615

6. Recalculate the data with centroid, but the next calculation uses the new centroid in table 4. By using the new centroid value, the distance to each centroid is calculated until the new table is the same as table 3 grouping the data and or until getting the last pattern not move.

Table 5 Results and Final Patterns of Iterations Between Data and Centroid

C1	C2	C3	C1	C2	C3
0	2	1	0	2	1
0	2	1	0	2	1
0	2	1	0	2	1
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
0	2	1	0	2	1
0	1	2	0	1	2
0	1	2	0	1	2
1	2	0	1	2	0
0	2	1	0	2	1
2	1	0	2	1	0
2	1	0	2	1	0
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
0	1	2	0	1	2
2	1	0	2	1	0
0	1	2	0	1	2
0	2	1	0	2	1

In the final results of Table 5 above, the data that has been grouped in cluster 1 is 3 sub-districts, in cluster 2 there are 7 sub-districts, and in cluster 3 there are 13 sub-districts. Done

7. Testing the system with applications that have been determined. The results of the analysis that have been carried out will be tested with RapidMiner software to ensure the compatibility of the results. Next Text View will display the cluster model. Cluster models obtained from the results of testing the data using the k-means algorithm.

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1414

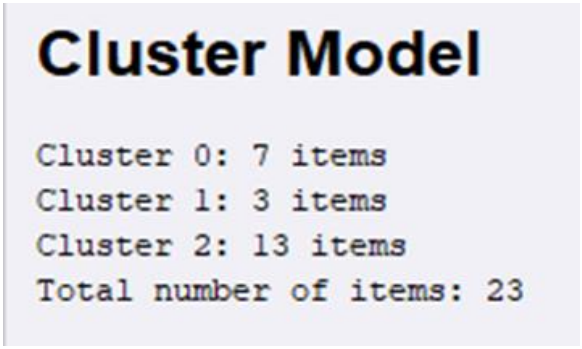

Figure 1 Cluster Model

It can be seen in Figure 1 above, the cluster model consists of cluster 0, namely 7 items, cluster 1 consists of 3 items, and cluster 2 consists of 13 items. Of the total number of 23 items. In the form of the root, it has 3 clusters namely folder clusters 0, cluster 1, and cluster 2. To make it easier to see the members that are owned by each cluster folder, consider Figure 2 below which will display each cluster member.

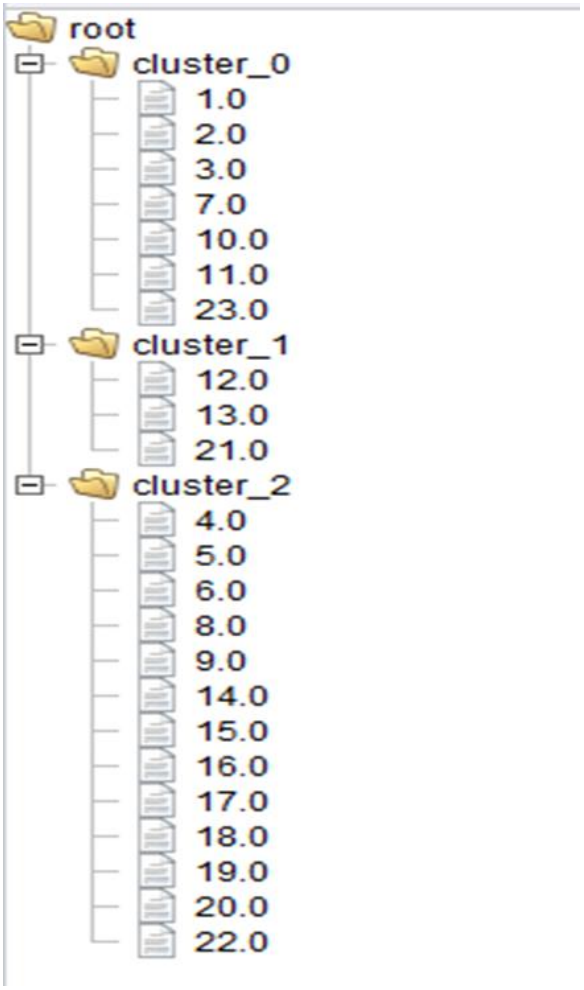

Figure 2. Display of Each Cluster Member

CONCLUSION

From the results of the analysis above, it can be concluded that the application of the K-Means method can map potential customers based on regions or sub-districts namely cluster 1 has 3 districts, cluster 2 has 7 districts and cluster 3 has 13 districts. These results are strengthened by RapidMiner software testing with data accuracy following the results of calculations from 23 data.

ANNOUNCEMENT

Thank you to the Deputy Minister for Research and Development Strengthening of the Ministry of Research and Technology / National Research and Innovation Ministry following the Research Contract for Fiscal Year 2020.

REFERENCE

[1] K. P. Arindra and F. Nursasari, "EKSPLORASI POLA BRALETTE DENGAN APLIKASI REKA LATAR," *ATRAT J. SENI RUPA*, vol. 5, no. 3, pp. 297–304, 2017.

[2] A. Fauzi, "Jual Beli Pakaian Bekas dalam Perspektif Fikih Muamalah Iqtishodiyah," *Iqtishodia J. Ekon. Syariah*, vol. 4, no. 2, pp. 235–267, 2019.

[3] A. Setyaningsih, N. Bahari, and D. T. Ardiyanto, "Kuasa Baju Bekas: Kode Kultural Fesyen Baju Bekas Dalam Ranah Industri Kreatif," *J. Ind. Kreat. dan Kewirausahaan*, vol. 1, no. 1, pp. 10–21, 2018.

[4] S. Kosasi, "Perancangan Sistem Electronic Customer Relationship Management Untuk Mempertahankan Loyalitas Pelanggan," *J. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 2, pp. 92–102, 2018.

[5] T. Alfina, B. Santosa, and A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri ITS)," *J. Tek. ITS*, vol. 1, no. 1, pp. 1–5, 2012.

[6] R. Nofitri and N. Irawati, "INTEGRASI METODE NEIVE BAYES DAN SOFTWARE RAPIDMINER DALAM ANALISIS HASIL USAHA PERUSAHAAN DAGANG," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 1, pp. 35–42, 2019.

[7] B. M. Metisen and H. L. Sari, "Analisis

clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

[8] S. Setiawan, "Pemanfaatan Metode K-Means Dalam Penentuan Persediaan Barang," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 6, no. 1, pp. 41–48, 2018.