THE STRATEGIES TO IMPROVE THE SALE OF BICYCLES USING K-MEANS METHODS

Adi Mas Afandi

Computer Technique, Sekolah Tinggi Manajemen Informatika Dan Komputer Royal, Indonesia

Corresponding author:
adimasafandi@gmail.com

Keywords:
clustering
data mining
k-means
sales data

ABSTRACT

Data mining is a term used to describe confidential information. A data mining technique is grouping. Clustering is grouping several objects into clusters, where available clusters have a high level. In this study the method used by K-Means is able to partition data in one or more clusters or it can be said to divide the data into several groups. By using the data mining technique, it can be used to classify bicycle sales data. Is the sale of bicycles a sale that is very salable, selling well less. The desired method of research using the K-Means method can produce sales patterns that can help shop owners to increase sales.

INTRODUCTION

Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods [1]. Mining data is a method used to extract predictive information hidden in a database and be able to predict trends, behave so that the company or agency is able to be more proactive and enriching knowledge [2]. The term data mining has its essence as a discipline whose main purpose is to find or mine knowledge from data or information owned [3].

One of the known data mining techniques is clustering. Cluster analysis is useful in summarizing data or a number of variables to become fewer [4]. Clustering is one of data mining methods that is without the referral (unsurvised). The K-Means algorithm is one of the non-hierarchical clustering data methods that attempt to partition data into a number of groups and work on numeric attributes [5].

Some previous authors have applied the technique of K-Means Clustering in his research using the method K-Means Clustering can assist students in determining the course to be taken [6]. Research using the K-Means Clustering algorithm to classify products categorized as the most sold and sold-out goods [7]. With a pilot of the K-Means Clustering method to determine the product's practices and underselling.[8]

Implemented the Fuzzy algorithm C-Means and K-Means to perform student data groupings where the algorithm K-Means with a total of 11 iterations to obtain clusters whereas the Fuzzy algorithm of C-Means requires as much as 35 iterations with complicated calculation processes and the results of less accurate student data clusters [9]. K-Means method in grouping sales of products resulted in two types of data groups

i.e. low sales data and high sales data. Where it affects the buildup of goods in the warehouse [10].

From the several research references already shown above, it was concluded that the K-Means Clustering algorithm could be used in bicycle sales data groupings. Bicycle sales Data will be processed and analyzed using the K-Means method to determine the level of consumer tendency in shopping. Where the influencing factor is the interest and budget that the conssumer has. From the processing of the data, it obtained a sales pattern in the form of sales are very selling and poorly sold. So that the sales pattern provides new knowledge for business owners in the provision of goods and increase profits (turnover)[11].

METHOD

To help prepare for research, a clear framework is needed. This framework is the steps that will be taken in solving problems. The following is the research framework carried out.
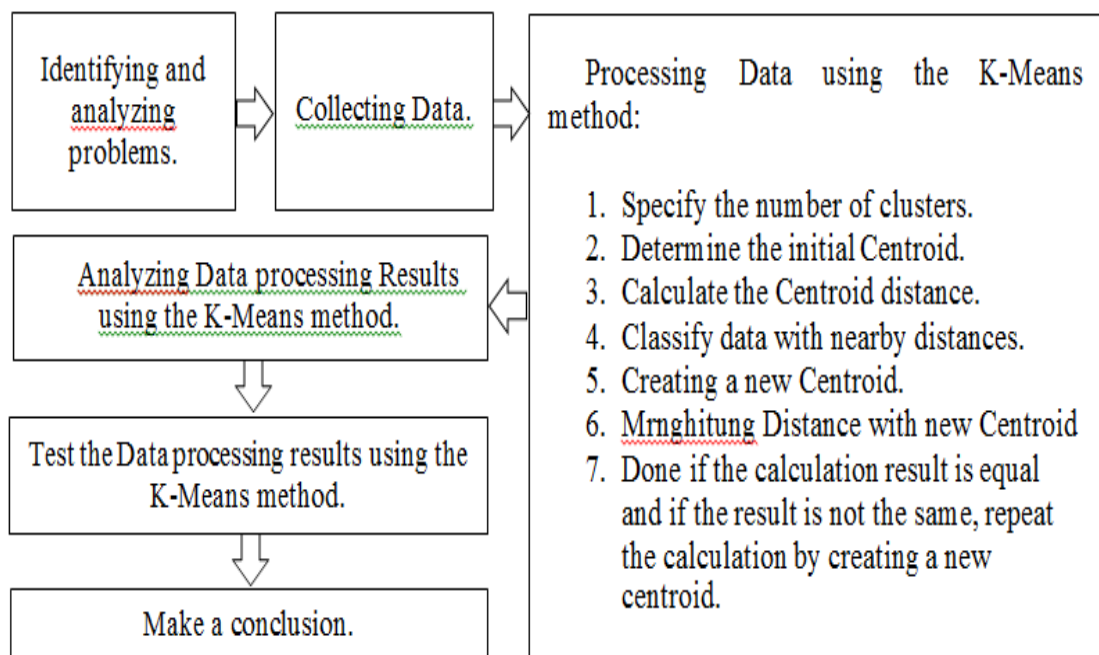

Image 1. Research frameworks

Image 1 is the stage or steps done in drafting a research report. The explanations of the above framework points are as follows:
  A. Identifying and analyzing problems

This stage is a fundamental step in the process of data mining. This phase explains the subject matter of the research by determining the boundaries of problems and scope so that the discussion does not deviate from research problems.

B. Collecting Data

Data obtained is sales data in the year 2018. In this study were taken 15 samples of bicycle sales data. The sales data used in the research can be seen in table 1.

Tabel 1. Sales Data

No	Bicycle Type	Unit	Price (Unit)
1	Family 943 Wheel 3	21	330.000
2	Family 943 H Wheel 3	19	360.000
3	Family 598 Wheel 3	8	300.000
4	Family 988 Wheel 3	8	310.000
5	Family 985 Wheel 3	8	290.000
6	Pakomon Wheel 3	12	280.000
7	Safari Musik Wheel 3	10	310.000
8	PMB Musik Wheel 3	13	300.000
9	Jetstar Nekel Wheel 3	13	310.000
10	Jetstar Eva Beautiful 12"	12	300.000
11	Jetstar Eva Boy 12"	12	285.000
12	Jetstar BMX Graff 12"	6	370.000
13	Jetstar Mini Princes 12"	6	400.000
14	Mini Trendy 12"	6	400.000
15	Mini Oka 12"	2	390.000

Table 1 is a data table for bicycle sales that is used in data processing using the K-Means Clustering method. Amount of data used by 15 bicycle sales data with variable type of bicycle, total sales and price (unit).

C. Processing Data using the K-Means method

The next step is to process the data using the K-Means method. The steps on

the K-Means method are as follows:

1. Specify the number of clusters
   The number of clusters specified to segment this sales data by 2 (two) clusters is very saleable and undersold.

2. Detemine Centroid
   The initial cluster Center (Centroid) is specified randomly or randomly from the bicycle sales data. As for the centroid on the bicycle sales data taken is as follows:
   a. Center C1 (Centroid 1) is taken from the 5th row with total sales of 8 and the price of 290,000.
   b. Center C2 (Centroid 2) is taken on the 13th line with total sales of 6 and the price of 400,000.

Table 2. First Centroid

Centroid	Bicycle Type	Unit	Price (Unit)
C1	Family 985 Roda 3	8	290.000
C2	Jetstar Mini Princes 12"	6	400.000

Table 2 above, center 1 centroid starting point is taken on the 5th row in the Sample data table Yaitutipe Bike: Family 985 wheels 3, total Sales: 8 and Price (unit): 290,000. While the centroid 2 starting point is taken on the 13th row of the Sample data table where the data is the type of bicycle: Jetstar Mini Princes 12 ", Total Sales: 6 and Price (unit): 400,000

3. Calculating the Centroid Distance
   Calculates the distance between the centroid points with the point of each object by using the Euclidian Distance formula. As for the calculation of Centroid manually with the formula below:

$$D(a, b) = \sqrt{((X1 - Y1)^2 + (Xn - Yn)^2)} \qquad (1)$$

Note:

   X = Cluster Center

   Y = Data

4. The next step is to group data based on closest distance between data with Centroid. The calculations are done as much as the sample data and the smallest value of each Cluster is coded "1". The results of the calculations are carried out as follows:

Table 3. Result of calculating spacing and iteration of data grouping 1

Bicycle Type	Unit	Price (Unit)	c1	c2
Family 943 wheel 3	21	330.000	40.000,0021	70.000,0016
Family 943 H wheel 3	19	360.000	70.000,0009	40.000,0021
Family 598 wheel 3	8	300.000	10.000	100.000
Family 988 wheel 3	8	310.000	20.000	90.000
Family 985 wheel 3	8	290.000	0	110.000
Pakomon wheel 3	12	280.000	10.000,0008	120.000,0002
Safari Musik wheel 3	10	310.000	20.000,0001	90.000,0001
PMB Musik wheel 3	13	300.000	10.000,0012	100.000,0002
Jetstar Nekel wheel 3	13	310.000	20.000,0006	90.000,0003
Jetstar Eva Beautiful 12"	12	300.000	10.000,0008	100.000,0002
Jetstar Eva Boy 12"	12	285.000	5.000,0016	115.000,0002
Jetstar BMX Graff 12"	6	370.000	80.000	30.000
Jetstar Mini Princes 12"	6	400.000	110.000	0

Mini Trendy 12"	6	400.000	110.000	0
Mini Oka 12"	2	390.000	100.000,0002	10000,0008

From table 3 above can be known iteration result as follows:
C1 = Cluster 1 (one) calculation Data has 10 members
C2 = Cluster 2 (two) calculated result Data has 5 members

5. The next step is to do 2 iterations, by specifying a new Centroid point by calculating the average value of the existing data in the same Centroid.

From the above calculations, the value of centroid 1 is total 11.7 and the price is 301,500

From the above calculations, then acquired centroid value 2 is the total sales 7.8 and the price of 384,000. The new centroid value obtained from the data processing process can be opened in table 4.

Table 4. New centroid

Centroid	Bicycle Type	Unit	Price (Unit)
C1	Family 985 Wheel 3	11,7	301.500
C2	Jetstar Mini Princes 12"	7,8	384.000

Table 4 above change the new centroid point to perform the 2nd iteration. Where the value C1 is the type of bicycle: Family 985 Wheels 3, total Sales: 11.7 and the Price (unit): 301,500. The value of C2 is the bicycle type: Jetstar Mini Princes 12 ", total Sales: 7.8 and Price (unit): 384,000.

6. Perform calculations by using the new Centroid point in the iteration process to 2.

Table 5. Result of calculation of data spacing and grouping of iteration 2

Bicycle Type	Unit	Price (Unit)	c1	c2
Family 943 Wheel 3	21	330.000	28.500,0015	54.000,0016
Family 943 H Wheel 3	19	360.000	58.500,0005	24.000,0026
Family 598 Wheel 3	8	300.000	1.500,0046	84.000
Family 988 Wheel 3	8	310.000	8.500,0008	74.000
Family 985 Wheel 3	8	290.000	11.500,0006	94.000
Pakomon Wheel 3	12	280.000	21.500	104.000,0001
Safari Musik Wheel 3	10	310.000	8.500,0002	74.000
PMB Musik Wheel 3	13	300.000	1.500,0006	84.000,0002
Jetstar Nekel Wheel 3	13	310.000	8.500,0001	74.000,0002
Jetstar Eva Beautiful 12"	12	300.000	1.500	84.000,0001
Jetstar Eva Boy 12"	12	285.000	16.500	99.000,0001
Jetstar BMX Graff 12"	6	370.000	68.500,0002	14.000,0001

Jetstar Mini Princes 12"	6	400.000	98.500,0002	16.000,0001
Mini Trendy 12"	6	400.000	98.500,0002	16.000,0001
Mini Oka 12"	2	390.000	88.500,0005	6.000,0028

From the iteration table above it can be seen that C1 has 10 members and C2 has 5 members. As the result of iteration 1 and Iteration 2 has the similarity of Cluster position, the process is terminated.

The conclusion of the calculated result begins in Iteration 1 and Iteration 2 is:

a. Cluster 1 (C1) has 10 members, of which C1 is the most saleable bicycle sales category, while

b. Cluster 2 (C2) has 5 members, of which C2 is a category of sales of bicycles undersold.

RESULT AND DISCUSSION

The research process that has been done on the family bike shop range goes well. Research on bicycle sales data with Data Mining technique using K-Means Clustering algorithm gives results according to research objectives. In the process of calculating with the K-Means algorithm is only two iterations. From the calculation process can be seen the result of the number of types of bicycles are best selling and less sales. The result of clusters of bicycle sales data on the bicycle family range store can be seen on table 6:

Table 6. Result of calculation of data spacing and grouping of iteration 2

Bicycle Type	Unit	Price (Unit)	C1	C2	Cluster
Family 943 Wheel 3	21	330.000	28.500,0015	54.000,0016	C1
Family 943 H Wheel 3	19	360.000	58.500,0005	24.000,0026	C2
Family 598 Wheel 3	8	300.000	1.500,0046	84.000	C1
Family 988 Wheel 3	8	310.000	8.500,0008	74.000	C1
Family 985 Wheel 3	8	290.000	11.500,0006	94.000	C1
Pakomon Wheel 3	12	280.000	21.500,0000	104.000,0001	C1
Safari Musik Wheel 3	10	310.000	8.500,0002	74000	C1
PMB Musik Wheel 3	13	300.000	1.500,0006	84.000,0002	C1
Jetstar Nekel Wheel 3	13	310.000	8.500,0001	74.000,0002	C1
Jetstar Eva Beautiful 12"	12	300.000	1.500	84.000,0001	C1
Jetstar Eva Boy 12"	12	285.000	16.500	99.000,0001	C1
Jetstar BMX Graff 12"	6	370.000	68.500,0002	14.000,0001	C2
Jetstar Mini Princes 12"	6	400.000	98.500,0002	16.000,0001	C2
Mini Trendy 12"	6	400.000	98.500,0002	16.000,0001	C2

Mini Oka 12"	2	390.000	88.500,0005	6.000,0028	C2

From the Bicycle Sales data table above can be seen clusters for the type of bike most practice there are 10 members and the type of bicycle less practice there are 5 members. As for the members of the very reliable bicycle type clusters can be seen in table 7:

Table 7. Result of calculation of data spacing and grouping of iteration 2

Bicycle Type	Unit	Price (Unit)	C1	C2
Family 943 Wheel 3	21	330.000	28500,0015	54000,0016
Family 598 Wheel 3	8	300.000	1500,0046	84000
Family 988 Wheel 3	8	310.000	8500,0008	74000
Family 985 Wheel 3	8	290.000	11500,0006	94000
Pakomon Wheel 3	12	280.000	21500	104000,0001
Safari Musik Wheel 3	10	310.000	8500,0002	74000
PMB Musik Wheel 3	13	300.000	1500,0006	84000,0002
Jetstar Nekel Wheel 3	13	310.000	8500,0001	74000,0002
Jetstar Eva Beautiful 12"	12	300.000	1500	84000,0001
Jetstar Eva Boy 12"	12	285.000	16500	99000,0001

As for bicycle type clusters that are less applicable can be seen in table 8:
Table 8. Result of calculation of data spacing and grouping of iteration 2

Bicycle Type	Unit	Price (Unit)	C1	C2
Family 943 H Wheel 3	19	360.000	58.500,0005	24.000,0026
Jetstar BMX Graff 12"	6	370.000	68.500,0002	14.000,0001
Jetstar Mini Princes 12"	6	400.000	98.500,0002	16.000,0001
Mini Trendy 12"	6	400.000	98.500,0002	16.000,0001
Mini Oka 12"	2	390.000	88.500,0005	6.000,0028

From table 7 and table 8, then the family bike range store can build new strategies in bicycle type inventory. So it will not happen again the buildup of the type of bicycle underselling that Mangakibatkan capital turnover becomes obstructed or not good.

CONCLUSION

Based on the research results that have been done by the authors in grouping bicycle sales data. The conclusion to be obtained is the method K-Means can be applied to the bicycle sales data, so this method very helpful business owners in grouping

bicycle sales patterns. Determination of the initial Centroid center greatly affects the outcome of the calculation and also affects the results of different clusters.

BIBLIOGRAPHY

[1]   Fina Nasari, Surya Darma. Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru. 2015

[2]   Y. A. M, F. Dristyan, and A. Syafnur, "Aplikasi Web Usage Mining Menggunakan Metode Association Rule Dengan Algoritma Fp-Growth Untuk Mengetahui Pola Browsing Pengunjung Website," Pros. Semin. Nas. Ris. Inf. Sci., vol. 1, p. 1060, Sep. 2019, doi: 10.30645/senaris.v1i0.117.

[3]   Nelisa, Aulia Fitrul Hadi. Perancangan Aplikasi Data Mining Transaksi Penjualan Untuk Mengetahui Pola Beli Konsumen Pada Toko Singgalang padang Menggunakan Algoritma Apriori Berbasis Web. 2018.

[4]   Koko Handoko. Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering. 2016

[5]   Abdul Haris. Model Segmentasi Pelanggan Dengan Kernel K-Means Clustering Berbasis Customer Relationship Management. 2016.

[6]   Ely Muningsih, Sri Kiswati. Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. 2015

[7]   Indah Permata Sari, Erfanti Fatkhiyah dan Joko Triyono. Data Mining Menggunakan Algoritma K-Means Clustering Untuk Pengelompokan Produk Yang Paling Tidak Laku Terjual Pada Koperasi Mahasiswa Universitas Negeri Yokyakarta (KOPMA UNY). 2017

[8]   Darmi Yulia, Setiawan Agus . Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. 2016.

[9]   Suprapto Totok. Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran.2016

[10]   Butarbutar Nelson, dkk. Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means Dalam Pengelompokan Data SIswa Berdasarkan Prestasi Nilai Akademik Siswa. 2016

[11]   Metisen Benri Melpa, Sari Herlina Latipa. Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila. 2015

[12]   M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," *Jurteksi*, vol. 4, no. 2, hal. 123–132, 2018, doi: 10.33330/jurteksi.v4i2.36.