Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of address recording management using the K-Means clustering classification algorithm in Kebayoran district, DKI Jakarta

Rusdiansyah[1)*], Harun Al Rasyid[2)], Suryanto Sosrowidigdo[3)]
[1)2)3)] Univeristas Bina Sarana Informatika, Indonesia
[1)]rusdiansyah.rds@bsi.ac.id, [2)]harun.har@bsi.ac.id, [3)]suryanto.sys@bsi.ac.id

Submitted : Jan 26, 2021 | **Accepted** : Feb 17, 2021 | **Published** : Apr 1, 2021

Abstract: The area is the center of problems in the administrative record management of Kebayoran District, because of its dense condition and it is difficult to determine land measurements due to the density of residential areas. The problem in Indonesia to this day is that the administrative boundaries of the kelurahan already exist, but the administrative boundaries for the Rukun Warga / Rukun Tetangga (RW / RT) do not yet exist. The local government of DKI already has a large scale map (1: 1,000) to map RW administrative boundaries. Large-scale mapping (Batas RW) is useful for accurate information on incidence of dengue fever or other diseases, thereby eliminating information bias due to the use of village boundary maps. Another benefit is the accuracy of address management for customers, for example PDAM customers, to facilitate verification of customer data with large-scale maps, especially those that only include RT / RW addresses, without mentioning street names and household numbers. The method used is data mining K-Means Clustering. By using this method, the data that has been obtained can be grouped into several clusters, where the application of the KMeans Clustering process uses Excel calculations. The processed data is divided into 3 clusters, namely: high cluster (C1), medium cluster (C2) and low cluster (C3). The iteration process of this research occurs 2 times so that an assessment is obtained in classifying the household / neighborhood unit based on the Kelurahan. The results obtained are that there is 1 neighborhood unit with the highest cluster (C1), there are 4 neighborhood units with 4 medium clusters (C2), and 5 neighborhood units with the lowest cluster (C3). This data can be input to the sub-district to disseminate information about dengue fever, health education, and for the accuracy of PDAM customer address management and others**.**

Keywords: RT / RW, Kecamatan, Data mining, Group, K-means Algorithm

INTRODUCTION

The author takes previous research references, uses the K-Means algorithm to group or cluster villages in Garoga District, North Tapanuli Regency based on similarities in regional characteristics in terms of health indicator values, namely infant and under-five mortality rates, infant and under-five morbidity rates, and nutritional status of infants and toddlers. Clustering is grouping data, observations, or cases into classes from similar objects(Sitohang & Rikki, 2019).

The concept of data mining is increasingly recognized as an important tool in information management because of the increasing amount of information (Alfina & Santosa, 2012). Data mining itself is often referred to as knowledge discovery in database (KDD), which is an activity that includes collecting, using historical data to find regularities, patterns of relationships in large data sets.(Santoso, Hariyadi, & Prayitno, 2016). The output from data mining can be used for future decision making (Tana, Marisa, & Wijaya, 2018).

That in an effort to increase the effectiveness of administering government services, development and community fostering in villages within the district area, it is deemed necessary to regulate the guidance and arrangement of Neighborhood Associations and Citizens' Rukun Tetangga and Rukun Warga as one of the social

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

institutions established through deliberations and / or elections in a better, orderly manner. and regular(Belanja, Apbdes, & Desa, 2020).

Rukun Tetangga (RT) is a division of territory in Indonesia under Rukun Warga. Rukun Tetangga is not part of the division of government administration, and its formation is through local community deliberations in the framework of community services determined by the Village or Kelurahan (Widodo, Utomo, & Miranti, 2009). Rukun Warga (RW) is a Community Institution formed through deliberation between the Rukun Tetangga (RT) administrators in their working area in the framework of government and community services that are recognized and fostered by the Regional Government as stipulated by the Village Head.(Andi IchsanN Syamsul, 2018). RT and RW that do not have a structure chart, institutions, regulations regarding RT and RW institutions, activity program planning, and do not have a RT and RW secretariat as a center and institutional administrative service, so generally the secretariat of these RT and RW institutions hitch a ride with Pos Ronda, a resident's house. and other public facilities(Samsudin, Djaeni, Idris, & Fathoni, 2017).

The problem in Indonesia to this day is that the administrative boundaries for the kelurahan already exist, but the administrative boundaries for the Rukun Warga / Rukun Tetangga (RW / RT) do not yet exist(Wibowo & Sudarmaji, 2010). The local government of DKI already has a large-scale map (1: 1,000) to map RW administrative boundaries. Large-scale mapping (Batas RW) is useful for accurate management of addresses of customers, for example PDAM customers, to facilitate verification of customer data with large-scale maps, especially those that only include RT / RW addresses, without mentioning street names and house numbers. The author uses the K-Means method with clustering techniques to classify the address data of the number of RT / RW, Kebayoran District. Clustering is grouping data, observations, or cases into classes from similar objects. Clustering is different from classification, in that there is no target clustering variable. Grouping does not attempt to classify estimates, or predict the value of the target variable. In contrast, the clustering algorithm seeks to find parts of the entire dataset into relatively homogeneous subgroups or groups, where the similarity of data within the cluster is maximized, and the similarity of data outside the cluster is minimized(Metisen & Sari, 2015).

LITERATURE REVIEW

Data Mining

Data mining is the mining or discovery of new information by looking for certain patterns or rules from a very large amount of data (Santoso et al., 2016). Data mining is also referred to as a series of processes to explore added value in the form of knowledge that has not been known manually from a data set (Ginting & Trinanda, 2013) Data mining, often referred to as knowledge discovery in database (KDD). KDD is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets(Panggabean, Buulolo, & Silalahi, 2020).

Clustering or clustering is a data mining technique used to analyze data to solve problems in grouping data or more precisely partitioning a dataset into subsets.(Ali, 2019). In the clustering technique the target is to distribute cases (objects, people, events and others) into a group, so that the degree of connection between members of the same cluster is strong and weak between different cluster members.(Sihombing, 2017). The cluster technique has two methods in grouping it, namely hierarchical clustering and non-hierarchical clustering. Hierarchical clustering is a method of grouping data that works by grouping two or more data that have similarities or similarities, then the process is continued to other objects that have two closeness, this process continues until the cluster forms a kind of tree where there is a clear hierarchy or level between objects from most similar to least similar(Ramadhan, Mustakim, & Efendi, 2017). However, logically, all objects in the end will only form a cluster (Ong, 2013).

K-Means Algorithm

K-Means is an algorithm used in partitioning that separates data into different groups - based. This algorithm is able to minimize the distance between the data to the cluster. Basically, the use of this algorithm in the clustering process depends on the data obtained and the conclusions to be reached at the end of the process(Gading Sadewo, Perdana Windarto, & Wanto, 2018). Basically, the k-means algorithm only takes a part of the number of components obtained and then becomes the initial cluster center, in determining the center of this cluster is randomly selected from the data population. Then the k-means algorithm will test each of each component in the data population and mark these components into one of the predefined cluster centers depending on the minimum distance between components and each cluster center.(Ali, 2020). Furthermore, the position of the center of the cluster will be recalculated until all data components are classified into each - each cluster and finally a new cluster will be formed. (Sihombing, 2017).

The K-Means algorithm basically performs 2 processes, namely the process of detecting the location of the center of the cluster and the process of searching for members of each cluster. The clustering process begins by identifying the data to be clustered, Xij (i = 1, ..., n; j = 1, ..., m) where n is the amount of data to be clustered and m is the number of variables(Maulida, 2018). At the beginning of the iteration, the center of each cluster is determined independently (arbitrarily), Ckj (k = 1, ..., k; j = 1, ..., m). Then the distance between each data and

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

each cluster center is calculated. To calculate the distance to the I-th data (xi) at the center of the k-th cluster (ck), named (dik), the Euclidean formula can be used. A data will be a member of the k-th cluster if the distance of the data to the center of the k-cluster is of the smallest value when compared to the distance to the center of another cluster.(Ali, 2020).

The following are the steps in the K-means algorithm (Khotimah, 2014).

1. Determine the number of clusters (k) in the data set
2. Determine the center value (centroid)

Determination of the value of the centroid at the initial stage is carried out randomly, while in the iteration stage the formula is used as in the equation (1) the following:

$$Vij = \frac{1}{Ni} \sum_{k=0}^{Ni} Xkj \quad (1)$$

Information :

Vij = centroid the I-th cluster average for the j-variable
Ni = The number of members of the ith cluster
i, k = index of the cluster
j = the index of the variable
Xkj = the k-th data value of the j-th variable for the cluster

3. On each record, calculate the distance closest to the centroid. The centroid distance used is Euclidean Distance, with the formula as in the equation (2) :

$$De = \sqrt{(xi - si)^2 + yi - ti)^2} \quad (2)$$

Information:

De = Euclidean Distance
i = The number of objects[2]
(x, y) = Object coordinates
(s, t) = Koordinat centroid

4. Group objects based on the distance to the nearest centroid
5. Repeat step 2, iterating until the centroid is optimal

The framework is an outline of the steps of the research being carried out, the frame of mind is used as a reference for carrying out the steps that are being carried out in the K-means algorithm research(Wardhani, Anindya Khrisna, 2016), shown in Figure 1 of the k-means framework.

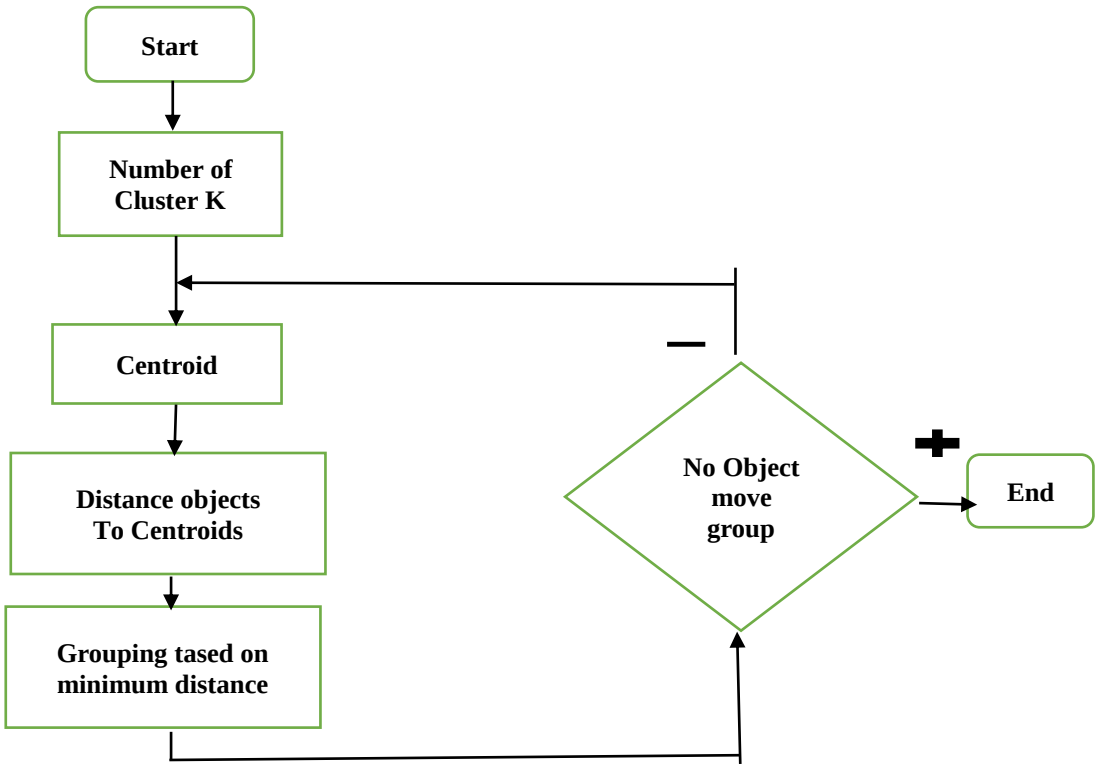

Fig.1 K-means framework

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Alternative Mixture Modeling Methods

Mixture modeling [is one type of clustering data where in modeling, data in a group is assumed to be distributed according to one type of statistical distribution that exists. Mixture Modeling is a method that has the same optimization method as K-Means through the optimization and maximization process. Unlike the Hard K-Means and Fuzzy K-Means methods, the comparison of the amount of data included in each cluster also affects the final result of the data clustering process. Comparison of the amount of data contained in each cluster is often termed relative abundance(Yudi Agusta, 2007).

METHOD

Data Collection Stage

Data collection for the Kebayoran district from data.jakarta.go.id. All data has been recorded at the address site (Dinas Komunikasi, Informatika, 2020), The author took the sample data of the Kebayoran district which consists of 10 villages. The variables used are the number of neighborhood associations (RT) and community associations (RW) according to the districts of DKI Jakarta Province. The data will be processed by clustering which is divided into 3 clusters, namely the cluster status of the number of RTs and the number of RWs with high-level clusters with medium levels and cluster status with low levels. The clustering method used in this study is the K-Means method

Data Processing Stage

The data that has been processed will be processed first to be clustered. In the previous stage, data from the Kebayoran District from the Number of RTs and the Number of RWs according to the DKI Jakarta Province in 2019, 10 Kelurahan by taking the average value of each number of RTs and the number of RWs so that at this stage the calculation of the value will be processed at the clustering stage.

A. Analysis Phase

At this stage, data analysis of Rukun Tangga and Rukun Warga according to Kebayoran District, DKi Jakartadan Province is carried out. The status has been determined to be clustered into 3 clusters, namely high-level clusters, medium-level clusters and low-level clusters. It is at this stage that the results will be analyzed in Table 1.

Table 1
Data The number of household RT / RW
Kebayoran District in 2019

Districts	Number of RT	Number of RW
MELAWAI	30	4
GUNUNG	68	7
KRAMATPELA	82	10
SELONG	33	4
RAWA BARAT	44	7
SENAYAN	19	3
PULO	46	6
PETOGOGAN	76	6
GANDARIA	147	15
CIPETE	103	11

RESULT

Implementing the K-Means algorithm using Microsoft Excel requires several stages of the process as follows:

Determine the Cluster Center Point

In table 1 above, the determination of the cluster center point or it can be called the centroid is used as a subtraction value for calculating the distance between the data to each cluster or called the distance. In this process, the determination of the cluster center point value can be determined using a random method as desired, provided that the centroid value is still included in the range of data values for each attribute. In addition, the determination of the cluster center point value can be determined using the average value in each attribute. In this process the authors determine the value of the cluster center point using a random method as desired as follows:

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A. Determination of the Initial Cluster Center

Take the 9th data as the center of the 1st cluster	147	15
The 8th data is taken as the second cluster center	76	6
The 6th data is taken as the center of the 3rd cluster	19	3

Based on Table 1 above, it can be explained that cluster 1 has a centroid value in the attribute number of RT 147 and number of RW 15, so cluster 1 is used as a place or container to accommodate data that has the closest value to the centroid value.

Calculating Data Distance to Each Cluster

After getting the central point value for each cluster, the next process is to calculate the data distance to each available cluster or it can be called distance. The results of calculations in this process are in table 2, as follows:

Table 2
Cluster Center Distance Calculation

Districts	Number of RT	Number of RW	C1	C2	C3	Shortest Distance
MELAWAI	30	4	1.175.159.564	4.604.345.773	1.104.536.102	1.104.536.102
GUNUNG	68	7	7.940.403.012	8.062.257.748	4.916.299.421	8.062.257.748
KRAMATPELA	82	10	6.519.202.405	7.211.102.551	6.338.769.597	7.211.102.551
SELONG	33	4	1.145.294.722	430.464.865	1.403.566.885	1.403.566.885
RAWA BARAT	44	7	1.033.102.125	3.201.562.119	253.179.778	253.179.778
SENAYAN	19	3	1.285.612.694	5.707.889.277	0	0
PULO	46	6	1.014.001.972	30	2.716.615.541	2.716.615.541
PETOGOGAN	76	6	7.156.814.934	0	5.707.889.277	0
GANDARIA	147	15	0	7.156.814.934	1.285.612.694	0
CIPETE	103	11	4.418.144.407	2.745.906.044	8.438.009.244	2.745.906.044

In table 2, after the process of calculating the distance of data to each cluster is complete, the next process is to allocate data into each formed cluster. The data allocation is based on the results of the distance between the data to each cluster, if the value of the distance between the first data to cluster 1 is smaller than the value of the distance between the first data to cluster 2 or cluster 3 then the first data goes into cluster 1.The results of the calculation of the shortest distance are formed in table 3. , where the data allocation is carried out in order to determine the new cluster center point in the next process. In this process the authors allocate data using the manual method, with the results in the following figure:

Tabel 3
1st data grouping

C1	C2	C3
		1
	1	
	1	
		1
		1
		1
		1
	1	
1		
	1	

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 5, Number 2, April 2021
DOI : https://doi.org/10.33395/sinkron.v5i2.10855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Determining the Center Point of the New Cluster

In table 3, after the data allocation process is complete, the next step is to determine the new cluster point. This new centroid determination uses a method similar to determining the cluster center point in the previous stage, what distinguishes the two stages is the amount of data used. In this stage, the centroid can be determined by calculating the total amount of data in each group divided by the amount of data and getting the following results:

Tabel 4
New Cluster Determination

District	Number of RT	Number of RT	New Cluster		
			C1	C2	C3
MELAWAI	30	4	147	82.25	34.4
GUNUNG	68	7	15	8.5	4.8
KRAMATPELA	82	10			
SELONG	33	4			
RAWA BARAT	44	7			
SENAYAN	19	3			
PULO	46	6			
PETOGOGAN	76	6			
GANDARIA	147	15			
CIPETE	103	11			

Determination of the initial center of the new cluster

New cluster 1st	147	15
New cluster 2nd	82.25	8.5
New cluster 3rd	34.4	4.8

Verification of the Cluster Center Points

Namely the process of verifying between the new cluster center point and the old cluster center point in Table 3 and Table 5, if the two cluster center points have different centroid values, the K-Means Clustering process will continue and will begin again in the 2nd process. namely (calculating the data distance to each cluster) by using the new cluster center point value. Meanwhile, if the two centroid values are the same, the K-Means Clustering process stops at that stage. This verification process is carried out in order to determine whether the K-Means Clustering process has been completed or if the process is still needed to repeat. In this process, the writer goes through 2 stages of the repetition process to get the value of the cluster center point or centroid that doesn't change anymore. The results of this application are as follows:

Tabel 5
2nd data grouping

No.	C1	C2	C3
1			**1**
2		**1**	
3		**1**	
4			**1**
5			**1**
6			**1**
7			**1**
8		**1**	
9	**1**		
10		**1**	

The iteration process of this research occurs 2 times so that an assessment is obtained in classifying the household / neighborhood unit based on the Kelurahan. The results of the verification of Table 3 and Table 5, the

e-ISSN : 2541-2019
p-ISSN : 2541-044X

calculation is complete because Table 3 and Table 5 have similarities. The results obtained are that there is 1 neighborhood unit with the highest cluster (C1), there are 4 neighborhood units with 4 medium clusters (C2), and 5 neighborhood units with the lowest cluster (C3). This data can be input to the sub-district to disseminate information about dengue fever, health education, and for the accuracy of PDAM customer address management and others.

DISCUSSIONS

In previous studies using the Cencus Mapping System (CMS) method from Small Area Statistics using the socialization of the program, namely Small Area Statistics. statistical mapping using the Densely Inhabited District (DID) system, namely mapping for statistics with the condition that buildings in the city are 5,000 units or more and are integrated into a government. The results of research and calculations using the K-means algorithm consist of 2 iterations so that the results of the grouping of the highest, medium and low clusters are the same and can be seen clearly in table 3 and table 5. Then in the Rukun Tetangga / Rukun Kawasan Tetangga with the highest cluster is Gandaria District. For medium clusters are Gunung, Kramatpela, Petogogan and Cipete villages and for the lowest clusters are Melawai, Selong, Rawa barat, Senayan and Pulo. This data can be used as input for the sub-district to disseminate information about dengue fever, health education, the accuracy of PDAM customer address management and others.

So the difference. Previous research used mapping data for statistics with the condition that buildings in the city amount to 5,000 units or more and are integrated into a government and in research with the K-Means Algorithm using data on the number of RT / RW as a clearer and more detailed data sample, so that the accuracy home address can be divided into high groups, medium groups and low groups and as a reference information for Kebayoran district.

CONCLUSION

From the results of the research that has been done, the authors can draw the conclusion that the calculation of the K-means algorithm occurs 2 times repetition of the grouping and there is a similarity in the grouping with the formation of the highest, medium and lowest cluster allocations with the Rukun Tetangga / Rukun Warga area with the highest cluster. is the Gandaria sub-district, for the medium cluster are Gunung, Kramatpela, Petogogan and Cipete villages and for the lowest cluster are Melawai, Selong, Rawa barat, Senayan and Pulo. This data can be input to the sub-district to disseminate information about dengue fever, health education, and for the accuracy of PDAM customer address management and others.

REFERENCES

Alfina, T., & Santosa, B. (2012). Comparative Analysis of Hierarchical Clustering Methods, K-Means and the Combination of Both in Forming Data Clusters (Case Study: Job Training Problems Department of Industrial Engineering ITS). *Analisa PerbandinganMetode Hierarchical Clustering, K-Means Dan Gabungan Keduanya Dalam Cluster Data*, *1*(1), 1–5.

Ali, A. (2019). Clustering of Patient Medical Record Data Using the K-Means Clustering Method at Anwar Medika Hospital Balong Bendo, Sidoarjo. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *19*(1), 186–195. https://doi.org/10.30812/matrik.v19i1.529

Ali, A. (2020). Toddler Anthropometric Data Clustering to Determine the Nutritional Status of Toddlers in Jumput Rejo Village, Sukodono, Sidoarjo. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(3), 395–407. https://doi.org/10.35957/jatisi.v7i3.530

Andi IchsanN Syamsul. (2018). Rukun Neighborhood Synergity With Rukun Rukga In Supervision Of Kost House In Kecamatan Tamalate Kota Makassar. *Journal of Business Ethics*, *14*(3), 37–45. Retrieved from https://www-jstor-org.libproxy.boisestate.edu/stable/25176555?Search=yes&resultItemClick=true&searchText=%28Choosing&searchText=the&searchText=best&searchText=research&searchText=design&searchText=for&searchText=each&searchText=question.%29&searchText=AND

Belanja, D. A. N., Apbdes, D., & Desa, D. I. (2020). *Analysis of village revenue and expenditure budget management (APBDes) in Lauru I Afulu Village, Afulu District, North Nias Regency*.

Dinas Komunikasi, Informatika, dan S. (2020). Jakarta Open Data.

Gading Sadewo, M., Perdana Windarto, A., & Wanto, A. (2018). *KOMIK (National Conference on Information Technology and Computers) The Implementation Of Clustering Algorithm In Grouping The Many Village / Village According To Anticipation / Mitigation Efforts Of Natural Disasters By Province With K-means*. *2*, 311–319. Retrieved from http://ejurnal.stmik-budidarma.ac.id/index.php/komik

Ginting, S. L. B., & Trinanda, R. P. (2013). Data Mining Technique Using Bayes Classifier Method To Optimize Search In Library Applications. *Universitas Pasundan*, *d*(Pencarian Informasi), 1–14.

Khotimah, T. (2014). Grouping of Surahs in the Qur'an Using the K-Means Algorithm. *Simetris: Jurnal Teknik*

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Mesin, Elektro Dan Ilmu Komputer, *5*(1), 83–88. https://doi.org/10.24176/simet.v5i1.141

Maulida, L. (2018). Application of Datamining in Grouping Tourist Visits to Leading Tourist Attractions in Prov. Dki Jakarta with K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *2*(3), 167. https://doi.org/10.14421/jiska.2018.23-06

Metisen, B. M., & Sari, H. L. (2015). Clustering analysis uses the K-Means method in grouping product sales at Fadhila Supermarkets. *Jurnal Media Infotama*, *11*(2), 110–118.

Ong, J. O. (2013). Implementation of Algotritma K-means clustering to determine the marketing strategy of the president university. *Jurnal Ilmiah Teknik Industri*, *vol.12*, *no*(juni), 10–20.

Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Application of Data Mining to Predict Tree Seed Orders with Multiple Linear Regression. *JURIKOM (Jurnal Riset Komputer)*, *7*(1), 56. https://doi.org/10.30865/jurikom.v7i1.1947

Ramadhan, A., Mustakim, & Efendi, Z. (2017). Comparison of K-Means and Fuzzy C-Means for Grouping User Knowledge Modeling Data. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 9*, 219–226.

Samsudin, A. M., Djaeni, M., Idris, & Fathoni, S. (2017). Proceedings of the National Seminar on Community Service "Community Service Contribution in Improving College Clusters." In *Penerapan Teknologi Tray Dryer Pada Pengeringan Dendeng Jantung Pisang Di Kelurahan Rowosari Kota Semarang*.

Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analysis of Product Purchase Patterns. *Teknik Informatika*, (1), 19–24. Retrieved from http://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/1267/1200

Sihombing, E. G. (2017). Data Mining Classification on Households According to Province and Ownership Status of Contract / Leased Houses Using the K-Means Clustering Method. *Computer Engineering, System and Science Journal*, *2*(2), 74–82.

Sitohang, D. W., & Rikki, A. (2019). Implementation of the K-Means Clustering Algorithm for Classifying Nutrition Data for Toddlers in the District of Garoga Tapanuli Ut. *Jl. Bilal Ujung No*, *24*, 80–92.

Tana, M. P., Marisa, F., & Wijaya, I. D. (2018). Application of Data Mining Market Basket Analysis Method to Product Sales Data at Oase Stores Using the Apriori Algorithm. *J I M P - Jurnal Informatika Merdeka Pasuruan*, *3*(2), 17–22. https://doi.org/10.37438/jimp.v3i2.167

Wardhani, Anindya Khrisna, W. (2016). Implementation of the K-Means Algorithm for Grouping Patient Diseases at Kajen Pekalongan Health Center. *Jurnal Transformatika*, *14*(1), 30–37.

Wibowo, A., & Sudarmaji, B. W. (2010). Peta Skala Besar (Batas RW) dan Manfaatnya: Studi Kasus di DKI Jakarta. *Jurnal Globe*, *12*(1), 82–88. Retrieved from http://jurnal.big.go.id/index.php/GL/article/view/119

Widodo, T., Utomo, W., & Miranti, B. (2009). Capacity Building of Rukun Tetangga / Rukun Warga as "Grassroots" Organizations in the Era of Wide Decentralization. *Jurnal Ilmu Administrasi*, *6*, 18–33. Retrieved from http://180.250.247.102/index.php/jia/article/view/344

Yudi Agusta. (2007). K-Means - Applications, Problems and Related Methods. *Jurnal Sistem Dan Informatika*, *3*(Februari), 47–60.