Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Clustering Of Polri Bintara Placement In North Sumatera Regional Police Using K-Means Algorithm

Ririn Pebrina Br. Marpaung[1], Hengki Tamando Sihotang[2]

[1,2]Informatics Engineering,
STMIK Pelita Nusantara, Jl. Iskanda Muda, Medan, Indonesia

E-mail: marpauungririn@gmail.com, hengki_tamando@yahoo.com

ARTICLE INFO

Article history:
Received: 19 Aug 2019
Revised: 27 Aug 2019
Accepted: 20 Sep 2019

***Keywords**: Regional Police, data mining, K-Means algorithm, Placement Value*

ABSTRACT

Data mining on process carried out to obtain information from a database or data that can be used to help solve the latest problems or solutions, data mining that is used in this paper is the process of merging by using K-Means solutions. K-Means is one of the techniques used to group non-hierarchical (bulk) data which is supported to provide existing data partition in the form of two or more groups. This method partitioned the data in groups so that the different characteristic data was grouped into other groups. The purpose of grouping this data is to support the objective functions arranged in the grouping process, which generally support variations between groups and take advantage of variations between groups. The agreed clustering was the grouping of non-commissioned police officers in the North Sumatra regional police, with the data collection used was the placement data within the North Sumater Regional Police HR. The procedure that is carried out in this research is the problem process to the design and testing of the program. The knowledge gained from the grouping of the Bintara Police of the National Police in the North Sumatra Regional Police HR is the 5th data Placement based on data collected related to the position of the Brig Ro Sarpras of the North Sumatra Regional Police, as well as related to the data analysis with the K-Intended Algorithm in the North Sumatra Police Brigade. Based on the analysis of the latest number of changes based on the calculation of the K-mean algorithm ie the value 79-100 Being the range for the First cluster, the range 70-78 becomes the second cluster and 60-69 is categorized as the Third cluster. To produce a new pattern, a data mining process is carried out with different data..

1. Introduction

Current technological developments need for accurate information is needed [there is a daily environment, so that part of the information will be a very important part in carrying out existing developments in society, but the high usefulness of information is sometimes not adjusted to the process of presenting information that is adequate and appropriate. The ability of the results of the development of information technology to ensure and store various types of data quite a lot and ensure the results obtained based on the ability to analyze, summarize and extract knowledge from the data. The traditional method for analyzing existing data cannot handle large amounts of data.

Utilizing existing data in the information technology environment to support activities in the presentation and decision making process or action, it is not enough to rely solely on operational data, we need a thought to conduct data analysis so that the potential information contained in the data pile can be found. Decision makers or stakeholders always try to utilize existing data warehouses to gather useful information to help retrieve the data needed, this encourages the emergence of new branches of science to

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

overcome the problem of extracting information or important or interesting patterns from large amounts of data, called with data mining. The use of data mining techniques is expected to provide knowledge previously hidden in the data warehouse so that it becomes valuable information.

Regional Police are tasked with carrying out the tasks of the National Police at the regional level level I such as the Province or Special Region. Polda is a direct extension of the National Police Headquarters. The Regional Police are led by the Head of the Indonesian National Police (Kapolda), who reports to the National Police Chief. One of the activities held by the North Sumatra Regional Police is the selection of prospective members of the National Police with the principles of Clean, Transparent, Accountable and Humanist. Through a long and time-consuming process, it will be announced who is eligible to take part in education and be appointed as a police officer. In education, improvements will be made to students about the tasks that will be done and carried out in the field.

During Panatia's education in this case the SPN will conduct periodic assessments of several aspects that will be used as a reference in determining the Best students, as well as the placement of a police officer based on their respective abilities and expertise. At the end of the education awaited by a student is the inauguration ceremony and placement, after the announcement is complete the data is stored and not used anymore for subsequent reception so that there is accumulation of data from year to year, in this thesis data mining is performed for placement in the field of human resources).

Related research proposed by fina nasari et al with the title "Application of the K-Means Clustering Algorithm for Grouping the Diarrhea Spread in Langkat District" explains that data mining with the K-Means black algorithm is capable of grouping according to the closest data so that it can be carried out periodically for the problem of the problem of diarrhea. there is, based on the description will be carried out the process of finding new information from data that has been left untouched and connecting with the placement of non-commissioned police officers with the procedure of conducting clustering and providing a new recommendation on the placement of police Bintara.

2. Basic Theory

2.1 Data Mining

According to tan (2006) in eco book Prasetyo (2017) gives a response that data mining as a process that can be utilized in producing information that is useful and useful for large database warehouses. Data mining can also be interpreted as part of the process of extracting or extracting new information taken from a collection of very large data piles so that it can help in decision making. The term data mining is often also called knoweledge discovery. One of the facilities created in data mining is how to search existing data to build and plan a model. The next step is to use the model in order to recognize other data patterns that are not in the stored database. The prediction section can also use data mining techniques. In data mining, data repair and grouping procedures can also be applied. The aim is to find and find out the universal pattern of available data.

2.2 Clustering Method

Berkhin (2006) said that one technique from several techniques that can be applied in KDD is clustering. Clustering is the occurrence of dividing data into groups that have the same object and characteristics. Garcia states clustering is a task that is used to group data items into groups that have similar attributes so that each group has something in common. (Asroni, 2015)

Clustering refers to grouping notes, observations, or cases into similar classes. A cluster is a collection of notes that are similar to each other and different from notes in another cluster. Clustering differs from classification in that there are no target variables for clustering. In contrast, the clustering algorithm searches for data segments throughout the set into relatively homogeneous subgroups or groups, where the similarity of notes in the cluster is maximized, and the similarity of notes outside the cluster is minimized (Daniel t. Larose, 2005) in Taslim's Posts, Fajar (2016).

2.3 K-Means algorithm

K-Means is one of several algorithms used for grouping nonhierarchical (bulk) data that attempts to provide existing data parcels into two or more groups. This method partitioned existing data into groups so that different characteristic data were grouped into other groups. Eko Prasetyo (2017). Based on the source and the book, in general the steps of K-Means by following the following procedure:

a. Determine the number of groups
b. Allocate data into groups randomly

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

c. Calculate the Group Center (Centroid / average) of the data in each group.
d. Allocate each data to the closest centroid / average
e. Doing repetition, if there is still data that moves groups or if there is a change in the centroid value above the specified threshold value, or if there is a change in the value of the objective function used is still above the specified threshold value.

In carrying out these steps by following some of the activities or formulas that have been presented, namely in the determination of the location of the centroid (center point) each group is taken from the mean (mean) of all data values on each feature must be recalculated, if M states the amount of data in a group, i represents the i-th feature in a group. And p states that the dimensions of the data are used to calculate the centroid of the i-th feature below

$$C_i = \frac{1}{M}\sum_{j=1}^{M} X_j \quad (1)$$

The formula is carried out as many as p dimensions so that i starts from 1 to p. There are several methods that can be used to measure the distance from the data to the center of the group
Measurement of distance in Euchdean distance space using formulas...

$$D(X_2, X_1)=\|(X_2.X_1)\|_2=\sqrt{\sum_{j=1}^{p} |X_{2j} - X_{1j}|^2} \quad (2)$$

D is the distance between x2 and x1 and |. | Data is the absolute value of distance measurement in Manhattan space using a formula.

$$D(X_2, X_1)=\|(X_2.X_1)\|_1=\sum_{j=1}^{p} |X_{2j} - X_{1j}| \quad (3)$$

Pengukuran jarak pada ruang jarak Minkowsky menggunakan formula

$$D(X_2, X_1)=\|(X_2.X_1)\|\alpha = \sqrt{\sum_{j=1}^{p} |X_{2j} - X_{1j}|^{\alpha}} \quad (4)$$

λ is the same distance parameter as Manhattan. If λ = 1, the spacing, the spacing at Minkowskyk Minkowsky is the same as Manhattan. If λ = 2, the distance will be the same as Euclidean; if λ = ∞, the distance will be the same distance as the Chebyshev space. However, the most widely used methods are Euclidean and Manhattan. Euclidean is an option if we want to give the shortest distance between two points (straight distance), while Manhattan gives the farthest distance on two data. Manhattan is also often used because of its ability to detect special circumstances, such as the presence of outliers, better.
(Agusta 2005)

Pada langkah 4, Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan sentroid setiap kelompok yang ada. Data dialokasiakan ulang secara tegas ke kelompok yang mempunyai sentroid dengan jarak terdekat dari data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut. *(MacQueen 1967)*

$$Ai|=\begin{cases}1 & d = \min[D(xi, C1)] \\ 0 & lainnya\end{cases} \quad (5)$$

A1 adalah nilai keanggotaan titik x ke pusat kelompok C_1 , d adalah jarak terpendek dari data x ke K Kelompok setelah dibandingkan dan C1 adalah sentroid (Pusat kelompok) ke – 1.
Fungsi objektif yang digunakan untuk K-means ditentukan berdasarkan jarak dan nilai keanggotaa data dalam kelompok. Fungsi objektif yang digunakan adalah sebagai berikut :

$$J=\sum_{i=1}^{N},\sum_{I=1}^{k} d(x_1, C_1)^2 \quad (6)$$

3. Research Methods

N is the amount of data, K is the number of Groups, ail is the value of membership of the data point x to the center of the Ci group, c1 is the Center of the 1st group and D (x1, C1) is the distance of the point x to the group C1 followed. a has a value of 0 or 1. If a data is a member of a group, aij value = 1, if not, then the value = 0.

In supporting the transparency of the process of making research by following the steps of the implementation from the analysis of the problem to the implementation and testing and drawing conclusions by following the flow below

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

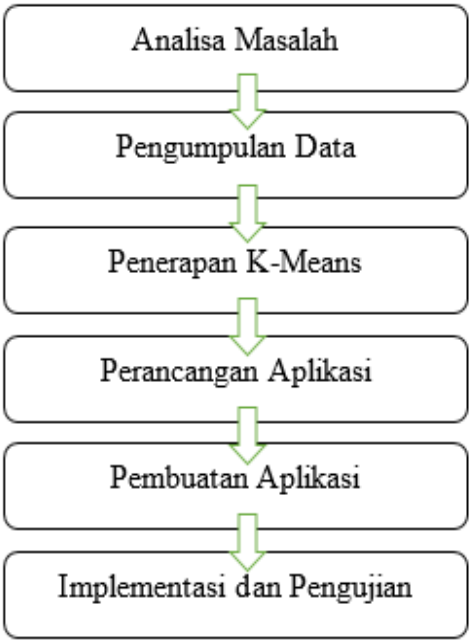

Figure 1. Research Work Steps

4. Results and Discussion

4.1 Data

The placement of non-commissioned police officers in the North Sumatra Regional Police is based on the needs of each field, as well as the HR Placement process by checking personnel needs and reporting to the committee to provide recommendations according to the required amount with the required range values, which are in the following table:

Table 1.
Placement Range

Range	Placement
80-100	BRIG RO SARPRAS POLDA SUMUT
70-79	BRIG RO OPS POLDA SUMUT
60-69	BRIG RO SDM POLDA SUMUT

The data used as testing in the application as much as 50 data and in testing manually using the K-Means Clustering algorithm as many as 12 Personnel, the data used is interpreted in table 2 below.

Table 2.
Placement of Police Bintara

Number	Name	Grade	Placement	Value
1	Deni hasibuan	Aiptu	Brig ro sarpras polda sumut	90
2	Saparuddin saragih	Aiptu	Brig ro sarpras polda sumut	94
3	Johan ginting	Aipda	Brig ro sarpras polda sumut	90
4	Faisal sagala	Bripka	Brig ro sarpras polda sumut	86
5	Tonggor h. Siregar	Brigadir	Brig ro sarpras polda sumut	80
6	Frans boyke s. Siagian	Briptu	Brig ro sarpras polda sumut	90
7	Eky wardana	Briptu	Brig ro sarpras polda sumut	85
8	Darpin simarmata	Brigadir	Brig ro ops polda sumut	79
9	Tomi yan pramadya	Brigadir	Brig ro ops polda sumut	78
10	Amanda amri hasibuan	Briptu	Brig ro ops polda sumut	76
11	Lilis setiawati	Aiptu	Brig ro sdm polda sumut	68
12	Novriyanti sidauruk, s.psi	Brigadir	Brig ro sdm polda sumut	65

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

4.2 Analysis of Clustering Algorithms by K-Means

After the data presented for the needs in the analysis then proceed with the K-Means process, the following is the process and calculation of the K-Means algorithm, in the K-Means algorithm expressed by the principle of rotation until the results are the same as the output described in step 5 of K-Means

a. Determine the Number of Groups

In this step the number of placement groups is determined according to the needs of the problem discussed, in this problem three groups (Clusters) are used with the highest value conditions in cluster I, the Middle Value as part of Cluster II and the Lowest value as Cluster III. Descriptions of the following groups:

a) Cluster I (C1) was taken from data 2 with a distribution of values of 94; 3
b) Cluster II (C2) is taken from the 9th data with a value distribution of 79; 2
c) Cluster III (C3) is taken from the 12th data with a 65: 1 Value Distribution

b. Random Data Allocation

At this stage is the formation of a new cluster table randomly from the data presented illustrated with the following table below:

Table 3.
Formation of Random Tables

Number	*Name*	*Value*	*Placement*	*C1*	*C2*	*C3*
1	Deni Hasibuan	90	3	C1		
2	Saparuddin Saragih	94	3	C1		
3	Johan Ginting	90	3	C1		
4	Faisal Sagala	86	3	C1		
5	Tonggor H. Siregar	80	3	C1		
6	Frans boyke S. Siagian	90	3	C1		
7	Eky Wardana	85	3		C2	
8	Darpin Simarmata	79	2		C2	
9	Tomi Yan Pramadya	78	2		C2	
10	Amanda Amri Hasibuan	76	2		C2	
11	Lilis Setiawati	68	1			C3
12	Novriyanti Sidauruk	65	1			C3

Calculating Group Centers

To measure the distance between the center of the Cluster, Euclidian Distance is used, then the distance matrix that is C1, C2 and C3 will be obtained using the following formula:

$D(X_2, X_1)=||(X_2.X_1)|| = \sqrt{\sum_{j=1}^{p} |X_{2j} - X_{1j}|^2}$

Pengujian Pada data I

C1. $d(90,3) =|(94,3||| = \sqrt{(90-94)^2 + (3-3)^2} = 4$

C2. $d(90,3)=|(79,2||| = \sqrt{(90-79)^2+(3-2)^2}$

$= \sqrt{(13)^2+(1)^2} = 11, 04536$

C.3 $d(90,3)=||(65,1)|| = \sqrt{(90-65)^2+(3-1)^2}$

$= \sqrt{(25)^2+(2)^2}$

$= \sqrt{(625) +(4)} = \sqrt{629} = 25.0799$

The closest number is 7.071068 on C2

So that the data generated for testing in the first data and being the shortest distance from the test is C1 with a value of 4 and the second data is the shortest distance is 0 so that in the same way will produce the data in the following table Table 4

Table 4.
Shortest Distance

Number	*Value*	*Placement*	*C1*	*C2*	*C3*	Shortest Distance
1	90	3	4	11,04536	25,07987	4
2	94	3	0	15,0333	29,06888	0
3	90	3	4	11,04536	25,07987	4

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Number	*Value*	*Placement*	*C1*	*C2*	*C3*	Shortest Distance
4	86	3	8	7,071068	21,09502	7,071067812
5	80	3	14	1,414214	15,13275	1,414213562
6	90	3	4	11,04536	25,07987	4
7	85	3	9	6,082763	20,09975	6,08276253
8	79	2	15,03329638	0	14,03567	0
9	78	2	16,03121954	1	13,0384	1
10	76	2	18,02775638	3	11,04536	3
11	68	1	26,07680962	11,04536	3	3
12	65	1	29,06888371	14,03567	0	0

Allocate each data to the centroid
Distance calculation results will be compared and selected the closest distance between the data with the center of the cluster, this distance indicates that the data is in one group with the closest cluster center. The following will display the grouping matrix data, the value of 1 means that the data is in a group (data group).

Table 5.
Grouping Iteration Data I

Number	*Value*	*C1*	*C2*	*C3*
1	90	1		
2	94	1		
3	90	1		
4	86		1	
5	80		1	
6	90	1		
7	85		1	
8	79		1	
9	78		1	
10	76		1	
11	68			1
12	65			1

In the same way, it is repeated until the value in the first process is the same as the current process, with the following results:

Table 6.
Grouping of new data in Iteration IV

No	*Value*	*C1*	*C2*	*C3*
1	90	1		
2	94	1		
3	90	1		
4	86	1		
5	80		1	
6	90	1		
7	85	1		
8	79		1	
9	78		1	
10	76		1	
11	68			1
12	65			1

From the grouping in the second round produced a new group of data 1,2,3,4,6 and 7 which were stated in cluster I, 5, 8,9,10 in Cluster II and data 11.12 in Cluster III position.
The testing process is complete because it is in accordance with the K-means Algorithm Principle if the Grouping Process is the same as the previous round then the process is stopped. So that the final results obtained are grouping in the last round, the following table is the result of the calculation process with a comparison of the initial placement and placement after using the K-Means Algoritam

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Table 7
Comparison of Placement with K-Means Clustering

Number	*Name*	*Initial Placement*	*Placement with K-Means*
1	Deni Hasibuan	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
2	Saparuddin Saragih	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
3	Johan Ginting	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
4	Faisal Sagala	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
5	Tonggor H. Siregar	Brig Ro Sarpras Polda Sumut	Brig Ro Ops Polda Sumut
6	Frans boyke S. Siagian	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
7	Eky Wardana	Brig Ro Sarpras Polda Sumut	Brig Ro Sarpras Polda Sumut
8	Darpin Simarmata	Brig Ro Ops Polda Sumut	Brig Ro Ops Polda Sumut
9	Tomi Yan Pramadya	Brig Ro Ops Polda Sumut	Brig Ro Ops Polda Sumut
10	Amanda Amri Hasibuan	Brig Ro Ops Polda Sumut	Brig Ro Ops Polda Sumut
11	Lilis Setiawati	Brig Ro Sdm Polda Sumut	Brig Ro Sdm Polda Sumut
12	Novriyanti Sidauruk, S.Psi	Brig Ro Sdm Polda Sumut	Brig Ro Sdm Polda Sumut

The knowledge gained from the analysis of the Police Grouping of the National Police in the North Sumatra Regional Police HR is the 5th data. . based on the analysis, there is a change in the latest range based on the calculation of the K-means algorithm, which is a value of 79-100. Being the range for the First cluster, the range 70-78 becomes the second cluster and 60-69 is categorized as the Third cluster. The knowledge generated from the analysis will be used for the placement of newly graduated with the process of adjusting the cumulative value generated and the data mining process is carried out again to generate new patterns on an ongoing basis.

4.3 Results

a. K-Means Process

Display k-means process is a process with the K-Means algorithm to group the placement of non-commissioned officers, with some procedures that must be performed is to display data will automatically display data stored in databases, while the type of cluster is used to select the number of clusters used and proceed by selecting the centroid data capture button and selecting the process. The process display is illustrated by the display below:

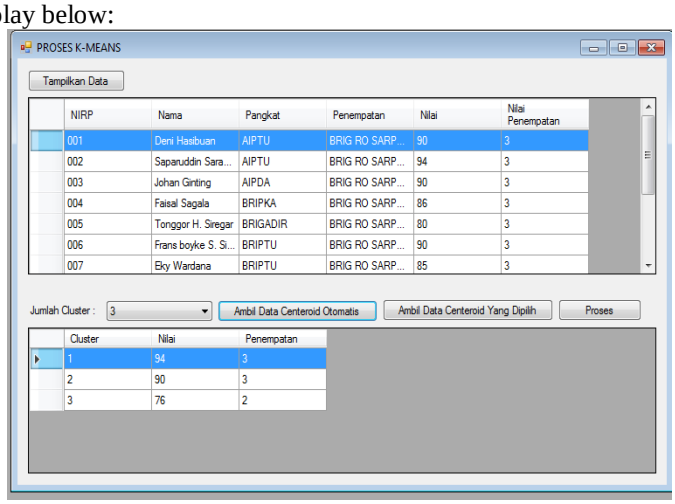

Figure 2. K-Means Process

b. Iteration Results

Cluster display is the position of the data closest to the selected cluster, grouping and proximity to the most recent iteration is the end of calculations, to further clarify the figure shown below:

Jurnal Mantik
Volume 3 Number 3, Nopember 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

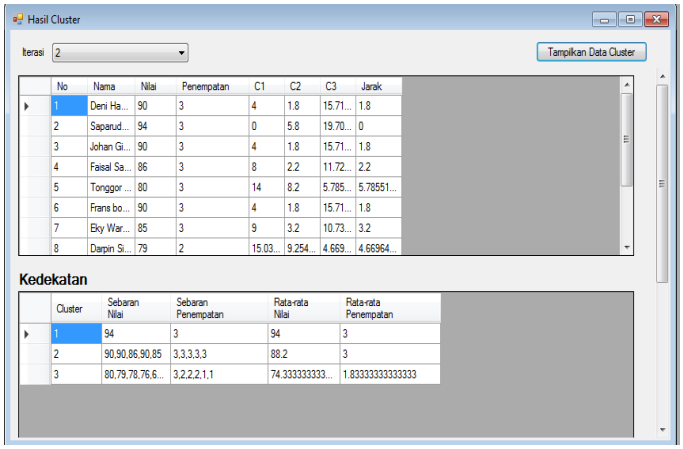

Figure 3. Display of Iteration Results

c. The final result

Display the final results of the application developed, with some specific buttons namely change the cluster name to support grouping, save to save new data and exit button to return to the main menu, the following is a display of the final results on placement.

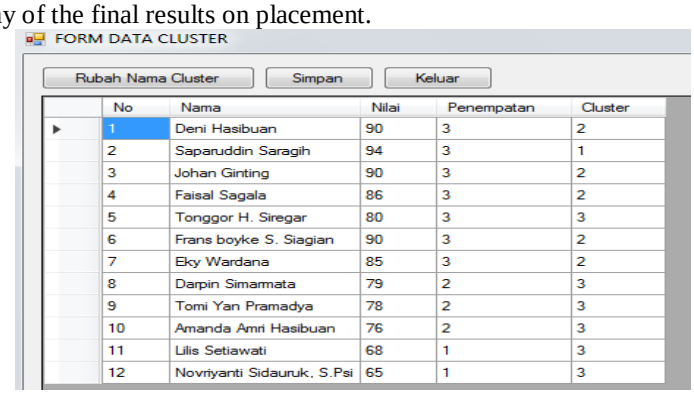

Figure 4. Results of Clustering

5. Conclusion

Based on what has been described in the introduction and the results obtained in the analysis and design, the conclusion is

1. The resulting pattern can be used as a non-commissioned police officer placement that will be placed on HR (Human Resources)
2. K-Means used in Grouping obtain results that are closer based on the calculation and proximity of an attribute.
3. The results of experiments conducted in this study indicate that the K-Means clustering algorithm performs grouping by dividing data into the specified number of k clusters, and utilizing distance calculations to measure the similarity between data. The choice of cluster number (k) greatly influences the quality of the cluster. The more the number of clusters (k) the results will be more accurate.

6. References

[1] Eko Prasetyo, *Data Mining : Konsep Dan Aplikasi Menggunakan Matlab*. 2013.
[2] F. Nur, M. Zarlis, and B. B. Nasution, "PENERAPAN ALGORITMA K-MEANS PADA SISWA BARU SEKOLAHMENENGAH KEJURUAN UNTUK CLUSTERING JURUSAN," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, 2017.
[3] F. Nasari and S. Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru," *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, 2015.
[4] A S Sinaga and A S Girsang, University Accreditation using Data Warehouse, Journal of Physics: Conference Series, Volume 801, Number 1 2017.

[5] T. Taslim and F. Fajrizal, "Penerapan algorithma k-mean untuk clustering data obat pada puskesmas rumbai," *Digit. Zo. J. Teknol. Inf. dan Komun.*, 2016.
[6] Ediyanto, M. N. Mara, and N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metod K-Means Cluster Analysis," *Bul. Ilm.*, 2013.
[7] I. A. Musdar and A. SN, "Metode RCE-Kmeans untuk Clustering Data," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, 2015.
[8] W. F. Astuti, I. Agusta, and M. Siwi, "The Impact of the Activities of Illegal Gold Mining for Household Welfare Gurandil," *J. Sains Komun. dan Pengemb. Masy. [JSKPM]*, 2017.
[9] Haviluddin, "Memahami Penggunaan UML (Unified Modelling Language)," *Memahami Pengguna. UML (Unified Model. Lang.*, 2011.
[10] 2016 Hendini Ade, "Pemodelan UML sistem informasi Monitoring Penjualan dan stok barang," *Pemodelan Uml Sist. Inf. Monit. Penjualan Dan Stok Barang (Studi Kasus Distro Zhezha Pontianak)*, 2016.
[11] Elmawati and S. Angga, "Perancangan Aplikasi Pemesanan Menu Makanan Dan Minuman Pada Cafe Living Room Bukittinggi," *J. Sains dan Teknol.*, 2017.
[12] U. F. Y. Muhammad Dahria, Ishak, "Pendukung Keputusan Seleksi Calon Polri Baru Di Polda Kota Medan Menggunakan Metode Multifactor Evaluation Process (MFEP)," *J. Ilm. Saintikom*, 2014.