

# Implementation of Data Mining Grouping Product Order Distribution Center In PT. Indomarco Prismatama Medan Branch

Nico Setiawan<sup>1</sup>, Post Marto Hasugian<sup>2</sup>

<sup>1,2</sup>Informatics Engineering Study Program, STMIK Pelita Nusantara, Jl. Iskandar Muda No. 1 Medan, North Sumatra, Indonesia 20154.

E-mail: setiawannico5@gmail.com

***Abstract**-Data mining is the technique of extracting previously unknown information in a set of data in the database. Data mining has been applied in various fields which require extracting information. One of them in groupings of data. Grouping is used to divide a set of data into several sections that are useful to more easily identify a class of data. Distribution companies can use groupings of one to determine the intensity of the volume of goods ordered. The study analyzes the application of data mining algorithms k-means clustering to elicit information from the data ordering goods contained in center PT distribution. Indomarco Prismatama Medan branch. That is by using a number of items and the total amount of the quantity of each item ordered.*

***Keywords:** Data mining, k-means algorithm, clustering*

## 1. Introduction

Data mining is an activity undertaken to explore the information-shaped pattern or grouping of data in a data set that has a large enough quantities to be taken of information that can be used to help make decisions. Data mining is also interpreted as an interesting pattern extraction of large amounts of data. In which a pattern is said to be interesting if the pattern is not implicit, previously unknown, but useful.

PT. Indomarco Prismatama is one of the largest franchised retail company in Indonesia. PT. Indomarco Prismatama have the space used to store merchandise while before they were distributed to retail stores called Distribution center. Distribution centers deliver goods based on data from the first order of goods sent by the store to the distribution center.

In the process of packing the goods to be delivered to the store frequently arise where frequent delays due to packing a large number of goods demand, but the number of personnel in the area are not qualified to do the packing at the request of such large items.

Based on the above problems, the researchers tried grouping goods order data with data mining. utilize goods order data sent by the store to find out what items are ordered by the store.

## 2. Theory

### 2.1. Data Mining

Data mining is the process to obtain useful information from large data base warehouse. Also defined as the extraction of new information, taken from a large a chunk of data that help in decision-making [1].

Data mining is the method used for extraction of hidden predictive information on the database [2]. The term data mining has several views, such as knowledge discovery or pattern recognition. the two terms actually has their accuracy, respectively. The term knowledge discovery or invention of appropriate knowledge for use purpose of data mining is to obtain knowledge that is still hidden inside a chunk of data. The term pattern recognition or pattern recognition also remains to be used because of the knowledge that was about to be extracted is shaped patterns that still need to be explored from a chunk of data that is facing [1].

### 2.2 K-Means Clustering

K-means clustering method is a method of group analysis that led to the object of observation of partitioning N into K groups (clusters) where each object observation is owned by a group with the closest mean

[1]. K-means clustering can be interpreted as one of the data mining tools aimed at grouping the objects into clusters [2]. Grouping the data with K-means clustering algorithm is generally performed in the following order:

- a. Define k as the number of clusters to be formed.
- b. Initialize k centroid (center point cluster) beginning at random.
- c. Allocate any data or object to the nearest cluster. The distance between the object and the distance between objects in the cluster. specific data with the specified distance between the center of the cluster. To calculate the distance of all data to each cluster center, can use distances theory ecuilidean defined as the following equation.

$$\{11\} \quad \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

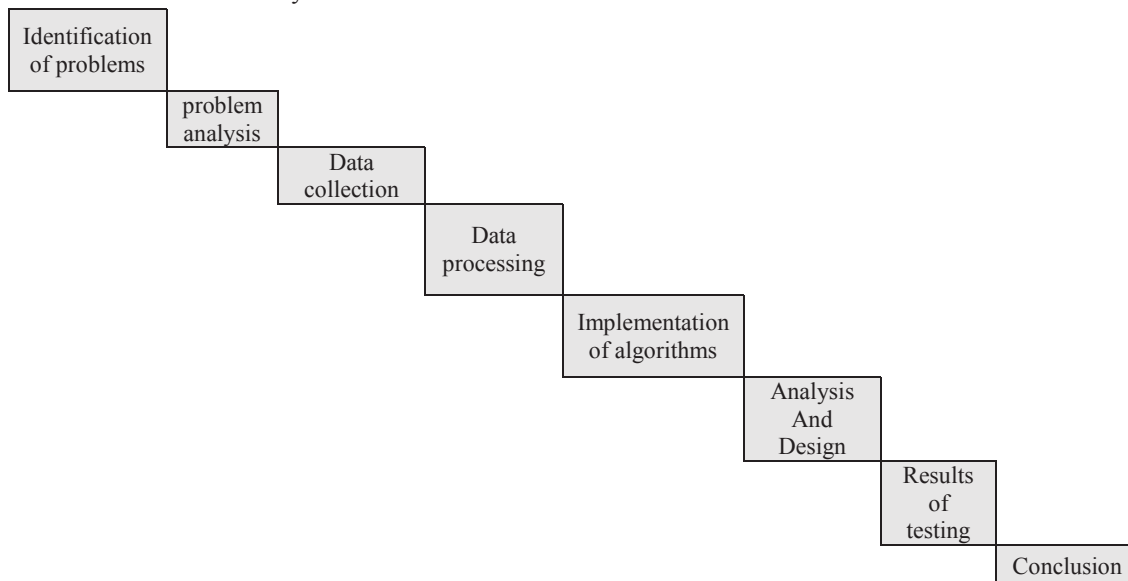
Distance to the center of the cluster is recalculated with the current cluster membership. Center of the cluster is the average of all the data or objects in a particular cluster, if desired can also use the median value of the cluster.

- d. Repeat step three until the results of iteration is worth the same as the previous iteration.

### 3. Research methods

Research in an attempt to obtain the truth. research must be based on scientific thinking process as outlined in the scientific method.

Research framework is procedural used in conducting research studies in order to run smoothly and systematically with prior calculation phases or activities to be undertaken while doing this penelitian. Berikut is the framework of this study.



**Fig 1.** Framework Research

### 4. Analysis and Design

In this study, there are 19 data to be grouped. This data is the reservation data that have been through the cleaning process data in accordance with the stages of data mining.

**Table 1.**

The data will be in the cluster

NO	ITEM	QTY
1	2,309	6.678
2	3,869	15.328

NO	ITEM	QTY
3	4.838	15,450
4	4,747	20,829
5	4,205	39,344
6	13,081	43,006
7	16,448	50,493
8	14,276	60,358
9	23,366	62,080
10	11,580	66,982
11	28,743	71,944
12	10,664	77,262
13	3,633	92,322
14	16,702	99,961
15	39,476	108,625
16	7,498	113,527
17	30,218	124,167
18	27,583	129,662
19	34,215	148,741

Step 1: determine how many clusters to be formed. In this research will dibetuk 4 cluster.

Step 2: determine the value of the first cluster centers randomly from the data that will be in the cluster

**Table 2.**

Center Clusterawal

cluster	item	Qty
C1	34,215	148,741
C2	10,664	77,262
C3	14,276	60,358
C4	2,309	6,678

Step 3: Allocate any data or object to the nearest cluster. The distance between the object and the distance between objects with a particular cluster is determined. the distance between data centers using the theory of distance ecuilidean Cluster. The first calculated distance data by the central cluster 1 first, followed by calculating the distance to the next cluster center.

$$a. \frac{2,309 - 34,215^2 + 6,678 - 148,741^2}{(15,450 - 148,741)^2} = 145\ 602$$

$$b. \frac{3,869 - 34,215^2 + (15,328 - 148,741)^2}{(20,829 - 148,741)^2} = 136\ 821$$

$$c. \frac{4,838 - 34,215^2 + (15,450 - 148,741)^2}{(15,450 - 148,741)^2} = 136\ 490$$

$$d. \frac{4,747 - 34,215^2 + (20,829 - 148,741)^2}{(20,829 - 148,741)^2} = 131\ 262$$

$$e. \frac{4,205 - 34,215^2 + (39,344 - 148,741)^2}{(39,344 - 148,741)^2} = 113\ 439$$

$$19. = 0 \frac{34,215 - 34,215^2 + (148,741 - 148,741)^2}{(148,741 - 148,741)^2}$$

Resulting in the following:

**Table 3.**

The distance of each data to a cluster 1

to-data	item	Qty	Distance to C1
1	2,309	6,678	145,602
2	3,869	15,328	136,821
3	4,838	15,450	136,490



to-data	item	Qty	Distance to C1
4	4,747	20.829	131.262
5	4,205	39.344	113.439
6	13.081	43.006	107.826
7	16.448	50.493	99.842
8	14.276	60.358	90.604
9	23.366	62.080	87.337
10	11.580	66.982	84.834
11	28.743	71.944	76.992
12	10.664	77.262	75.259
13	3,633	92.322	64.174
14	16.702	99.961	51.829
15	39.476	108.625	40.460
16	7.498	113.527	44.202
17	30.218	124.167	24.897
18	27.583	129.662	20.199
19	34.215	148.741	-

Do the same calculation for the entire center of the cluster that had been predetermined.

Step 4: calculate the shortest distance between the data with each cluster center. The data are part of a cluster center clusterapabila distance with the smallest distance between clusters of data more

**Table 3.**  
Each member of the Cluster

Cluster members 1				Cluster member 2			
to-data	item	Qty	cluster	to-data	item	Qty	cluster
15	39.476	108.625	1	12	10.664	77.262	2
17	30.218	124.167	1	13	3,633	92.322	2
18	27.583	129.662	1	14	16.702	99.961	2
19	34.215	148.741	1	16	7.498	113.527	2
<b>Total</b>	<b>131.492</b>	<b>511.195</b>		<b>Total</b>	<b>38.497</b>	<b>383.072</b>	

Cluster members 3				Cluster member 4			
to-data	item	Qty	cluster	to-data	item	Qty	cluster
5	4,205	39.344	3	1	2,309	6.678	4
6	13.081	43.006	3	2	3,869	15.328	4
7	16.448	50.493	3	3	4.838	15,450	4
8	14.276	60.358	3	4	4,747	20.829	4
9	23.366	62.080	3	<b>Total</b>	<b>15.763</b>	<b>58.285</b>	
10	11.580	66.982	3				
11	28.743	71.944	3				
<b>Total</b>	<b>111.699</b>	<b>394.207</b>					

Step 5: Calculate new cluster centers based on the new data is generated, by adding data on each column of data and then divided by the data incorporated in each cluster

**Table 4.**  
Determination New Cluster

cluster	amount			New Cluster Center	
	item	Qty	Member	item	Qty
cluster 1	131.492	511.195	4	32.873	127.799
cluster 2	38.497	383.072	4	9.624	95.768
cluster 3	111.699	394.207	7	15.957	56.315
cluster 4	15.763	58.285	4	3.941	14.571

Step 6: Repeat step three to back calculate the distance of each data center newly formed cluster

Step 7: compare the results obtained, whether there are data to be migrated. If there are data to move the cluster then proceed to the next iteration calculation. If there is no data to move the cluster then the calculation is completed, and the latest iteration is the result of clustering.

In this study, the calculation stopped at iteration 4, with the result

**Table 5.**  
 The final result Clustering

to-data	item	Qty	cluster
1	2,309	6.678	4
2	3,869	15.328	4
3	4.838	15,450	4
4	4,747	20.829	4
5	4,205	39.344	4
6	13.081	43.006	3
7	16.448	50.493	3
8	14.276	60.358	3
9	23.366	62.080	3
10	11.580	66.982	3
11	28.743	71.944	3
12	10.664	77.262	2
13	3,633	92.322	2
14	16.702	99.961	2
15	39.476	108.625	1
16	7.498	113.527	2
17	30.218	124.167	1
18	27.583	129.662	1
19	34.215	148.741	1

The above table is a table of the results of the processing of k-means cluster is indicative of a column in the cluster where the data is located.

## 5. System implementation

Implementation of the system is the process of implementation of the system in a real situation. In the implementation phase of the system are faced with real problems as the system is designed, in this case the system is used to perform the data clustering ordering goods at PT. Indomarco Prismatama Cab. Field. on analysis *K-Mean* there is a menu that can be used in order to report the results of the analysis of K-Means can be displayed in the form of iteration-iteration. Here are the results of K-Means analysis report can be seen in Figure 2.



**Figure 2.** Display Reports Analysis K-Means

## 6. Conclusion

Based on the results of research and experimentation in the implementation of data mining system to categorize the data ordering goods at the distribution center PT. Indomarco Prismatama Medan branch can be concluded as follows:

- a. Presentation of data clustering results are presented in the form of iteration displayed on web pages. And displayed throughout the iterations that are included in the calculation of clustering. This iteration shaped calculation results can be used to perform an analysis of clustered data.
- b. Application of k-means clustering method dalam pengelompokan goods order data can help analyze the data that existed before ordering goods. Information gained from the implementation of k-means clustering method can be used to help make decisions to address the problems that occurred previously.
- c. Application design data grouping based website is very well done for the implementation of the website is easy to use and relatively lightweight application be a distinct advantage.

## 7. Reference

- [1] Prasetyo, E. 2017. *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*, Yogyakarta: C.V Andi Offset.
- [2] Rosa, A.S, Salahuddin, M. 2018. *Rekayasa Perangkat Lunak-Terstruktur Dan Berorientasi Objek*. Bandung: INFORMATIKA.
- [3] Vulandari, R. T. 2018. *Data Mining: Teori dan Aplikasi Rapidminer*, Yogyakarta: Gava media.
- [4] Destiningrum, M. & Adrian, J. Q. *Sistem Informasi Penjadwalan Dokter Berbasis Web dengan Menggunakan Framework codeigniter (Studi Kasus : Rumah Sakit Yukum Medical Center)*, Jurnal Teknoinfo, 2017, 11(2), 3.
- [5] Darmi, Y. & Setiawan, A. *Penerapan Metode Clustering K-means dalam Pengelompokan Penjualan Produk*, Jurnal Media Infotama, 2016, 12(2), 149-150.
- [6] Aprianto, B. W., Niswatin, R. K., Hariri, F. R. *Sistem Pengelompokan dan Rekomendasi Pengadaan Barang dengan Menggunakan Metode K-Means*, Simki-Techsain, 2018, 2(3), 5.
- [8] Bastian, A., Sujadi H., Febrianto. G. *Penerapan Algoritma K-means Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)*, Jurnal Sistem Informasi, 2018, 14(1), 28-29.
- [9] Harminingtyas, R. *Analisis Layanan Website Sebagai Media Promosi, Media Transaksi dan Media Informasi dan Pengaruhnya Terhadap Brand Image Perusahaan Pada Hotel Ciputra di Kota Semarang*, Jurnal STIE Semarang, 2014, 6(3), 42.
- [10] Prayitno, A.& Safitri, Y. *Pemanfaatan Sistem Informasi Perpustakaan Digital Berbasis Website Untuk Para Penulis*, IJSE-Indonesian Journal on Software Engineering, 2015, 1(1),2.
- [11] Raval, U. R., &Jani. C. *Implementing & Improvisation of K-means Clustering Algorithm*, International Journal of Computer Science and Mobile Computing, 2016, 5(5), 193.