Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2022. Vol. 1. No. 2

Implementation of the K-Means Method in Grouping Merchandise Locations at the Market Service

Syawalluddin Pohan[1]*, Sumarno[2], Ika Okta Kirana[3], Dedy Hartama[4], Jalaluddin[5]

[1,2,3,4,5]*STIKOM Tunas Bangsa Pematangsiantar, North Sumatra, Indonesia*
**syawalluddinpohan@gmail.com*

Abstract

One of the strategies to increase sales in traditional markets is the strategy of placing the selling location. This is done so that the products marketed are in accordance with the type so that consumers will feel comfortable with the ease of shopping. In this study, observations were made on the traditional market of Horas market in the market service area of Pematangsiantar City. At this time the arrangement of selling locations has not been well organized so that there is little interest in the community to shop. This of course will affect the economic turnover of traders. These problems still occur today and there is no solution because market managers do not have a model that can be simulated. One of the computer science approaches to this problem is the K-Means algorithm data mining so that it is hoped that this research can help the market department in classifying merchandise locations in order to attract people's interest to shop at traditional markets so that there is an increase in the community's economy.

Keywords: *Location, Datamining, K-Means Algorithm, Interests*

1. Introduction

Data mining is a method for finding knowledge in a large enough pile of data by digging and analyzing a very large amount of data in order to obtain something true, new and useful so that a pattern or pattern can be found in the data [1],[2]. Datamining is a branch of artificial intelligence in extracting patterns to turn data into information[3]. Data mining is a process that employs one or more computer learning techniques to analyze and extract knowledge (knowladge) automatically[4],[5]. The term data mining has been known since 1990, The emergence of data mining is based on the amount of data stored in databases that is getting bigger.

Traditional markets are a pillar of economic activity in Indonesia. The market is a place that can connect sellers and buyers to make transactions for goods or services[6]. The market is also defined as a place where the price determination process takes place and the market plays an important role in driving the community's economy. The market not only moves the economy of the community but also helps in improving the country's economy. With the increasing economy society, the purchasing power of the people also increases. This can be seen from the many innovations in the production of goods and services that are in great demand by the public. Along with the development of the era of traditional markets began to be replaced by the modern market. The emergence of this modern market has begun to marginalize the existence of traditional markets in driving the community's economy. People began to prefer shopping in modern markets because it is more systematic and clean, and not all goods or services offered can be obtained from traditional markets. The rapid growth of this modern market is felt to be less coordinated and less synergized with small and medium traders which has resulted in the displacement of traditional market traders by the practicality of shopping in modern markets. To overcome the negative impact of the emergence of modern markets, as well as as a protective measure for traditional market traders, traditional markets need to be further developed both in terms of layout and management of market traders so that they can continue to exist in the midst of the onslaught of modern markets.

Horas market is one of the traditional markets in Pematangsiantar City, North Sumatra. Horas market is a market that is quite active until now. However, the layout of the location of the merchandise in this traditional market is still less structured and quite messy. In this case, the market office of Pematangsiantar City plans to regroup the locations of the existing merchandise in the Horas Market of Pematangsiantar City in order to attract people's interest to shop for necessities in traditional markets. To facilitate the grouping of existing merchandise locations in the Pematangsiantar Horas market, this study uses the data mining method. One of the main tasks of data mining is clustering[7],[8]. Clustering is the process of organizing objects into classes/groups by looking for similarities in each object[9],[10]. The potential of clustering is that it can be used to find out structures in data that can be used further in a wide variety of applications such as classification, image processing, and pattern recognition[11]. One of the algorithms in clustering is the K-Means Algorithm. In this study, the author uses the K-Means Algorithm in classifying the location of merchandise at the market service of Pematangsiantar City.

The K-Means algorithm is an iterative clustering algorithm that partitions the data set into a number of K clusters that have been set at the beginning[12],[13]. The K-Means method tries to group existing data into several groups, where data in one group has the

same characteristics as each other and has different characteristics from data in other groups[14]. The K-Means algorithm is well known for its ease and ability to cluster large data and outlier data very quickly[15]. In several studies that have been carried out by researchers related to datamining methods using the K-Means Algorithm[16] explains the K-Means Algorithm in classifying students based on academic scores and the results are 4 clusters. Students with GPA = 0.5167 for cluster 0, 9 students, students with GPA = 3.4143 for cluster 1, 28 students, students with GPA = 3.3092 for cluster 2, 40 students, students with GPA = 3.8991 for cluster 3, 47 Student. The K-Means algorithm can be used to group students based on their GPA and some course attributes. [17]also uses the K-Means Algorithm in determining the nutritional status of toddlers which results from the use of the K-Means Algorithm, namely grouping the nutritional status of toddlers into 5 clusters, cluster 1 - poor nutrition, cluster 2 - poor nutrition, cluster 3 - good nutrition , cluster 4 - overweight, cluster 5 - obesity.

2. Research Methodology

The data in this study were obtained by means of observation, namely observing and studying problems that exist in the field related to the object under study and a Literature Study was carried out by looking for materials that support the definition of problems through journals, books, papers, the internet which are closely related to the object of the problem.Data analysis used quantitative data with data analysis techniques that used descriptive statistics. The data obtained is then processed with RapidMiner to find the accuracy of the data. The research data used are as follows:

Table 1. Research Data

No	Trade Type	G1	G2	G3	Total
1	accessories	3	4	1	8
2	Fruits	4	5	5	14
3	Books	1	0	1	2
4	Crops	3	8	8	19
5	Fish & Meat	2	2	10	14
6	O'clock	7	2	0	9
7	Shoes	1	0	0	1
8	Food & Drink	2	3	6	11
9	Textile	0	2	3	5
10	Children toys	0	1	0	1
11	Clothes	6	3	0	9
12	Used Clothes	0	18	8	26
13	Mobile	3	0	0	3
14	Cigarette	0	3	0	3
15	Shoes & Sandals	8	8	0	16
16	Processed soybeans	0	1	1	2
17	Bag	2	0	0	2
18	VCD	0	2	0	2

This research flow chart consists of problem analysis, namely analyzing problems related to grouping market merchandise locations and determining what parameters or criteria are used. Studying Literature is a reference used to obtain information in research. Define Method to solve the problem. In this study, the method used is Data Mining with the K-means algorithm. Collecting data is collecting data by taking data that already exists in related agencies, managing data managing data using the k-means algorithm. The data processing here uses Ms. Excel 2010. Testing the data is done using the RapidMiner application. Making conclusions in grouping the location of market merchandise. The flow chart of this research can be seen in Figure 1 below:

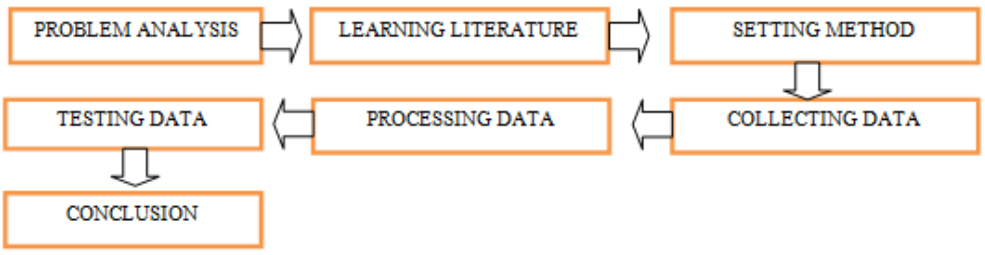

Figure 1. Research Framework

This research flow chart consists of problem analysis, namely analyzing problems related to grouping market merchandise locations and determining what parameters or criteria are used. Studying Literature is a reference used to obtain information in research. Define Method to solve the problem. In this study, the method used is Data Mining with the K-means algorithm. Collecting data is collecting data by taking data that already exists in related agencies, managing data managing data using the k-means algorithm. The data processing here uses Ms. Excel 2010. Testing the data is done using the RapidMiner application. Make conclusions in grouping the location of market merchandise.

3. Results and Discussion

3.1. Solution with KMeans Algorithm

The research data consists of 18 types of merchandise consisting of 3 buildings, then calculate the total merchandise from 3 buildings as shown in table 3.1. Then enter the calculation stage using the following k-means method: Determination of the Initial Cluster Center, namely the determination of the cluster point is done by taking the largest (maximum) value for the highest type of merchandise (C1), the smallest value for the type of merchandise (C2), namely High Cluster (C1) = 26 and Low Cluster (C2) = 1. Calculation of Cluster Distance, namely To calculate the distance between the data and the center of the cluster using equation (2).

$d_{ik}\sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$ ..... (2)

C1=$\sqrt{(8-26)^2}$=18

C2=$\sqrt{(8-1)^2}$=7

the results of the calculation of the cluster distance:

Table 2. Iteration Results

No	Trade Type	Total	C1	C2	Distance
1	accessories	8	18	7	7
2	Fruits	14	12	13	12
3	Books	2	24	1	1
4	Crops	19	7	18	7
5	Fish & Meat	14	12	13	12
6	O'clock	9	17	8	8
7	Shoes	1	25	0	0
8	Food & Drink	11	15	10	10
9	Textile	5	21	4	4
10	Children toys	1	25	0	0
11	Clothes	9	17	8	8
12	Used Clothes	26	0	25	0
13	Mobile	3	23	2	2
14	Cigarette	3	23	2	2
15	Shoes & Sandals	16	10	15	10
16	Processed soybeans	2	24	1	1
17	Bag	2	24	1	1
18	VCD	2	24	1	1

Based on table 2, the data from iteration 1 can be grouped as follows:

Table 3. Grouping of data iteration 1

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
6		1
7		1
8		1
9		1
10		1
11		1
12	1	
13		1
14		1
15	1	
16		1
17		1
18		1

After getting the results from each cluster, then the new cluster center is calculated based on the member data of each cluster that has been obtained using the formula that corresponds to the center member cluster as follows:

$C_{kj}\frac{\sum_{i=1}^{p} X_{ij}}{p}$

C1=(14+19+14+26+16)/5=17,8

C2=4,4615

The next iteration is carried out in the same way until there is no change in the data in a cluster.
Iteration 2:

$$d_{ik}\sqrt{\sum_{j}^{m}(C_{ij}-C_{kj})^2}$$

C1=$\sqrt{(8-17{,}8)^2}$=9,8

C2 =$\sqrt{(8-4{,}4615)^2}$=3,358

Here is the result of iteration 2:

Table 3. Calculation Results of Iteration 2

No	Trade Type	Total	C1	C2	Distance
1	accessories	8	9,800	3,538	3,538
2	Fruits	14	3,800	9,538	3,800
3	Books	2	15,800	2,462	2,462
4	Crops	19	1,200	14,538	1,200
5	Fish & Meat	14	3,800	9,538	3,800
6	O'clock	9	8,800	4,538	4,538
7	Shoes	1	16,800	3,462	3,462
8	Food & Drink	11	6,800	6,538	6,538
9	Textile	5	12,800	0,538	0,538
10	Children toys	1	16,800	3,462	3,462
11	Clothes	9	8,800	4,538	4,538
12	Used Clothes	26	8,200	21,538	8,200
13	Mobile	3	14,800	1,462	1,462
14	Cigarette	3	14,800	1,462	1,462
15	Shoes & Sandals	16	1,800	11,538	1,800
16	Processed soybeans	2	15,800	2,462	2,462
17	Bag	2	15,800	2,462	2,462
18	VCD	2	15,800	2,462	2,462

Based on table 3, the data from iteration 2 can be grouped as follows:

Table 4. Grouping of iteration data 2

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
6		1
7		1
8		1
9		1
10		1
11		1
12	1	
13		1
14		1
15	1	
16		1
17		1
18		1

From the results of iteration 2, it can be seen that the results obtained are the same as iteration 1, so the calculation is stopped until iteration 2. For the types of merchandise in cluster C1 (the highest), namely Fruit, Agricultural Products, Fish and Meat, Used Clothing, Shoes and Sandals. The types of merchandise in cluster C2 (low) are Accessories, Books, Watches, Sewing Shoes, Food, Fabrics, Children's Toys, Clothing, Cellphones, Cigarettes, Tofu, Bags, and VCDs.

3.2. Modeling with RapidMiner

Tests using the K-Means Clustering algorithm using Rapidminer can be seen in Figure 2:

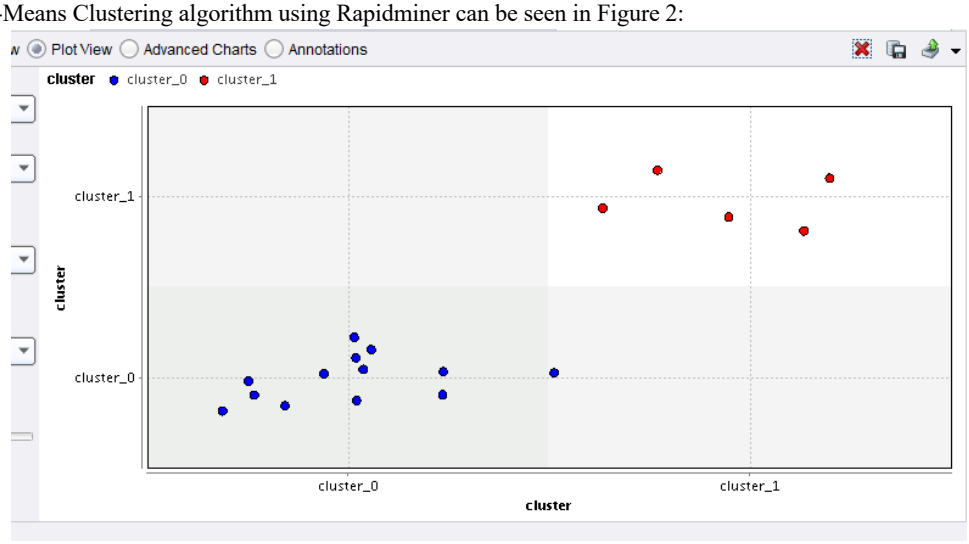

Figure 2. Graph of Cluster Results

Based on Figure 2, it can be seen that cluster 0 is a low cluster and cluster 1 is a high cluster. From Figure 4.4 it can also be seen that there are 5 alternatives in the high cluster and 13 alternatives in the low cluster as shown in Figure 3.

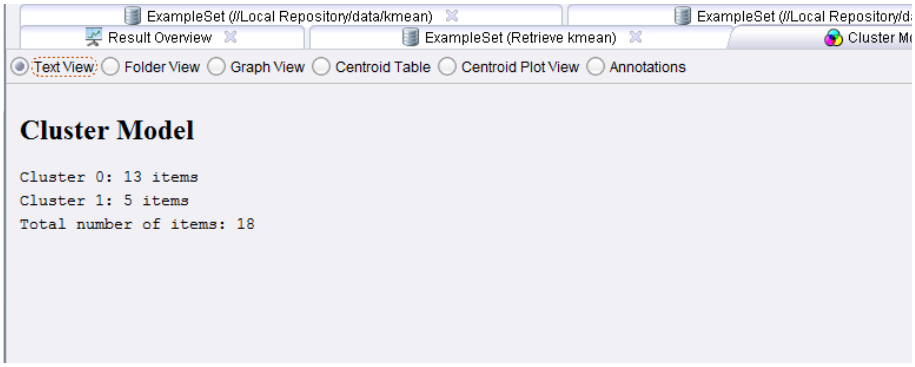

Figure 3. Cluster Model

on data validation, the algorithm calculation produces the final result of grouping with 2 clusters, and the data used is valid data and is the same as that used in Rapidminer tools. The following shows the results obtained from algorithm calculations and tests on rapidminer.

Table 5. Comparison of Results of Algorithms and RapidMiner

No	Trade Type	Algorithm Calculation	*RapidMiner* Calculation
1	accessories	Cluster 0	Cluster 0
2	Fruits	Cluster 1	Cluster 1
3	Books	Cluster 0	Cluster 0
4	Crops	Cluster 1	Cluster 1
5	Fish & Meat	Cluster 1	Cluster 1
6	O'clock	Cluster 0	Cluster 0
7	Shoes	Cluster 0	Cluster 0
8	Food & Drink	Cluster 0	Cluster 0
9	Textile	Cluster 0	Cluster 0
10	Children toys	Cluster 0	Cluster 0
11	Clothes	Cluster 0	Cluster 0
12	Used Clothes	Cluster 1	Cluster 1
13	Mobile	Cluster 0	Cluster 0
14	Cigarette	Cluster 0	Cluster 0
15	Shoes & Sandals	Cluster 1	Cluster 1
16	Processed soybeans	Cluster 0	Cluster 0
17	Bag	Cluster 0	Cluster 0
18	VCD	Cluster 0	Cluster 0

Based on the comparison table above, it can be seen that the results of manual calculations are the same as those of rapidminer. Where the high cluster consists of 5 alternatives, namely: Fruits, Agricultural Products, Fish and Meat, Used

Clothing, Shoes and Sandals. For the low cluster there are 13 alternatives, namely: accessories, books, clocks, sewing shoes, food and beverages, cloth, children's toys, clothes, cell phones, cigarettes, tofu, bags, VCD.

4. Conclusion

Determination of the centroid (central point) in the early stages of the K-Means Algorithm is very influential on the results of the cluster as in the results of tests carried out using 18 datasets with different centroids producing the same cluster results. The process of testing the K-Means Algorithm using Rapidminer 5.0. in grouping the location of merchandise can be divided into 5 processes, New process, Open Recent Process, Open Process, Open Template and Online Tutorial.

Acknowledgement

Infinite thanks to the supervisory team and the testing team who are lecturers at AMIK and STIKOM Tunas Bangsa so that this research can be completed. The discussion in this study is still on a small scale, so we really hope that constructive suggestions and suggestions will make this research useful and become a reference for other researchers.

References

[1] B. Baldassari, "Mining Software Engineering Datafor Useful Knowledge," 2014.

[2] R. W. Nasution, I. O. Kirana, I. Gunawan, and I. P. Sari, "Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT . Jalur Nugraha Ekakurir (JNE) Pematangsiantar," vol. 1, no. 4, pp. 274–281, 2021.

[3] F. Gullo, "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do," *Phys. Procedia*, vol. 62, pp. 18–22, Dec. 2015, doi: 10.1016/j.phpro.2015.02.005.

[4] U. Ependi and A. Putra, "Solusi Prediksi Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang)," *J. Edukasi dan Penelit. Inform.*, vol. 5, p. 139, Aug. 2019, doi: 10.26418/jp.v5i2.32648.

[5] A. E. Lubis and P. M. Hasugian, "Implementation Of Data Mining On Suzuki Motorcycle Sales In Gemilang Motor Prosperous With Apriori Algorithm Method," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 2, no. 1, pp. 23–29, 2020, doi: 10.47709/cnapc.v2i1.353.

[6] G. Ontario, "E-COMMERCE : PURCHASING Why E-Commerce for Small Businesses ?," *E-bus. Toolkit*, p. 1, 2013.

[7] S. Budiman, D. Safitri, and D. Ispriyanti, "Perbandingan Metode K-Means Dan Metode Dbscan Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang," *None*, vol. 5, no. 4, pp. 757–762, 2016.

[8] A. Ali Hussein and A. Oluwaseun, "Data Mining Application Using Clustering Techniques (K-Means Algorithm) In The Analysis Of Student's Result," pp. 2458–2925, May 2019.

[9] M. Omran, A. Engelbrecht, and A. Salman, "An overview of clustering methods," *Intell. Data Anal.*, vol. 11, pp. 583–605, Nov. 2007, doi: 10.3233/IDA-2007-11602.

[10] O. Abu Abbas, "Comparisons Between Data Clustering Algorithms," *Int. Arab J. Inf. Technol.*, vol. 5, pp. 320–325, Jul. 2008.

[11] R. Xu and D. Wunsch, "Survey of Clustering Algorithms," *Neural Networks, IEEE Trans.*, vol. 16, pp. 645–678, Jun. 2005, doi: 10.1109/TNN.2005.845141.

[12] N. Wakhidah, "Clustering Menggunakan K-Means Algorithm," *J. Transform.*, vol. 8, no. 1, p. 33, 2010, doi: 10.26623/transformatika.v8i1.45.

[13] A. M. H. Pardede *et al.*, "Implementation of Data Mining to Classify the Consumer's Complaints of Electricity Usage Based on Consumer's Locations Using Clustering Method," 2019, doi: 10.1088/1742-6596/1363/1/012079.

[14] H. Alashwal, M. El Halaby, J. J. Crouse, A. Abdalla, and A. A. Moustafa, "The Application of Unsupervised Clustering Methods to Alzheimer's Disease," *Front. Comput. Neurosci.*, vol. 13, p. 31, May 2019, doi: 10.3389/fncom.2019.00031.

[15] A. Fahim, A.-B. M.Salem, F. Torkey, and M. Ramadan, "Efficient enhanced k-means clustering algorithm," *J. Zhejiang Univ. Sci. A*, vol. 7, pp. 1626–1633, Jan. 2006, doi: 10.1631/jzus.2006.A1626.

[16] A. Roni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang (Implementation Method for K-Means Clustering Based Student Value with Weka Interface a Case Study of Department of Information UMM Magelang)," pp. 76–82, May 2015.

[17] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.