JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 6, No. 2, December, pp. 161~169
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 36/E/KPT/2019
DOI: 10.36378/jtos.v6i2.3157

The Relationship Between Age, Parity, Ideal Weight, and Blood Pressure in Diagnosing Hypertension in Pregnant Women Using The K-Means Algorithm

Hendry Fonda[1], Uci Rahmalisa[2]
[1,2]Information Systems, Faculty of Computer, Hang Tuah Pekanbaru University

Article Info

Article history:

Received 07 27, 2023
Revised 09 14, 2023
Accepted 12 15, 2023

Keywords:

Data Minin,
Prediction,
Clustering,
K-Means
Hypertension

ABSTRACT

Hypertension is one of the health problems that often arise during pregnancy and can cause complications in 2-3% of pregnancies. Hypertension In Pregnancy (HDK) is defined as a blood pressure of ≥140/90 mmHg in two or more measurements. Data mining is a combination of a number of computer science disciplines that is defined as the process of discovering new patterns from very large data sets. By looking at records on Age, IMT, Parity / Gravidity, and Blood Pressure and analysis with K-Means clustering, it can be seen that the similarity of values of the above variables ultimately forms patterns related to hypertension in pregnant women. The clustering process using 5 clusters according to the elbow chart analysis. In this study, it was seen that the variable Blood Pressure is the same pattern and often appears in each cluster. While hypertension occurs in 1 cluster out of 5 existing clusters.

Corresponding Author:

Hendry Fonda
Information Systems,
Faculty of computer,
Hang Tuah Pekanbaru University
Email : fonda_hendry@yahoo.com
@ Author(s) 2023

1. Introduction

Health is the most important aspect in every life, especially for pregnant women. Hypertension is one of the health problems that often arise during pregnancy and can cause complications in 2-3% of pregnancies. Hypertension is a "*silent killer*" that causes the iceberg phenomenon. The prevalence of hypertension increases with age. This pathological condition if not treated quickly and early will aggravate the risk.[1]

Hypertension In Pregnancy (HDK) is defined as a blood pressure of ≥140/90 mmHg in two or more measurements. [2] Berdasarkan *International Society for the Study of Hypertension in Pregnancy* (ISSHP) There are 4 categories of hypertension in pregnancy, namely preeclampsia-eclampsia, gestational hypertension, chronic hypertension and superimpose chronic hypertensive preeclampsia.[3] The incidence of hypertension in pregnancy can be influenced by several factors (*multiple causation*). Maternal age (<20 or ≥35 years), primigravida, nulliparity and increased Body Mass Index (IMT) are predisposing factors for the occurrence of hypertension in pregnancy.[4][5]

Bangkinang City Community Health Center is a place for community health center services to facilitate health tests, both in the form of direct health checks and hospitalization. Not only that, the Bangkinang City Health Center also conducts training and development aimed at pregnant women in the Kampar area.Currently, data on pregnant women from 2017-2021 is 568 data at the Bangkinang Health Center in Kampar District. However, the data still cannot be used as specific data. This is because the data is not

intended for the public but as an ordinary report that is still manual (recording is done on the general ledger). To determine whether the pregnant woman has hypertension or not during pregnancy, the data will be studied based on age, IMT, blood pressure, and Parity / Gravidity.[6]

Based on several variables used to test for pregnancy hypertension, one of them is age. The age studied was pregnant women aged < 20 years were adolescents, ages 20-35 years Adult I, and age > 35 years adult II. Next. Furthermore, blood pressure is generally determined, normal blood pressure for adults (> 18 years) is 120/80, the number 120 is called systolic pressure, and the number 80 is called diastolic pressure, Furthermore, IMT, IMT results are obtained from the division between body weight and height. Last is Parity / Gravidity, Gravidity is the number of pregnancies regardless of gestational age.[7][8]

To determine the relationship between IMT, blood pressure, parity and hypertension in pregnant women, it is necessary to do mapping using the clustering method. On this occasion the K-Means Algorithm is used. K-means is one of the algorithms that is unsupervised learning. K-Means has a function to group data into cluster data.[9] [10].

2. Research Method

In this research methodology describes the process of applying the research framework. Then you can understand some of the stages described in the form of a design framework, starting from the needs analysis process to the results of this study. Stages of conducting research, ie. data needs analysis, data collection process, data analysis using K-Means algorithm.[11]

2.1 Data mining methods

The method used is in data mining is to use the method CRISP-DM.[12] *Cross-Industry Standard Process for Data Mining* or CRISP-DM is one of the datamining process models *(datamining framework)* which was originally (1996) built by 5 companies namely Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation and OHRA. This framework was later developed by hundreds of organizations and companies in Europe to become a *non-proprietary standard methodology for data mining.[13]*

Figure 1: CRISP-DM

Stages of the Method *CRISP-DM*[14]

1. Business Understanding Phase
   Business understanding, The purpose of the project (research) to find out whether pregnant women experience gestational hypertension. The need for this business objective (research) to find out based on the variables of age, IMT, parity/gravidity, and blood pressure, which is more decisive.
2. Data Understanding Phase
   The collection of data on pregnant women and samples of pregnant women data are cleaned target data with the aim of obtaining consistent data.
3. Data Preparation Phase
   This data processing creates a final dataset that will be applied into modeling tools.

4. Modeling Phase
   In this modeling phase, researchers selected data mining techniques and clustering methods using K-Means.
5. Evaluation Phase
   This evaluation phase will evaluate and examine to ensure that the modeling phase of the pregnant women's data used meets the objectives of the study.
6. Deployment Phase
   Deployment is the final stage in making a report on the results of processed pregnant women's data. Final report containing the knowledge gained or pattern recognition on data in the process of data mining.

2.2 Algoritma K-Means

The K-Means algorithm takes an unlabeled dataset as input, then divides the dataset into k clusters, and repeats the process until it doesn't find the best cluster.[15] The K-Means Algorithm process can be seen in the flowchart of this image:[16]

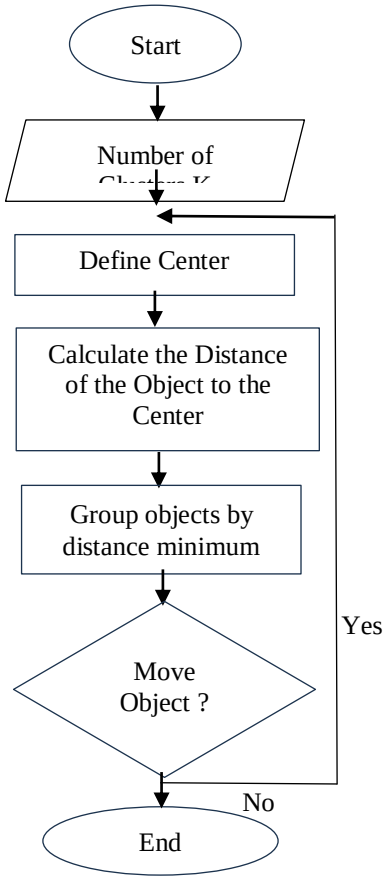

Figure 2. K-Means algorithm

3. Result and Discussion

3.1 Data Preprocessing

This study used data on pregnant women as shown in table 1. Data was taken from the Bangkinang City Health Center with a total record of 200.

Table 1. Dataset of Pregnant Women at Bangkinang City Health Center

No.	RHT	Age (mother)	Parity	IMT	Blood pressure
1.	1	2	1	2	2
2.	1	2	1	2	1
3.	1	2	1	2	2
4.	1	2	1	2	2
5.	1	2	1	3	1
6.	1	2	1	2	2
7.	1	2	1	3	2
8.	1	2	1	3	3
9.	2	2	1	2	2
10.	1	1	1	2	2
…	…	…	…	…	…
197	2	2	1	1	3
198	1	2	3	2	2
199	1	2	3	2	1
200	2	3	2	3	2

The data above is to use the name of pregnant women at the Bangkinang City Health Center, the following table below groups the division of each variable.

Table 2. Division of variables and data classes

Field Name	Division of Variables and Data Classes
RHT	1. No Hypertension 2. Hypertension
Age	1. Age < 20 Years "Teenager", 2. Age 20-35 Years "Adult I", and 3. Age 35> Years "Adult II".
IMT	IMT = BB/(TB)$^{2.}$ Thin

Class	IMT	Nutritional Status	Information
1	<17.0	Undernutrition	Very skinny
2	17.0 – 18.4	Undernutrition	That
3	18.5 – 25.0	Good Nutrition	Usual
4	25.1 – 27.0	More Nutrition	Fat
5	>27.0	More Nutrition	Very fat

Parity/Gravidity	1. Pregnancy 1, 2. Pregnancy 2, 3. Pregnancy 3, 4. Pregnancy 4, 5. Pregnancy 5
Blood pressure	1. 110/75 mmHg "Low", 2. 120/80 mmHg "Normal", 3. 130/85 mmHg "High".

Data on age, IMT, and blood pressure, the results are taken from some literature on age, IMT, and blood pressure, height is good according to WHO (*World Health Organization*), JNC 7 (*Joint National*

Committee 7),[17] and the Ministry of Health of the Republic of Indonesia. Then summarized and adjusted to the data owned.[18]

3.2 Modeling

3.2.1 Determination of Optimum K

The ideal number of clusters for K-Means analysis is to use elbow charts. Determination of the optimal number of clusters is determined using the Elbow method.[19] This method uses the Sum of Square Error (SSE) value of each cluster count. The greater the number of clusters, the SSE will continue to shrink.[20]

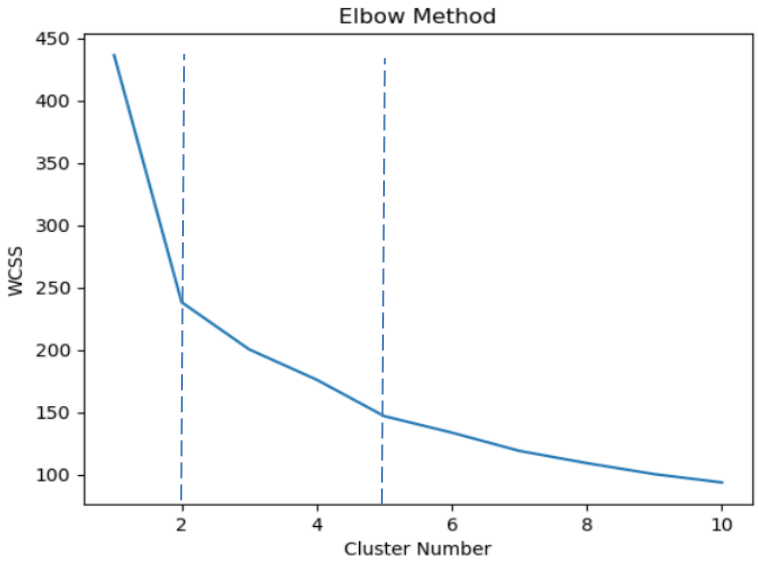

Figure 3: Elbow Chart

From the graph of figure 3 above, it can be seen that the ideal number of clusters needed for the modeling process using K_Means are 2 and 5. In this study, 2 midpoints (centroid) and 5 midpoints (centroid) were used.[21]

3.2.2 K-Means Testing

2 centroid

Tests were performed on each clustering RHT, age, parity, IMT and blood pressure. K-Means Done by calculating the accuracy, recall, sensitivity, specificity values of the confusion matrix of each experiment.[22] From the test results of 200 data columns, clustering results were obtained as table 3.

Table 3: Number of items by each cluster

Cluster	Number of Items
Cluster 0	146
Cluster 1	54

For the centroid table can be seen in the following table:

Table 4: Centroid Distance in Each Cluster

Atribute	Cluster_0	Cluster_1

Age (mother)	2.075	2.33
Parity	1.260	3.481
IMT	2.171	2.259
Blood pressure	1.911	1.407
RHT	1.267	1.407

From table 4, it can be seen that grouping occurs at value 2 in cluster 0 with a differentiator at parity with a value of 1. In cluster 1 the parity value shows the number 3 with each attribute valued at 2.

The shape of the plot graph in the K-Means process can be seen in the graph below:

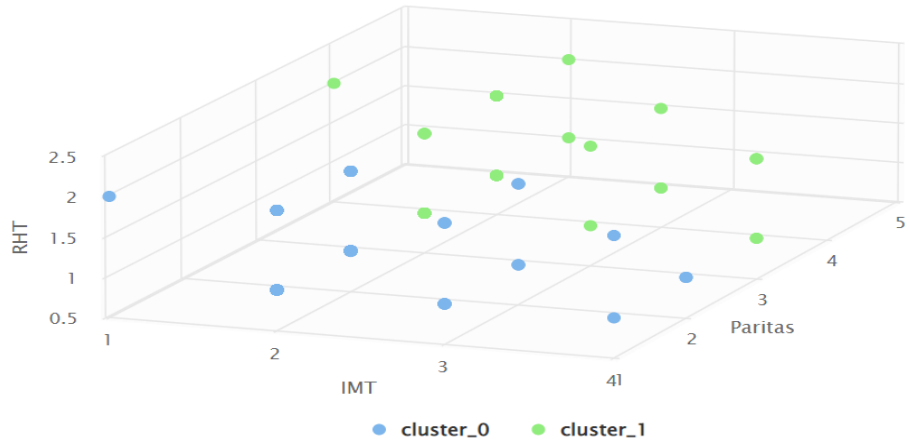

Figure 4: K-Means plot diagram

5 centroid

Tests were performed on clustering RHT, age, parity, IMT and blood pressure. K-Means is done by calculating the accuracy, recall, sensitivity, specificity values of the confusion matrix of each experiment.[22] From the test results of 200 data columns, clustering results were obtained as table 3.

Table 5: Number of items by each cluster

Cluster	Number of Items
Cluster 0	96
Cluster 1	44
Cluster 2	20
Cluster 3	30
Cluster 4	10

For the centroid table can be seen in the following table:

Table 6: Centroid Distance in Each Cluster

Atribute	Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Age (mother)	1.948	2.250	2.050	2.500	2.700
Parity	1.271	3.523	1.400	1.133	3.300
IMT	2	1.977	3.200	2.033	3.500
Blood pressure	1.802	1.864	1.900	2.267	2.400
RHT	1.031	1.386	1.300	2	1.500

From the table above, it can be seen that all clusters show that blood pressure is 2 (normal) resulting in 4 clusters (clusters 0, 1, 2 and 4) with proximity values RHT 1.

The RHT for clusters 0, 1, 2 and 4 is 1, while for cluster 3 the RHT is 2. This shows that the proximity of values in each attribute is 2 with the exception of parity value 1 resulting in a RHT of 2. Cluster 0 describes the value of each attribute has something in common, namely at value 2 and at parity value 1. However, in cluster 1 parity is worth 3. The proximity of each attribute in cluster 2 parity is 1 and IMT is 3 with IMT still valued at 1.

The shape of the plot graph in the K-Means process can be seen in the graph below:

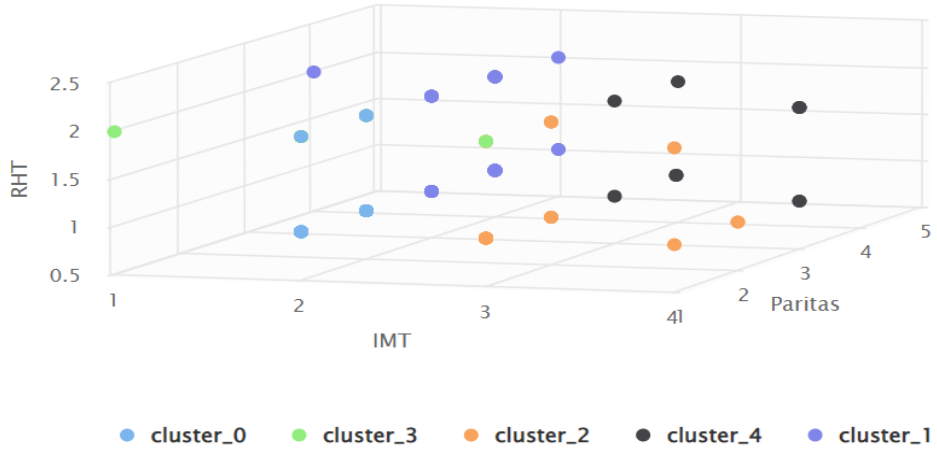

Figure 5: K-Means plot diagram

3.3 Evaluation of K-Means Using Silhouette Score

Silhouette analysis can be used to study the separation distances between the resulting clusters. Silhouette plots display a measure of how close each point in a cluster is to points in neighboring clusters and thus provide a way to visually assess parameters such as cluster counts.[23] This size has a range [-1, 1]. [24]

The Silhouette Score results on clustering K-means that have been analyzed produce a roughed value of 0.5.[25] This shows that grouping using K-Means is appropriate and optimal. Silhouette Score scores and graphs are as follows:

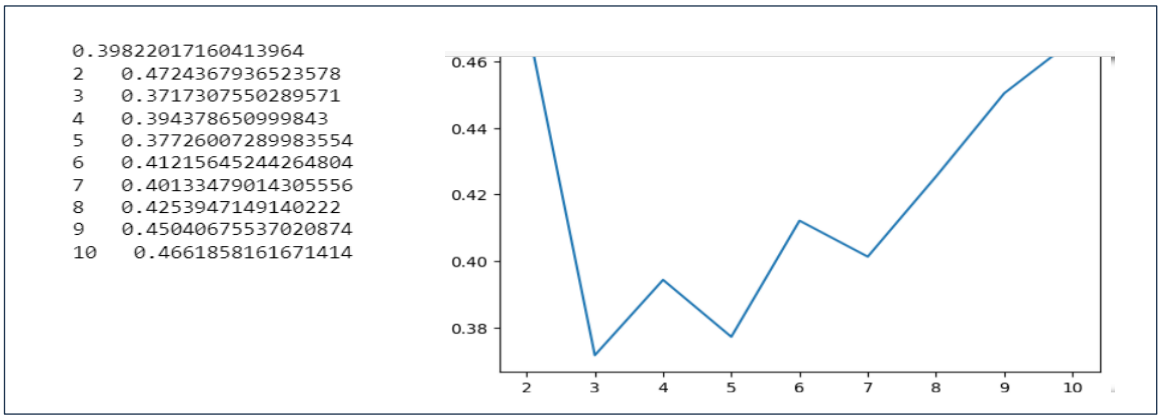

Figure 5: Evaluation of K-Means with Silhouette Score

4. CONCLUSION

The results of the analysis of maternal age, parity, ideal weight and blood pressure showed that in clustering with 2 centroids, grouping produced a differentiator in parity values. This indicates that the hypertension of pregnant women will always be normal.

Analysis using 5 centroids shows that pregnant women will experience hypertension in the first maternal pregnancy.

References

[1] N. Masruroh, A. P. R. Santoso, W. Thoyyibah, and R. A. Laloda, "Hubungan Body Mass Index dan Protein Urine Terhadap Kejadian Hipertensi Pada Ibu Hamil Trimester III," *JIDAN (Jurnal Ilmiah Bidan)*, vol. 8, no. 1, 2021, doi: 10.47718/jib.v8i1.1200.

[2] N. Schellack and N. Padayachee, "Hypertension," *SA Pharmaceutical Journal*, vol. 89, no. 1. 2022. doi: 10.29309/tpmj/2016.23.12.1795.

[3] L. A. Magee and P. Von Dadelszen, "Management of hypertension in pregnancy," *Maternal-Fetal Medicine*, vol. 3, no. 2. 2021. doi: 10.1097/FM9.0000000000000095.

[4] T. Arikah, T. B. W. Rahardjo, and S. Widodo, "Kejadian Hipertensi pada Ibu Hamil," *Jurnal Penelitian dan Pengembangan Kesehatan Masyarakat Indonesia*, vol. 1, no. 2, 2020, doi: 10.15294/jppkmi.v1i2.40329.

[5] Eneng Emi Saputri, H.A.Y.G Wibisono, Septy Ariani, and Fenie Waty, "Hubungan Perilaku Gaya Hidup Dengan Hipertensi Ibu Hamil Di Rsu Kota Tangerang Selatan Tahun 2019," *Jurnal Kesehatan*, vol. 8, no. 2, 2019, doi: 10.37048/kesehatan.v8i2.135.

[6] Dinas Kesehatan Kabupaten Kampar, "Renstra kesehatan kampar 2022," 2022.

[7] J. M. Flack and B. Adekola, "Blood pressure and the new ACC/AHA hypertension guidelines," *Trends in Cardiovascular Medicine*, vol. 30, no. 3. 2020. doi: 10.1016/j.tcm.2019.05.003.

[8] T. Unger *et al.*, "2020 International society of hypertension global hypertension practice guidelines," *Journal of Hypertension*, vol. 38, no. 6. 2020. doi: 10.1097/HJH.0000000000002453.

[9] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2988796.

[10] A. Janßen and P. Wan, "k-means clustering of extremes," *Electron J Stat*, vol. 14, no. 1, 2020, doi: 10.1214/20-ejs1689.

[11] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8. 2020. doi: 10.3390/electronics9081295.

[12] M. Gbededo, M. Stone, P. Beaver, and Y. J. Lee, "Assessing current and potential basin assets using an innovative predictive analytics workflow," in *SPE Western Regional Meeting Proceedings*, 2017. doi: 10.2118/185673-ms.

[13] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, 2021. doi: 10.1016/j.procs.2021.01.199.

[14] T. H. A. S. Siriweera, I. Paik, B. T. G. S. Kumara, and K. R. C. Koswatta, "Intelligent Big Data Analysis Architecture Based on Automatic Service Composition," in *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015*, 2015. doi: 10.1109/BigDataCongress.2015.46.

[15] S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, "Two improved k-means algorithms," *Applied Soft Computing Journal*, vol. 68, 2018, doi: 10.1016/j.asoc.2017.08.032.

[16] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit*, vol. 36, no. 2, 2003, doi: 10.1016/S0031-3203(02)00060-2.

[17] J. Y. Islam, M. M. Zaman, S. A. Haq, S. Ahmed, and Z. Al- Quadir, "Epidemiology of hypertension among Bangladeshi adults using the 2017 ACC/AHA Hypertension Clinical Practice Guidelines and Joint National Committee 7 Guidelines," *J Hum Hypertens*, vol. 32, no. 10, 2018, doi: 10.1038/s41371-018-0087-5.

[18] Dinas Kesehatan Provinsi Riau, "Profil Kesehatan Provinsi Riau Tahun 2020," 2021.

[19] B. A. Jaafar, M. T. Gaata, and M. N. Jasim, "Home appliances recommendation system based on weather information using combined modified k-means and elbow algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 3, 2020, doi: 10.11591/ijeecs.v19.i3.pp1635-1642.

[20] P. Bholowalia and A. Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *Int J Comput Appl*, vol. 105, no. 9, 2014.

[21] Mustakim, "Centroid k-means clustering optimization using eigenvector principal component analysis," *J Theor Appl Inf Technol*, vol. 95, no. 15, 2017.

[22] F. Azmi, K. Utama, O. T. Gurning, and S. Ndraha, "Initial Centroid Optimization of K-Means Algorithm Using Cosine Similarity," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 3, no. 2, 2020, doi: 10.31289/jite.v3i2.3211.

[23] E. Aytaç, "Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey," *International Soil and Water Conservation Research*, vol. 8, no. 3, 2020, doi: 10.1016/j.iswcr.2020.05.002.

[24] F. A. Setiawan, M. Sadikin, and E. R. Kaburuan, "Analisis Permasalahan Perangkat Pencetak Menggunakan Metode Algoritma K-Means dan K-Medoids," *Teknika*, vol. 11, no. 2, 2022, doi: 10.34148/teknika.v11i2.471.

[25] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 2020. doi: 10.1109/DSAA49011.2020.00096.