BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset

Andysah Putera Utama Siahaan , Nur Azizah Harahap, Rahma Yuni Simanullang, Khairunnisa' , Puspita Wanny, Utari

Faculty of Postgraduate, Master Of Information Technology Study Program, Pembangunan Panca Budi University, Medan, Indonesia
Email: [1]andiesiahaan@gmail.com, [2,*]nazizahhrp@gmail.com, [3] rahmayunisimanullang2009@gmail.com, [4] Nisak030720@gmail.com, puspitawanny142@gmail.com, [5]Utariiijaya1999@gmail.com
Correspondence Author Email: nazizahhrp@gmail.com

Abstract−This study aims to analyze inpatient data using the K-Means Clustering method on a simulated dataset. The dataset includes various patient-related attributes such as age, billing amount, length of stay, medical condition, and type of admission. Several preprocessing steps were applied, including date conversion, duration calculation, numerical normalization, and one-hot encoding for categorical attributes. The Elbow Method was used to determine the optimal number of clusters, and clustering quality was evaluated using both the Silhouette Score and Davies-Bouldin Index. The analysis results show that the patients can be segmented into three major clusters, each exhibiting distinct characteristics—for example, younger patients with short and low-cost stays, and elderly patients with prolonged and more expensive hospitalizations. The resulting Silhouette Score of 0.14 and Davies-Bouldin Index of 1.74 reflect a moderate clustering performance, yet the model remains informative and meaningful. These clusters provide actionable insights that hospitals can use to optimize their service strategies, improve resource allocation, and enhance operational efficiency. Moreover, the study illustrates the practical application of unsupervised learning techniques in healthcare settings, contributing to data-driven decision-making practices and offering a foundation for further research into patient segmentation.

Keywords: Inpatient; clustering; K-Means; data analysis; cluster evaluation

1. INTRODUCTION

The advancement of digital technology has brought significant changes in the health sector, especially in the use of data to support fast and accurate decision making. Health data generated from various sources such as electronic medical records, wearable sensors, and laboratory reports are now increasingly large and complex. Therefore, an analysis method is needed that is able to explore hidden patterns in the data. One technique that is widely applied is unsupervised learning [1], [2]. especially the K-Means Clustering algorithm [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. This algorithm is considered effective for clustering data based on feature similarities, both numeric and categorical, without requiring previous labels. This is very relevant in the health sector, where patient data is often not clearly categorized but still needs to be analyzed to find homogeneous groups[17].

Other studies also emphasize the importance of big data analytics in improving the efficiency and resilience of healthcare systems [17]. On another front, Basile et al. (2024) [18] found that large-scale data utilization can enhance healthcare quality through process optimization and personalized patient care.

In the context of hospital management, clustering methods can also be used to design more adaptive policies. In a systematic review, Chao et al. (2023) [19] noted that big data analytics has been applied in various public health policies, including population risk mapping and more efficient resource allocation.

In patient data analysis, attributes such as age, type of disease, and length of hospitalization often provide critical insights into treatment patterns and patient needs [20]. By employing clustering algorithms like K-Means, patient data management can be optimized, enabling the identification of patient groups requiring special attention or specific medical interventions [21] .Other studies indicate that this algorithm is also suitable for application in medical imaging data [22].Additionally, other approaches, such as combining K-Means with optimization algorithms, have been developed to improve segmentation accuracy [3].

Although K-Means is known for its simplicity, one of its main challenges is determining the optimal number of clusters (k) and managing the various data attributes [14], [23], [24], [25]. Determining the right number of clusters is very important because it will affect the quality of the clustering results; if the number of clusters is too small, important information can be lost, while if there are too many, the results can be less meaningful. Therefore, various approaches have been developed to help determine the optimal value of k. Two of the most commonly used methods are the Elbow Method and the Silhouette Score, which have proven to be reliable approaches to evaluate clustering quality [14], [25], [26].

The Elbow Method is used to see the point at which increasing the number of clusters no longer provides a significant decrease in variance within the cluster, while the Silhouette Score measures how similar a data is to its cluster compared to other clusters, thus providing a more intuitive measure of compactness and separation between clusters. The fundamental objective of this study is to cluster inpatient data based on similar characteristics using the K-Means algorithm. In the context of healthcare, clustering like this is very useful because it can help hospitals identify important patterns in patient characteristics that were previously unseen. The clustering results are expected to provide valuable insights for hospital management, both in formulating more personalized and efficient service strategies, planning bed

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

allocation according to the needs trends of each patient group, and developing more targeted data-based decision support systems. Thus, cluster analysis not only improves the quality of service but also supports evidence-based decision making.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The research framework explains the sequence of stages that must be carried out in the research process. Each step is arranged systematically and interrelated. The arrangement of these stages aims to support the smooth implementation of research so that it runs more effectively and efficiently. The image of the research stages can be seen in image 1 below:

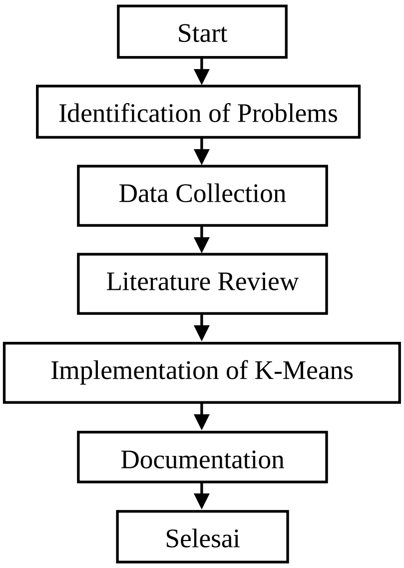

Figure 1. Research Framework

The following is an explanation of the research stages in the image above:

a. Problem Identification
At this stage, the researcher identifies the problems faced, namely how to group inpatient data based on certain characteristics to find hidden patterns or trends in the data. In addition, the objectives and scope of the research to be conducted are also determined.

b. Data collection
At this stage, relevant data collection is carried out in accordance with the research objectives, namely inpatient data from the simulation dataset. The data obtained will go through a selection process and initial processing to ensure completeness and suitability before further analysis is carried out. The collected data is an important basis for the grouping process and subsequent analysis.

c. Literature Study
This stage is carried out by tracing various scientific references that discuss the cluster analysis method and its application in grouping patient data or medical data. Through this stage, researchers gain an understanding of the basic concepts, appropriate methods, and related previous research results, so that they can be used as references and theoretical foundations in this study.

d. Application of K-Means Algorithm
This stage includes the process of designing and implementing cluster analysis algorithms on the collected inpatient data. The algorithm is used to group data into several categories based on the similarity of characteristics between the data, so that patterns or tendencies that are useful for the purpose of patient data analysis can be identified.

e. Conclusion
This stage is the drawing of conclusions based on the results of data analysis, literature studies, and implementation of the cluster analysis algorithm that has been carried out. Researchers draw conclusions based on research findings, evaluate whether the research objectives have been achieved, and interpret the significance of the results obtained. This conclusion serves as a summary of the entire research process and as a basis for further research development

2.2 Data Mining

Data mining is the process of utilizing computational methods on large data to find new information that is significant, valuable, and useful. This process includes automatically identifying hidden patterns by utilizing statistical techniques, machine learning, and databases, with the aim of building descriptive and predictive models from the analyzed

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

data[26][27][28]. Data mining is also an important part of the Knowledge Discovery in Databases (KDD) process, which consists of data pre-processing, modeling, evaluation, and interpretation of results. Through this iterative process, researchers are able to obtain accurate, useful, and applicable knowledge from large-scale data sets. Data mining is a data processing technique that aims to extract patterns, relationships, and important information from large and complex data sets. This process utilizes various analysis methods such as statistics, artificial intelligence algorithms, and database techniques to produce hidden information that was previously not directly detected. In the context of this study, data mining plays an important role in processing and analyzing inpatient data using the Cluster Analysis method. By applying this technique to a simulated dataset, this study aims to group patients based on certain characteristics, so that it can assist hospitals in making decisions related to medical services and patient care management[29][30].

2.3 K-Means Algorithm

The K-Means algorithm is one of the methods in data mining that is used to group data into several clusters based on similar characteristics. This algorithm was first introduced by MacQueen in 1967 and is still widely used today because of its simplicity and efficiency in handling large data[31][32][33].

The purpose of the K-Means algorithm is to divide a set of data into a number of clusters according to a predetermined number. Each data will be entered into the cluster that has the closest distance to the center point of the cluster or centroid, usually calculated using the Euclidean distance method [3][4]. In this study, the K-Means algorithm was applied to group inpatient data based on certain attributes such as age, type of disease, length of treatment, and cost, using a simulation dataset. The stages in the K-Means algorithm include:

1. Determine the number of clusters (k).
2. Select k initial centroids randomly.
3. Calculate the distance between each data to all centroids.
4. Groups data into clusters with the closest centroid.
5. Recalculate the centroid position based on the average of the data in each cluster.
6. Repeat the process until the centroid position is stable or the maximum number of iterations is reached.

The distance calculation generally uses the Euclidean Distance formula as follows:

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

Information:

The notation d(x, y) represents the distance between the data and the centroid, where x_i is the value of the i-th attribute of the data, while y_i is the value of the i-th attribute of the centroid. The number of attributes taken into account in calculating this distance is denoted by n, which indicates the number of dimensions of the data attributes.

2.4 Inpatients

An inpatient is someone who receives medical services in a hospital or similar health facility with the provision that they must undergo treatment for at least one night or more, according to the medical considerations of professional health workers. This service is intended for patients who require intensive monitoring, further medical action, ongoing supervision, or special care that cannot be done in a short time or on an outpatient basis. In its implementation, inpatient services involve various medical procedures, such as routine physical examinations, administration of drugs, nursing actions, and surgery if necessary. During the period of hospitalization, the patient's condition is monitored periodically by doctors and nursing staff, and recorded in medical records to monitor the development of their health. Patients who undergo inpatient care come from various age groups, from infants to the elderly, with various types of diseases, both acute and chronic. The decision to hospitalize a patient is generally influenced by the severity of the condition, potential complications, and the need for treatment that requires intensive supervision. Inpatient data plays an important role in supporting the management of hospital services, facility planning, and evaluation of the quality of health services to the community.

3. RESULT AND DISCUSSION

This study aims to cluster inpatients based on specific characteristics using the K-Means Clustering algorithm. The research stages include collecting simulated inpatient data, preprocessing, determining the optimal number of clusters, applying the algorithm, and evaluating and interpreting the results. These processes were described in detail in Chapter 2. This chapter presents the results of those stages and the subsequent discussion.

The dataset consists of numerical and categorical attributes such as age, billing amount, admission and discharge dates, medical condition, and type of admission. After preprocessing, including date conversion, creation of a new feature (Duration of Stay), encoding, and normalization, the data was ready for clustering. The first step was determining the optimal number of clusters using the Elbow method. Figure 1 shows the Elbow graph, where a clear elbow is observed at k = 5. This suggests that five is the optimal number of clusters for the dataset.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

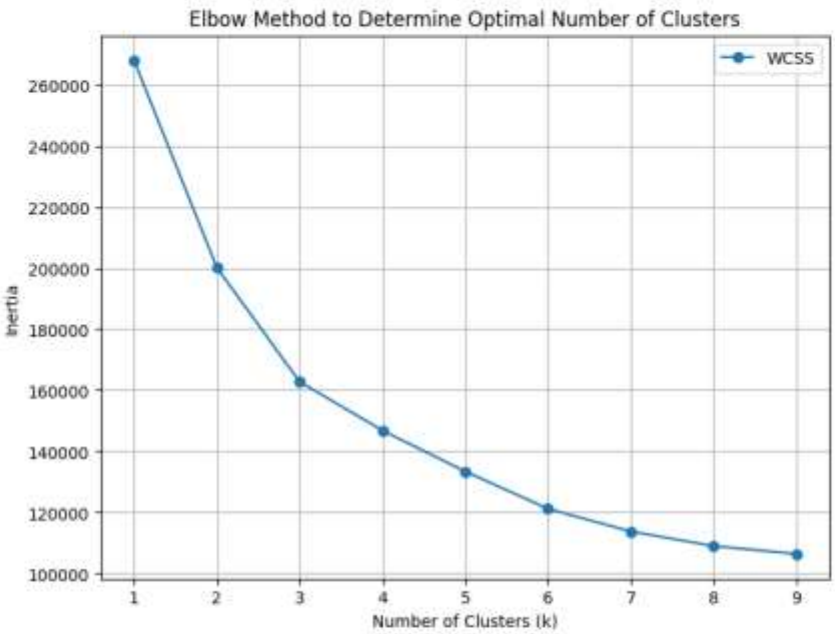

Figure 1. Elbow Plot for Determining Optimal Number of Clusters

After selecting the value of k, the K-Means algorithm was applied to the preprocessed data. The clustering procedure yielded five distinct patient groups, designated as clusters 0 through 4. To assess the quality of the resulting clusters, two principal evaluation metrics were employed: the Silhouette Score and the Davies-Bouldin Index. A Silhouette Score of 0.14 reflects a moderate degree of cluster separation, suggesting that although the segmentation is not optimal, it still reveals meaningful and interpretable patterns.Meanwhile, the Davies-Bouldin Index produced a value of 1.74 suggests some overlap between clusters, but still shows interpretable cluster structures. These values are common in complex, high-dimensional data such as patient datasets.

To better understand the characteristics of each cluster, a pairplot visualization was generated (Figure 2). This visualization illustrates the distribution among numerical variables and how each cluster is distributed across combinations of features.

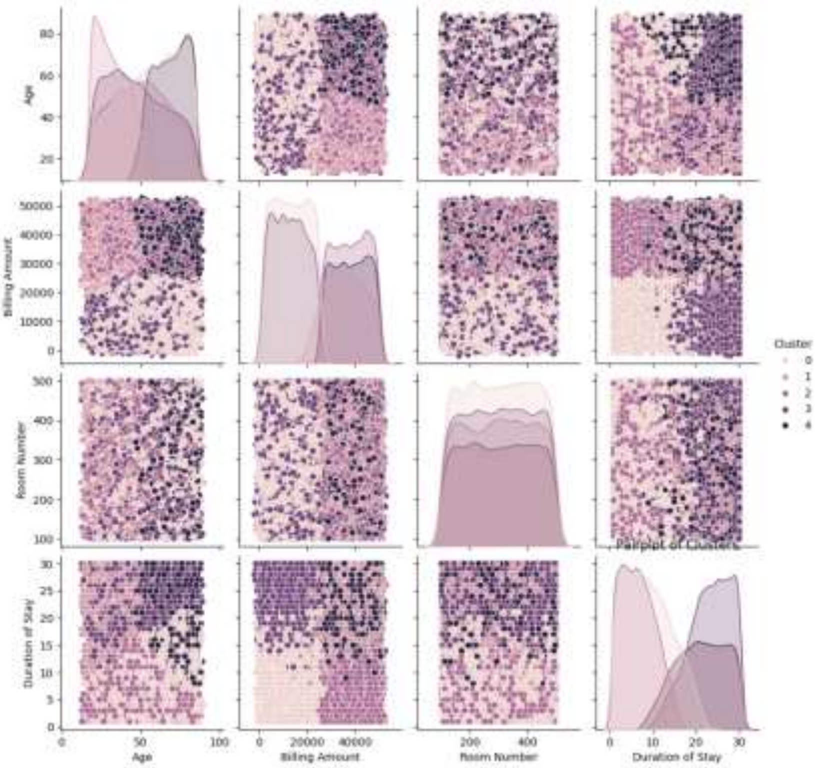

Figure 2. Pairplot Visualization of Patient Distribution by Cluster

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

Quantitative characteristics of each cluster are summarized in Table 1 This table presents the average values for each numerical attribute per cluster. Cluster 0 has an average age of 56.42 years, a billing amount of 13,150.60, and an average duration of stay of 9.04 days. Cluster 1 contains patients with an average age of 31.55 years, a relatively high billing amount of 35,593.99, and a longer average length of stay at 21.12 days. Cluster 2 consists of patients with an average age of 50.26 years, the highest billing amount among all clusters at 38,041.78, and the shortest duration of stay at 6.67 days. Cluster 3 includes patients with an average age of 47.80 years, the lowest billing amount at 12,617.20, and the longest average hospital stay of 23.21 days. Cluster 4 represents the oldest group, with an average age of 68.63 years, a high billing amount of 38,099.94, and a relatively long average duration of stay at 21.11 days.

Table 1. Average of Numerical Attributes per Cluster

Cluster	Age	Billing Amount	Duration of Stay
0	56.42	13150.60	9.04
1	31.55	35593.99	21.12
2	50.26	38041.78	6.67
3	50.26	38041.78	6.67
4	68.63	38099.94	21.11

Based on these results, Cluster 0 consists of middle-aged patients with low average treatment costs and relatively short hospital stays. This cluster likely represents patients with mild cases or routine care. Cluster 1 includes young patients with high treatment costs and relatively long hospital stays, possibly indicating complex cases or intensive procedures. Cluster 2 comprises middle-aged patients with the highest treatment costs but the shortest length of stay, suggesting patients who undergo costly short-term procedures. Cluster 3 consists of working-age patients with very low treatment costs but long hospital stays, which may reflect chronic conditions or extended conservative treatment. Finally, Cluster 4 contains elderly patients with high costs and long stays, indicating a possible need for long-term, intensive care.

From these results, it can be concluded that the clustering successfully differentiates patient groups based on a combination of age, cost, duration of stay. For example, Clusters 1 and 4 include patients with high treatment costs and long stays, while Clusters 0 and 3 include patients with lower costs, though Cluster 3 has extended stays.

These findings are useful for hospital management to design service strategies based on the resulting segmentation, such as providing specialized services for elderly patients with longer stays and allocating resources more efficiently. Nevertheless, the moderate evaluation metric values suggest there is room for model improvement, such as by adding more detailed clinical features or applying alternative clustering algorithms like DBSCAN or Hierarchical Clustering.

Overall, the K-Means Clustering method has proven capable of meaningfully segmenting inpatients into five distinct groups, offering a foundation for data-driven strategic decision-making in hospital settings.

The complete source code used in this study is available upon request by contacting the author via email for academic or development purposes.

These findings will be summarized in Chapter IV, which presents the conclusion and recommendations for future research directions.

4. CONCLUSION

This study aimed to apply the K-Means Clustering method to analyze inpatient data based on a simulated dataset that reflects common hospital attributes such as age, treatment cost, length of stay, medical condition, and type of service. Based on the analytical process carried out, three main clusters were successfully identified, each with distinct patient characteristics. The stages of analysis included data preprocessing, determining the optimal number of clusters using the Elbow method, and evaluating clustering results using the Silhouette Score and Davies-Bouldin Index. A Silhouette Score of 0.14 and a Davies-Bouldin Index of 1.74 indicate a moderate level of clustering quality, yet still informative and relevant. The main findings of this study demonstrate that grouping patients based on certain attributes can reveal patterns useful for supporting decision-making in hospital management, such as identifying patient groups with high length of stay or significant treatment costs. The key contribution of this research lies in the application of unsupervised learning methods in the healthcare domain, with the integration of quantitative evaluation metrics as a validation approach, and the use of simulated data that reflects real-world healthcare scenarios. The implication of this study is that hospitals can utilize the clustering results as a basis to improve operational efficiency, allocate resources more appropriately, and design service strategies tailored to the characteristics of each patient group. However, this study also has limitations, particularly the use of a simulated dataset that does not fully reflect the complexity of real-world data, and the clustering evaluation values remain within a moderate range. For future work, further research can be conducted using real data from larger and more diverse healthcare institutions, adding clinical attributes such as disease history or laboratory results, and

Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

exploring other clustering algorithms such as DBSCAN, Hierarchical Clustering, or deep learning-based approaches to enhance the accuracy of patient segmentation. Additionally, integrating clustering results into the hospital information system as part of a decision support system also represents a promising area for future development.

REFERENCES

[1] V. V. Baligodugula and F. Amsaad, "Unsupervised Learning: Comparative Analysis of Clustering Techniques on High-Dimensional Data," Mar. 2025, [Online]. Available: http://arxiv.org/abs/2503.23215

[2] H. Byeon, P. Kumar, I. R. Khan, F. Y. Alghayadh, M. A. Rusho, and M. Soni, "Unsupervised Single Valued Neutrosophic Sets Approach for Cloud Clustering," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 3580–3589. doi: 10.1016/j.procs.2025.04.613.

[3] S. Anam, Z. Fitriah, N. Hidayat, H. Akbar, and A. Maulana, "Classification Model for Diabetes Mellitus Diagnosis based on K-Means Clustering Algorithm Optimized with Bat Algorithm." [Online]. Available: www.ijacsa.thesai.org

[4] W. Aulia, A. Putera Utama Siahaan, L. Marlina, and M. Iqbal, "K-Means Clustering Algorithm Analysis For Grouping Patient Medical Record Data Based On Disease Type-Wina Aulia et.al K-Means Clustering Algorithm Analysis For Grouping Patient Medical Record Data Based On Disease Type," *Informatika dan Sains*, vol. 14, no. 04, p. 2024, doi: 10.54209/infosains.v14i04.

[5] I. D. Borlea, R. E. Precup, and A. B. Borlea, "Improvement of K-means Cluster Quality by Post Processing Resulted Clusters," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 63–70. doi: 10.1016/j.procs.2022.01.009.

[6] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, pp. 30–38, Mar. 2021, doi: 10.37148/bios.v2i1.19.

[7] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[8] I. K. Khan *et al.*, "Standardization of expected value in gap statistic using Gaussian distribution for optimal number of clusters selection in K-means," *Egyptian Informatics Journal*, vol. 30, Jun. 2025, doi: 10.1016/j.eij.2025.100701.

[9] Y. Li and H. Zhang, "Big data technology for teaching quality monitoring and improvement in higher education - joint K-means clustering algorithm and Apriori algorithm," *Systems and Soft Computing*, vol. 6, Dec. 2024, doi: 10.1016/j.sasc.2024.200125.

[10] M. R. Nahoujy, "Applying a K-means model to TSD data to find categories for the structural assessment of flexible pavements," *Transportation Engineering*, vol. 20, Jun. 2025, doi: 10.1016/j.treng.2025.100342.

[11] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *International Journal of Advances in Scientific Research and Engineering*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.

[12] Y. Pang and D. Nie, "Regional economic development level assessment based on K-means clustering algorithm," *Procedia Comput Sci*, vol. 262, pp. 1137–1143, 2025, doi: 10.1016/j.procs.2025.05.152.

[13] W. A. Prastyabudi, A. N. Alifah, and A. Nurdin, "Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 96–105. doi: 10.1016/j.procs.2024.02.156.

[14] M. Puli *et al.*, "Exploring Optimal Cluster Quality in Health Care Data (HCD): Comparative Analysis utilizing k-means Elbow and Silhouette Analysis," 2024. [Online]. Available: www.iscientific.org/Journal.html

[15] W. Zhang, L. Wu, and S. Zhang, "Clinical phenotype of ARDS based on K-means cluster analysis: A study from the eICU database," *Heliyon*, vol. 10, no. 20, Oct. 2024, doi: 10.1016/j.heliyon.2024.e39198.

[16] N. Alajmi, N. Ali, A. L. Ajmi, and M. Sabihaksoy, "A Review Of Big Data Analytic In Healthcare," 2021. [Online]. Available: https://www.researchgate.net/publication/354365827

[17] D. Sumrit, "Unveiling the effects of big data analytic capability on improving healthcare supply chain resilience: An integrated MCDM with spherical fuzzy information approach," *Results in Engineering*, vol. 25, Mar. 2025, doi: 10.1016/j.rineng.2025.104499.

[18] L. J. Basile, N. Carbonara, U. Panniello, and R. Pellegrino, "The role of big data analytics in improving the quality of healthcare services in the Italian context: The mediating role of risk management," *Technovation*, vol. 133, May 2024, doi: 10.1016/j.technovation.2024.103010.

[19] K. Chao, M. N. I. Sarker, I. Ali, R. B. R. Firdaus, A. Azman, and M. M. Shaed, "Big data-driven public health policy making: Potential for the healthcare industry," *Heliyon*, vol. 9, no. 9, Sep. 2023, doi: 10.1016/j.heliyon.2023.e19681.

[20] K. N. Singh and J. K. Mantri, "A clinical decision support system using rough set theory and machine learning for disease prediction," *Intelligent Medicine*, Aug. 2024, doi: 10.1016/j.imed.2023.08.002.

[21] V. Yfantis, A. Wagner, and M. Ruskowski, "Federated K-means clustering via dual decomposition-based distributed optimization," *Franklin Open*, vol. 10, Mar. 2025, doi: 10.1016/j.fraope.2024.100204.

[22] E. Durom *et al.*, "Quantification of 129Xe MRI Ventilation-defect-percent Using Binary-threshold, Gaussian Linear-Binning and K-means Methods: Differences in Asthma and COPD," *Acad Radiol*, 2025, doi: 10.1016/j.acra.2025.04.030.

[23] S. Ilbeigipour, A. Albadvi, and E. Akhondzadeh Noughabi, "Cluster-based analysis of COVID-19 cases using self-organizing map neural network and K-means methods to improve medical decision-making," *Inform Med Unlocked*, vol. 32, Jan. 2022, doi: 10.1016/j.imu.2022.101005.

[24] S. J. Maceachern and N. D. Forkert, "Machine learning for precision medicine," 2021, *Canadian Science Publishing*. doi: 10.1139/gen-2020-0131.

[25] B. Zhou, B. Lu, and S. Saeidlou, "A Hybrid Clustering Method Based on the Several Diverse Basic Clustering and Meta-Clustering Aggregation Technique," *Cybern Syst*, vol. 55, no. 1, pp. 203–229, 2024, doi: 10.1080/01969722.2022.2110682.

[26] F. Harahap, N. E. Saragih, E. T. Siregar, and H. Sariangsah, "Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat," *Jurnal Ilmiah Informatika*, vol. 9, no. 01, pp. 19–23, 2021, doi: 10.33884/jif.v9i01.3702.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 1, Maret 2025, Hal. 33 - 39
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1830
https://journal.fkpt.org/index.php/BIT

[27] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4. 5," *Journal of Computer and Information Systems Ampera*, vol. 4, no. 2, pp. 55–65, 2023, doi: 10.51519/journalcisa.v4i2.393.

[28] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus: Makan Barbeque Sepuasnya," *Jurnal Ilmiah Informatika*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[29] C. Hardjono and S. M. Isa, "Implementation of Data Mining for Churn Prediction in Music Streaming Company Using 2020 Dataset," *Journal on Education*, vol. 5, no. 1, pp. 1189–1197, 2022, doi: 10.31004/joe.v5i1.740.

[30] D. Marlina and M. Bakri, "Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4. 5," *Jurnal Teknologi Dan Sistem Informasi*, vol. 2, no. 1, pp. 23–28, 2021, doi: 10.33365/jtsi.v2i1.627.

[31] S. N. B. Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, pp. 31–40, 2022, doi: 10.53513/jursi.v1i1.4784.

[32] A. Yudistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.

[33] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.