@is The Best :
Accounting Information Systems and Information Technology Business Enterprise
Volume 8, Issue 2 (2023) Pages 121-135
ISSN: 2252-9853 (Print) | ISSN: 2656-808X (Online)
https://ojs.unikom.ac.id/index.php/aisthebest/index

Accredited Ranking 4th, SK Number: 105/E/KPT/2022
DOI: https://doi.org/10.34010/aisthebest.v8i2.11597

Enhancing Promotional Strategy Mapping Using the K-Means Clustering Algorithm to Raise Sales

Ahmad Fathurrozi[1], Tri Ginanjar Laksana[2]
[1, 2] Bhayangkara Jakarta Raya University, Harsono RM Street No 67, Ragunan, Pasar Minggu, Jakarta, 17121, Indonesia
Email: fathur@dsn.ubharajaya.ac.id*

ABSTRACT
To enhance sales, organizations must improve the alignment of their promotional tactics. Enterprises have the ability to promote their goods in locations where there is demand for them. Facilitating the delivery of the goods would enhance the ease with which clients can carry out their purchases and sales transactions. A corporation's ability to strategically allocate its goods enables it to expand its operations. Prospective clients have a greater array of choices at their disposal than the total number of enterprises operating within the same sector. This is accomplished by using a diverse range of promotional media to enhance the sales of products and services. Optimizing promotional strategies is the first and critical stage in presenting items to clients, as it directly impacts the benefits that the firm will get. So far, the sales process has not been affected by the promotional method. The objective of this research was to use the K-Means Clustering algorithm in a data mining procedure to optimize the categorization of customer data, CRISP-DM is used for the purpose of comprehending and preparing data, constructing models, evaluating them, and deploying them. The CRISP-DM method is employed specifically for the construction of clusters. A non-hierarchical clustering technique called K-Means divides data into many groups according on how similar they are. The program facilitates the determination of appropriate location mapping for promotional purposes. The study results may serve as a foundation for decision-making in order to maximize promotional techniques, using the generated clusters.

Keywords: Algorithms, CRISP-DM, Data Mining, Evaluation, K-Means Clustering

Introduction

Promotional strategy optimization might be helpful if it is backed by organized planning[1],[2]. The launch of products or services to the public is marked by promotional methods, which are crucial since they tie into the advantages that the business will experience[3],[4]. Planning that is well-organized will help promotional initiatives work to their fullest potential. Every aspect of life is seeing a rapid development in information technology[5]. Sophisticated information technology generates a great deal of data, originating from industries, economics, science, technology, and other spheres of life. Information technology used in the business sector can generate a wealth of data about customers and the transactions that are made [6].

Clustering is an unsupervised data mining technique that does not rely on a teacher or training and does not need an output aim. Data mining employs two distinct clustering techniques for data grouping[14],[15],[16]. Delivery services are essential to e-commerce because they make it simple for individuals and businesses to send goods[7]. In addition, even when sales and purchases do not immediately coincide, it can facilitate shopping. That is why goods delivery services are increasingly needed[8],[9]. This can be seen in many companies expanding their wings in goods delivery[10]. Private companies and Pos Indonesia, a state-owned company, are also improving and implementing several conveniences in goods delivery services. This is because of the high level of competition that occurs between expedition companies in Indonesia. Clustering

is a technique used to look for and group data that have similar traits or similarities with other data points [11],[12],[13].

Clustering, also known as grouping, is a fundamental technique in data analysis and machine learning. Its purpose is to detect patterns and correlations within data. The K-Means technique is a widely used clustering method. This technique operates by partitioning data into many groups or clusters, determined by the closeness of the data points to the centroid of a certain cluster. This paper examines the concept and use of the K-Means algorithm in the domain of data analysis and clustering. The K-Means technique consists of iterative steps with the goal of reducing the variance or distance between the data points and their cluster centroids. This methodology produces clusters that precisely depict the patterns and structures inherent in the data. This research also investigates the advantages, limitations, and factors that influence the results of the K-Means algorithm. Moreover, this paper evaluates the use of the K-Means algorithm in many domains, including image processing, sentiment analysis, and large-scale data clustering. The experimental results illustrate the effectiveness and capacity to handle datasets of varying sizes and degrees of complexity of the K-Means approach. This research explores methods to improve the effectiveness of the K-Means algorithm within the context of sustainability. This includes investigating the use of intelligent initialization or incorporating other clustering algorithms. To summarise, the K-Means method is a potent instrument in data analysis and clustering. However, it is crucial to carefully pick parameters and thoroughly comprehend the available data in order to get the most favourable outcomes. This study offers profound understanding of the most recent applications and advancements in K-Means clustering.

The goal of the research is to analyze customer data in order to identify customers who have the potential to be credit customers or who are interested in credit offers made by the bank, according to research past in his journal article "Data Mining Design for Analysis of Criteria for Potential Credit Customers and Their Benefits for Banking Customer Relationship Management." When analyzing customer data, marketing targets are typically created by grouping or classifying all of the consumers who have paid their credit installments. High marketing operating costs are thus a result of this approach. Since Bank X is located in Singaraja City, Buleleng Regency, Bali province, the problem in this paper is how to design a data mining application that predicts the criteria for credit customers who have the potential to borrow (credit) or who might be interested in credit offers made by the funds section of Bank X[17].

To find the requirements for possible clients who would be interested in a bank's credit offer, a data mining design idea was created specifically for the banking sector. To implement the research concept, Bank X's current system was enhanced and a new, data-mining system was installed in its place[18]. Several conclusions can be drawn from the system analysis and discussion that have been presented. Specifically, data mining is a tool that organizations and entities can use for operational purposes as well as to help make appropriate decisions when determining marketing strategies for both products and services. Customers will be able to access banking, particularly with regard to credit issues. Forming client groups (also known as customer profiling) and assisting the marketing and customer service departments are two advantages of data mining for the banking customer relationship management concept. It is possible to suggest that every institution or organization should employ data mining to make database analysis easier so that management can make the most use of the data they have. Research ideas for the future might cover the significance of data mining for sectors besides banking. We can comprehend data mining's role in business as a whole by talking about data mining for sectors other than banking.

In their journal article "Determining Promotional Strategies for New Promotion Admissions," Tria Titiani Chasanah and Widiyono report that the K-Means Clustering Algorithm indicates that STMIK Widya Pratama Pekalongan's new promotionadmissions are an annual

event. Every installation faces a number of challenges and issues. Since three years ago, the decrease in potential new promotionapplicants has been the issue. A number of assessments were conducted in an effort to boost the quantity of fresh promotionapplications[19]. The need for more interest in particular study programs is one of the other issues, as seen by the low number of candidates for the study program in computer accounting. In this research, the K-Means method will be used to analyse the new promotion registration data. The aim is to extract information by categorising the data into specified groups, using the registration data from the previous year as a reference. The promotional strategies for the new STMIK Widya Pratama Pekalongan promotionadmissions will be based on the clustering outcomes. The data will be divided into three segments with unique variable values by the use of K-Means clustering. The majority of students in the Informatics Engineering Study Programme from Pemalang mostly originate from State Vocational Schools, whereas the majority from Pekalongan City predominantly come from Private Vocational Schools and State High Schools. The majority of registrations occurred during the second wave. The marketing promotion strategy for admissions at STMIK Widya Pratama may be determined by using a simplified K-Means clustering methodology. This technique offers an excellent opportunity to enhance the enrollment rate of STMIK Widya Pratama Pekalongan's new campaign.

Nielza Athina and Lizda Iswari's earlier work on unsupervised data mining was published in the publication Clustering Population Health Data to Determine Ranges of Regional Health Degrees. "Algorithms are one method or way of providing solutions to problems found (want to solve), using tiered solutions (levels or hierarchies) in solving the problem, with reasonable reasoning," the K-Means approach was used to explain.. Nielza Athina and Lizda Iswari's earlier work on unsupervised data mining was published in the publication Clustering Population Health Data to Determine Ranges of Regional Health Degrees. "Algorithms are one method or way of providing solutions to problems found (want to solve), using tiered solutions (levels or hierarchies) in solving the problem, with reasonable reasoning," the K-Means approach was used to explain[22].

Tutik Khotimah's research, titled "Grouping Letters in the Al-Qur'an," builds upon previous works in the field of data mining and the k-means clustering technique. The K-Means Algorithm is used to illustrate that clustering is an unsupervised branch of data mining. Clustering is the process of categorising or grouping data based on its level of uncertainty. The clustering process groups like data together and separates unrelated data into different clusters[20].

Data grouping studies previously published by Wahyudi and Mario Anggara in their journal article, Selection of distance measurements in K-Means Alvaro Fitness personnel clustering and promotiongrouping The following was clarified using the K-Means Algorithm: "Unlike classification, which examines object data classes with labels. By clustering data objects, known class labels are not examined during analysis. In the training data, class labels are present. as a result of ignorance previously. Assembling a collection of closely related objects is called clustering. The process of clustering begins with a random selection of K, the number of clusters you wish to create from the input data. One feature of the clustering technique is its unsupervised nature. The goal of both the hierarchical cluster and the non-hierarchical cluster, which are the two divisions of the clustering method, is to arrange things into a single cluster that shares the same degree of similarity with other objects in the cluster. The identical cluster, yet markedly different from other cluster objects[21].

Compiling tables and figures in table 1. Amount Location Data Marketing:

Table 1. Amount Location Data Marketing

No	Month	Customer data	Regions
1	March	485	Plumbon
2	April	323	Source
3	May	367	Plered
4	June	282	Interweaving
5	July	325	Belah Stone
6	August	475	Martapura
7	September	308	Weru
8	October	250	Pabedilan
9	November	314	Pabuaran
10	December	218	Wangon
11	January	255	Gunung Sari
12	February	329	Talun

Data Source, 2021

Based on table 1 regarding customer data, it explains that in March there were 485 customers with the majority of customers from the plumbon area, in April there were 323 customers with the majority of customers from the source area, in May there were 367 customers with the majority of customers from the plumbon area, in June there were 282 customers with the majority of customers from the Watu Belah area, in August there were 475 customers with the majority of customers from the Martapura area, in September there were 308 customers with the majority of customers from the Weru area, in October there were 250 customers with the majority of customers from the Pabedilan area, in November there were 314 customers with the majority of customers from the Pabuaran area, in December there were 218 customers with the majority of customers from the Wangon area, in January there were 255 customers with the majority of customers from the Gunung Sari area, in February there were 329 customers with the majority of customers from the talun area.

Competition in the goods delivery service industry is getting tighter. This is because the number of companies operating in this field is increasing. One of the goods delivery service companies, namely company X,Y,Z, claims to have its own strategy in facing competition between delivery services. The most important thing is to maintain our quality by ensuring that the goods sent arrive at their destination in good condition and on time.

The root of the problem is the large number of competitors and cheap shipping costs from competitors, which has an impact on reducing the number of customers for the Company. One of the causes of the decline in the number of consumers in May was the lack of regular and appropriate analysis and processing of consumer data based on historical data by the PT admin. Company. This can influence decision making in determining the right target promotion area. Consumer data processing should be carried out in order to determine the right targeted promotional areas so that there is no decrease in the number of consumers in the following month.

ISSN: 2252-9853 (Print) | ISSN: 2656-808X (Online)

Method

Below is a study flowchart that will be executed by the researcher:.

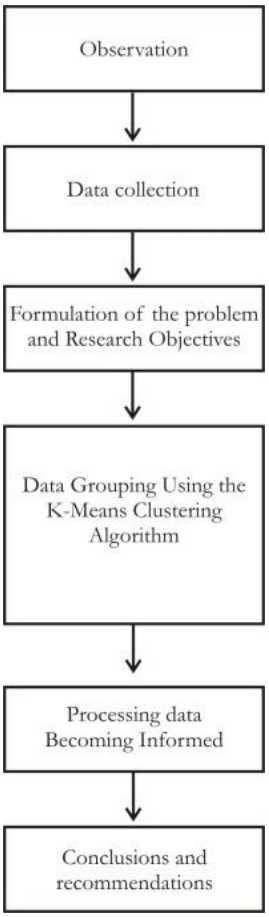


Figure 1. Research Stages

1. Observations

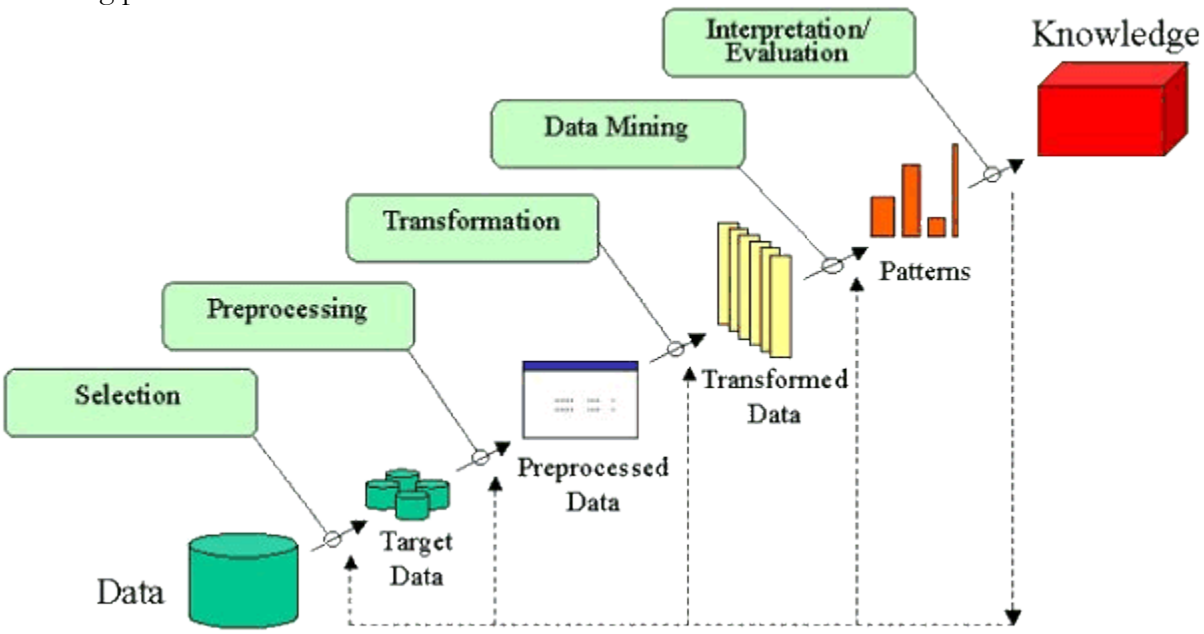
This observation was conducted by on-site fieldwork at the Company by physically visiting the Company. The researcher promptly conducted observations and sought authorization from the company's policy owner (leadership) to access the necessary data for categorising the company's customer data. This data will be used for analysis in order to determine suitable promotional media using mapping techniques.

2. Data Collection

This research uses the method of collecting the necessary data through data collection interviews in the form of questions and answers with sources to identify problems in the research. The data collected in the research is from companies that have collaborated with various promotional methods when offering products and marketing the products offered. In its application, structured interviews are carried out; in the case of interviews conducted with sources, the author uses research instruments such as recording devices, cameras, or other instruments. The target sources are customer service, consumers, and company leaders who use the company's goods delivery services.

3. Data Analysis

The data analysis conducted at this stage involves gathering data from interviews and observations. This data is then comprehensively examined and researched to identify specific concerns that will be further investigated in this study. The research methodology used in this study involves the utilisation of the K-means clustering algorithm, a technique often used in the field of Data Mining. Data Mining is a component of the Knowledge Discovery in Database (KDD) series. KDD pertains to the amalgamation of methodologies and scientific exploration, comprehension, and depiction of data patterns. The sequence of procedures comprises the following phases:



Data Resource : [23]

Figure 2. Stages of Knowledge Discovery in Databases (KDD)

a. Data cleaning (to remove consistent data and noise).
b. Data integration (combining data from several sources).
c. Data transformation (data is converted into a form suitable for mining).
d. Utilizing Data Mining methods involves the extraction of patterns from available data.
e. Evaluation of discovered patterns (the procedure of translating patterns into actionable knowledge to aid in decision-making).
f. Presentation of knowledge (with visualization techniques).

This stage is part of the data knowledge search process which includes checking whether the patterns or information found contradict previously existing facts or hypotheses. The final step of KDD is to present knowledge in a form that is easy for users to understand. Figure 3.1 shows the KDD process or stages.

4. Problem formulation and research objectives

After carrying out the data analysis stage, the author began to formulate the problems and research objectives to be achieved as per the results of the data analysis that had been carried out previously. And processing data into metadata to produce keywords and input data that can be processed and produce information.

5. Application of Method

In applying this method, systematic and directed steps will be described using a K-means Clustering methods.

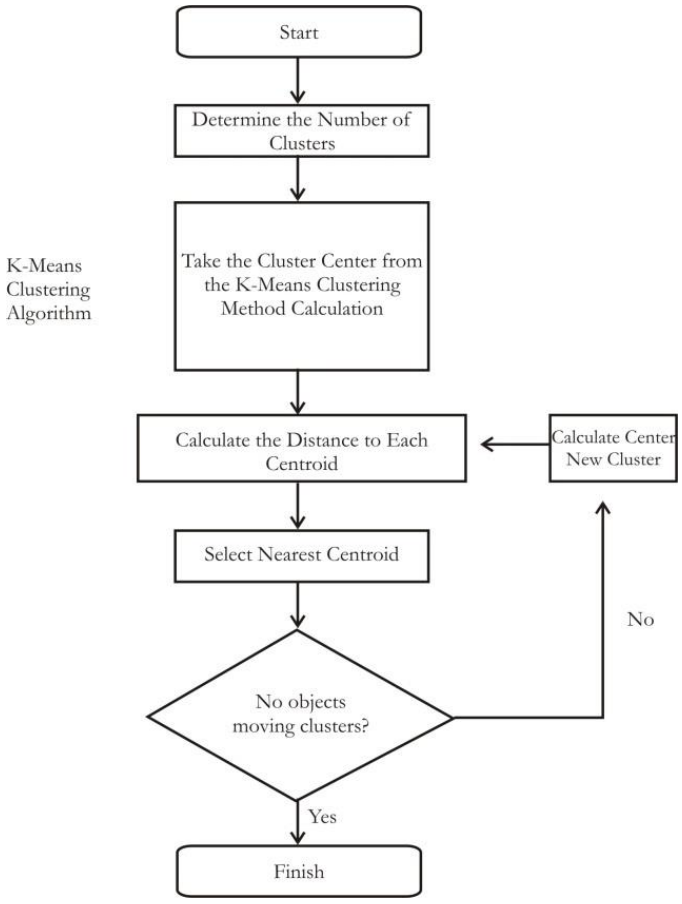


Figure 3. K-means Algorithm Flowchart

Results and Discussion

At this stage, data is grouped using a combination of K-Means clustering, namely hierarchical clustering and the k-means method. This hierarchical clustering algorithm is used to determine the cluster center. Next, the cluster centers obtained from hierarchical clustering are used for the data grouping process using the k-means method. From the picture above we can see the sequence of clustering methods with hierarchical clustering from the initial to the final stages.

Determine the condition k = 1, then continue with the k-means method:

1. Starting with determining the number of k clusters to be formed, then continuing with randomly determining the initial cluster center.
2. Because this method is a combination of hierarchical clustering and k-means, determining the cluster center for the k-means method is determined by finding the average of the data in a cluster resulting from hierarchical k-means. So at this stage, the cluster center of the k-means method can be directly determined.
3. Calculate the distance of the cluster members to each centroid.

4. After the results are obtained, the cluster members are put into the cluster that has the closest distance to the centroid. Iteration in K-means will stop when all the data in a particular cluster does not move to another cluster. For example: the Data processing stage into Metadata from all the data obtained, keywords are selected that can represent the data grouping of users of the Company's goods delivery services.

Table 2. Example of Keyword List for Metadata

No.	Keywords
1.	Promotion Strategy
2.	Products
3.	Regional Distribution

Based on Table 2, there are 3 keywords that will be used to form the metadata matrix columns. Meanwhile, the amount of data used as input is 2265 data which forms the metadata matrix rows. Next, the data is grouped using a combination of K-Means clustering as calculated above.

Research Results and Discussion

In this research, a system will be built that can be used to optimize the grouping of relevant promotional strategies that have been taken first. The physical condition data used is promotional strategies, products and regional distribution. It is assumed that the data taken is customer data which is the main target of the promotion. After the data is obtained, calculations are then carried out to find the size of each existing data.

After getting the target area and frame value for each data, the next step is to carry out the data clustering process using the K-Means method. If you want to classify data into 3 classes, then it can be determined that the K value that will be used in the K Means classification process is:

1. Data Testing

Examples of initial data before calculations are carried out to find the BMI value and frame size for each data can be seen in Table 3.

Table 3. Promotion Strategy Data, Products and Regional Distribution

Promotion	SP	Prd	Wil
1	163	59	14
2	170	125	19
3	164	53	15
4	166	58	16
5	167	50	13
6	168	50	12
7	173	56	15
8	168	73	18
9	177	60	15
10	168	52	15
11	159	58	15
12	167	75	16
13	170	73	16
14	172	68	15

Promotion	SP	Prd	Wil
15	165	73	18
16	169.5	55	14
17	160	54	15
18	173	56	14
19	162	54	15
20	169	79	17

Note: Tb:
Promotion Strategy (SP): Prd. Product: Wil: Region. The data in Table 2.1 is then used to calculate BMI values and frame size using equations (1) and (2). The results of the calculations can be seen in Table 4.

Table 4. Results of calculating BMI values and measuring framework for each data

Promotion	SP	Prd
1	22,21	11.64
2	43.25	8.95
3	19.71	10.93
4	21.05	10.38
5	17.93	12.85
6	17.72	12
7	18.71	11.53
8	25.86	9.33
9	19.15	11.8
10	18.42	11.2
11	22.94	10.6
12	26.89	10.44
13	24.91	10.63
14	22.99	11.47
15	26.82	9.17
16	19.14	12,11
17	21.09	10.67
18	18.71	12.36
19	20.58	10.8
20	27.66	9.94

2. Training Results

Following this, the K-Means classification algorithm will be employed to categorize the available data into three groups. The data will be organized into three groups using the following steps:

$$d(x,y) = ||xy|| \, {}^{2}\sum_{i=1}^{n}(xi - yi)2$$

a. Determine the cluster center randomly, for example we determine C1 = (20.9); C2=(23,10); and C3=(27,11)
b. Compute the distance from every existing data point to each respective cluster center. For instance, in determining the distance between the initial promotional data and the centroid of the initial cluster, the process involves:

$$d_{11} = \sqrt{(22{,}21 - 20)^2 + (11{,}64 - 9)^2} = 4{,}97$$

Distance between the first promotional data and the second cluster center:

$$d_{12} = \sqrt{(22{,}21 - 23)^2 + (11{,}64 - 10)^2} = 2{,}04$$

Distance between the first promotional data and the third cluster center:

$$d_{13} = \sqrt{(22{,}21 - 27)^2 + (11{,}64 - 11)^2} = 1{,}99$$

Table 5. Results of calculating BMI values and frame size for each data

SP	Prd	UK	C1	C2	C3
1	22,21	11.64	4.97	2.04	1.91
2	43.25	8.95	25.25	22.28	19.36
3	19.71	10.93	2.58	1.6	4.29
4	21.04	10.38	3.34	0.38	3.02
5	17.93	12.85	3.85	4.19	6.35
6	17.72	12	3.01	3.85	6.,36
7	18.71	11.53	2.63	2.76	5.32
8	25.86	9.33	7.87	4.91	2.5
9	19.15	11.8	3.03	2.58	4.91
10	18.42	11.2	2.24	2.84	5.58
11	22.94	10.6	5.19	2.03	1.13
12	26.89	10.44	9.01	5.91	2.95
13	24.91	10.63	7.1	3.96	0.99
14	22.99	11.47	5.56	2.47	1.12
13	26.81	9.17	8.82	5.87	3.36
16	19.14	12,11	3.31	2.81	4.98
17	21.09	10.67	3.51	0.67	2.93
18	18.71	12.36	3.43	3.29	5.46
19	20.58	10.8	3.14	0.91	3.43
20	27.66	9.94	9.71	6.66	3.81

c. A data will belong to a cluster with the shortest distance to the cluster center. For instance, for the initial data, the smallest distance is found in the third cluster; thus, the first data will be assigned to the third cluster. The complete cluster position can be seen.

Table 6. Cluster Positions in the First Iteration

SP	Prd	UK	C1	C2	C3
1	22,21	11.64			*
2	43.25	8.95			*
3	19.71	10.93		*	
4	21.04	10.38		*	
5	17.93	12.85	*		
6	17.72	12	*		
7	18.71	11.53	*		
8	25.86	9.33			*
9	19.15	11.8		*	
10	18.42	11.2	*		
11	22.94	10.6			*
12	26.89	10.44			*

SP	Prd	UK	C1	C2	C3
13	24.91	10.63			*
14	22.99	11.47			*
13	26.81	9.17			*
16	19.14	12,11		*	
17	21.09	10.67		*	
18	18.71	12.36		*	
19	20.58	10.8		*	
20	27.66	9.94			*

d. Calculate the Center of the new cluster. For the first cluster there are 4 data, namely data e 5, 6, 7, and 10, so:

C_{11} = (17.93 + 17.72 + 18.71 + 18.42) / 4 = 18.19

C_{12} = (12.85 + 12.00 + 11.53 + 11.20) / 4 = 11.89

For the second cluster, there are 7 data, namely data 3, 4, 9, 16, 17, 18, and 19:

$$C21 = \frac{(19{,}71 + 21{,}05 + 19{,}15 + 19{,}14 + 21{,}09 + 18{,}71 + 20{,}58)}{7}$$

C21 = 19 .92

$$C22 = \frac{(10{,}98 + 10{,}38 + 11{,}80 + 12{,}11 + 10{,}67 + 12{,}36 + 10{,}8)}{7}$$

C22 = 11 .29

For the third cluster, there are 9 data, namely data 1, 2, 8, 11, 12, 13, 14, 15, 20:

$$C31 = \frac{(22{,}21 + 43{,}25 + 25{,}86 + 22{,}94 + 26{,}89 + 24{,}91 + 22{,}99 + 26{,}81 + 27{,}66)}{9}$$

C31 = 27 .06

$$C32 = \frac{(11{,}64 + 8{,}96 + 9{,}33 + 10{,}6 + 10{,}44 + 10{,}63 + 11{,}47 + 9{,}17 + 9{,}94)}{9}$$

C32 = 10 .24

e. Repeat step 2 until the data position has not changed.

Table 7. Cluster positions in the 2nd iteration

SP	Prd	UK	C1	C2	C3
1	22,21	11.64		*	
2	43.25	8.95			*
3	19.71	10.93		*	
4	21.04	10.38		*	
5	17.93	12.85	*		
6	17.72	12	*		
7	18.71	11.53	*		
8	25.86	9.33			*
9	19.15	11.8		*	
10	18.42	11.2	*		
11	22.94	10.6		*	
12	26.89	10.44			*
13	24.91	10.63			*

SP	Prd	UK	C1	C2	C3
14	22.99	11.47		*	
13	26.81	9.17			*
16	19.14	12,11	*		
17	21.09	10.67		*	
18	18.71	12.36	*		
19	20.58	10.8		*	
20	27.66	9.94			*

Table 8. Cluster positions in the 8th iteration

SP	Prd	UK	C1	C2	C3
1	22,21	11.64		*	
2	43.25	8.95			*
3	19.71	10.93		*	
4	21.04	10.38		*	
5	17.93	12.85	*		
6	17.72	12	*		
7	18.71	11.53		*	
8	25.86	9.33			*
9	19.15	11.8	*		
10	18.42	11.2	*		
11	22.94	10.6	*		
12	26.89	10.44			*
13	24.91	10.63		*	
14	22.99	11.47		*	
13	26.81	9.17			*
16	19.14	12,11	*		
17	21.09	10.67		*	
18	18.71	12.36	*		
19	20.58	10.8		*	
20	27.66	9.94			*

Table 9. Cluster positions in the 8th iteration

SP	Prd	UK	C1	C2	C3
1	22,21	11.64	*		
2	43.25	8.95			*
3	19.71	10.93	*		
4	21.04	10.38	*		
5	17.93	12.85	*		
6	17.72	12	*		
7	18.71	11.53	*		
8	25.86	9.33		*	
9	19.15	11.8	*		
10	18.42	11.2	*		
11	22.94	10.6		*	

SP	Prd	UK	C1	C2	C3
12	26.89	10.44		*	
13	24.91	10.63		*	
14	22.99	11.47		*	
13	26.81	9.17		*	
16	19.14	12,11	*		
17	21.09	10.67	*		
18	18.71	12.36	*		
19	20.58	10.8	*		
20	27.66	9.94		*	

Table 10. Cluster positions in the 9th iteration

SP	Prd	UK	C1	C2	C3
1	22,21	11.64	*		
2	43.25	8.95			*
3	19.71	10.93	*		
4	21.04	10.38	*		
5	17.93	12.85	*		
6	17.72	12	*		
7	18.71	11.53	*		
8	25.86	9.33		*	
9	19.15	11.8	*		
10	18.42	11.2	*		
11	22.94	10.6		*	
12	26.89	10.44		*	
13	24.91	10.63		*	
14	22.99	11.47		*	
13	26.81	9.17		*	
16	19.14	12,11	*		
17	21.09	10.67	*		
18	18.71	12.36	*		
19	20.58	10.8	*		
20	27.66	9.94		*	

The iteration was concluded and the final outcome of 3 clusters was achieved as the cluster position table remained unchanged in both the 8th and 9th iterations.

i. The first cluster has a center (19.53; 11.52) which can be interpreted as a promotional strategy group with a Normal cluster and a large framework.
ii. The second cluster has a center of (25.44; 10.22) which can be interpreted as a regional group with medium clusters and medium framework.
iii. The third cluster has a center (43.25; 8.95) which can be interpreted as a group of heavy cluster areas and small frames

Tria Titiani Chasanah states that Widiyono, in his publication titled "Optimising Promotion Strategy Determination," The K-Means Clustering Algorithm is a data mining technique that operates in an unsupervised manner, meaning it does not rely on training or a teacher and does not need an output objective. Data mining employs two distinct clustering

techniques for categorising data: hierarchical clustering and non-hierarchical clustering. The user's text is "(12)". Clustering is a technique used to look for and group data that have similar traits or similarities with one another.

Conclusion

After conducting trials and analysing the results, many conclusions were drawn from this research: The analysis of consumer data in the Majalengka region indicates a total of 2265 customers. By using the k-means algorithm and data mining methods, a total of three clusters were established. Cluster one has 1896 customers, or 72.6% of the total, and is determined by geographical location. Cluster two consists of 717 customers, or 27.4% of the total, and is determined by the category of items bought. Cluster three comprises 162 customers, representing 20% of the overall total, and is characterised by its reliance on promotional strategies. Based on the study's results, it is advised to use a marketing strategy for prospective consumer users that involves sending a set of brochures with customised marketing approaches to each location, taking into account the clusters generated. The plan should be in accordance with the highly desired items delivery programme and include promotions that specifically cater to the unique requirements of customers in each location.

Acknowledgement

The writing team would like to express their appreciation to LPPMP Bhayangkara University, Greater Jakarta, for their assistance in enabling the research to proceed. We extend our appreciation to the research team from Telkom University's Jakarta campus and the University of Malaysia, as well as the student staff who have collaborated with us. This research will function as a standard by which future academics might measure their work.

References

[1] M. C. Thrun, "Swarm intelligence for self-organized clustering," *Artificial Intelligence*, vol. 290, 2021, doi: 10.1016/j.artint.2020.103237.

[2] J. Zhang, "Customizing the promotion strategies of integrated air-bus service based on passenger satisfaction," *Transportation Research Part D: Transport and Environment*, vol. 109, 2022, doi: 10.1016/j.trd.2022.103385.

[3] L. Zhang, "The promotion of the concept of sustainable development to the reform of enterprise human resource management," *International Journal of Knowledge-Based Development*, vol. 13, no. 2, pp. 344–362, 2023, doi: 10.1504/IJKBD.2023.133334.

[4] O. Valentim, "The relation between lifestyles and positive mental health in Portuguese higher education students," *Global Health Promotion*, vol. 30, no. 1, pp. 23–32, 2023, doi: 10.1177/17579759221112552.

[5] G. H. Krause, "Computational thinking in a bilingual kindergarten classroom: Emergent ideas for teaching across content areas," *Education and Information Technologies*, vol. 28, no. 8, pp. 9767–9782, 2023, doi: 10.1007/s10639-022-11454-1.

[6] G. R. Gagan, "Business Promotion through Instagram using Semi-Supervised Topic Modelling," *2023 IEEE 8th International Conference for Convergence in Technology, I2CT 2023*. 2023. doi: 10.1109/I2CT57861.2023.10126267.

[7] C. S. Reddy, "A Comparative Survey on K-Means and Hierarchical Clustering in E-Commerce Systems," *IDCIoT 2023 - International Conference on Intelligent Data Communication Technologies and Internet of Things, Proceedings*. pp. 805–811, 2023. doi: 10.1109/IDCIoT56793.2023.10053472.

[8] Y. Zhang, "Spatial relationships between ecosystem services and socioecological drivers across a large-scale region: A case study in the Yellow River Basin," *Science of the Total*

Environment, vol. 766, 2021, doi: 10.1016/j.scitotenv.2020.142480.

[9] S. Dong, "Network Abnormal Traffic Detection Model Based on Semi-Supervised Deep Reinforcement Learning," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4197–4212, 2021, doi: 10.1109/TNSM.2021.3120804.

[10] R. D. Firdaus, T. G. Laksana, and R. D. Ramadani, "Applying the K-Means method and Hierarchical Clustering Single Linkage to categorise medicine inventory data," *Inista*, vol. 8106, pp. 33–48, 2019.

[11] X. Ran, "A novel K-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Applied Sciences (Switzerland)*, vol. 11, no. 23, 2021, doi: 10.3390/app112311202.

[12] H. Yu, "A three-way density peak clustering method based on evidence theory," *Knowledge-Based Systems*, vol. 211, 2021, doi: 10.1016/j.knosys.2020.106532.

[13] M. Ghobaei-Arani, "A workload clustering based resource provisioning mechanism using Biogeography based optimization technique in the cloud based systems," *Soft Computing*, vol. 25, no. 5, pp. 3813–3830, 2021, doi: 10.1007/s00500-020-05409-2.

[14] M. Jabi, "Deep Clustering: On the Link between Discriminative Models and K-Means," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1887–1896, 2021, doi: 10.1109/TPAMI.2019.2962683.

[15] T. M. Ghazal, "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[16] S. Yang, "Sports Health Analysis and Promotion Countermeasures of Weak Constitution College Students based on K-Means Clustering: Take a University in Beijing as an Example," *ACM International Conference Proceeding Series*. pp. 138–143, 2022. doi: 10.1145/3568364.3568385.

[17] T. Mohith, "A Unified Analysis of Bank Customer Using Machine Learning," *Lecture Notes in Electrical Engineering*, vol. 690. pp. 457–463, 2021. doi: 10.1007/978-981-15-7504-4_44.

[18] S. E. A. BUANANTA, "Bi dashboard to support decision making on product promotion for payment/purchase transactions on e-banking," *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 15, pp. 3713–3724, 2021, [Online]. Available: https://api.elsevier.com/content/abstract/scopus_id/85112557641

[19] F. Liu, "Determine the Number of Unknown Targets in Open World Based on Elbow Method," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 986–995, 2021, doi: 10.1109/TFUZZ.2020.2966182.

[20] S. Mishra, "A Decisive Metaheuristic Attribute Selector Enabled Combined Unsupervised-Supervised Model for Chronic Disease Risk Assessment," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi: 10.1155/2022/8749353.

[21] R. Qaddoura, "A multi-stage classification approach for iot intrusion detection based on clustering with oversampling," *Applied Sciences (Switzerland)*, vol. 11, no. 7, 2021, doi: 10.3390/app11073022.

[22] W. Van Gansbeke, "Unsupervised Semantic Segmentation by Contrasting Object Mask Proposals," *Proceedings of the IEEE International Conference on Computer Vision*. pp. 10032–10042, 2021. doi: 10.1109/ICCV48922.2021.00990.

[23] M. Tomasso, "Cluster boosting and data discovery in social networks," *Proceedings of the ACM Symposium on Applied Computing*. pp. 1801–1803, 2022. doi: 10.1145/3477314.3507243.