

Analysis of Delivery Data by Medical Staff Using The K-Means Algorithm in Sleman District

Sefitriani Khasanah¹, Adinda Rifi Yanti¹, Dwi Puspa Sari¹, Imam Tahyudin^{1*}, Dhanar Intan Surya Saputra²

¹Department of Information System, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia

²Department of Informatics, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia

*Corresponding author E-mail: imam.tahyudin@amikompurwokerto.ac.id

Manuscript received 19 July 2022; revised 1 Sept 2022; accepted 15 Sept 2022. Date of publication 4 Nov 2022

Abstract

The process of childbirth has many factors that result in the death of the baby and the mother during the delivery process, namely the lack of medical staff or health workers (midwife, doctor, or another paramedic). There needs to be an analysis of the delivery process assisted by medical staff. This analysis maps the readiness of medical staff with the needs in the field. Both natural and cesarean births have the same main goal, to make labor run smoothly and ensure that the mother and baby are safe. Deliveries assisted by health workers use secure, clean, and sterile equipment to prevent infection and other health hazards. The hope is to minimize the number of mothers who are not helped during childbirth. This study aims to analyze data on deliveries assisted by medical staff for case studies in Sleman District, Province of Yogyakarta Special Administrative Region, Indonesia, with the K-Means Algorithm. K-means is an unsupervised learning algorithm that has a function to group data into data clusters. This algorithm can accept data without any category labels, the learning process requires a relatively fast time, is quite easy to understand and implement, and the algorithm is quite popular. The research used 13,869 data in 2018. In 2019, the decrease in the number of mothers giving birth from 13,470 who were rescued was 13,469. The 2018 data produced 3 (three) clusters. In 2019 data produced 4 (four) clusters. With different levels of levels assisted by medical staff starting from the high, medium, and low groups.

Keywords: Childbirth, Clustering, Data Mining, Delivery, K-Means Algorithm.

1. Introduction

Childbirth is a physiological process experienced by women in general. In this process, a series of changes occur in the mother to be able to give birth to her fetus through the birth canal. For the delivery process itself, medical staff is needed so that the delivery process runs smoothly and safely for the mother and baby. Attention to pregnant women needs to be given intensively, and pregnant women must routinely consult about the importance of maintaining food intake and meeting nutritional needs during pregnancy. Pregnant women should check their pregnancies regularly until the first 1000 days, the period from when the child is in the womb until the child is two years old [1]. Problems during this period can be dealt with immediately so that the baby's health is maintained and stunting can be overcome immediately [2], [3]. Giving birth to inadequate health facilities and assistance from experts can help mothers, and babies get the best treatment they need. Additionally, awareness to routinely checks the obstetrician the doctor also needs to be increased. That way, plans to fulfill nutrition for mothers and babies can run well as recommended. Death during childbirth often occurs due to a lack of action from medical staff [4]. Data from the Central Statistics Agency (in Bahasa Indonesia: Badan Pusat Statistik/ BPS) records the percentage of deliveries in Indonesia from several provinces, including Papua (69.41%), North Maluku (81.34%), West Papua (82.5%), East Nusa Tenggara (83.19%) and the lowest percentage of deliveries assisted by health personnel in 2019 was occupied by Maluku with a rate (68.34%). The five provinces are below the national average of 94.17% [5]–[7].

The Sleman Health Office recorded data on deliveries assisted by medical staff from several sub-districts in Sleman Regency in 2018 - 2019. From the amount of delivery data collected, a method is needed to help gain insight and information in decision-making and solving problems [8]. One technique for managing data mining is using the K-means (clustering) algorithm. Clustering itself is an algorithm that produces attribute groupings. In addition, this method is known to be the most popular because it is simple, easy to implement, has clustering capabilities for extensive data, and is widely used in various fields [9].

A previous study discussed the K-Means Algorithm entitled Optimization of the C4.5 Algorithm Using the K-Means Algorithm and Selection of Particle Swarm Optimization Features in Breast Cancer Diagnosis. This study implements the K-Means Algorithm, C4.5 Algorithm, and the PSO (Particle Swarm Optimization) algorithm. It is based on clustering calculations using C4.5 classification and selected with the highest accuracy. From the experiments carried out, it was found that the number of clusters = 3 could provide a better accuracy of 1.169%. And it was implementing PSO results in feature selection. The feature selection result is the attribute based on the highest



gain value. The proposed method provides an average accuracy of 97.894%, thus providing better accuracy than the C4.5 algorithm, which is 94.152% [8].

In the previous study entitled Analysis of Data Mining Using the K-Means Clustering Algorithm for Product Grouping, applying data mining clustering in the product grouping process based on the most desirable and less desirable products with the results of cluster technique calculations, it can be seen that with a sample of 10 data consists of 2 clusters, namely cluster1 and cluster2. In cluster_1, it is known to have four less popular products, and in cluster_2, six products are most in demand [10].

The previous study entitled Decision Support System using Data Mining Method for a Cross-Selling Strategy in Retail Stores. This research uses the Apriori Algorithm method to collect transaction data that is used as a reference for new cross-selling sales strategies. Based on calculations, the Association Rule is implemented by applying a Confidence value = 0.8 while the Support value = 0.1 from the minimum value determined; the total result is 77 rules [11].

The previous study also discussed using the K-Means Algorithm with the title Population Density Cluster Analysis in DKI Jakarta Province, using the K-Means Algorithm with the results of grouping starting from data collection, data understanding, data processing to improve data quality and attribute selection, assessment of cluster tendencies, cluster formation to cluster analysis, the results of the study there are 3 clusters formed cluster 0 (not densely populated), cluster 1 (moderate population density), and cluster 2 (densely populated) [12].

Based on the description above, the researcher intends to research cases of childbirth assisted by medical staff using the K-Means (clustering) algorithm. This study compares data on the number of deliveries assisted by medical staff in 2018 and 2019.

2. Literature Review

2.1. Data Collection

The advancement of information technology, particularly the internet, has resulted in the loss of a large amount of data. Data analysis is an effort to systematically search for and organize records of observations, interviews [13], and others to increase the researcher's understanding of the cases studied and present them as findings for others [14]. The definition of analysis is the activity of thinking to break down a whole into components so that you can recognize the signs of the components, their relationship to each other, and their respective functions in an integrated whole [15]. The definition of analysis is solving or decomposing a unit into the smallest unit. From the above opinion, analysis is an activity of thinking to describe or solve a problem from unit to smallest unit [16], [17].

2.2. Data Mining

Data mining is collecting raw data from various databases that have yet to be processed so the data can be grouped later. Data mining is finding interesting patterns and knowledge from large amounts of data. Search and analysis of very large amounts of data and aims to find the meaning of patterns and rules. Data mining is a process of extracting or extracting data that was not previously known but understandable and useful from large databases and used to make a very important business decision. Data mining is an activity that describes an analytical process that occurs iteratively on a large database, intending to extract accurate and potentially useful information and knowledge for knowledge workers related to decision-making and problem-solving [18]. Data mining is a field of several scientific fields that combines techniques from machine learning, recognition, patterns, and retrieval of information from large databases [19].

2.3. K-Means Algorithm

One of the clustering algorithms in data mining is the K-Means Clustering algorithm, which produces groups with the same attributes. K-Means is the most popular clustering method widely used in various fields because it is simple, easy to implement, has extensive data clustering capabilities, and can handle outlier data [20]. Meant to divide data into one or more groups, with similarities in one group and disparities in another [21].

The K-Means algorithm is one of the unsupervised learning algorithms and follows the most commonly used partition method for grouping because of its simplicity and efficiency [22], [23]. The K-Means algorithm consists of two phases: the first phase calculates k centroids, and the second phase brings each point to the cluster that has the closest center of the adjacent data points [24]. The K-means algorithm is a simple iterative clustering algorithm. Using distance as the metric and given the class M in the data set, calculate the average distance, given a starting centroid, with each class described by a centroid. For a given data set X containing n multidimensional data points and M categories to be split, the Euclidean distance is chosen as the similarity index. The clustering target minimizes the sum of squares of different types, i.e., minimizing.

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (1)$$

2.4. Clustering

Data clustering is an unsupervised and arithmetic data analysis procedure. Cluster analysis is used to group large data sets into subsets called clusters. Each cluster is a collection of data objects that are similar to each other in the same place in the same cluster but different from objects in other clusters [25].

Clustering is a method for grouping data into classes with identical characteristics where similarities between classes are maximized or minimized. Data clustering is a procedure that divides a data collection into groups so that data points within a group are very similar to one another but very distinct to those in other groups [26]. For example, if there is a set of objects, the first process can be clustered into several class sets and then become an ordered set to be derived based on certain classification groups [27]. Clustering itself is also called unsupervised classification because clustering is more to be studied with attention. Cluster analysis is partitioning a set of data objects into subsets. Each set of parts is a cluster, so the objects in the cluster are similar and differ from objects in other clusters.

3. Methods

3.1. Principles of the K-Means Method

The K-Means algorithm is classified as learning without supervision. This algorithm determines the distance between the centroid and the training data and the number of centroids based on the desired number. Meanwhile, the initialization centroid is generated randomly by considering the training data. Each training data will have a centroid distance. At the same time, cluster members are shown with the smallest space. The iteration will be stopped if the cluster members do not change [28].

3.2. Data Source

This research has a dataset consisting of the same four attributes in 2018 and 2019. The 2018 maternity data contains the number of mothers giving birth, namely 13869, the same as the number of characteristics assisted by medical staff, 13869. In 2019 there are 4 of the same qualities. With delivery data for 2018, however, the number of mothers giving birth decreased to 13,470, and the number assisted by medical staff was 13,469, Figure 1 and Figure 2.

1	1.0	Moyudan	349.0	349.0	1	1.0	Moyudan	333.0	333.0
2	2.0	Minggir	432.0	432.0	2	2.0	Minggir	402.0	402.0
3	3.0	Seyegan	715.0	715.0	3	3.0	Seyegan	732.0	732.0
4	4.0	Godean	925.0	925.0	4	4.0	Godean	838.0	838.0
5	5.0	Gamping	1319.0	1319.0	5	5.0	Gamping	1263.0	1263.0
6	6.0	Mlati	1255.0	1255.0	6	6.0	Mlati	1187.0	1187.0
7	7.0	Depok	1488.0	1488.0	7	7.0	Depok	1493.0	1493.0
8	8.0	Berbah	853.0	853.0	8	8.0	Berbah	747.0	747.0
9	9.0	Prambanan	738.0	738.0	9	9.0	Prambanan	719.0	719.0
10	10.0	Kalasan	1005.0	1005.0	10	10.0	Kalasan	872.0	872.0
11	11.0	Ngemplak	774.0	774.0	11	11.0	Ngemplak	802.0	802.0
12	12.0	Ngaglik	1167.0	1167.0	12	12.0	Ngaglik	1227.0	1226.0
13	13.0	Sleman	851.0	851.0	13	13.0	Sleman	930.0	930.0
14	14.0	Tempel	692.0	692.0	14	14.0	Tempel	626.0	626.0
15	15.0	Turi	463.0	463.0	15	15.0	Turi	462.0	462.0
16	16.0	Pakem	420.0	420.0	16	16.0	Pakem	436.0	436.0
17	17.0	Cangkringan	423.0	423.0	17	17.0	Cangkringan	401.0	401.0
18	18.0	Total	13869.0	13869.0	18	18.0	Total	13470.0	13469.0

Fig 1. (a) Childbirth Data for 2018 and (b) Childbirth Data for 2019

This study used the k-means algorithm with a clustering process to compare data on deliveries assisted by medical staff in Sleman Yogyakarta District in 2018 and 2019. This method was used to classify the suitable clusters, namely the highest to the lowest sets. In managing this data, researchers use software assistance, namely Weka. The data obtained in this study were from the Sleman Health Office via the www.data.slemankab.go.id website. This data set consists of 17 sub-districts in Sleman Regency and has four attributes (id, sub-district, number, assisted by medical staff).

3.3. Flow of K-Means

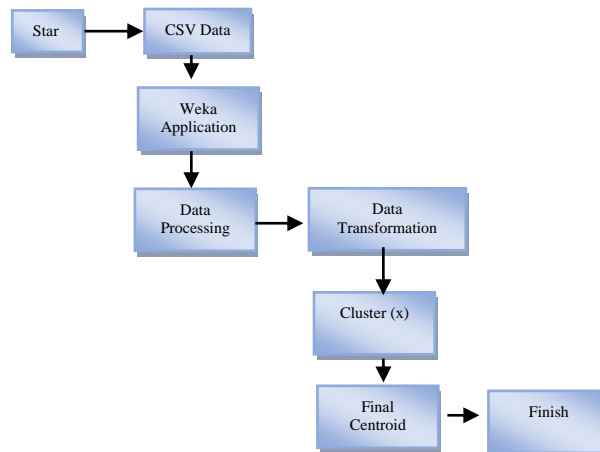


Fig 2. K-Means flowchart

3.3.1. CSV Data

Numerical values in the form of a matrix are generated from preprocessing coding that was carried out and stored in .csv so that they are efficient and easy to analyze in Weka later. Then open the Weka application to perform data processing Using WEKA. The output generated from the preprocessing set is processed at this stage using the WEKA application.

3.3.2 Data Transformation

Data transformation is an effort made with the primary objective of changing the scale of measurement of the original data into another form so that the data can meet the assumptions underlying the analysis of variance. Clustering is used to group data and explain the relationships between the data and maximize the similarity between one class/ cluster but minimize the similarities between classes/ clusters.

3.3.3. Clustering

Clustering is used to analyze data and is expected to produce a representation representing a pattern formed due to the relationships between data. (the clustering process is used to identify the grouping of several events in the dataset to produce information that the user can analyze. There are several options in the Weka cluster sub-menu, including using a training set, supplied test set percentage split, and classes to cluster evaluation used for compares how well the data are compared without being assigned a type between data).

3.3.4. Final Centroids

The arithmetic means value of an object shape from all points in the object that have been processed.

4. Results and Discussion

4.1. Data Processing

The next step is the data processing of the K-Means clustering algorithm. Determine the K value of the cluster you want to create. The clusters created are 3 clusters. Determine the initial center point for each cluster. In this study, the initial center point was determined by the random method obtained from each set.

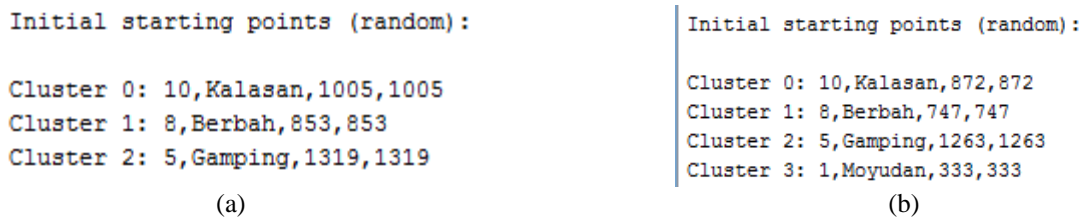


Fig 3. (a) Cluster Value in 2018; (b) Cluster Value in 2019

4.2. Allocating Data To Clusters

Using simple k-means from Weka, it is known that the data belonging to the cluster has the smallest distance from the cluster center. The iterations that were stopped and the data did not change for the 3rd iteration of 2018 data are the same as the 2019 data.

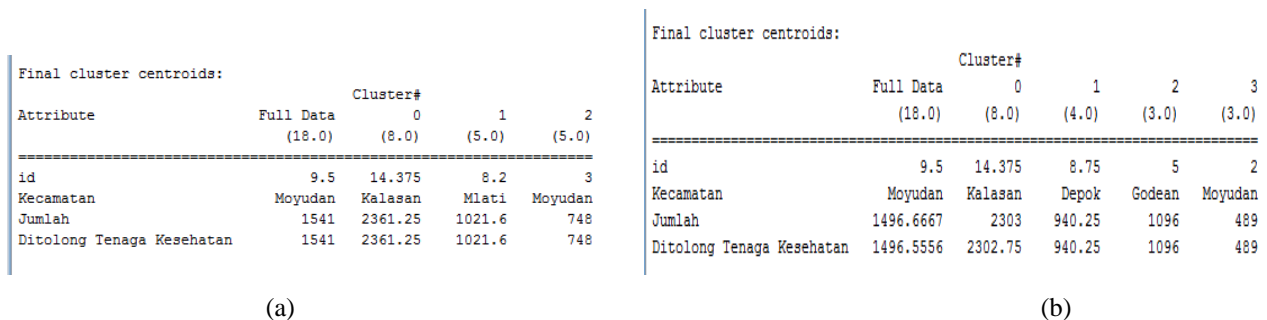


Fig 4. (a) Final Centroid Clusters in 2018; (b) Final Centroid Clusters in 2019

From the application of K-means clustering in this study, the results obtained were clusters of 3 groups with the percentage of members in the table. From the results of research on cluster 0.

Clustered Instances	
0	8 (44%)
1	5 (28%)
2	5 (28%)

Fig 5. Cluster Instances in 2018

From the Figure 8, in 2018, from 13,869 produced cluster 0, there were 1,005 deliveries assisted by health personnel (moderate) in Kalasan sub-district, cluster 1, there were 853 deliveries assisted by health personnel (low) in the Berbah sub-district, cluster 2, there were 1,319 deliveries assisted by health personnel (high) in Gamping sub-district.

Clustered Instances	
0	8 (44%)
1	4 (22%)
2	3 (17%)
3	3 (17%)

Fig 6. Cluster Instances in 2019

From the Figure 9, in 2019, out of 13,469 produced in cluster 0, there were 872 deliveries assisted by health personnel (high) in Kalasan sub-district, cluster 1, there were 747 deliveries assisted by health personnel (moderate) in Berbah sub-district, cluster 2, there were 1,263 deliveries assisted by health personnel (very high) in the Gamping sub-district, cluster 3, there were 333 deliveries assisted by health personnel (low) in the Moyudan sub-district.

5. Conclusion

Based on the research conducted, it can be concluded that the K-Means algorithm can be used to classify birth data assisted by medical staff. Based on four attributes of delivery data in 2018 and 2019. The clustering results can support information for conducting work evaluations for medical staff in the delivery process.

References

- [1] E. Isaevska *et al.*, "Exposure to ambient air pollution in the first 1000 days of life and alterations in the DNA methylome and telomere length in children: A systematic review," *Environ. Res.*, vol. 193, p. 110504, 2021.
- [2] E. Darnila, M. Maryana, K. Mawardi, M. Sinambela, and I. Pahendra, "Supervised models to predict the Stunting in East Aceh," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 33–39, 2022.
- [3] T. N. Farisni and others, "The Relationship between Nutritional Health Services of Pregnant Women and the Incidence of Stunting in Lhok Bot, Aceh Jaya," *Clin. Med. Insights*, vol. 3, no. 1, pp. 250–256, 2022.
- [4] T. S. Price, C. Andries, M. Palokas, and R. Christian, "Factors contributing to a positive childbirth experience for women in a hospital setting: an umbrella review protocol," *JBI Evid. Synth.*, vol. 20, no. 3, pp. 854–859, 2022.
- [5] Badan Pusat Statistik (BPS), "Angka Kelahiran Total Menurut Provinsi," <https://www.bps.go.id/>, 2022. <https://www.bps.go.id/indicator/30/1399/1/angka-kelahiran-total-menurut-provinsi.html> (accessed Oct. 20, 2022).
- [6] A. D. Laksono, R. D. Wulandari, and R. Rukmini, "The determinant of healthcare childbirth among young people in Indonesia," *J. Public Health Res.*, vol. 10, no. 1, p. jphr--2021, 2021.
- [7] A. D. Laksono and R. D. Wulandari, "Regional disparities of facility-based childbirth in Indonesia," *Trends Sci.*, vol. 18, no. 21, p. 387, 2021.
- [8] A. A. Septiantina and E. Sugiharti, "Optimization of C4. 5 Algorithm Using K-Means Algorithm and Particle Swarm Optimization Feature Selection on Breast Cancer Diagnosis," *J. Adv. Inf. Syst. Technol.*, vol. 2, no. 1, pp. 51–60, 2020.
- [9] A. Sarker, S. M. Shamim, M. S. Zama, and M. M. Rahman, "Employee's performance analysis and prediction using K-means clustering & decision tree algorithm," *Glob. J. Comput. Sci. Technol.*, 2018.
- [10] S. Mishra, P. K. Mallick, H. K. Tripathy, A. K. Bhoi, and A. González-Briones, "Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier," *Appl. Sci.*, vol. 10, no. 22, p. 8137, 2020.
- [11] I. Tahyudin, M. Imron, and S. A. Solikhatin, "Decision Support System using Data Mining Method for a Cross Selling Strategy in Retail Stores," *Int. J. Informatics Commun. Technol.*, vol. 3, no. 3, p. 171, 2014, doi: 10.11591/ijict.v3i3.pp171-177.
- [12] A. A. Arifiyanti, F. S. Darusman, and B. W. Trenggono, "Population Density Cluster Analysis in DKI Jakarta Province Using K-Means Algorithm," *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 772–783, 2022.
- [13] T. Li, J. P. T. Higgins, and J. J. Deeks, "Collecting data," *Cochrane Handb. Syst. Rev. Interv.*, pp. 109–141, 2019.
- [14] M. K. Alam, "A systematic qualitative case study: questions, data collection, NVivo analysis and saturation," *Qual. Res. Organ. Manag. An Int. J.*, 2020.
- [15] S. Kumari and A. Harikrishnan, "Importance of Financial literacy For Sustainable Future Environment: A Research Among People In Rural Areas With Special Reference To Mandi District, Himachal Pradesh," *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 15–19, 2021.
- [16] S. Wang, "An Exploration into Research on Critical Thinking and Its Cultivation: An Overview.," *Theory Pract. Lang. Stud.*, vol. 7, no. 12, 2017.
- [17] L. E. W. Fajari, Sarwanto, and Chumdari, "Enhancement of Students Critical Thinking Skills Through Problem-Based Learning Multimedia," in *3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019)*, 2019, vol. 397, no. Icliqe 2019, pp. 976–987. doi: 10.2991/assehr.k.200129.121.
- [18] D. van den Bergh *et al.*, "A tutorial on Bayesian multi-model linear regression with BAS and JASP," *Behav. Res. Methods*, pp. 1–21, 2021.
- [19] D. A. Mutiara, A. Susli, D. Suhartono, D. Arifudin, and I. Tahyudin, "Data Mining Method to Determine a Fisherman's Sailing Schedule Using Website," *Telematika*, vol. 14, no. 2, pp. 122–132, 2021.
- [20] D. Ardiada, P. A. Ariawan, and M. Sudarma, "Evaluation of Supporting Work Quality Using K-Means Algorithm," *IJEET Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 52–55, 2018.
- [21] M. Faisal, N. Nurdin, F. Fajriana, and Z. Fitri, "Information and Communication Technology Competencies Clustering For Students For Vocational High School Students Using K-Means Clustering Algorithm," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 111–120, 2022.
- [22] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Appl. Sci.*, vol. 11, no. 23, p. 11202, 2021.
- [23] E. Zhang, H. Li, Y. Huang, S. Hong, L. Zhao, and C. Ji, "Practical multi-party private collaborative k-means clustering," *Neurocomputing*, vol. 467, pp. 256–265, 2022.
- [24] X. Wang, Z. Wang, M. Sheng, Q. Li, and W. Sheng, "An adaptive and opposite K-means operation based memetic algorithm for data clustering," *Neurocomputing*, vol. 437, pp. 131–142, 2021.
- [25] I. P. Nasution and A. S. Sitio, "Decision Support System Budget Allocation Fund Village With AHP method (Analytic Hierarchy Process) In the village of beehive," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 2, no. 1, pp. 53–62, 2020.
- [26] R. T. Adek, R. K. Dinata, and A. Ditha, "Online Newspaper Clustering in Aceh using the Agglomerative Hierarchical Clustering Method," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 1, pp. 70–75, 2022.
- [27] Q. Jin, N. Lin, and Y. Zhang, "K-means clustering algorithm based on chaotic adaptive artificial bee colony," *Algorithms*, vol. 14, no. 2, p. 53, 2021.

- [28] A. F. O. Gafar, I. Tahyudin, and others, "Comparison between k-means and fuzzy c-means clustering in network traffic activities," in *International Conference on Management Science and Engineering Management*, 2017, pp. 300–310.