Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

ISSN 2723-4509 (Online)

K-MEANS CLUSTERING HWI PRODUCTS
(Case Study: HWI Kisaran Distributor)

Risnawati[1*], Rohminatin[2]
[1]Computer Systems, Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Indonesia
[2]Information Systems, Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Indonesia

Corresponding author:
rhisnawati716@gmail.com

Keywords:
cluster
goods stock
HWI products
k-means clustering

ABSTRACT

HWI products are types of products in the form of food and health drinks and beauty products which are marketed and have been certified by DINKES, BPPOM, MUI and others, and provide excellent quality at relatively affordable prices and have re- leased 34 products. This HWI product itself can be obtained from an official HWI Distributor that has been verified. Of the many consumers and the many types of products registered with HWI distributors, the distributor makes a product supply without limit- ing the supply, but the large stock of products avail- able at the distributor does not guarantee that the consumer needs it, but it causes loss and income in- stability. In this study the method used is the K- Means Clustering method of 15 samples of product types. The data used and taken from the last 5 years namely 2015, 2016, 2017, 2018, and 2019 to do the grouping of data with the provisions of 3 clusters are Very in demand, in demand, and in demand. The results of the study and group- ing data by the K- Means Clustering method of 3 clusters determined there are 2 products with very best-selling clus- ters, 6 products with best-selling clusters, and 7 products with less-selling clusters.

INTRODUCTION

The very rapid development of technology in the current era of globalization has provided many benefits in the progress in various social aspects. The use of technology by humans in helping to finish work is a necessity in life, even the development of tech- nology like this must also be followed by developments in Human Resources (HR)[1]. The technology developed at this time is internet technology, especially in the business world, selling and buying online and in e-commerce that provides benefits for each company so that it becomes a competitive company [2]. So also with HWI Distributors who must be ready to serve buyers to retail their products. Distributor is a marketing product that can be done by entrepreneur helpers, another meaning distributor is the way entrepreneurs who usually work from home to sell products to end consumers, recruit, motivate, and educate new distributors to do the same thing [3].

HWI Kisaran Distributor is one of the official verified distributors or product sales centers and registration of new HWI members located at Jln. Sanusi Pane Mutiara

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

Kisaran. This HWI distributor has been active since 7 years ago and already has many new members for the sale of HWI products themselves.

HWI stands for Health Wealth International, a subsidiary of Guna Cipta Group, a native of Indonesia, which has been running for 3 generations as a market leader in spices, pasta flavorings, food coloring with the brands of people-koepoe and two grouse chili products that are well known to foreign countries such as America, the Netherlands, Australia, and Taiwan. In 2009 HWI products released new health food products, health drinks and beauty products which were marketed which were highly tested and had FDA, GMP, ISO, USDA Organic, DINKS, BPPOM, MUI and other certifications, and provided excellent quality at relatively affordable prices.

HWI products which are in the form of food and health drinks and beauty products that have been issued by PT. HWI there are 34 types of products, of the 34 types of HWI products that already exist have their respective benefits. To find out the many types of HWI products, they are only available at verified authorized distributors in their respective regions. The many types of HWI products that will be sold every day the distributor also has a large stock of products without limitation to serve consumers who need, but with the large stock of products not necessarily consumers also need these products, resulting in losses to the distributor itself. Moreover, the process of supervision and checking is still done by writing on a piece of paper and still using Microsoft Excel, really very inefficient. By not estimating the stock of products, HWI distributors experienced problems in determining which products must be sold every day. This happens because it is less precise to make a decision to determine the type of product whatever is more needed by consumers. Stacking stock of products and uncertain sales every day can lead to income instability.

To overcome these problems, HWI Distributors do a proper analysis by grouping the types of products to improve product stock performance correctly to find out what products are needed more by consumers with the provisions of the best-selling, best-selling, and less-selling clusters. This study uses 15 product samples by counting from the last 5 years of the HWI Distributor's lifetime. One way is to use data mining technology to group these products with the K-Means Clustering technique. Data mining is a field of science for the process of gaining knowledge or patterns from a database [4]-[5]. Data Mining is also a computational methodology for obtaining useful information from large data to find out patterns of behavior that will be used in analysis and prediction [3]. K-Means Clustering is a data mining technique that provides a cluster description of a product, so the simulation results can provide knowledge of the types of products that will be the most stock to increase company revenue [6], [7]. Clustering is also a process of grouping a set of objects inward with similar characteristics [8], [9].

METHOD

Data Mining is the process of finding relationships in data that are unknown to the user and presenting them in a way that can be understood so that these relationships can be the basis for decision making [10]. Data Mining is an analysis of data to deter-

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

mine clear relationships and conclude what has not been known before by the current way understood and useful for the owner of the data [5]. Data mining from another meaning is a process that finds meaningful relationships such as patterns and tendencies in examining a large set of stored data using pattern recognition techniques, statistics, and mathematics [11]. Whereas from another meaning, data mining is the process of finding interesting patterns from large amounts of data [7], [12].

K-Means is a non-hierarchical (grouped) data grouping method that attempts to partition existing data into two or more groups. This method of partitioning data into groups so that data with the same characteristics are entered into the same group and data with different characteristics are grouped into other groups. The purpose of this data grouping is to minimize the objective functions arranged in the grouping process, which generally tries to minimize variations within a group and maximize variation between groups [13]. K-Means algorithm is an algorithm in the clustering or grouping function. Clustering refers to the grouping of data, observations or cases based on the similarity of the object under study [4].

Clustering is part of the science of Data Mining that is without direction (unsupervised). Clustering is the process of sharing data into classes or clusters based on their level of similarity. In clustering, data that has similarities are included in the same cluster, while those that do not have similarities are entered in different clusters [14][15]. Clustering will do the grouping of data into a number of cluster groups based on the similarity of the characteristics of each data in the existing groups [6].

K-Means Clustering is the process of grouping dataset elements into k clusters. The process of grouping is done by calculating the euclidean distance between elements and centroids where centroids are uniform density elements [7].

In general, grouping data clustering using the K-Means method is done with the first basic algorithm to determine the number of clusters to be formed, secondly to allocate data into random clusters for the selection of the initial random center number of the cluster must be in the order of the data, the fourth calculate the center groups of each data group. From each centroid, the average value will be taken. If the average is expressed as a group, then i is a feature, p is the dimension of the data whose equation is to calculate the centroid of the feature i used equation 1 with the example formula as follows:

$$Ci = \frac{1}{M}\sum_{j=1}^{M} Xj \qquad (1)$$

Equation 1 is done as p dimensions i = 1 up to i = p [13]

The fourth is to allocate each data to the closest centroid by measuring the distance of the space that can be found in the equation with the example formula as follows:

$$d = \sqrt{(x_1 + x_2)^2 + (y_1 + xy_2)^2} \qquad (2)$$

The fifth to allocate data to each K-means method based on the comparison of the distance between centroids to each group [13], with the example formula as follows:

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

ISSN 2723-4509 (Online)

$$ai1\begin{cases} 1 & d = \min\{D(x1, C1)\} \\ 0 & lainnya \end{cases} \quad (3)$$

ai1 is the value of point xi members to the center of group c1, d is the shortest distance from data xi to the group after comparison, and c1 is the center of group 1.

$$j = \sum_{i=1}^{n} \sum_{i=1}^{k} aic\ D(Xi, C1)2 \quad (4)$$

n is the number of data, k is the number of groups, ai1 is the value of data point member xi to the ci group followed. a has a value of 0 or 1, the data is a member of the group. The sixth recalculates if there is still a change in the data to completion. A characteristic of the K-Means algorithm is the determination of the initial cluster center point because K-Means can generate the initial cluster center point randomly

RESULT AND DISCUSSION

The process is carried out to help facilitate HWI distributors in making HWI products that will be sold using the K-Means Clustering Algorithm method with the provisions of the cluster not selling, selling, and very in demand. The clusters that are made are based on the stock of HWI products that accumulate which ensures consumers buy or not the product so that it can harm the distributor itself. 15 types of samples from 34 types of products, including:

Table 1. HWI Product Data

NO	NAMA PRODUK	TAHUN				
		2015	2016	2017	2018	2019
1	CMP	600	840	1020	1200	1440
2	WMP	120	180	360	660	1080
3	GLUCELLA	120	180	240	600	780
4	NES-V	420	600	900	960	1440
5	VITAMALE	180	336	420	540	660
6	ILAV LIPSTIK	240	420	420	540	660
7	T2T SERUM	120	180	180	300	480
8	T2T CREAM	120	180	180	300	480
9	PSHBL	120	180	240	360	600
10	HABA OLIVE	96	180	240	456	600
11	VINMORI	36	132	240	300	540
12	SPIRUTHIN	180	240	360	480	660
13	CALSEABONE	96	180	180	300	540

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

14	MR-PRO	60	120	288	600	780
15	GICAFE	120	180	216	300	540

The following are examples of data that will be calculated using K-Means Clustering, namely the Number of Clusters = 5, Number of Data = 15, Number of Attributes = 3.

The manual calculation sample is the number of products taken from the last 5 years namely from 2015, 2016, 2017, 2018, and 2019.

Determination of the initial cluster center:
The first data is taken as the center of the first cluster 600 840 1020 1200 1440
The 9th data is taken as the center of the 2nd cluster 120 180 240 360 600
The 13th data is taken as the center of the 3rd cluster 36 132 240 300 540

Following is the calculation of the distance of the first data taken from the HWI product with the specified cluster:
d = √ (600-600) ¬2 + (840-840) 2+ (1020 + 1020) 2+ (1200-1200) 2+ (1440-1440) 2 = 0.00

After evaluating the initial cluster center, the next step is to calculate the cluster center distance using the equation to produce C1, C2, C3 as follows:

Table 2. Results of Cluster Center Calculation Results Using 15 Data

NO	PRODUK HWI	TAHUN					C1	C2	C3	Jarak Terpendek
		2015	2016	2017	2018	2019				
1	CMP	600	840	1020	1200	1440	0,00	1638,78	1745,78	0,00
2	WMP	120	180	360	660	1080	1234,02	578,62	667,05	578,62
3	GLUCELLA	120	180	240	600	780	1438,75	300,00	396,18	300,00
4	NES-V	420	600	900	960	1440	402,49	1329,51	1430,97	402,49
5	VITAMALE	180	336	420	540	660	1354,41	310,38	408,35	310,38
6	ILAV LIPSTIK	240	420	420	540	660	1307,67	374,70	478,50	374,70
7	T2T SERUM	120	180	180	300	480	1761,59	146,97	128,69	128,69
8	T2T CREAM	120	180	180	300	480	1761,59	146,97	128,69	128,69
9	PSHBL	120	180	240	360	600	1638,78	0,00	128,69	0,00
10	HABA OLIVE	96	180	240	456	600	1599,11	98,95	183,96	98,95
11	CALSEABONE	96	180	180	300	540	1736,44	106,66	97,49	97,49
12	SPIRUTHIN	180	240	360	480	660	1448,72	199,00	305,94	199,00
13	VINMORI	36	132	240	300	540	1745,78	128,69	0,00	0,00
14	MR-PRO	60	120	288	600	780	1463,36	315,44	388,10	315,44
15	GICAFE	120	180	216	300	540	1712,43	88,18	99,68	88,18

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

After calculating the cluster center distance to find the closest value, then the group grouping matrix will be performed which will be marked with 1 meaning that the data is in the grouping data group.

Table 3. Grouping Data 1

No.	C1	C2	C3
1	1	0	0
2	0	1	0
3	0	1	0
4	1	0	0
5	0	1	0
6	0	1	0
7	0	0	1
8	0	0	1
9	0	1	0
10	0	1	0
11	0	0	1
12	0	1	0
13	0	0	1
14	0	1	0
15	0	1	0

After obtaining the results of grouping data 1, it will enter the calculation based on the members of each cluster according to the formula as follows:

The first new cluster:

$$\frac{600+420}{2} = 510$$

$$\frac{1020+900}{2} = 960$$

$$\frac{1200+960}{2} = 1080$$

$$\frac{1440+1440}{2} = 1440$$

The second new cluster:

$$\frac{120+120+180+240+120+96+180+60+120}{9} = 137,33$$

$$\frac{180+180+336+420+180+180+240+120+180}{9} = 224,00$$

$$\frac{360+240+420+420+240+240+360+288+216}{9} = 309,33$$

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

$$\frac{660+600+540+540+360+456+480+600+300}{9} = 504{,}00$$

$$\frac{1080+780+660+660+600+600+600+780+540}{9} = 706{,}67$$

The third new cluster:

$$\frac{120+120+96+36}{4} = 93$$

$$\frac{180+180+180+132}{4} = 168$$

$$\frac{180+180+180+240}{4} = 195$$

$$\frac{300+300+300+300}{4} = 300$$

$$\frac{480+480+540+540}{4} = 510$$

The initial center determination of the new cluster is:
The new 1 st cluster is 510.00 720.00 960.00 1,080.00 1,440.00
The second new cluster is 137.33 224.00 309.33 504.00 706.67
The new cluster is 93.00 168.00 195.00 300.00 510.00

After that, keep doing repeated calculations until the data group has the same value as the previous group value from the clustering results. From the results of the calculation iteration obtained stops at the 4th iteration, like the table below:

Table 4. Results of the 4th Iteration Clustering Calculation

NO	PRODUK HWI	TAHUN					C1	C2	C3	Jarak Terpendek
		2015	2016	2017	2018	2019				
1	CMP	600	840	1020	1200	1440	201,25	1361,15	1706,95	201,25
2	WMP	120	180	360	660	1080	1053,42	331,06	650,06	331,06
3	GLUCELLA	120	180	240	600	780	1275,97	133,87	362,34	133,87
4	NES-V	420	600	900	960	1440	201,25	1050,68	1401,65	201,25
5	VITAMALE	180	336	420	540	660	1203,31	164,88	366,91	164,88
6	ILAV LIPSTIK	240	420	420	540	660	1163,83	237,82	426,73	237,82
7	T2T SERUM	120	180	180	300	480	1606,89	436,44	76,86	76,86
8	T2T CREAM	120	180	180	300	480	1606,89	436,44	76,86	76,86
9	PSHBL	120	180	240	360	600	1478,55	299,87	75,51	75,51
10	HABA OLIVE	96	180	240	456	600	1440,96	246,64	142,07	142,07

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT
ISSN 2723-4509 (Online)

11	CALSEABONE	96	180	180	300	540	1577,91	401,62	44,47	44,47
12	SPIRUTHIN	180	240	360	480	660	1291,39	145,88	263,77	145,88
13	VINMORI	36	132	240	300	540	1583,42	404,30	87,96	87,96
14	MR-PRO	60	120	288	600	780	1296,18	169,04	374,84	169,04
15	GICAFE	120	180	216	300	540	1554,23	385,33	37,16	37,16

After repeated calculations and stop at the 4th iteration, it can be seen in the grouping of data as below:

Table 5. Grouping Iteration Data to 4

No.	C1	C2	C3
1	1	0	0
2	0	1	0
3	0	1	0
4	1	0	0
5	0	1	0
6	0	1	0
7	0	0	1
8	0	0	1
9	0	0	1
10	0	0	1
11	0	0	1
12	0	1	0
13	0	0	1
14	0	1	0
15	0	0	1

From table 5 above, the results of the grouping of data that have been calculated from the past 5 years are that there are 2 HWI products with very best-selling clusters, 6 HWI products with best-selling clusters, and 7 HWI products with poor-selling clusters. From 15 samples of HWI products that the clusters have determined are as follows:

Table 6. Types of HWI Products that Appropriate with Cluster Results

C1	C2	C3
1. CMP	1. WMP	1. T2T SERUM
2. NES-V	2. GLUCELLA	2. T2T CREAM
	3. VITAMALE	3. PSHBL
	4. ILAV LIPSTIK	4. HABA OLIVE
	5. SPIRUTHIN	5. CALSEA BONE
	6. MR-PRO	6. VINMORI
		7. GICAFE

Proceeding
International Conference
on Social, Sciences and Information Technology
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

ISSN 2723-4509 (Online)

CONCLUSION

From the discussion of the research that has been done, the conclusion obtained from this study is the K-Means Clustering Method applied in the Data Mining technique used to calculate the existing stock of products at HWI distributors can be done and successfully applied so as to facilitate the HWI distributor for product supply.

The K-Means Clustering method with the Data Mining technique can be used as a reference for distributors to facilitate and assist in the provision of HWI products.

From the grouping of data, there are 3 types of clusters, which are Very Hot Selling, Selling, and Not Selling. Of the 3 clusters specified there are 2 products with very best-selling clusters, 6 products with best-selling clusters, and 7 products with less-selling clusters.

BIBLIOGRAPHY

[1] H. Agusvianto, "Sistem Informasi Inventori Gudang Untuk Mengontrol Persediaan Barang," *Journal Infrmation Engineering and Educational Technology*, vol. 01, no. 012017. pp. 40–46, 2017.

[2] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang," *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.

[3] N. S. Represa, A. Fernández-Sarría, A. Porta, and J. Palomar-Vázquez, "Data Mining Paradigm in the Study of Air Quality," *Environ. Process.*, vol. 7, no. 1, 2020, doi: 10.1007/s40710-019-00407-5.

[4] A. Rohman and M. Rochcham, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 5, no. 1, pp. 23–29, 2019, doi: 10.37760/neoteknika.v5i1.1379.

[5] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

[6] A. Sani, "Penerapan Metode K-Means Clustering Pada Perusahaan," *J. Ilm. Teknol. Inf.*, no. 353, pp. 1–7, 2018.

[7] S. Wijayanti, Azahari, and R. Andrea, "K-Means cluster analysis for students graduation (case study: STMIK widya cipta dharma)," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1296, pp. 20–23, 2017, doi: 10.1145/3108421.3108430.

[8] N. Midha and V. Singh, "Classification of e-commerce products using reptree and k-means hybrid approach," *Adv. Intell. Syst. Comput.*, vol. 654, pp. 265–273, 2018, doi: 10.1007/978-981-10-6620-7_26.

[9] V. Soundarya, "Recommendation System for Criminal Behavioral Analysis on

ISSN 2723-4509 (Online)
Kisaran, August 19th, 2020, hlm. 27 - 36
DOI: https://doi.org/10.33330/icossit.v1i1.824
Available online at https://jurnal.stmikroyal.ac.id/index.php/ICdoSSIT

Social Network using Genetic Weighted K-Means Clustering," *J. Comput.*, vol. 12, no. 3, pp. 212–220, 2017, doi: 10.17706/jcp.12.3.212-220.

[10] A. H. Putri, R. C. W, and M. A. Fauzi, "Implementasi Algoritma Improved K-Means Pada Portal Jurnal International," *https://www.researchgate.net/publication/324561707*, no. April 2018, pp. 1–9, 2017.

[11] N. Rahmadani and E. Kurniawan, "Implementasi Metode K-Means Clustering Tunggakan Rekening Listrik pada PT . PLN (Persero) Gardu Induk Kisaran," *J-SISKO TECH*, vol. 3, no. 1, pp. 103–117, 2020.

[12] yulya M. Anita, "Penentuan Tingkat Minat Belanja Online Melalui Media Sosial Menggunakan Metode Clustering K-Means," *Rang Tek. J. http//joernal.umsb.ac.id/index.php/RANGTEKNIKJOURNAL*, vol. 2, no. 2, pp. 6–11, 2018.

[13] L. R. Ananda, "Clustering Untuk Menentukan Calon Mahasiswa Berprestasi," *Jiti*, vol. 1, no. 2, pp. 16–19, 2018.

[14] C. Groß and D. Vriens, "The Role of the Distributor Network in the Persistence of Legal and Ethical Problems of Multi-level Marketing Companies," *J. Bus. Ethics*, vol. 156, no. 2, pp. 333–355, 2019, doi: 10.1007/s10551-017-3556-9.