

# Application of K-Means Clustering Data Mining in the Extension of the Contract of Work at PT. Afixkogyo Indonesia

Austin Almayda<sup>1</sup>, Sudin Saepudin<sup>2</sup>

<sup>1 2</sup>Program Studi Sistem Informasi,

<sup>12</sup>Universitas Nusaputra, Sukabumi Jl. Raya Cibatua Cisaat No.21, Cibolang Kaler, Kecamatan. Cisaat, Sukabumi Regency, Jawa Barat 43155, Telp (0266) 210594, Indonesia

E-mail: [austin.almayda\\_si18@nusaputra.ac.id](mailto:austin.almayda_si18@nusaputra.ac.id)<sup>1</sup>, [sudinsaepudin@nusaputra.ac.id](mailto:sudinsaepudin@nusaputra.ac.id)<sup>2</sup>

## ARTICLE INFO

## ABSTRACT

### Article history:

Received: September 19, 2021

Revised: October 15, 2021

Accepted: November 30, 2021

### Keywords:

Contract employee;

K-Means;

Cluster;

Employment Contract/Employment Agreement according to Law No.13/2003 concerning Manpower is an agreement between a worker/labourer and an entrepreneur or employer that contains the terms of employment, rights and obligations of the parties. The omnibus law on employment, revised the provisions of the PKWT (certain time work agreement) in Law No. 13 of 2003 through amendments, deletions and additions to articles. However, the rules regarding contract employees are only described in outline in the Job Creation Law No. 11 of 2020. The law stipulates the maximum period of time for companies to carry out PKWT contracts for a maximum of 5 years.[1] The purpose of this study is to assist superiors or personnel in assessing contract employees in the company Pt.Afixkogyo Indonesia based on the results of employee assessments or the needs of the company itself. In the selection process for contract employees at Pt. Afixkogyo Indonesia is still using manual assessment in the field where the assessment of its employees is assessed by the head of the department, the purpose of this study is to apply and evaluate the Data Mining Clustering method to get an extension of the work contract at this company. The author uses one of the data mining methods Clustering/clustering because this method is very common in use so it is easy to implement because it uses simple principles that can be explained in non-statistical terms.

Copyright © 2021 Jurnal Mantik.  
All rights reserved.

## 1. Introduction

Technological developments have contributed greatly to human life, one of which is the development of technology in artificial intelligence (Artificial Intelligence), this is evidenced by the emergence of new technologies that can help human life and even some of them are able to replace the role of humans in everyday life day. [2]

Data mining is a branch of science from artificial intelligence that is usually used by humans as a tool in processing large data sets to obtain information. The definition of data mining according to Mujib ridwan et al. (2013: 60) "Data mining is a process that uses statistical techniques, mathematics, artificial intelligence and machine learning to extract and identify useful information and related knowledge from various large databases". One example of the use of data mining is by analyzing data or carrying out a modeling process without supervision (unsupervised / to draw conclusions from datasets) and is one method of grouping data with a partition system.[10]

K-means clustering is a non-hierarchical cluster analysis method that seeks to partition existing objects into one or more clusters or groups of objects based on their characteristics, so that objects with the same characteristics are grouped into the same cluster and objects with different characteristics. grouped into another cluster.[3]

Pt. Afixkogyo Indonesia is one of the companies in Indonesia that produces Automotive Stickers, heavy equipment and electronics. The selection process for the employment contract system in the company currently applies to every contract employee whose work contract period has expired, the selection process carried out by humans may result in errors resulting in inaccurate results and of course this will have a negative impact on the company because it retains employees. who have poor quality or even lose employees who have a good record of performance and potential.

Based on these problems, it is necessary to find the right solution for the problems at hand. The solution used to overcome these problems needs to involve data mining in the assessment process, which is expected to be able to overcome the problem of inaccuracies in the selection process for contract employee extensions. Therefore, the author conducted a study entitled "The application of meaning cluster data to obtain an extension of the work contract at Pt. Afixkogyo Indonesia".

## 2. Method

### 2.1

#### ustering

Clustering According to [4] grouping is a method for finding and classifying data that have characteristics (similarities) between one data and another. The main purpose of the clustering method is to group the number of data or objects into clusters (groups) so that each cluster can contain as much data as possible. In this clustering method, we try to place similar objects (closely distanced) in one cluster and make the distance between clusters as far as possible. This means that objects in one cluster are very similar to each other and different from objects in other clusters [4]. In data mining there are two types of clustering methods used in data grouping, namely hierarchical clustering and non-hierarchical clustering. The non-hierarchical (partitioning) clustering method begins by determining in advance the desired number of clusters (two clusters, three clusters, or so on). After the number of clusters is known, then the cluster process is carried out. This method is commonly called K-Means Clustering [5].

#### 2.2 K-Means Algorit

Algorithm According to [6], the K-Means Algorithm is an iterative grouping algorithm that partitions the data set into a number of K clusters that have been set at the beginning. The K-Means algorithm is simple to implement and run, relatively fast, adaptable, and commonly used in practice. Historically, K-Means has been one of the most important algorithms in the field of data mining [7].

The K-means algorithm is the best algorithm in the partitioning clustering algorithm and the most frequently used among other clustering algorithms because of its simplicity and efficiency. The advantage of the K-means algorithm itself, according to K. Arai and A. R. Barakbah, is the simplest and most common clustering algorithm, this is because K-means has the ability to group data in large enough quantities with relatively fast and efficient computation time. However, K-means has a weakness due to the determination of the initial center of the cluster [8].

The K-Means algorithm starts with a random selection of K, K here is the number of clusters to be formed. Then set the K values randomly, temporarily the value becomes the center of the cluster or commonly called the centroid at random, the mean or "means". Calculate the distance of each existing data to each centroid using the Euclidean formula until the closest distance of each data to the centroid is found. Classify each data based on its proximity to the centroid. Perform these steps until the centroid value does not change [9]. The Euclidean distance is formulated as follows:

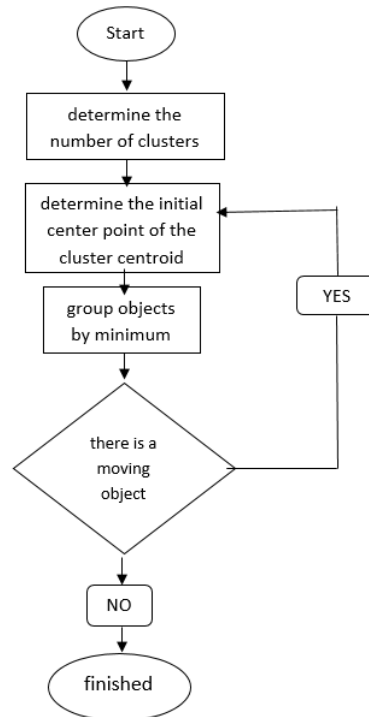
$$D(i,j)=\sqrt{((x1j-x1i)^2+(x2j-x2i)^2+\dots+(xkj-xki)^2)} \dots(2)$$

D(i,j) = Distance of data from i to cluster center j

Xki = Data to i on attribute data to k

Xkj = Data to j on attribute data to k

The stages carried out in the research are as shown in the image below:



**Fig 1.** Carried Out Of the Research

The picture above shows the steps for determining clustering with the K-Means method according to [4] which consists of several stages as follows:

The proximity of two objects is determined based on the distance between the two objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data to each cluster center. The closest distance between one data and a certain cluster will determine which data belongs to which cluster. To calculate the distance of all data to each cluster center point, you can use the Euclidean distance which is formulated as follows:

1. Determine the number of clusters
2. Determine the starting point of the cluster randomly
3. Calculate the center of the group from the data in each group. Location of each centroid

The group is taken from the average of all data values for each feature. If M represents the amount of data in a group, *i* represents the *i*-th feature in a group, and p represents the dimension data, then the equation to calculate the centroid of the *i*-th feature is used the equation [11]

$$c^1 = \frac{1}{M} \sum_{j=1}^M X_j$$

Equation (11) is carried out with p dimensions from *i*= 1 to = p

4. Group objects based on minimum distance. There are several ways you can carry out to measure the distance of the data to the center of the group, including the Euclidean measurement distance in distance space. Euclidean can be found using equation [12].

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

The reallocation of data into each group in the K-Means method is based on the comparison of the distance between the data and the centroid of each existing group. Data allocated repeat firmly to the group that has the centroid with the closest distance from the data. The allocation of this data according to MacQueen (1967) can be determined using equation [13].

$$a_{i1} = \begin{cases} 1, & d = \min\{D(X_i, C_i)\} \\ 0, & \text{lainnya} \end{cases}$$

$a_{i1}$  is the membership value of the point  $x_i$  to the center of the group  $Cl$ ,  $d$ , is the shortest distance from the data  $x_i$  to  $K$  group after comparison, and  $Cl$  is the 1st centroid. The objective function used for K-Means method is determined based on the distance and the value of data membership in the group. Function objective according to MacQueen (1967) can be determined using equation (14)

5. Return to Step 2, if there is still data that moves clusters or if the centroid value changes, some are above the specified threshold value or if the change in the value of the objective function used is above the specified threshold value.  $x_1 - x_2$

Using the Euclidean Distance Method

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 \dots (1)}$$

Description :

- $d$  = Distance
- $x_1$  = Koordinat Latitude 1
- $x_2$  = Koordinat Latitude 2
- $y_1$  = Koordinat Longitude1
- $y_2$  = Koordinat Longitude 2

The latitude and longitude coordinates can be used as variables to calculate the distance between two location points if a straight line is stretched between them [15]

### 3. Result and Analysis

#### 3.1 Data Retrieval

Determine the number of Clusters The data analyzed are data from 56 QC and printing employee data

**Table 1.**  
Number of Clusters the Data Analyzed Are Data from 56 QC and Printing Employee Data

NO	Nama	Parameter	
		KINERJA	ABSENSI
1	Yoga Septi	4,30	0,472
2	Muhamad Fahad	4,30	0,900
3	Muhammad Ageng Baskara	4,30	0,675
4	Muhamad Dikri Permana	4,30	0,900
5	Seli Audianti Sholehah	4,15	0,219
6	Sri Sania Aryati	4,30	0,000
7	Adilla Fitri M	4,30	0,000
8	Agung Gunawan	4,30	0,600
9	Eman Sulaeman	4,25	1,350
10	Ade Juwita P	4,30	0,125
11	Dian Nurdiana	4,30	0,000
12	Didit Radiansyah	4,30	0,000
13	Dira Dwicandra	4,30	0,225
14	Erna Rosmiati	4,30	2,025
15	Herlan Ruslandi	4,30	2,025
16	Imas Sulastri	4,30	2,025
17	Indra Sunandar	4,30	2,025
18	Jeni Surahman	4,30	2,025
19	Kamaludin	4,30	2,025
20	Leman Slamet	4,30	2,025
21	Lilis Lisnawati	4,30	2,025
22	Lina Oktafiana	4,30	2,025
23	Mardian Susanto	4,30	2,025
24	Mulyawan	4,30	2,025
25	Ujang Supriyadi	4,25	0,000
26	Yudi Wahyudi	4,35	0,900
27	Hasanudin	3,00	1,025

NO	Nama	Parameter	
		KINERJA	ABSENSI
28	Leni Marlina	3,00	1,025
29	Siti Maryam	3,00	1,025
30	Fajar Sugesti	4,30	2,230
31	Barkah Mustabah	4,30	1,580
32	Dani Hardiman	4,18	0,000
33	Dede Supriatna	4,18	0,740
34	Deni Febrian	4,30	0,440
35	Deni Wijaya	4,18	0,040
36	Hotib	4,18	0,090
37	Indra Andriawan	4,18	1,580
38	Iwan Irawan	4,30	0,580
39	Kurniawan	4,35	0,000
40	Mulyadi	4,30	0,380
41	Ohim	3,80	0,900
42	Pardiansyah	4,35	0,23
43	Rudi Suryawan	4,35	0,00
44	Sri Harsoyo	4,45	0,00
45	Susminto	4,30	0,18
46	Sutiawan	4,30	0,00
47	Ujang Wawan	4,30	0,00
48	Moch.Rahman Alfarizi	4,30	0,00
49	Bintar Alang Kibar	4,30	0,00
50	Ramdhoni Nur Akbar	4,30	0,00
51	M.Insani Haryadi	4,30	0,00
52	Teguh Pamungkas	4,30	0,00
53	Yusuf Rahman Hidayat	4,30	0,00
54	M.Syahrul Muksin	4,30	0,00
55	GenisMulyahati	4,30	0,00
	Subandi	4,30	0,00
56	Reynaldi Pamungkas	4,30	0,00

**3.2 Determine the Starting Point of the Cluster Randomly**

Based on the data from the table above, we will determine the number of clusters formed by three (k=3). The number of clusters must be smaller than the number of data (k<n).

**Table 2**

Number of Clusters Formed By Three (K=3)

	Jumlah Cluster	3	
	Pusat Cluster	(random)	
C1	Yudi Wahyudi	4,35	0,9
C2	Lina Oktafiana	4,30	2,025
C3	Siti Maryam	3,00	1,025

**Table 3**

Kriteria

	KRITERIA	RANGE
4	KURANG	<5
5	CUKUP	
6	BAIK	

The initialization of the dataset centroid in the dataset table above is C1 = {4,35, 0,09}, C2 = {4,30, 2,025} and C3 = {3,00, 1,025} . Centroid initialization can be determined manually or randomly. For the next iteration (1st repetition until completion), the new centroid is calculated by calculating the average value of the data in each cluster. If the new centroid is different from the previous centroid, then the process continues to the next step. However, if the new centroid is calculated the same as the previous centroid, then the clustering process is complete.

**3.3 Calculate the Distance to Each Centroid and Group Objects Based on the Minimum Distance**



After obtaining the initial center point of the cluster, then calculating the Euclidian distance, and grouping based on the smallest distance, then a new centroid value will be obtained for the next calculation reference until the centroid values before and after are the same. Calculation of Euclidean distance in iteration (1),

a. The following are some results of the calculation of the Euclidean Distance, at the center point of Cluster 1 (4,35 , 0,9)

$$1) d(x1,c1) = \sqrt{(a1 - c1a)^2} + \sqrt{(b1 - c1b)^2} = \sqrt{(4,30 - 4,35)^2} + \sqrt{(0,472 - 0,9)^2} = 0,431$$

$$2) d(x2,c1) = \sqrt{(a2 - c1a)^2} + \sqrt{(b2 - c1b)^2} = \sqrt{(4,30 - 4,35)^2} + \sqrt{(0,9 - 0,09)^2} = 0,050$$

$$3) d(x3,c1) = \sqrt{(a3 - c1a)^2} + \sqrt{(b3 - c1b)^2} = \sqrt{(4,30 - 4,35)^2} + \sqrt{(0,675 - 0,09)^2} = 0,230$$

Do the same calculation until the data to 56

b. The following are some results of the calculation of the Euclidean Distance, at the center point of Cluster 2 (4,30, 2,025)

$$1) d(x1,c1) = \sqrt{(a1 - c1a)^2} + \sqrt{(b1 - c1b)^2} = \sqrt{(4,30 - 4,30)^2} + \sqrt{(0,472 - 2,025)^2} = 1,553$$

$$2) d(x2,c1) = \sqrt{(a2 - c1a)^2} + \sqrt{(b2 - c1b)^2} = \sqrt{(4,30 - 4,30)^2} + \sqrt{(0,9 - 2,025)^2} = 1,125$$

$$3) d(x3,c1) = \sqrt{(a3 - c1a)^2} + \sqrt{(b3 - c1b)^2} = \sqrt{(4,30 - 4,30)^2} + \sqrt{(0,675 - 2,025)^2} = 1,350$$

Do the same calculation until the data to 56

c. The following are some results of the calculation of the Euclidean Distance, at the center point of Cluster 3 (3,00 , 1,025)

$$1) d(x1,c1) = \sqrt{(a1 - c1a)^2} + \sqrt{(b1 - c1b)^2} = \sqrt{(4,30 - 3,00)^2} + \sqrt{(0,472 - 1,025)^2} = 1,413$$

$$2) d(x2,c1) = \sqrt{(a2 - c1a)^2} + \sqrt{(b2 - c1b)^2} = \sqrt{(4,30 - 3,00)^2} + \sqrt{(2,025 - 1,025)^2} = 1,306$$

$$3) d(x3,c1) = \sqrt{(a3 - c1a)^2} + \sqrt{(b3 - c1b)^2} = \sqrt{(4,30 - 3,00)^2} + \sqrt{(2,025 - 1,025)^2} = 1,346$$

Do the same calculation until the data to 56

For so on, calculate the distance in each row of data, and the results are as in the table below.

**Table 4**  
Calculate the Distance in Each Row of Data, and the Results

NO	Nama	Parameter				
		KINERJA	ABSENSI	C1	C2	C3
1	Yoga Septi	4,30	0,472	0,431	1,553	1,413
2	Muhamad Fahad	4,30	0,900	0,050	1,125	1,306
3	Muhammad Ageng Baskara	4,30	0,675	0,230	1,350	1,346
4	Muhamad Dikri Permana	4,30	0,900	0,050	1,125	1,306
5	Seli Audianti Sholehah	4,15	0,219	0,710	1,812	1,404
6	Sri Sania Aryati	4,30	0,000	0,901	2,025	1,655
7	Adilla Fitri M	4,30	0,000	0,901	2,025	1,655
8	Agung Gunawan	4,30	0,600	0,304	1,425	1,368
9	Eman Sulaeman	4,25	1,350	0,461	0,677	1,292
10	Ade Juwita P	4,30	0,125	0,777	1,900	1,581
11	Dian Nurdiana	4,30	0,000	0,901	2,025	1,655
12	Didit Radiansyah	4,30	0,000	0,901	2,025	1,655
13	Dira Dwicandra	4,30	0,225	0,677	1,800	1,526
14	Erna Rosmiati	4,30	2,025	1,126	0,000	1,640
15	Herlan Ruslandi	4,30	2,025	1,126	0,000	1,640
16	Imas Sulastri	4,30	2,025	1,126	0,000	1,640
17	Indra Sunandar	4,30	2,025	1,126	0,000	1,640
18	Jeni Surahman	4,30	2,025	1,126	0,000	1,640
19	Kamaludin	4,30	2,025	1,126	0,000	1,640
20	Leman Slamet	4,30	2,025	1,126	0,000	1,640
21	Lilis Lisnawati	4,30	2,025	1,126	0,000	1,640
22	Lina Oktafiana	4,30	2,025	1,126	0,000	1,640
23	Mardian Susanto	4,30	2,025	1,126	0,000	1,640
24	Mulyawan	4,30	2,025	1,126	0,000	1,640
25	Ujang Supriyadi	4,25	0,000	0,906	2,026	1,617
26	Yudi Wahyudi	4,35	0,900	0,000	1,126	1,356

N0	Nama	Parameter				
		KINERJA	ABSENSI	C1	C2	C3
27	Hasanudin	3,00	1,025	1,356	1,640	0,000
28	Leni Marlina	3,00	1,025	1,356	1,640	0,000
29	Siti Maryam	3,00	1,025	1,356	1,640	0,000
30	Fajar Sugesti	4,30	2,230	1,331	0,205	1,773
31	Barkah Mustabah	4,30	1,580	0,682	0,445	1,414
32	Dani Hardiman	4,18	0,000	0,916	2,029	1,563
33	Dede Supriatna	4,18	0,740	0,233	1,291	1,214
34	Deni Febrian	4,30	0,440	0,463	1,585	1,426
35	Deni Wijaya	4,18	0,040	0,877	1,989	1,537
36	Hotib	4,18	0,090	0,828	1,939	1,506
37	Indra Andriawan	4,18	1,580	0,701	0,461	1,304
38	Iwan Irawan	4,30	0,580	0,324	1,445	1,374
39	Kurniawan	4,35	0,000	0,900	2,026	1,695
40	Mulyadi	4,30	0,380	0,522	1,645	1,451
41	Ohim	3,80	0,900	0,550	1,231	0,810
42	Pardiansyah	4,35	0,23	0,670	1,796	1,567
43	Rudi Suryawan	4,35	0,00	0,900	2,026	1,695
44	Sri Harsoyo	4,45	0,00	0,906	2,031	1,776
45	Susminto	4,30	0,18	0,722	1,845	1,550
46	Sutiawan	4,30	0,00	0,901	2,025	1,655
47	Ujang Wawan	4,30	0,00	0,901	2,025	1,655
48	Moch.Rahman Alfarizi	4,30	0,00	0,901	2,025	1,655
49	Bintar Alang Kibar	4,30	0,00	0,901	2,025	1,655
50	Ramdhoni Nur Akbar	4,30	0,00	0,901	2,025	1,655
51	M.Insani Haryadi	4,30	0,00	0,901	2,025	1,655
52	Teguh Pamungkas	4,30	0,00	0,901	2,025	1,655
53	Yusuf Rahman Hidayat	4,30	0,00	0,901	2,025	1,655
54	M.Syahrl Muksin	4,30	0,00	0,901	2,025	1,655
55	Genis Mulyahati Subandi	4,30	0,00	0,901	2,025	1,655
56	Reynaldi Pamungkas	4,30	0,00	0,901	2,025	1,655

**Table 5**

New Cluster

N0	Pusat Cluster Baru					
	C1		C2		C3	
	Kinerja	Absensi	Kinerja	Absensi	Kinerja	Absensi
1	4,300	0,472				
2	4,300	0,900				
3	4,300	0,675				
4	4,300	0,900				
5	4,150	0,219				
6	4,300	0,000				
7	4,300	0,000				
8	4,300	0,600				
9	4,250	1,350				
10	4,300	0,125				
11	4,300	0,000				
12	4,300	0,000				
13	4,300	0,225				
14			4,300	2,025		
15			4,300	2,025		
16			4,300	2,025		
17			4,300	2,025		
18			4,300	2,025		
19			4,300	2,025		
20			4,300	2,025		
21			4,300	2,025		
22			4,300	2,025		
23			4,300	2,025		



NO	Pusat Cluster Baru					
	C1		C2		C3	
	Kinerja	Absensi	Kinerja	Absensi	Kinerja	Absensi
24			4,300	2,025		
25	4,250	0,000				
26	4,350	0,900				
27					3,000	1,025
28					3,000	1,025
29					3,000	1,025
30			4,300	2,230		
31			4,300	1,580		
32	4,180	0,000				
33	4,180	0,740				
34	4,300	0,440				
35	4,180	0,040				
36	4,180	0,090				
37			4,180	1,580		
38	4,300	0,580				
39	4,350	0,000				
40	4,300	0,380				
41	3,800	0,900				
42	4,350	0,230				
43	4,350	0,000				
44	4,450	0,000				
45	4,300	0,180				
46	4,300	0,000				
47	4,300	0,000				
48	4,300	0,000				
49	4,300	0,000				
50	4,300	0,000				
51	4,300	0,000				
52	4,300	0,000				
53	4,300	0,000				
54	4,300	0,000				
55	4,300	0,000				
56	4,300	0,000				

Group the data according to the cluster, which is the data that has the shortest distance. Example; because  $dc3 < dc2 < dc1$  then  $x1$  is included in cluster 3. In the table above, data  $N = 1$  is included in cluster 1 because  $dc1 < dc3 < dc2$ , while  $N = 14 - 24$  is included in cluster 2 because  $dc2 < dc1 < dc3$ . After getting the cluster label for each data  $n=1, - 56$ , the average value is sought by adding up all the members of each cluster and dividing by the number of members.

Pusat Cluster	KINERJA	ABSENCE
C1	4,277	0,255
C2	4,291	1,976
C3	3,000	1,025

From the table of the results of iteration 1, it can be seen that the new centroid generated from iteration 1 with the initial center point is not the same, it is necessary to do a second iteration. Calculation of the Euclidean distance in iteration 2 The following are some of the results of the calculation of the Euclidean distance. After getting the cluster label for each data  $n=1, - 56$ , the average value is sought by adding up all the members of each cluster and dividing by the number of members.

Pusat Cluster	KINERJA	ABSENSI
C1	4,278	0,226
C2	4,289	1,934
C3	3,000	1,025

From the table of the results of iteration 2, it can be seen that the new centroid generated from iteration 2 with the initial center point is not the same, it is necessary to do a second iteration. Calculation of the Euclidean distance in iteration 3 The following are some of the results of the calculation of the Euclidean distance.

After getting the cluster label for each data  $n=1, - 56$  the average value is sought by adding up all the members of each cluster and dividing by the number of members.

Pusat Cluster	KINERJA	ABSENCE
C1	4,291	0,208
C2	4,289	1,934
C3	3,200	0,994

From the table of the results of iteration 3, it can be seen that the new centroid generated from iteration 3 with the initial center point is not the same, it is necessary to do a second iteration. Calculation of the Euclidean distance in iteration 4 The following are some of the results of the calculation of the Euclidean distance. After getting the cluster label for each data  $n=1, - 56$ , the average value is sought by adding up all the members of each cluster and dividing by the number of members.

Pusat Cluster	KINERJA	ABSENCE
C1	4,291	0,208
C2	4,289	1,934
C3	3,200	0,994

Because the Centroid has not changed (same as the previous centroid), the clustering process is complete.

#### 4. Conclusion

Based on this research, the following conclusions were obtained:

- a. It has been successfully grouped data on various types of smartphones using the K-means Clustering algorithm into 3 groups, namely for group 1 totaling 37 data, group 2 totaling 16 data, and group 3 totaling 3 data.
- b. Determination of the initial center point (centroid) is very influential on the number of iterations to be calculated.

#### 5. References

- [1] Admin DSLA. (n.d.). Surat Perjanjian Kerja. <https://www.dslalawfirm.com/surat-perjanjian-kerja/>
- [2] Muhammad Robith Adani. (n.d.). Artificial Intelligence: Definisi, Sejarah, Contoh, dan Manfaat. Januari 13, 2021. <https://www.sekawanmedia.co.id/apa-itu-artificial-intelligence/>
- [3] Syafnidawaty. (2020). K-means Clustering - UNIVERSITAS RAHARJA (pp. 1–1). <https://raharja.ac.id/2020/04/19/k-means-clustering/>
- [4] Santosa, B. (2007). Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- [5] Santoso, S. (2010). Statistik Multivariat. Jakarta: Elex Media Komputindo.
- [6] Wu, X. and Kumar, V. (2009). The Top Ten Algorithms in Data Mining. London: CRC Press Taylor & Francis Group.
- [7] Prasetyo, Eko. (2014). Data Mining - Mengolah Data Menjadi Informasi menggunakan Matlab. Yogyakarta: CV. Andi Offset.
- [8] Tahta, Budi, dan Ali. 2013. Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Custer Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS). Jurnal Skripsi Teknik Industri ITS. Surabaya: Teknik Industri, Institut Teknologi Sepuluh November (ITS).
- [9] Witten, Ian H. dan Frank, Eibe. 2005. Data Mining Practical Machine Learning Tools and Techniques, Second Edition. Morgan Kaufmann, San Fransisco
- [10] Agusta, Y. (n.d.). K-Means Rehat with Yudi Agusta. <https://yudiagusta.wordpress.com/k-means/>
- [11] Abdyo and H. Permadi, Metode Statistika Praktis, Malang: Universitas Negeri Malang, 1999.
- [12] B. P. S. P. Maluku, “Indikator Indeks Pembangunan Manusia 2010-2014,” Badan Pusat Statistik Provinsi Maluku, Juli 2014. [Online]. Available: <https://maluku.bps.go.id/subject/26/indeks-pembangunan-manusia.html>. [Accessed Juli 2014].



- [13] B. P. S. P. Maluku, “Indikator Kesejahteraan Rakyat Provinsi Maluku 2014,” Badan Pusat Statistik Provinsi Maluku,[Online].Available:  
<https://maluku.bps.go.id/publication/2015/09/30/6cc18acb861407726a19addc/indikator-kesejahteraan-rakyat-provinsi-maluku-2014.html>.
- [14] Windha Mega P. D., “Clustering Menggunakan Metode K -Means Untuk Menentukan Status Gizi Balita,” *Jurnal Informatika*, vol. 15, no. No.2, pp. 160-174, 2015.
- [15] Yulianto, Ramadiani, & Kridalaksana, A. H. (n.d.). (PDF) Penerapan Formula Haversine Pada Sistem Informasi Geografis Pencarian Jarak Terdekat Lokasi Lapangan Futsal.
- [16] [https://www.researchgate.net/publication/324727225\\_Penerapan Formula Haversine Pada Sistem Informasi Geografis Pencarian Jarak Terdekat Lokasi Lapangan Futsal](https://www.researchgate.net/publication/324727225_Penerapan_Formula_Haversine_Pada_Sistem_Informasi_Geografis_Pencarian_Jarak_Terdekat_Lokasi_Lapangan_Futsal)