

Grouping Building Material Sales for Management Optimization with the K-Means Algorithm

Nofet Putri*, Abulwafa Muhammad, Raja Ayu Mahessya

Universitas Putra Indonesia YPTK Padang Jl. Raya Lubuk Begalung Padang, Sumatera Barat - 25221 Indonesi

* nofetputri@gmail.com

Abstract

Sales transactions at the Padi Building Store are recorded in a daily notebook provided by the store, so it cannot be immediately known what types of products are selling well or not selling well. To overcome this problem, sales data analysis was carried out using the K-Means algorithm. The K-Means algorithm is an algorithm that requires as many input parameters and divides a set of objects into clusters so that the level of similarity between members in one cluster is high, while the level of similarity with members in other clusters is very low. This research aims to find out what types of products are selling well and which are not selling well. The data used in this analysis were 100 data samples with 2 clusters formed, namely goods that were not selling well (Cluster 1) and goods that were selling well (Cluster 2). With the analysis of the K-Means algorithm, it produces 4 items of data that are not selling well (Cluster 1), while the data of items that are selling is 96 items data (Cluster 2). With the results of the K-Means Algorithm analysis, a sales application was developed for the Padi Building Store which was managed using the K-Means Algorithm analysis. The application makes it easier for sellers to manage data to form a sales process for selling and unselling goods.

Keywords: Clustering, K-Means Algorithm, Sales, Products, Objects.

JCSITech is licensed under a Creative Commons 4.0 International License.

1. Introduction

Currently, the sales process is experiencing very rapid development along with technological developments. Entrepreneurs use various ways and methods to increase sales of goods, including sales of building materials. For this reason, an analysis is needed in the sales business [1] [2]. Sales is an important part, sales can be a factor in growing a business to become bigger. Technology can be used and utilized efficiently, various data processing can be done by computerization, starting from calculations, data storage, to making reports that can produce accurate information, whether needed by individuals or business entities [3][4] .

Padi building shop is a materials shop where we can buy all building needs, from bricks, cement, iron, concrete, plywood, nails, and so on. All the materials or materials needed to build a house or carry out renovations can usually be found here. In the padi building shop there are still processes that are carried out manually so that errors often occur in recording existing data and there is also a lack of time efficiency required. As is the case in the Padi Building shop.

Currently the Padi Building store is still processing its sales data manually. Apart from that, storing the results of feedback or messages with customers is also needed, to help the shop in making decisions in product marketing and customer service in the future.

One way to overcome this problem is that a large transaction historical data processing process is needed, namely data mining techniques which will create a Web application created using the K-means algorithm method to process incoming transaction data so that it will make things easier for Padi building shop owners.

The K-Means algorithm is the simplest clustering algorithm compared to other clustering algorithms [5]. This algorithm has the advantages of being easy to implement and run, relatively fast, easy to adapt, and most widely practiced in data mining tasks [6].

The K-Means algorithm is an algorithm that requires as many input parameters and divides a set of objects into clusters so that the level of similarity between members in one cluster is high, while the level of similarity with members in other clusters is very low [7]. The similarity of members to a cluster is measured by the closeness of the object to the mean value in the cluster or what is called the cluster centroid [8].

The K-Means algorithm is a Clustering algorithm that is included in the Unsupervised learning group which is used to group data into several groups with a partition system [9]. K-means uses the Euclidean distance measure and iteratively determines each record from the original cluster. The procedure begins by selecting k initial records as the cluster center (initial seed) and determining each record that is closest to the cluster. New records are added to the cluster, cluster centers are recalculated to reflect their new members. This iterative

procedure is repeated until it converges and migrating records by cluster no longer filters the solution[10]

Previous research on the application of the k-means algorithm in determining the level of spread of the Covid-19 pandemic in Indonesia obtained results. This grouping will produce central points for the spread of COVID-19 cases. One of the Clustering algorithms is K-Means which uses several groups to place some data with a partition system. Data without class labels is accepted by this algorithm. Due to the global pandemic that has occurred, many parties are trying to take part in overcoming it. This research was conducted for application in the spread of the Covid-19 pandemic in Indonesia. In this research, the K-Means algorithm was used to determine the level of spread of Covid-19 in regions in Indonesia [11].

K-Means is a technique in the field of data mining that is used to group data in an unsupervised manner using a partition approach, where data is grouped based on their similarities without any guidance or supervision in the modeling process [12] .

This research aims to help increase sales of the Pedi Building Store through management optimization, which can differentiate between what is selling well and what is not selling well. To help Pedi Building stores know what stock will be produced , so they don't run out of stock. So researchers used the K-means algorithm method to make it easier for shop owners to group items that were selling well and not selling well.

2. Research methodology

The research framework is the sequence that exists in a study. Where the aim of this framework is to create a design so as not to cross the boundaries of the subject of discussion and to make it easier to understand the content of the research. The sequence of these steps is created into a framework that will make it easier to complete this research. The framework for the research is as follows:

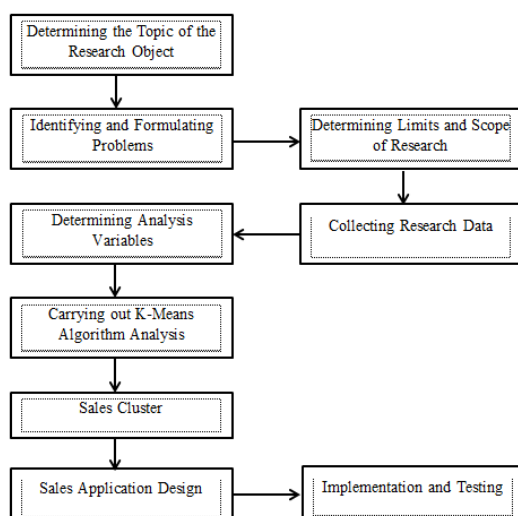


Figure 1. Research Framework

Research stages are steps that must be taken to make conducting research easier. The stages of this research are as follows:

2.1 Determining the Topic and Research Object

The first thing to do is determine the topic of the problem that will be raised and determine the research object. After observing, the researcher determined the topic of this research, namely: sales grouping using the K -means algorithm method, the Pedi building shop was chosen as the research object for the final project.

2.2 Identifying and Formulating Problems.

Identifying and formulating the problem along with the scope or boundaries of the problem in this research, what is obtained from this stage is in the form of a problem formulation from the final assignment.

2.3 Determining Limits and Scope of Research

Determining the boundaries and scope of the research aims to clarify the essay that is the target of the research to be carried out and so that the research is focused and does not deviate from the background that the author described previously.

2.4 Collect research data

In collecting data, the author obtained data from various sources. The data in this research was obtained from journals and obtained from other references. Data collection was carried out by applying the library study method.

2.5 Determining Analysis Variables

Data analysis begins by analyzing all system requirements for the proposed website system, so that findings are obtained based on the focus or problem to be answered. In the analysis and design stage of this system, the UML method will be used and processing data regarding grouping and sales. Before processing data, there are several phases to analyze and design the system

2.6 Performing K-Means Algorithm Analysis

K-Means is included in the data mining partitioning clustering method, that is, every data must enter a cluster and it is possible for every data included in a particular cluster at one stage of the process, at the next stage to move to another cluster. K-Means separates data into K regions and is famous for its ability to classify large data and outliers very quickly.

2.7 Sales Cluster

In this research, sales data can be grouped using the K-Means algorithm, there are 109 data samples at the Pedi Building Store. In this research, the initial center point is determined randomly. The existing data will be grouped into two clusters, namely the first cluster (C1) of building materials that are not selling well and the second cluster (C2) of building materials that are selling well.

2.8 Sales Application Design

In this design, there are two types of design stages that will be carried out, including the following: Model design includes use case diagrams, class diagrams, sequence diagrams, collaboration, etc. Interface design, namely interface design or interface design, is a form of temporary appearance design of system design. This design was created to be able to provide information or explanations about the appearance faced by actors when using the system, so that it can make it easier to implement applications and make it easier to develop applications.

2.9 Implementation and testing

The implementation stage is carried out to find out the computer specifications for running the program and what software is needed. At this implementation stage the researcher used the PHP programming language and MySQL database. After the implementation stage was carried out, the next stage was carrying out the testing process, where the testing process was carried out on the program that had been designed to find out whether the program was running correctly and in accordance with the design carried out.

3. Results and Discussion

3.1 Analysis and Design Stages

The system analysis process in developing an information system is a system for grouping sales of building materials for optimizing pedi building shop management with the K-Means algorithm. Process Analysis.

3.2 Data analysis

Data analysis is everything that is obtained through observation or from a view. At this stage, data processing is carried out in accordance with the provisions of the formula that has been determined. The data for this research came from the Pedi Building Store, where the data used was 20 sales data. The expected output is to produce 2 clusters, namely (C1) sales data that is not selling well, and (C2) sales data that is selling well. The variables or attributes used in grouping sales data consist of price and Sold, as seen in Table 3.1 below:

No	Building material	Sold	Price
1	Iron Armature 12 ksty thick	125	140000
2	Armature iron 10 ksty thick	125	100000
3	Iron Armature 8 ksty thick	125	70000
4	Armature iron 6 ksty thick	125	40000
5	Ceramics 20x25/box	40	73000
6	Ceramics 25x25/box	40	68000
7	Ceramics 25x40/box	40	73000
8	Ceramic 40x40 plain/boxed	40	70000

No	Building material	Sold	Price
9	Ceramic 40x40 motifs/box	40	95000
10	Ceramic 50x50 plain/box	40	95000
11	Ceramic 50x50 motifs/box	40	130000
12	Concrete Wire/kg	50	25000
13	Granite	135	150000
14	Ceramic flour/wrapping	10	20000
15	Ceramic bon (10bh/m)/ box	10	110000
16	Lestelo (4bh/m)/sheet	10	13000
17	Diamond Elephant Tin Roof	10	1300000
18	Zinc Roof SB	10	1380000
19	HN Zinc Roof	10	1300000
20	Roof Tiles/sheets	50	72000

3.3 K-Means Algorithm Calculation

Step 1 Determine Clusters

Determine the number of *clusters* used in building material sales data, 2 clusters including selling and not selling well based on sales data in 1 year.

Step 2 Determine the Centroid

Determination of the initial center of the cluster (centroid) is determined randomly taken from existing data, for initial determination of clusters the following assumptions are made:

Cluster center 1: 29 and 216360

Cluster center 2: 21 and 138320

Step 3 Calculate Distance from Centroid

Calculate the distance between the centroid point and the point of each object using Euclidian Distance. As for calculating the initial centroid manually, with the following formula:

$$D_e = \sqrt{((x_i - s_s)^2 + (y_i - t_s)^2)}$$

Where D_e is the Euclidean Distance, i is the number of objects, (x, y) are the object coordinates (s, t) .

Calculate the *Euclidian Distance* from all data to each first center point:

$$D_{11} = \sqrt{(125 - 29)^2 + (140000 - 216360)^2} = 76360.06035$$

$$D_{12} = \sqrt{(125 - 29)^2 + (100000 - 216360)^2} = 116360.03960$$

$$D_{13} = \sqrt{(125 - 29)^2 + (70000 - 216360)^2} = 146360.03148$$

$$D_{14} = \sqrt{(40 - 29)^2 + (40000 - 216360)^2} = 176360.026123$$

$$D_{15} = \sqrt{(40 - 29)^2 + (73000 - 216360)^2} \\ = 143360.00042$$

$$D_{16} = \sqrt{(40 - 29)^2 + (68000 - 216360)^2} \\ = 148360.00041$$

$$D_{17} = \sqrt{(40 - 29)^2 + (73000 - 216360)^2} \\ = 143360.00042$$

$$D_{18} = \sqrt{(40 - 29)^2 + (70000 - 216360)^2} \\ = 121360.00050$$

$$D_{19} = \sqrt{(40 - 29)^2 + (95000 - 216360)^2} \\ = 86360.00070$$

$$D_{110} = \sqrt{(40 - 29)^2 + (95000 - 216360)^2} \\ = 86360.00070$$

$$D_{111} = \sqrt{(40 - 29)^2 + (130000 - 216360)^2} \\ = 191360.00115$$

$$D_{112} \sqrt{(50 - 29)^2 + (25000 - 216360)^2} \\ = 66360.08466$$

$$D_{113} \sqrt{(135 - 29)^2 + (150000 - 216360)^2} \\ = 196360.00092$$

$$D_{114} \sqrt{(10 - 29)^2 + (20000 - 216360)^2} \\ = 106360.00170$$

$$D_{115} \sqrt{(10 - 29)^2 + (110000 - 216360)^2} \\ = 106360.00170$$

$$D_{116} \sqrt{(10 - 29)^2 + (13000 - 216360)^2} \\ = 203360.00089$$

$$D_{117} \sqrt{(10 - 29)^2 + (130000 - 216360)^2} \\ = 1083640.00017$$

$$D_{118} \sqrt{(10 - 29)^2 + (1380000 - 216360)^2} \\ = 1163640.00016$$

$$D_{119} \sqrt{(10 - 29)^2 + (130000 - 216360)^2} \\ = 1083640.00017$$

$$D_{120} \sqrt{(10 - 29)^2 + (72000 - 216360)^2} \\ = 10144360.00153$$

In the same way calculate the distance of each point to the second center point. From the results of the *Euclidian Distance calculation*, the iteration results obtained are as shown in Table 2 below:

Table 2 . Distance Calculation and Data Grouping 1st Iteration

Code	D1	D2	C1	C2
M1	76360.06035	1683.21597		OK
M2	116360.03960	38320.14113		OK
M3	146360.03148	68320.07916		OK
M4	176360.02613	98320.05500		OK
M5	143360.00042	65320.00276		OK
M6	148360.00041	70320.00257		OK
M7	143360.00042	65320.00276		OK
M8	146360.00041	68320.00264		OK
M9	121360.00050	43320.00417		OK
M10	121360.00050	43320.00417		OK
M11	86360.00070	8320.02169		OK
M12	191360.00115	113320.00371		OK
M13	66360.08466	11680.55632		OK
M14	196360.00092	118320.00051		OK
M15	106360.00170	28320.00214		OK
M16	203360.00089	125320.00048		OK
M17	1083640.00017	1161680.00005		OK
M18	1163640.00016	1241680.00005	OK	
M19	1083640.00017	1161680.00005	OK	
M20	144360.00153	66320.00634	OK	

C1 has 3 data which means that the first group is the category of building materials that are not selling well, namely {M18 , M19, M20 } : Member C1

C2 has 17 data which means that the second group is the category of building materials that are sold, namely {M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15, M16, M17} : C2 Member

From the results of the calculations above, a new cluster center is obtained. Determine the position of the new centroid (Ck) by calculating the average value of the data at the same centroid.

$$Ck = \left(\frac{1}{nk}\right) \sum di$$

3.4 System Testing

System testing and implementation aims to see whether the designed system is in accordance with what is desired or not . After testing and implementation, the quality of a system will be visible [15]. The program display is a sub-chapter that explains the process in the program, both the program input process and the output execution of the program being run. The following is the program display that has been built, including an

admin login page which is the initial display when the website is accessed. On this page display there are two forms for username/email and password. As in Figure 4 below:

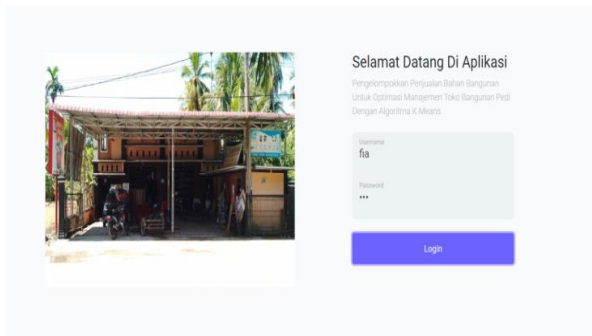


Figure 2. Admin Login Page

Then after successful login, the admin home page will appear as in Figure 3

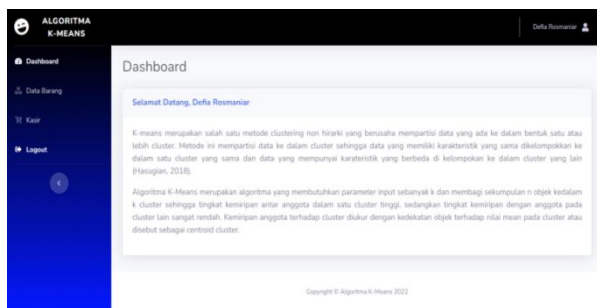


Figure 3. Home page Admin

displays several menus, namely dashboard , item data , centroid , k-means results, transactions, k-means reports, sales reports, user, logout as in Figure 4 .

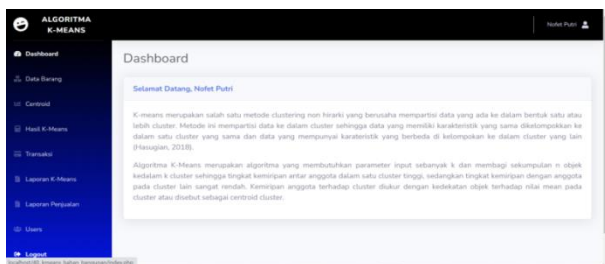


Figure 4 . Home page admin k-means

On this page there is a centroid data page which displays added centroid data. As in Figure 5 .

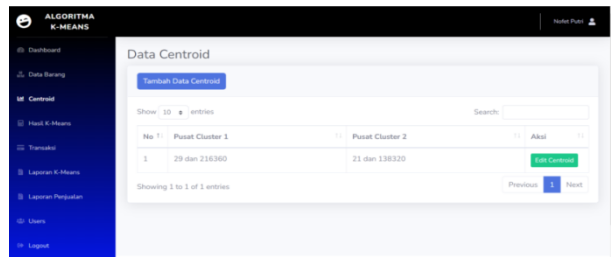


Figure 5. Centroid Page

The k-means results page displays the k-means algorithm calculations, namely k-means calculations and k-means results, as in Figure 6 .

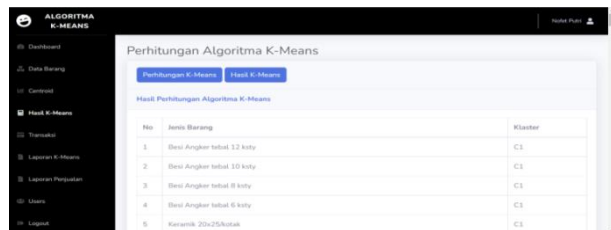


Figure 6 . K-Means Results Page

The transaction report page is printed from the home menu at the cashier. As in Figure 7 :

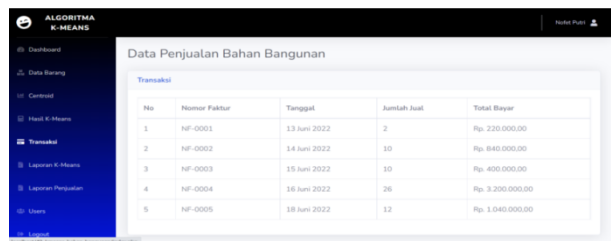


Figure 7. Transaction Page

On the sales report page displays print the buyer's report. Like Figure 8 _

TOKO BANGUNAN PADI				
Laporan Pembelian Pertama				
No	Tanggal Penjualan	No Faktur	Jumlah	Pembelian
1	13 Juni 2022	NF-0001	2	Rp. 220.000,00
2	14 Juni 2022	NF-0002	10	Rp. 840.000,00
3	15 Juni 2022	NF-0003	10	Rp. 400.000,00
4	16 Juni 2022	NF-0004	26	Rp. 3.200.000,00
5	18 Juni 2022	NF-0005	12	Rp. 1.040.000,00
Total Pembelian :				Rp. 5.700.000,00

Figure 8. Sales Report Page

4. Conclusion

Based on research conducted on grouping sales of building materials for management optimization using the K-Means algorithm, the results obtained were grouping goods at the Pedi Building Store using the K-Means Algorithm, where there was a total of 100 items of data, then divided into two clusters where there were 4 items. which is included in the category of building materials that are not selling well and 96 items are included in the category of building materials that are selling well. To facilitate the sales performance of building materials based on the results of the K-Means Algorithm, a sales application was built using the Web-based K-Means Algorithm.

References

- [1] Hadi, F., & Diana, Y. (2020). Building Material Sales Clustering Using the K-Means Algorithm. JOISIE (Journal Inf. Syst. Informatics Eng, 4(1), 22.
- [2] Sudaryono, S., Rahwanto, E., & Komala, R. (2020). E-Commerce Boosts the Indonesian Economy, During the Covid 19 Pandemic as a Modern Entrepreneur and Its Influence on Offline Business. Journal of Management and Business, 2(02), 110-124. <https://doi.org/10.47080/10.47080/vol1no02/jumanis>
- [3] Siregar, MH (2018). Data Mining Clustering Sales of Building Equipment Using the K-Means Method (Case Study at the Adi Bangunan Store). Journal of Technology and Open Source, 1(2), 83-91. <https://doi.org/10.36378/jtos.v1i2.24>
- [4] Rizal, MA, & Misriati, T. (2018). Design of a Web-Based Clothing Sales Information System at Uj Outlet Stores. Sisfokom Journal (Information and Computer Systems), 7(1), 9-15. <https://doi.org/10.32736/sisfokom.v7i1.281>
- [5] Ramadhan, A., Efendi, Z., & Mustakim, M. (2017). Comparison of K-Means and Fuzzy C-Means for User Knowledge Modeling Data Grouping. In National Seminar on Communication Information Technology and Industry (pp. 219-226).
- [6] Fizzy Andriani, SE, Surianto, F., Fil, S., Kesos, M., Mbon, BHP, Jenderal, S., & Indonesia, KPR Study of Monitoring & Evaluation Models for Public Communication Programs.
- [7] Rismayadi, AA, Fatonah, NN, & Junianto, E. (2021). K-Means Clustering Algorithm to Determine Marketing Strategy at CV. Construction Integrates. Responsive Journal: Science and Informatics Research, 3(1), 30-36. <https://doi.org/10.51977/jti.v3i1.393>
- [8] Yunita, F. (2018). Application of Data Mining Using the K-Means Clustering Algorithm in New Student Admissions. Systemization: Journal of Information Systems, 7(3), 238-249.
- [9] Nabila, Z., Isnain, AR, Permata, P., & Abidin, Z. (2021). Data mining analysis for clustering Covid-19 cases in Lampung Province using the k-means algorithm. Journal of Information Technology and Systems, 2(2), 100-108. <https://doi.org/10.33365/jtsi.v2i2.868>
- [10] Aditya, A., Jovian, I., & Sari, BN (2020). Implementation of K-Means Clustering for the National Junior High School Examination in Indonesia 2018/2019. Budidarma Media Informatics Journal, 4(1), 51-58. <http://dx.doi.org/10.30865/mib.v4i1.1784>
- [11] Dwitri, N., Tampubolon, JA, Prayoga, S., Zer, FIR, & Hartama, D. (2020). Application of the K-Means algorithm in determining the level of spread of the COVID-19 pandemic in Indonesia. (JurTI) Journal of Information Technology, 4(1), 128-132. <https://doi.org/10.36294/jurti.v4i1.1266>
- [12] Indriyani, F., & Irfiani, E. (2019). Clustering Sales Data at Outdoor Equipment Stores Using the K-Means Method. JUITA: Journal of Informatics, 7(2), 109-113. <https://doi.org/10.30595/juita.v7i2.5529>