Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

ISSN 2723-4509 (Online)

LAPTOP SALES LEVEL USING THE K-MEANS CLUSTERING METHOD

Hari Jalsa Marpaung
Computer System, Sekolah Tinggi Menegemen Informatika dan Komputer Royal, Indonesia

***Corresponding author*:**
hari.marpaung@gmail.com

Keywords:
clustering
k-means
laptop
rapid miner

ABSTRACT

Laptops are electronic products that have increased every month. The number of brands or types of product supply can affect people's purchasing power. In this case the shop must be selective in providing products. By utilizing the previous sales data history and using the K-means Clustering algorithm as a cluster or grouping of laptop sales, new knowledge is obtained in determining the brands or types that are in demand, selling well and not selling well.

INTRODUCTION

The amount of competition in the business world, especially in the industry of selling electronic products. Laptops are part of the needs that are used in various fields of work, college assignments, even in daily communication where a laptop is needed. many brands and types of laptops sold in the market, of course, the price varies too, making it difficult for users to determine what suits their needs, to determine the laptop stock and avoid disappointment because of the unavailability of the type of laptop needed by the customer. Inaccurate and careless stock management will cause high and uneconomic saving costs.

In this case, implementing Data Mining and classifying laptop sales needs to use previous sales data, Data Mining is a process that uses static, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases[1], for basic materials to be processed for new knowledge the supply of laptops called KDD (knowledge discovery in a database). KDD is an activity that includes data collection, data selection, data processing, data transformation, data usage, history to find large data regularity[2]. for basic materials to be processed for new knowledge the supply of laptops called KDD (knowledge discovery in a database). KDD is an activity that includes data collection, data selection, data processing, data transformation, data usage, history to find large data regularity[2]. In this study the results of existing sales can be used the K-Means Clustering method and implemented in the RapidMiner application, Clustering is a data method used in Data Mining that works by finding and grouping data with other data[3].

K-Means cluster method is a grouping method that aims to group individuals so that the distance of each individual to the center of the group in one group is a minimum[4], K-Means algorithm is an algorithm that partitioned data into clusters so that similar data are on the same cluster and data that has an inequality is in another cluster[5]. Clustering K-Means is used for grouping. The K-Means method seeks to group data that is in several unique groups, where data in one group has the same

Proceeding **ISSN 2723-4509 (Online)**
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

characteristics with each other and has different characteristics from the data in other groups[6]. Before conducting further analysis, the data standardization is done first, because there are striking unit differences so that it can result in calculations in the cluster analysis being invalid[7]. The stages in this research are by studying the literature, collecting data, analyzing data, analyzing the K-Means method, testing the results, implementing, analyzing the results, then the results can be used as materials or new knowledge base in the supply of laptop stock. Some previous research has been done on the supply of stock products and how customers in determining the choice of purchasing a laptop with several methods.

In the sale of electronic products in previous studies of research on Data Mining Implementation using the Apriori algorithm[8]. As for determining the stock of goods from previous studies the application of the K-Means Clustering Method for online shop products[9].

The purpose of this study is how to determine the supply of laptop stock with the K-Means Clustering method so that the supply of laptop stock has an accuracy and is more economical to avoid saving costs.

METHOD

Data mining methods, it is difficult to choose the best algorithm for a given application, moreover, surely the algorithm has many parameters to be aligned, and performance is vulnerable to a large variety[10]. In this study for the supply of laptop stock, from the collection of existing data, the data can be tested with K-Means Clustering and determine the centroid center point in the initial iteration, after the first iteration process and the results of the first iteration become a new centroid if the new centroid results are the same with the initial centroid the iteration is complete.

1. Knowledge Discovery in Databases

   Knowledge Discovery in Databases (KDD) and the term Data mining are often used interchangeably to explain the process of extracting hidden information in a large database.

   a. Data Selection

      Operational data needs to be done before the information gathering stage in KDD begins.

   b. Pre-processing: Data Mining

      Can be done, it needs to be done the process of cleaning up data duplication, data preprocessing in which there is data cleaning activity from data that experiences inconsistent data, missing data, confusing data to produce a valid knowledge, data such as accessories and other hardware categories are not used due to the many inconsistent data on goods sales from laptop sales bonuses.

   c. Transformation: coding

      This has been chosen, so that the data according to KDD Data Mining coding is a creative process and is very dependent on the type or pattern of information that will be sought in the database.

Proceeding **ISSN 2723-4509 (Online)**
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

d. Data Mining
Interesting information in selected data using certain techniques or methods.

e. Interpretation / Evaluation
Data mining is displayed in a form that is easily understood by interested parties. This stage is part of the KDD process[11].

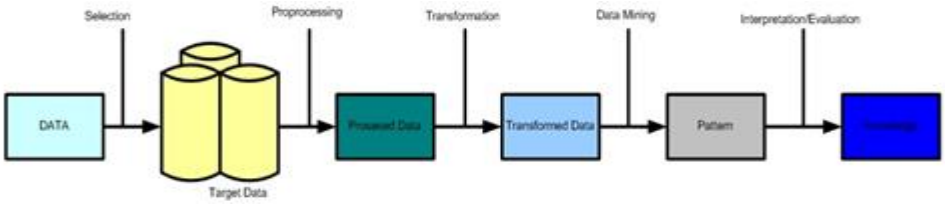

Figure 1. Information Flow in Data Mining.

In Image 1 can be seen the process of KDD, data sources, then select the data to be processed (selection data), then do the data/data cleaning process, data transformation if the data obtained is in the form of data, there is no need to transform data, then calculate with the data mining process, the results of the data mining process become a new source of knowledge.

In this study, the data to be processed is laptop sales data which can be seen in Table 1.

Table 1. Dataset

NO.	CODE	TYPE LAPTOP	SALES	STOCK
1.	NB-AC-E5473-	Notebook Acer E5-473 I3 VGA	13	15
2.	NB-AC-E5475-I3	Notebook Acer E5-475 I3	9	9
3.	NB-AC-E5475-I5	Notebook Acer E5-475 I5	4	5
4.	NB-AC-ES1-132	Notebook Acer ES1-132	18	20
5.	NB-AC-E5475-I3-1T	Notebook Acer E5-475 I3 1TB	15	17
6.	NB-LN-IP310S	Notebook Lenovo IP310S 11.6 DC	24	34

From Table 1 it can be seen that 6 dataset items have gone through data selection, data processing, and then will be processed by the data mining method because the data received is in the form of numbers, then data transformation is no longer done.

RESULT AND DISCUSSION

In this section, the results of the analysis and discussion using the K-means method which will be implemented into RapidMiner aims to test the results of manual counts with the results of Rapidminer studio v 7.3.

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

1. Analyzing the K-Means Algorithm Method

K-Means is one of the most popular clustering algorithms. One of the reasons for the popularity of K-Means is because it is easy and simple when implemented[12][13]. As for the algorithm contained in K-Means are as follows:

a. Determine the Number of Clusters
The number of Clusters in the deception of laptop sales data is given by three Clusters namely Selling, Very Selling, and Not Selling.

b. Determine Cluster Center Points
The centroid center point is taken randomly from the dataset. The centroid points taken can be seen in Table 2.

Table 2 Center / Centroid

NO.	CODE	TYPE LAPTOP	SALES	STOCK
1	NB-AC-E5473-5	Notebook Acer E5-473 I3 VGA	13	15
2	NB-AC-T643-5	Notebook Acer Travelmate 643 I5	9	10
3	NB-LN-IP310S	Notebook Lenovo IP310S 11.6 DC	24	34

Cluster center point determination is done randomly, where the Centroid 1 data is taken from the 1st row in the dataset, the 2nd centroid data is taken from the 5th row and the 3rd Centroid is taken from the 6th row.

c. Calculate Distance from Centroid
The formula used to calculate the distance between Centroid points and points of each object is using Euclidian Distance. Perform the calculation of the distance to the centroid as much as in the dataset, The initial Centroid calculation manually using the formula:

$$D(i,f) = \sqrt{(X_{i1} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + ...+(X_{ni} - X_{nj})^2} \quad (1)$$

1. Calculation of the distance from the 1st data to the center of the Cluster
   DC1 = $\sqrt{(13-13)^2 + (15-15)^2}$ = 0
   DC2 = $\sqrt{(13-9)^2 + (15-10)^2}$ = 6,403124
   DC3 = $\sqrt{(13-24)^2 + (15-34)^2}$ = 21,9545
2. Calculation of the distance from the 2st data to the center of the Cluster
   DC1 = $\sqrt{(9-13)^2 + (10-15)^2}$ = 7,211103
   DC2 = $\sqrt{(9-9)^2 + (10-10)^2}$ = 1
   DC3 = $\sqrt{(9-24)^2 + (10-34)^2}$ = 2915476

Perform the distance calculation from the 1.2 n data ... n to the center of the Cluster until a distance matrix is obtained

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

d. Cluster Grouping
After the calculation results of the dataset with Centroid data are obtained then allocating or placing cluster members based on the smallest value. Cluster member allocations are grouped by giving the code "C1" if the number of the results of the closest distance calculation lies in group C1, giving a code "C2" is given if the smallest number is in the second group and giving the code "C3" is given to the smallest group if it is in the third group.

Table 3. Grouping of Cluster Members by closest distance

NO.	CODE	SOUL	STOCK	DC1	DC2	DC3	CLUSTER
1	NB-AC-E5473-	13	15	0	6,403124	21,9545	C1
2	NB-AC-E5475-I3	9	9	7,211103	1	2915476	C2
3	NB-AC-E5475-I5	4	5	13,45362	7,0771968	35,22783	C2
4	NB-AC-ES1-132	18	20	7,071068	13,45362	15,23155	C1
5	NB-HP-AM0-I5-8	5	5	12,8062	6,403124	34,66987	C2
6	NB-LNIP310S	24	34	21,80625	6,40312	0	C3

e. New Centroid
Each cluster member has been grouped, then iterations will be carried out by generating a new centroid value based on the total number of Cluster members divided by the total Cluster members, with the following equation:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}; i = 1,2,3 \quad ...n \qquad (2)$$

Table 4. Calculations get new Centroid values

CLSTR	SOUL	STOCK
C1	(13+18)/2	(15+20)/2
C2	(9+4+5)/3	(9+5+5)/3
C3	(24)/1	(34)/1

From the above calculation results obtained a new cluster center like Table 5 as follows:

Table 5. First Iteration New Centroid

CLUSTER	C1	C2
1	15.5	17.5
2	6	6,3333
3	24	34

The 2nd iteration then performs another calculation using the new Centroid point Iteration 2. The results of the 2nd iteration calculation can be seen in Table 6

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

Table 6 Calculation of Distance and Classification of Second Iteration Data

NO.	KODE	DC1	DC2	DC3	CLUSTER
1	NB-AC-E5473-	2,25347	8,877442	28,296545	C1
2	NB-AC-E5475-I3	9,411595	2,072364	26,889176	C2
3	NB-AC-E5475-I5	15,55886	4,577318	33,176715	C2
4	NB-AC-ES1-132	5,131094	15,94483	12,676531	C1
5	NB-HP-AM0-I5-8	12,8062	6,403124	34,66987	C2
6	NB-LNIP310S	19,7440[illegible]	30,6828[illegible]	4,6218082	C3

After the calculation of the second iteration is obtained the same Cluster members as in the first iteration, the iteration process is stopped, namely:
1. C1 consists of 2 cluster members as Clusters in the best-selling category.
2. C2 consists of 3 cluster members as Clusters in the best-selling category.
3. C3 consists of 1 cluster member as a cluster with the category of not selling.

Based on the results of the grouping analysis, it was concluded that the supply of laptops in the Cluster is less in demand can be minimized, while in the Cluster is very in demand the supply of laptop stock can be increased.

RapidMiner Implementation

1 Following is data processing using K-Means on RapidMiner:

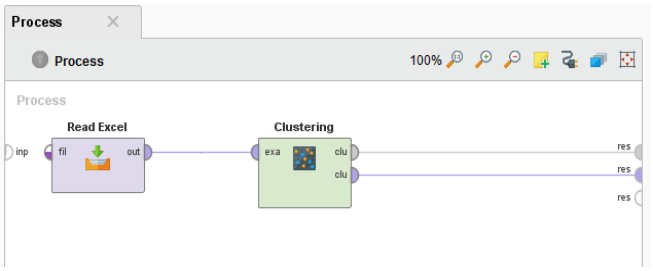

Figure 2. Display for the K-Means Process

Image 2. is a step in the K-Means process on Rapidminer whose process results can be seen in Image 3. Cluster Model.

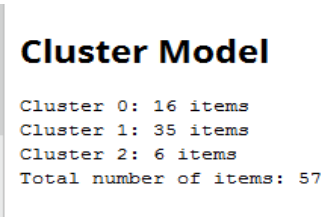

Figure 3. Display Text View

By using K-Means modeling as shown in Image 2 above, with this the number of Clusters is 3 units according to the number of Clusters 0 or referred to as the first

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 201 - 208
DOI: https://doi.org/10.33330/icossit.v1i1.722
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

ISSN 2723-4509 (Online)

Clusters there are 16, Clusters 1 or meant by the second Clusters there are 35, and Clusters 2 or referred to Clusters 3 there are 6 with a total of 57 items, the Cluster model can be seen the results of the Cluster. To see the results of a visual cluster of charts by selecting the chart menu so that it can Figure 4 Display Chart Histogram Style.

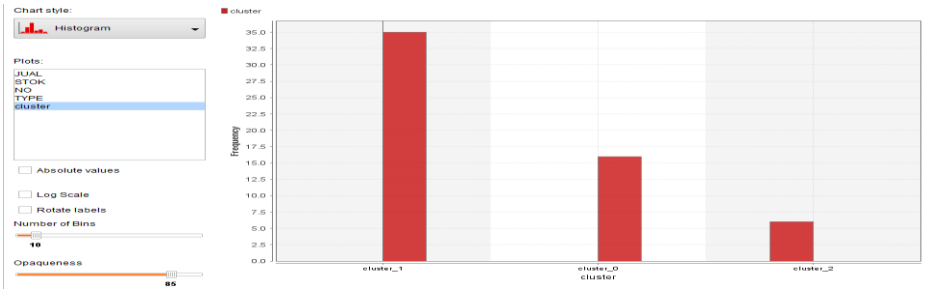

Figure 4. Display Chart Histogram Style

In Image 4 you can see the histogram chart style display of 3 Clusters and display from the highest order, Cluster 2 is a very best-selling Cluster found in the highest red display, Cluster 1 which can be seen in the red display with the best-selling category and the least Cluster 3 namely less desirable.

CONCLUSION

In providing laptop stock at an ordinary store, it still uses the manual method, which is the calculation of all indicators or based on distribution, the processing of the data indicator data also still uses basic static techniques, this results in less than optimal output and has consistent data problems. To determine the consistent data of laptop stock data mining techniques can be used that can dig up hidden information from the data set that has been obtained, besides extracting data that is connected with other data can also be done by the well-known Data Mining technique is the K-Means method.

BIBLIOGRAPHY

[1]. Nelisa dan Hadi A.F, "Perencangan Aplikasi *Data Mining* Transaksi Penjualan Untuk Mengetahui Pola Beli Konsumen Pada Toko Singgalang Padang Menggunakan Algoritma Apriori Berbasis Web." Majalah Ilmiah, vol. 25, No.1, 2018.

[2]. Putra P.P., Chan A.S, "Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Mengunakan Algoritma C 4.5 Sebagai Bagian Dari Sistem Pengambilan Keputusan(Studi Kasus di Saudara Motor."Jurnal Inovtek Polbeng- Seri Informatika, Vol. 3 No.1, 2018.

[3]. Wardhani A.K, "Implementasi Algoritma *K-Means* Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan." Jurnal Transformatika, Vol. 14, No1, 2016).

[4]. Rosadi R., Akmal.,Hidayat A., Kharismawan,"Aplikasi *K-Means Clustering* Untuk Mengelompokan Data Kinerja Akademik Mahasiswa," Senter, 26- 27 November 2016 ISBN: 978-602-60581-0-2.
[5]. Rohmawati N.W., Defiyanti S., Jajuli M., "Implementasi Algoritma *K-Means* Dalam Pengklasteran Mahasiswa Pelamar Mahasiswa." Jurnal Ilmiah Teknologi Informasi Terapan. Vol. 1, No 2, 2015,
[6]. Vhallah I., Sumijan., Santony J., "Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means." Jurnal Resti. DOI : https://doi.org/10.29207/resti.v2i2.308. Vol. 2 No. 2 (2018) 572-577
[7]. Silvi R, "Analisa *Cluster* dengan Data outlier menggunakan Centroid Linkage dan K-means *Clustering* untuk Pengelompokan Indikator *Hiv/Aids* di Indonesia," Jurnal Matematika vol. 4 n0.1, 2018.
[8]. Lie X., Liu H., "*Greedy Optimization for K-Means-Based Consensus Clustering." Tsinghua Science And Technology*. DOI : 10.26599 / TST.2018.9010063. Vol. 23, Number 2, April 2018.
[9]. Syafiti N.A., Sutardi., Dewi A.P,. "Penerapan Metode Weighted Produk Dalam Sistem Pendukung Keputusan Pemilihan Laptop Berbasis Web." Vol. 2, No. 1, 2016
[10].Muningsih E., Kiswati S., "Penerapan Metode *K-Means Clustering* Untuk Produk Online Shop Dalam Penentuan Stok Barang." Jurnal Bianglala Informatika. Vol.3 No 1 Maret 2015
[11].Kurniawansyah A. S., "Implementasi Metode *Artificial Neural Network* Dalam Memprediksi Hasil Ujian Kompetensi Kebidanan (Studi Kasus: Akademi Kebidanan Dehasen Bengkulu," Jurnal Pseudocode, vol. 5, No 1, 2018.
[12].Pratama I.P.A., Harjoko A., "Penerapan Algoritma Invasive Weed Optimnization untuk penentuan titik pusat Klaster Pada K-Means." IJCCS, Vol.9 No. January 2015.