Jurnal Mantik
Volume 4, Number 1, May 2020, pp. 718-722 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Customer Segmentation Using K-Means Algorithm As A Basis For A Marketing Strategy In The Store Rumah Tua VAPE

Rahmat Fadly[1] Iwan Wahyudin[2] Winarsih[3]

Universitas Nasional
Email: Vexarious4@gmail.com, iwan_wyd@yahoo.com, winarsih@civitas.unas.ac.id

ARTICLE INFO

Article history:
Received: 04/04/2020
Revised: 20/04/2020
Accepted: 30/05/2020

Keywords:
Algorithms,
Clustering,
K-means,
Customer segmentation

ABSTRACT

The development of the industrial system at this time more and more. Competition in the business world is increasingly stringent and requires business people to create good marketing strategies, one of which utilizes transaction data from Old Vape store companies is a company engaged in the field of electronic cigarettes. But the lack of observation in choosing customer data is less effective. And to overcome this problem can use data clustering analysis using k-means algorithm. And clustering techniques are functional in data mining, data mining algorithms are grouping a lot of data into several data groups or clusters. The data used is taken from the Vape Old House Store customer data, which results in a type of data grouping, namely customer segmentation for basic strategies in the Vape Old House Store so that customer data is not in vain.

1. Introduction

In the old vape shop that has been in the vape business from 2015 until now, it has too many customer transaction data, and the old vape shop now has branches in every major city and to face competitors who In this modern era, customers are the first priority for an old vape home shop[1]. For customers, old vape home stores give a new and comfortable impression when making transactions when purchasing at old vape shop houses because of customer segmentation to prioritize customers who frequently make transactions[2].

2. Literature review

Data collection on old vape shop houses is very important for customers or the store to provide comfort and benefit so that both parties are equally happy[1]. customer data helps the old vape home shop to segment to determine the best marketing strategy for the old vape shop house[2]. Transaction data obtained from an old vape home store that will be processed and become information to assist the store in conducting marketing strategies that will be used[3]. Old vape shop houses that are building marketing strategies like to not know the results of designing information on customer data collection they have because the system they have is not very complicated so old vape shop houses find it difficult to understand the results of data collection on customers and ultimately there is no reciprocity from oko old house vape to customers who are thirsty become an important role in customer segmentation[4]. Loyal customer is one who has an important role for an old vape shop, information about customers who often make transactions because it makes customers can be served well in an old vape shop so that it becomes a loyal customer[5]. All research that has been done by customer data, can segment with different data to get the right information about transactions conducted by customers in the old vape shop[6]. Customer data is used to segment customers using the k-means method, this method is used because it is very effective[7]. If customer data is not used properly it will cause problems with customer data collection[8].

3. Research methods

3.1 Discussion of k-means

Kmeans is one of the clustering algorithms, the purpose of this algorithm is to divide the data into several groups which are fixed or existed since the beginning. The input received is the data or object and the desired

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

cluster and then this algorithm will group the data or cluster, in each cluster there is a center point or centeroid that represents the cluster.

3.2 Discussion of RFM

Recency, frequency, and monetary stands for RFM. This is a useful method for selecting customers that must be prioritized based on the results of customer data transactions in old vape home stores. Customers who often make purchases and customers who make higher transactions. The rfm method is very much used as a customer data management in an old vape home store to find customers who frequently conduct transactions and also collect a lot of customer data collection at this time.

3.3 Data collection

Data can be obtained from bookkeeping that will be processed using sales that will be processed in order to produce a good business strategy. Initial data that can be in the form of sales lists from random memorandum notes has not been selected into data that will be used in research.

3.4 Data selection

In the data of items obtained from old vape shop houses, a selection is made after the selection is made on the memorandum based on the customer's name.

3.5 Preproccesing / cleaning

After the data is selected, the researcher conducts a preprocessing / cleaning process, namely the selection of attributes that will be used in the data mining process and selecting data attributes that can be used. In the memorandum there are several attributes used including the number of many purchases, total expenditure, and the last transaction date in the transaction date attribute rounding off in the day to make it easy to carry out the data mining process.

3.6 Data transformation

At this stage the data transformation is done by entering the sales transaction data that has been obtained right into the data mining application

3.7 Data design

In designing the data after the data transformation is complete. The following is the data design and its clusters.

Table 1

Name	F	M	R
Eddie	14	500000	23
Rafi	20	350000	7
Rahman	11	50000	16
Fadly	4	200000	26
Jaya	30	600000	6

Customer data taken from the old vape shop

Table 2

	F	M	R
cluster 1	15	400000	20
cluster2	3	200000	6

The cluster values that have been set in the old vape home shop

4. Results and Discussion

In the calculation of this customer data collection uses two clusters. At this stage, it is done by way of collecting customer data from old vape shop houses.

Table 3

Initial Calculation

Name	F	M	R	in day
Eddie	14	500000	May 7 2020	23
Rafi	20	350000	May 23, 2020	7
Rahman	11	50000	May 14 2020	16
Fadly	4	200000	May 4 2020	26
Jaya	30	600000	May 14 2020	6

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Sales transaction data that will be input into the data mining application,

Table 4

	F	M	R
cluster 1	15	400000	20
cluster2	3	200000	6

cluster value that has been specified.

$$d(1.1) = \sqrt{(14-15)^2 + (500.000-400.000)^2 + (23+20)^2} = 100.000.00$$

$$d(1.2) = \sqrt{(14-3)^2 + (500.000-200.000)^2 + (23-6)^2} = 300.000.00$$

$$d(2.1) = \sqrt{(20-15)^2 + (350.000-400.000)^2 + (7-20)^2} = 50.000.00$$

$$d(2.2) = \sqrt{(20-3)^2 + (350.000-200.000)^2 + (7-6)^2} = 150.000.00$$

$$d(3.1) = \sqrt{(11-15)^2 + (50.000-400.000)^2 + (16-20)^2} = 350.000.00$$

$$d(3.2) = \sqrt{(11-3)^2 + (50.000-200.000)^2 + (7-6)^2} = 150.000.00$$

$$d(4.1) = \sqrt{(4-15)^2 + (200.000-400.000)^2 + (26-20)^2} = 200.000.00$$

$$d(4.2) = \sqrt{(4-3)^2 + (200.000-200.000)^2 + (26-6)^2} = 20.20$$

$$d(5.1) = \sqrt{(30-15)^2 + (600.000-400.000)^2 + (6-20)^2} = 200.000.00$$

$$d(5.2) = \sqrt{(30-3)^2 + (600.000-200.000)^2 + (6-6)^2} = 400.000.00$$

Calculations from customer data

Table 5

Name	F	M	in day	C1	C2	Results
edi	14	500000	23	100000.00	300000.00	C1
rafi	20	350000	7	50000.00	150000.00	C1

User data uses the k-means cluster method

C1 (f) = (14 + 20 + 30)/3 = 21.33333

C1 (m) = (500.000 + 350.000 + 600.000)/3 = 483333.33

C1 (r) = (23 + 7 + 6)/3 = 12.00

C2 (f) = (4 + 11)/2 = 7.50

C2 (m) = (500.000 + 200.000)/3 = 125.000

C2 (r) = (16 + 26)/3 = 21.00

the results of calculations from customer data with the k-means clustering method. Then after finding the cluster, the results of the clustering will be entered into the cluster table

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Table 6

	F	M	R
cluster 1	21,333	483333.33	12.00
cluster2	7.50	125000.00	21:00

Results of average clustering in customer data, using the k-means method and if the results of the calculation are further different, it will continue to repeat until the results of the clustering are the same as the results before it

$$d(1.1) = \sqrt{(14 - 21.33333333)^2 + (500.000 - 483333.33)^2 + (23 + 12.00)^2} = 16666.67$$

$$d(1.2) = \sqrt{(14 - 7.50)^2 + (500.000 - 125000.00)^2 + (23 - 21.00)^2} = 375000.00$$

$$d(2.1) = \sqrt{(20 - 21.33333333)^2 + (350.000 - 483333.33)^2 + (7 - 12.00)^2} = 133333.33$$

$$d(2.2) = \sqrt{(20 - 7.50)^2 + (350.000 - 125000.00)^2 + (7 - 21.00)^2} = 225000.00$$

$$d(3.1) = \sqrt{(11 - 21.33333333)^2 + (50.000 - 483333.33)^2 + (16 - 12.00)^2} = 433333.33$$

$$d(3.2) = \sqrt{(11 - 7.50)^2 + (50.000 - 125000.00)^2 + (7 - 21.00)^2} = 75000.00$$

$$d(4.1) = \sqrt{(4 - 21.33333333)^2 + (200.000 - 483333.33)^2 + (26 - 12.00)^2} = 283333.33$$

$$d(4.2) = \sqrt{(4 - 7.50)^2 + (200.000 - 125000.00)^2 + (26 - 21.00)^2} = 75000.00$$

$$d(5.1) = \sqrt{(30 - 21.33333333)^2 + (600.000 - 483333.33)^2 + (6 - 12.00)^2} = 116666.67$$

$$d(5.2) = \sqrt{(30 - 7.50)^2 + (600.000 - 125000.00)^2 + (6 - 21.00)^2} = 475000.00$$

Calculation of repetition of the average clustering value. In the 3rd picture is the calculation of the 2nd k-means iteration from the previous calculation which is described in the first and second figures.

C1 (f) = (14 + 20 + 30)/3 = 21.33333

C1 (m) = (500.000 + 350.000 + 600.000)/3 = 483333.33

C1 (r) = (23 + 7 + 6)/3 = 12.00

C2 (f) = (4 + 11)/2 = 7.50

C2 (m) = (500.000 + 200.000)/3 = 125.000

C2 (r) = (16 + 26)/3 = 21.00

in the 4th picture is an average calculation to determine the value of the first cluster that has been obtained after going through the calculation of the 2nd k-means which is described in the 3rd image.

In the following table is the result of clastering that has been determined through calculations.

Table 8

Name	F	M	R	C1	C2	Results
edi	14	500000	23	100000.00	300000.00	C1
rafi	20	350000	7	50000.00	150000.00	C1

Table 9

	F	M	R

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

cluster 1	21,333	483333.33	12.00
cluster2	7.50	125000.00	21:00

In table 8 the researcher gets the clastering value data that has been determined in the 7th table. Judging from the 8th data table, researchers found the similarity of values in the 6th table. Here we get the final result of the k-means calculation to determine the nalia of each cluster.

5. Conclusion

From this study it can be concluded as follows

a. From the calculation of selling using the K-means method to get consistent results, so that the criteria data can be used.
b. From the clustering calculations applied to the program, the results are in accordance with the criteria data that has been inputted.
C. From the calculation of the level of concordance between the two clusters it can be concluded that in C1 there are 60% and in C2 there are 40%

6. Reference

[1] A. D. Savitri, f. A. Bachtiar, and n. Y. Setiawan, "segmentasi pelanggan menggunakan metode k-means clustering berdasarkan model rfm pada klinik kecantikan (studi kasus : belle crown malang)," j. Pengemb. Teknol. Inf. Dan ilmu komput. Univ. Brawijaya, vol. 2, no. 9, pp. 2957–2966, 2018.
[2] A. H. Lubis, "model segmentasi pelanggan dengan kernel k-means clustering berbasis customer relationship management," j. Penelit. Tek. Inform., vol. 1, pp. 36–41, 2016.
[3] A. Hadi, "segmentasi pelanggan internet service provider (isp) berbasis pillar k-means," j. Ilm. Teknol. Inf. Asia, vol. 13, no. 2, p. 151, 2019, doi: 10.32815/jitika.v13i2.413.
[4] D. P. Hidayatullah, r. I. Rokhmawati, and a. R. Perdanakusuma, "analisis pemetaan pelanggan potensial menggunakan algoritma k-means dan lrfm model untuk mendukung strategi pengelolaan pelanggan (studi pada maninjau center kota malang)," j. Pengemb. Teknol. Inf. Dan ilmu komput., vol. 2, no. 8, pp. 2406–2415, 2018.
[5] E. Muningsih and a. B. S. I. Yogyakarta, "optimasi jumlah cluster k-means dengan metode elbow untuk pemetaan pelanggan," pros. Semin. Nas. Elinvo, no. September, pp. 105–114, 2017.
[7] T. I. Hermanto, "implementasi algoritma association rule dan k-means sebagai sistem rekomendasi produk pada website penjualan online," pp. 70–73.
[8] T. W. Rusdiawan, a. Alamsyah, m. Manajemen, f. Ekonomi, d. Bisnis, and u. Telkom, "eksplorasi data pelanggan , untuk kontekstual marketing voice over long term evolution pt . Telkomsel menggunakan metode clustering k-means customer data exploration , for contextual marketing voice over long term evolution pt . Telkomsel using k-means cl," vol. 6, no. 1, pp. 537–544, 2019.
[9] Z. G. Prastyawan, m. Ridho bagaskara, and d. Fitriati, "segmentasi pelanggan restoran menggunakan metode clustering simple k-means (studi kasus xyz)," inform. Dan keamanan siber jakarta-indonesia, 2018.