

**APPLICATION OF DATA MINING WITH *K-MEANS CLUSTERING*
ALGORITHM FOR HYPERTENSION DISEASE CLASSIFICATION
AT PUSKESMAS LUBUK BUAYA PADANG IN 2024**

Anisa Mafera^{1*}, Herman Susilo², Ilma Nuria Sulrieni³
¹²³University of Syedza Saintika, Padang, West Sumatra, Indonesia

Article Info

Article history:

Received October 08, 2024
Revision October 12, 2024
Accepted December 28, 2024

Keywords:

Data Mining
K-Means Clustering Algorithm
Hypertension
Classification

ABSTRACT

Lubuk Buaya Health Center faces a high challenge in treating hypertension with an increasing prevalence of cases. This study aims to group patients based on certain characteristics using the *K-Means Clustering* algorithm in *data mining*, in order to provide deeper insights and support more effective medical interventions. The type of research used is descriptive quantitative research with observation method. The population of this study was medical record documents of hypertension patients with a total population of 1,897 data. The sample used was 330 data selected using the *Slovin* formula. Data collection was done through observation and interviews with officers. The results showed that the number of severe hypertension patients was 73 patients, moderate level was 12 patients, and low level was 248 patients. From these results, it can be concluded that the *k-means clustering* method successfully classifies the level of hypertension in Lubuk Buaya Health Center.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Anisa Mafera
Health Information Management Study Program, University of Syedza Saintika
Address : Street Prof. Dr. Hamka No. 228 Air Tawar Timur, Padang, West Sumatra, Indonesia
E-mail: anisamaghfirajaya@gmail.com

1. INTRODUCTION

Puskesmas is a health service center that provides basic health services to the community at the lowest level, especially at the village or kelurahan level. Puskesmas Lubuk Buaya Padang has experienced an increase in the number of patient visitors who come to the health center, patients with hypertension disease are mostly experienced by every patient who comes to Puskesmas Lubuk Buaya. In the context of Informatics, the problem that arises is the complexity and heterogeneity of patient health data that requires sophisticated analytical approaches to collect disease data periodically to increase data accumulation.

Medical records are an important part of health center services because they include a complete record of the patient's identity, history of examination, actions, and treatment received by the patient, and serve as a continuous legal document to support the quality of health services.

Current information provision has not fully met the need for adequate information. Therefore, this study aims to cluster hypertension patients and identify patient groups with severe, moderate, mild hypertension classifications using data mining techniques. The *k-means clustering* method is applied to group patients based on the characteristics present in the data. This method was chosen because of its ability to identify homogeneous groups from large data sets, thus providing a deeper understanding of

the variations and patterns that may appear in the hypertension patient population. This study used hypertension patient data from Lubuk Buaya Health Center. Cluster analysis with the *k-means* method was conducted using *RapidMiner* software. The results of the analysis will be reviewed in depth to identify classifications and patterns that appear in each group of patients.

Some previous studies in the context of cluster analysis of patient data have provided valuable insights, but still provide opportunities for further research. The study of the application of the K-means Clustering method for mapping the distribution of Acute Respiratory Tract Infection (ARI) disease. The purpose of this study is to apply the concept of data mining using the k-means algorithm (clustering) to get the results of the algorithm in mapping the spread of ARI disease into 2 clusters, where cluster 1 provides high recommendations totaling 10 districts, cluster 2 provides low recommendations totaling 2 districts.

Based on the results of the researcher's survey at the Lubuk Buaya Health Center, it was recorded that the second highest number of patients with hypertension in Padang City. Hypertensive disease is one of the significant public health problems and has a serious impact on quality of life. In order to address this problem, early detection and classification of hypertensive disease is crucial to provide timely medical intervention.

Based on the description of the background above, the researcher has conducted a study entitled "Application of Data Mining with the K-Means Clustering Algorithm for the Classification of Hypertension Diseases at the Lubuk Buaya Padang Health Center in 2024".

2. METHODS

This research method uses a quantitative descriptive design and literature study with K-Means cluster analysis of hypertension cases at the Lubuk Buaya Padang Health Center. The study was conducted from January to June 2024 with a population of 1,897 hypertension patient data and a sample of 330 cases randomly selected using the Slovin formula. The research attributes include age, gender, symptoms, and hypertension level.

Data was collected through observation, interviews, and literature studies, and processed using RapidMiner and Microsoft Excel. Data processing involved description, normalization, selection of distance measure (Euclidean Distance), determination of the number of clusters (using the elbow method), K-Means clustering, and output interpretation. Analysis was conducted by describing the data, characterizing the groups using scatter plots, and generating operational definitions for each variable measured.

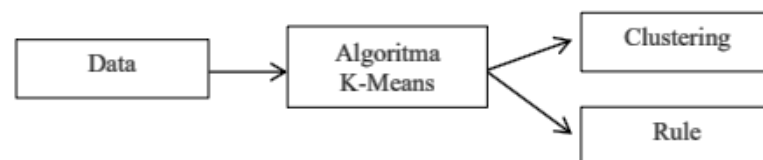


Figure 1. Data Processing

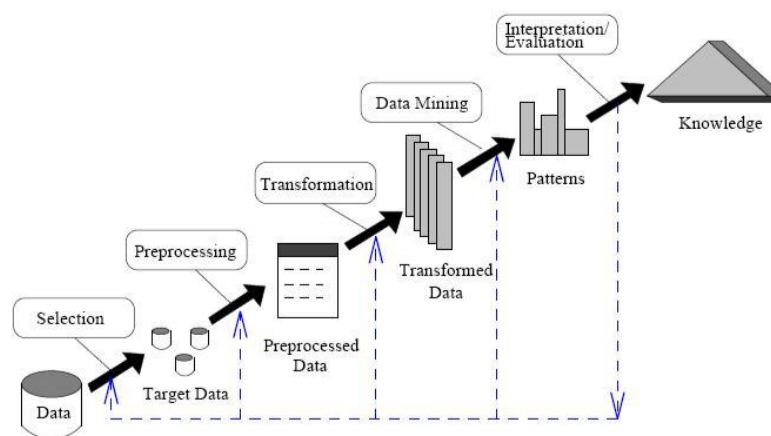


Figure 2. Knowledge Discovery Process in Database

3. RESULTS AND DISCUSSION

This research uses the K-Means algorithm in the application of data mining methods. As for knowing the clustering results regarding hypertension disease data. In this research has several stages are as follows:

3.1. Dataset

In this study, the dataset used has several attributes, namely age / year, gender, systolic and diastolic. The following is the dataset table used.

Table 1. Dataset

No.	Sequence to	Gender	Age	Systolic	Diastolic
1	1152	Female	75	130	70
2	282	Female	58	120	70
3	1534	Female	74	140	80
4	1519	Female	68	130	70
...
...

After getting the dataset, the next step is the data import process with the Rapid Miner software.

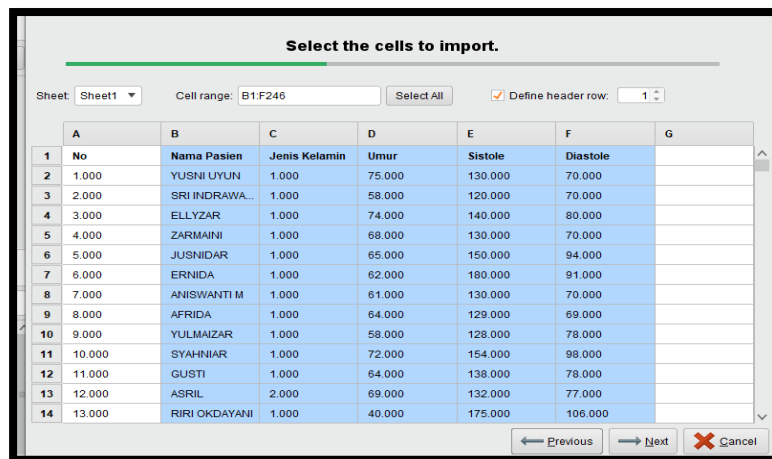


Figure 3. Import Data

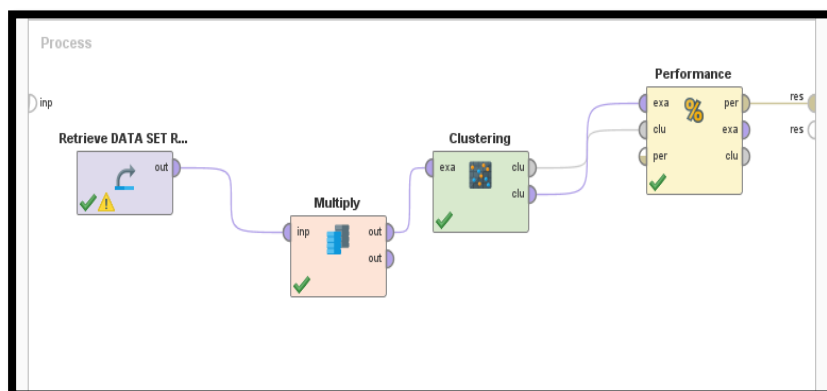


Figure 4. Import Data

3.2. RapidMiner Performance View

The image below is a view of the dataset that has been imported into Rapid Miner, then the dataset will be managed with the K-Means Algorithm with the number of clusters consisting of two clusters.

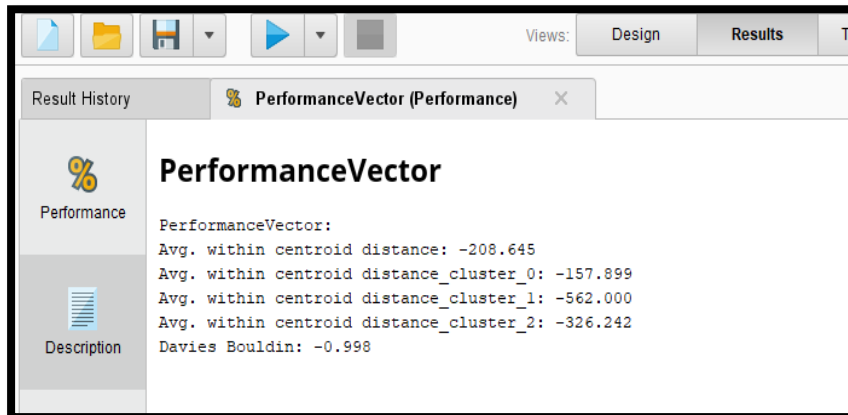


Figure 5. Performance page in RapidMiner

3.3. Data Classification Results with K-Means Clustering

This page is the clustering results page from the previous data importing process when we do a run as on the clustering process page, the clustering results page appears as shown below, where the number of clusters formed becomes two clusters, namely cluster 0, cluster 1, and cluster 2. Cluster 0 (Severe Hypertension), cluster 1 (Moderate Hypertension), and cluster 2 (Mild Hypertension).

Row No.	id	Nama Pasien	cluster	Jenis Kelamin	Umur	Sistole	Diastole
1	1	YUSNI UYUN	cluster_0	1	75	130	70
2	2	SRI INDRAM...	cluster_0	1	58	120	70
3	3	ELLYZAR	cluster_0	1	74	140	80
4	4	ZARMAINI	cluster_0	1	68	130	70
5	5	JUSNIDAR	cluster_2	1	65	150	94
6	6	ERNIDA	cluster_1	1	62	180	91
7	7	ANISWANTI M	cluster_0	1	61	130	70
8	8	AFRIDA	cluster_0	1	64	129	69
9	9	YULMAIZAR	cluster_0	1	58	128	78
10	10	SYAHNIAR	cluster_2	1	72	154	98

Data Clustering Results with K-Means

4. CONCLUSIONS

The application of the K-Means Method for clustering Hypertension disease data concluded that this algorithm is able to make hypertension disease data clustering. In other words, through the rapidminer application it is the same as the results applied with the number of clusters consisting of three clusters which clusters start from cluster 0, cluster 1, cluster 2. After clustering and testing, it can be concluded that of the three clusters, namely cluster C1 (Severe Hypertension) with 73 members, cluster C2 (Moderate Hypertension) with 12 members, cluster C3 (Mild Hypertension) with 245 members. It can be seen that the largest number of Hypertension patients are patients with mild Hypertension symptoms.

REFERENCES

- [1] Ade Bastian et al, 2018, Application of K-Means Clustering Analysis Algorithm on Human Infectious Diseases (Case Study of Majalengka Regency). *Journal of Information Systems*, Vol. 14, No. 1, April 2018, 26-32.
- [2] Budiraharjo, K. Rahardjo, B. Umar Basuki, M. & Handoyo, G. (2019) *Database System Design and Implementation*.
- [3] Padang City Health Office. Annual Report of Padang City Health Office. Padang: Padang City Health Office; 2019.
- [4] Padang City Health Office. Annual Report of Padang City Health Office. Padang: Padang City Health Office; 2020
- [5] Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering COVID-19 Data. *Journal of Industrial Engineering*, 10(3), 275-282. DOI: <https://doi.org/10.25105/jti.v10i3.8428>
- [6] Iriadi, Nandang, et al (2021), *Application of Data Mining with Rapid Miner*, Yogyakarta: Graha Ilmu, ISBN: 978-623-228-513-2
- [7] Kamila, I., Khairunnisa, U. & Mustakim. 2019. Comparison of K-Means and K-Medoids Algorithms for Grouping Loading and Unloading Transaction Data in Riau Province. *Scientific Journal of Information Systems Engineering and Management* Vol. 5, No. 1, February 2019.
- [8] Kristanto, Andri., 2018, *Information System Design and its Applications*,. Vol. 1, Ed.Revision, Yogyakarta.
- [9] Kumar, S. & Ganesh, R. T., 2016. Forensic Odontology: The Investigative Branch of Dentistry - A Review. *International Journal of Oral Health Dentistry*, 2(1), pp. 29-34.
- [10] Maharani, Adisya Putri. (2022). Redesign of Tracer (Outtguide) on Medical Record Document Storage at Mitra Husada Magetan Inpatient Clinic. Malang. *Indonesian Journal of Medical Records and Health Information (Jurmiki)*.
- [11] Manochandar, S., Punniyamorthy, M., & Jeyachitra, R. K. (2020). Development of new seed with modified validity measures for k-means clustering. *Computers & Industrial Engineering*, 141, 106290. DOI: <https://doi.org/10.1016/j.cie.2020.106290>
- [12] Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process. *IEEE Transactions on Network and Service Management*, 17(4), 2451-2479. DOI: <https://doi.org/10.1109/TNSM.2020.3016246>
- [13] Nasution, L. S., Maya, W. R., Halim, J., & Marsono, M. (2020). Data Mining to Analyze Silver Purchase Patterns Using the Fp-Growth Algorithm at Adi Saputra Tanjung Gold and Silver Shop. *Journal of Information Systems Technology and Computer Systems TGD*, 3(2), 96-107. DOI : <https://doi.org/10.53513/jsk.v3i2.2039>
- [14] Nugraha, U. (2021). Classifying Event Participants in Universities and Industries Using Knowledge Discovery in Databases. *Review of International Geographical Education Online*, 11(1), 526-542. DOI: 10.48047/rigeo.11.1.36
- [15] Purba, Nirina. et al. (2021). Application of K-Means Clustering Algorithm on the Spread of Acute Respiratory Infection (ARI) Disease in Riau Province. Pematang Siantar. *STIKOM Tunas Bangsa. Journal of Information System Research (JOSH)*.
- [16] Rahmadani, N., & Kurniawan, E. (2020). Implementation of the K-Means Clustering Method for Delinquent Electricity Accounts at PT PLN (Persero) Substation Kisaran. *Journal of Information Systems Technology and Computer Systems TGD*, 3(1), 103-117. DOI : <https://doi.org/10.53513/jsk.v3i1.201>
- [17] Suntoro, Joko (2019). *Data Mining Algorithms and Implementation with PHP Programming*. Jakarta:
- [18] Syahril, M., Erwansyah, K., & Yetri, M. (2020). Application of Data Mining to Determine Sales Patterns of School Equipment at Brand Wigglo Using the Apriori Algorithm. *Journal of Information Systems Technology and Computer Systems Tgd*, 3(1), 118-136.