JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

OPTIMIZATION OF K-MEANS AND K-MEDOIDS CLUSTERING USING DBI SILHOUETTE ELBOW ON STUDENT DATA

Dedy Hartama[1*], Selli Oktaviani[1]
[1]Informatics Engineering, STIKOM Tunas Bangsa
email: *dedyhartama@amiktunasbangsa.ac.id

Abstract: Clustering methods such as K-Means and K-Medoids are often used to analyze data, including student data, due to their efficiency. However, this method has weaknesses, such as sensitivity to selecting cluster centers (centroids) and cluster results that depend on medoid data. Clustering, an essential technique in data analysis, aims to reveal the natural structure of the data, even in the absence of labeled information. The study, conducted with complete objectivity, compared the performance of two popular clustering methods, K-Means, and K-Medoids, on student data. Three evaluation metrics, namely the Davies-Bouldin Index (DBI), silhouette score, and elbow method, were used to compare clustering and determine the ideal number of clusters for the two algorithms. The data taken in this study are in the form of names, attendance, assignments, formative, midterm exams, final exams, and quality numbers. Based on the existing optimization results, it can be concluded that the K-Means method excels in grouping Student Data. The best results were obtained from the K-Means Algorithm with the Silhouette Coefficient Method with a value of 0.7509 in cluster 2, and the Elbow Method with a value of 1428076.08 in cluster 2, DBI K-Medoids with a value of 0.7413 in cluster 3. So, the best cluster lies in 3 clusters.

Keywords: clustering; davies-bouldin indek; elbow method; k-means; k-medoids; silhouette score;

Abstrak : Metode clustering seperti K-Means dan K-Medoids sering digunakan untuk menganalisis data, termasuk data siswa, karena efisiensinya. Namun, metode ini memiliki kelemahan, seperti sensitivitas terhadap pemilihan pusat klaster (centroids) dan hasil klaster yang bergantung pada data medoid. Clustering, sebuah teknik penting dalam analisis data, bertujuan untuk mengungkapkan struktur alami dari data, bahkan tanpa adanya informasi berlabel. Penelitian ini, yang dilakukan dengan objektivitas penuh, membandingkan kinerja dua metode clustering populer, yaitu K-Means dan K-Medoids, pada data mahasiswa. Tiga metrik evaluasi, yaitu Davies-Bouldin Index (D.B.I.), silhouette score, dan metode elbow, digunakan untuk membandingkan clustering dan menentukan jumlah cluster yang ideal untuk kedua algoritma tersebut. data yang diambil dalam penelitian ini berupa nama, kehadiran, tugas, formatif, ujian tengah semester, ujian akhir semester, angka mutu. Berdasarkan hasil optimasi yang ada, dapat disimpulkan bahwasannya metode K-Means unggul dalam pengelompokkan Data Mahasiswa. Sehingga di peroleh hasil terbaik dari Algoritma K-Means dengan Metode Silhouette Coefficient dengan nilai 0,7509 di cluster 2, dan Elbow Method dengan nilai 1428076,08 di cluster 2, DBI K-Medoids dengan nilai 0,7413 di cluster 3. Sehingga cluster terbaik terletak pada 3 cluster.

Kata kunci: klasterisasi; davies-bouldin indek; elbow method; k-means; k-medoids; silhouette score;

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

In the increasingly advanced digital era, the industry's need for computer science graduates with high technical skill competence is increasing. However, additional characteristics, including practical skills, job experience, and interest in extracurricular activities, also affect computer science students' employment success and academic accomplishments [1] , educational institutions and students alike must have a thorough understanding of the elements that impact students' professional success to produce graduates who are equipped to confront the challenges of the working world.

Classification is one of the standard methods used in machine learning to understand the relationship between variables that affect students' careers. Among the most frequently used algorithms are Decision Tree, ID3, and Random Forest. The Decision Tree algorithm is known to be easy to understand and implement. Still, it is prone to overfitting, especially when the resulting decision tree is too complex or the dataset is limited [2] , [3] . ID3 is a variant of the Decision Tree that uses information gain to select the best attributes at each node. However, ID3 is often less accurate when applied to large, complex datasets [4] .

On the other hand, Random Forest is an algorithm that enhances prediction accuracy and reduces the chance of overfitting by combining numerous decision trees. Different subsets of the data are used to train each decision tree, so the final result is an average of the predictions of multiple trees [5] . Although Random Forest is more accurate, it requires more computation time than other algorithms, mainly due to the large number of decision trees used [6] .

Emphasize the importance of using clustering in education, especially in grouping student data based on academic attributes, economic background, or participation level. The study shows how clustering can identify hidden patterns and segment students, which can help provide more focused services according to their needs. The study serves as a reference for understanding the effectiveness of clustering algorithms in education [7] .

The clustering results show that this cluster encompasses various services, trade, and food and beverages sectors. This segmentation can support data-driven decision-making at the village level. Although this research shows promising results, expanding the quantity and variety of data and considering external factors affecting MSME performance is recommended. Thus, this study makes a valuable contribution to understanding the business characteristics of MSMEs in Sampang District [8] .

The various outcomes are the primary motivations for continuing to create and develop applications. This research aims to make an application that can evaluate cluster data using the K-Medoids method, which can be further optimized using the Davies Bouldin Index (DBI). Because the target application is students and lecturers who use it in learning and observers of the cluster field, it can be accessible through a browser to make it easier to use. The program is available on desktop and mobile platforms for ease of use. Through separately created applications, it is intended that this research will give an alternative to clustering and optimization [9] .

The outcomes encompassed

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

identifying clusters of MSMEs based on their closeness in the feature space within a specific region. Optimizing the clustering outcomes involved modifying algorithm parameters like epsilon and minimum points for DBSCAN and the number of clusters for K-Means. Furthermore, this study attained a deeper understanding of the arrangement and characteristics of MSME clusters in the region through a comparative analysis of the two methodologies [10] .

Therefore, this research attempts to contrast the performance of the three algorithms in classifying factors that influence the careers of computer science students. This study will also analyze the dominant factors that play a role in the success of student careers based on the classification results of the three algorithms. The results of this study are expected to provide deeper insights for educational institutions and students in designing more effective career development strategies.

METHOD

Literature Review

The following is a review of the literature used in this study.

Clustering and Data Grouping

Clustering is a data analysis method that groups objects based on the similarity of their attributes without the need for previously defined data labels [11] . According to Singh, Pandey, & Dubey [12] , clustering allows researchers to identify hidden patterns in data, making it an essential technique in fields such as education, business, and health. In the educational context, clustering can be used to group students based on various parameters, such as academic performance, economic background, or level of campus participation [13] .

K-Means and K-Medoids Algorithm

Two clustering techniques often employed in various applications are the K-Means and K-Medoids algorithms. According to the Euclidean distance between each data point and the centroid, K-Means divides the data into k groups [14] . In many different applications, two clustering approaches that are often used are the K-Means and K-Medoids algorithms. K-Means organizes the data into k groups based on the Euclidean distance between each data point and the centroid [15] . On the other hand, K-Medoids uses medoids as cluster centers, which are points in the dataset itself, making it more robust to outliers and non-normal data [16] . Although both methods are similar, K-Medoids is more suitable for datasets containing many outliers or noise [17] .

Clustering Quality Evaluation

The K-Means and K-Medoids algorithms are two clustering techniques often utilized in several applications. Based on the Euclidean distance between each data point and the centroid, K-Means divides the data into k groups [18] , [19] . D.B.I. measures each cluster's compactness and degree of separation from one another; it was created by Davies & Bouldin in 2019. The clustering quality is more excellent, and the D.B.I. value is lower. Rousseeuw proposed the Silhouette Score in 2020 to gauge how consistent items in a cluster are. The better the items are grouped, the higher the score for Silhouette. The Elbow approach looks at the sum of squared distances between the data points and the centroid or the decrease in the degree of distortions to find the ideal number of clusters [20] , [21] .

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

K-Means and K-Medoids Optimization

One of the main challenges in using K-Means and K-medoids is choosing the optimal number of clusters. According to Syakur et al. (2018), the elbow method is often used to determine the optimal number of clusters by analyzing the sum of squared distances graph [22] . In addition, the Silhouette Score can be used to validate the quality of the clustering results after the number of clusters is determined. The selection of initial centroids in K-Means can also be optimized to improve the clustering results [23] . In this study, this approach will be used to optimize the clustering process so that the results obtained are more accurate and relevant in grouping student data.

Application of Clustering in the Field of Education

Clustering in education, especially in student grouping, has various practical applications. Majeed & Ali (2022) explained that student grouping can help understand academic trends, identify groups that need more attention, and personalize educational approaches. With proper grouping, universities or educational institutions can make more effective policies regarding resource management, improving the quality of education and supporting students who need special academic assistance [24] .

Research Methods

The following is the research methodology used in this study.

Research Design

The following is a research design used to calculate K-Means and K-Medoids Clustering Optimization Using D.B.I., Silhouette, and Elbow on Student Data.

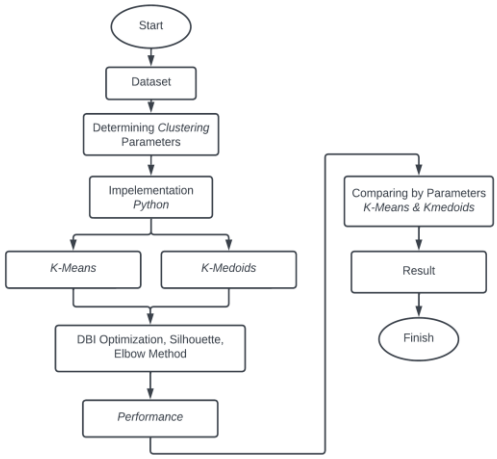

Image 1. Research Design

Below is a description of the research design image:

1. Start - Start the research process.
2. Dataset - Collect or select the dataset to be used.
3. Determining Clustering Parameters - Determines clustering parameters for grouping data.
4. Implementation in Python - Implementing clustering methods using Python.
5. K-Means & K-Medoids - Perform clustering with two algorithms, namely K-Means and K-Medoids.
6. D.B.I. Optimization, Silhouette, Elbow Method - Perform D.B.I. optimization and evaluation using Silhouette and Elbow methods to assess clustering results.
7. Performance - Measures the performance of clustering results.
8. Comparing by Parameters (K-Means & K-Medoids) - Comparing clustering results based on the parameters used between K-Means and K-Medoids.
9. Result - Get comparison results and conclusions.
10. Finish - Research is complete.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 1. Raw Data

Name	Present	Task	Task	Formative	UTS	UAS	...	Quality Score
Person 1	86	70	70	75	71	0	...	48
Person 2	7	0	0	0	0	0	...	1
Person 3	90	80	80	80	80	85	...	83
Person 4	90	80	80	80	85	85	...	84
Person 5	71	0	0	80	0	0	...	15
Person 6	36	0	0	0	70	0	...	21
Person 7	36	0	0	0	70	0	...	21
Person 8	50	0	0	0	70	0	...	22
...	...	...	...	...	...	...	...	...
Person 1556	93	79	79	81	82	83	...	83

Raw Data

The following is the raw data that will be used to calculate the optimization of K-Means and K-Medoids Clustering Using D.B.I., Silhouette, Elbow on Student Data.

D.B.I. (Davies-Bouldin Index)

The Davies-Bouldin Index (D.B.I.) formula measures how well the clusters are compacted. A lower D.B.I. value indicates better cluster quality. The following is the Davies-Bouldin Index (D.B.I.) formula used to calculate optimization and find out the optimal cluster:

$$DBI = \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j}\left(\frac{S_i + S_j}{d_{ij}}\right) \quad (1)$$

Information :

n : number of clusters

S_i : average distance between each point in the cluster i with the cluster center (centroid), which represents the cluster dispersion i

d_{ij} : the distance between the cluster center i and the cluster center j , which measures the distance between clusters.

$\max_{i \neq j}$: max value of the comparison between 2 clusters i and j.

Silhouette

The silhouette score formula measures how well objects are in the correct cluster. Silhouette scores range from -1 to 1, with higher values indicating that objects are well clustered. Here is the formula for the silhouette score :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Information :

$S(i)$: *Silhouette Score* for data points i

$a(i)$: average distance between a data point i and all other points in the same cluster (intra-cluster dispersion)

$b(i)$: average distance between a data point i and all other points in different clusters (intra-cluster dispersion)

$\max(a(i), b(i))$: maximum value between $(a(i)$ and $b(i)$

Elbow

The Elbow method determines the optimal number of clusters in the K-Means algorithm by analyzing the sum of squared errors (S.S.E.), also called the within-cluster sum of squares (W.C.S.S.). Although there is no specific formula for

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

the Elbow method, the approach is based on calculating S.S.E. for various numbers of clusters. The S.S.E. (Sum of Squared Errors) formula :

$$SSE = \sum_{i=1}^{n} \sum_{x \in C_i} \square x - \mu_i \square^2 \quad (3)$$

Information :

n : number of clusters

c_i : cluster i

x : data points in a cluster c_i

μ_i : centroid or cluster center c_i

$\square x - \mu_i \square^2$: squared Euclidean distance between data points x and the centroid μ_i

RESULT AND DISCUSSION

K-Means Optimization

Below is a table of cluster optimization results from the K-Means algorithm optimization of D.B.I. (Davies Boulden Index), S.C. (Silhouette Coefficient) and W.C.C.S. (Elbow Method) results.

Table 2. K-Means Optimization Calculation

Clusters	D.B.I. Results	SC Results	Elbow Results
2	1.0920	0.7255	1428076
3	0.9003	0.7509	1062113
4	0.7473	0.7430	850003
5	0.7609	0.7493	746389
6	0.7823	0.7508	654519
7	0.9694	0.2978	564521
8	0.9144	0.3727	510548
9	0.8724	0.3765	452518

Below is a plot from Google Colab in Python using the K-Medoids algorithm.

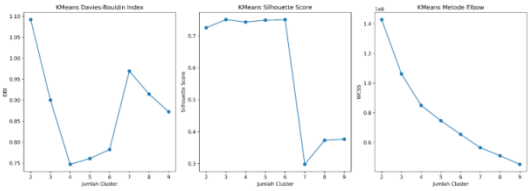

Image 2. K-Means Plot

K-Medoids Optimization

Below is the optimization table of the K-Medoids algorithm optimization cluster from D.B.I. (Davies Boulden Index), S.C. (Silhouette Coefficient) results, W.C.C.S. (Elbow Method) results.

Table 3. K-Medoids Optimization Calculation

Clusters	D.B.I. Results	SC Results	Elbow Results
2	0.7460	0.7232	32446.03
3	0.7413	0.7458	28588.85
4	1.0145	0.7337	27506.68
5	0.9570	0.7368	26032.24
6	0.7840	0.7501	24509.85
7	0.9946	0.3005	21889.27
8	0.9861	0.3000	21317.48
9	0.9318	0.2998	20571.18

The cluster starts from 2 because the minimum cluster range starts from 2. Subject clusters have categories or binomials, such as "Good" or "Not Good," or Polynomials, such as "Very Good," "Normal," or "Not Good."

Below is a plot image from Google Colab using Python language and the K-Medoids algorithm.

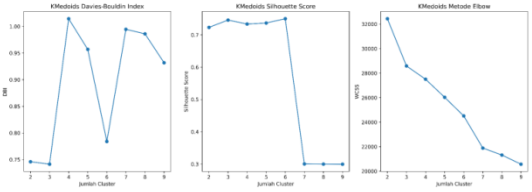

Image 3. K-Medoids Plot

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 289 – 296 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3531
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

CONCLUSION

Based on the optimization results, the K-Means method excels in grouping Student Data. So, the best results are obtained from the K-Means Algorithm with the Silhouette Coefficient Method with a value of 0.7509 in cluster 2 and the Elbow Method with 1428076.08 in cluster 2, D.B.I. K-Medoids with a value of 0.7413 in cluster 3. So that the best cluster is located in 3 clusters. The results of the clustering evaluation show that dividing data into three clusters produces the best grouping. This is usually determined based on evaluation metrics such as the Silhouette Index or the Davis-Bouldin Index (DBI), which measure how well the data in each cluster is similar (homogeneity) and how different they are between clusters. In other words, three clusters are considered optimal for representing the data accurately.

BIBLIOGRAPHY

[1] N. Majeed and S. Ali, "Application of clustering in higher education for student performance analysis," Journal of Educational Data Mining, vol. 14, no. 1, pp. 24-35, 2022.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68-73.

[3] G. Eason, B. Noble, and IN Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, Apr. 2020.

[4] IS Jacobs and CP Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, GT Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, Aug. 2021 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 2022].

[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2020.

[7] M. Singh, R. Pandey, and R. Dubey, "Clustering in educational data mining: A review," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 5, pp. 33-39, 2020.

[8] Firmansyah, Muhammad Iqbal, Yeni Kustiyahningsih, Eza Rahmanita, and Mochammad Syahrul Abidin. 2025. "Optimization of MSMEs Clustering in Sampang District Using K- Medoids Method and Silhouette Coefficient Method" 14 (March): 1–8. https://doi.org/10.34148/teknika.v14i1.1116.

[9] Christian, Ryan, and Deny Jollyta. 2022. "Web-Based Cluster Optimization Using K-Medoids and Davies Bouldin Index." *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)* 9 (1): 109–16. https://doi.org/10.33330/jurteksi.v9i1.1855.

[10] Sutramiani, Ni Putu, I. Made Teguh Arthana, Pramayota Fane'A Lampung, Shana Aurelia, Muhammad Fauzi, and I. Wayan Agus Surya Darma. 2024. "The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping Based on Asset Value and Turno-

ver." Journal of Information Systems Engineering and Business Intelligence 10 (1): 13–24. https://doi.org/10.20473/jisebi.10.1.13-24.

[11] A. Kumar and P. Goyal, "Analysis of clustering techniques: K-means and K-medoids," Journal of Computer Science and Applications, vol. 8, no. 2, pp. 45-50, 2020.

[12] M. Singh, R. Pandey, and R. Dubey, "Clustering in educational data mining: A review," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 5, pp. 33-39, 2020.

[13] N. Majeed and S. Ali, "Application of clustering in higher education for student performance analysis," Journal of Educational Data Mining, vol. 14, no. 1, pp. 24-35, 2022.

[14] S. Pratama and W. Pardede, "Comparative analysis of K-Means and K-Medoids in clustering education data," Procedia Computer Science, vol. 172, pp. 367-374, 2020.

[15] P. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, 2nd ed. Boston: Pearson, 2020.

[16] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA: Morgan Kaufmann, 2020.

[17] R. Aggarwal and R. Agrawal, "Data clustering: A review," International Journal of Data Science and Analytics, vol. 6, no. 1, pp. 37-56, 2021.

[18] R. Davies and D. Bouldin, "A cluster separation measure," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224-227, 2021.

[19] J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 2020.

[20] R. Ali and S. Ali, "Evaluating clustering methods with Silhouette and Elbow approaches," International Journal of Advanced Computer Science and Applications, vol. 12, no. 4, pp. 159-166, 2021.

[21] S. Syakur, F. Khotimah, E. Rochman, and B. Satoto, "Determining the number of clusters in K-Means and K-Medoids clustering using the elbow method," Journal of Physics: Conference Series, vol. 1028, no. 1, p. 012075, 2020.

[22] D. Liu and Y. Wang, "Optimizing clustering algorithms using hybrid approaches," IEEE Access, vol. 10, pp. 8754-8765, 2021.

[23] A. Ng, "Machine learning and data clustering in educational research," Machine Learning Journal, vol. 16, pp. 132-141, 2020.

[24] K. Pearson, "Principles of clustering and dimensional reduction," Journal of Computational Analysis, vol. 5, pp. 12-23, 2021.