Scientific Journal of Informatics
Vol. 11, No. 2, May 2024

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Comparative Analysis of K-Medoids and Purity K-Medoids Methods for Identifying Accident-Prone Areas in North Aceh Regency

Novia Hasdyna[1*], Rozzi Kesuma Dinata[2]

[1]Department of Informatics, Faculty of Computer and Multimedia, Universitas Islam Kebangsaan Indonesia, Indonesia
[2]Department of Informatics Engineering, Faculty of Engineering, Universitas Malikussaleh, Indonesia

Abstract.
Purpose: This study aimed to conduct a comprehensive comparative analysis between K-Medoids and Purity K-Medoids clustering methods for identifying accident-prone areas in North Aceh Regency. The analysis was carried out to provide valuable insights for policymakers and stakeholders to implement targeted interventions as well as improve road safety measures in the region.
Methods: This study compared the performance of K-Medoids and Purity K-Medoids, on accident-prone area data in North Aceh Regency. The algorithm performance was measured using the Davies-Bouldin Index (DBI) method, where a low value signifies superior performance. Additionally, the number of iterations produced by K-Medoids and Purity K-Medoid methods were compared, with lower iterations indicating better performance.
Result: The results showed that Purity K-Medoids had superior performance with an average of 2 iterations and DBI value of 0.7847 across 10 testing runs, while K-Medoids obtained 13.4 iterations and 1.5128, respectively.
Novelty: The study offers valuable insights into the effectiveness and efficiency of clustering methods for identifying accident-prone areas, as guides for policymakers and stakeholders in implementing targeted interventions to improve road safety measures. Additionally, the results provide a methodological framework for evaluating clustering algorithms in similar geographical contexts, enhancing the understanding of their applicability and performance in real-world scenarios.

Keywords: Clustering, K-medoid, Purity, Accident-prone, North Aceh
Received April 2024 / **Revised** April 2024 / **Accepted** May 2024

INTRODUCTION

North Aceh Regency is located in the province of Aceh, Indonesia, between 04° 43' - 05° 16' North Latitude and 96° 47' - 97° 31' East Longitude. Administratively, the regency comprises 27 districts and 852 villages, with a total area of approximately 3,296.86 km². According to the Central Statistics Agency of North Aceh Regency, the population was 572,961 in 2010, with a relatively high population density of 174 people per km². The district of Dewantara is the most densely populated area, with an average of 1,202 inhabitants per km², while the lowest is found in the Paya Bakong district, comprising 33 inhabitants per km² [1]. Data from the Traffic Directorate of the North Aceh Regional Police Department shows that from January to December 2022, there were 168 traffic accidents, where 148 cases were resolved. Generally, traffic accidents are common issues in major cities in Indonesia, occurring across various locations due to several factors, posing challenges in identifying accident-prone areas [2]. For the public, the information is useful for taking preventive measures and increasing surveillance in accident-prone areas by providing warning signs for traffic police. Meanwhile, for the government, it helps in policymaking to improve road infrastructure [3].

K-Medoids method also recognized as the Partitioning Around Medoid (PAM) algorithm, was conceived by Leonard Kaufman and Peter J. Rousseeuw. This method is similar to K-Means algorithm [4]–[6] as both are partitioning methods aimed at segmenting datasets into clusters but their divergence depends on the strategy of determining cluster centers. K-Means relies on the mean value of each cluster, while K-Medoids applies data objects as representatives, termed as medoids, for the cluster center [7], [8], [9]. This divergence is crucial, enabling K-Medoids to address the susceptibility of K-Means to outliers. Since

*Corresponding author.
Email addresses: noviahasdyna@gmail.com (Hasdyna)
DOI: [10.15294/sji.v11i2.3433](https://doi.org/10.15294/sji.v11i2.3433)

outliers can distort the mean value of a cluster, K-Medoids offer a more potential alternative by using medoids as representatives of the cluster center [10].

Several studies have applied K-Medoids method, each focusing on distinct objectives. Utomo [11] compared two data clustering algorithms to analyze patterns and determine the best processing method for evaluating COVID-19 spread in Indonesia. The comparison was carried out using data from the Indonesian Ministry of Health regarding confirmed cases and deaths. From the data analysis and processing using the K-Means and K-Medoids methods, the results showed that K-Means method was superior for clustering the spread of the COVID-19 outbreak. Similarly, Samudi et al. [12] addressed the significant health threat posed by the widespread COVID-19 disease, which affected various organs through airborne transmission or direct contact. To mitigate the spread of the virus, the central government implemented social distancing measures, altering daily routines. This includes transitioning from traditional classroom learning to online education using various applications.

During the COVID-19 pandemic, the use of online learning applications has gained significant attention, including K-Medoids Clustering Algorithm. Oktarina et al. [13] found that the optimal number of clusters differs between the k-means and k-medoids methods. Furthermore, Ushakov et al. [14] conducted experiments including clustering large-scale collections of face images into several thousand clusters. The results showed that the methods surpassed parallel improved versions of the most popular k-medoids clustering algorithms, achieving approximately linear parallel speedup. Lund et al. [15] also observed a significant increase in the number of LIS studies using a cluster analytic method, increasing from approximately 5 per year in the early 2000s to an average of 35 studies per year in the mid- and late-2010s. Among the journals, Scientometrics had the highest number of articles published within LIS using cluster analysis, with 102 studies. Additionally, Scientometrics was found as the most common subject area using a cluster analytic method, with 152 studies. The results suggested that cluster analysis has the potential to enhance accessibility to LIS study by offering an innovative and insightful method of knowledge discovery.

The comparative analysis considers both the effectiveness and efficiency of clustering methods in accurately identifying regions with a high frequency of accidents. Specifically, the Davies-Bouldin Index (DBI) method is used to evaluate the clustering performance, with a value indicating superior clustering quality. Additionally, the number of iterations required for convergence is compared between the K-Medoids and Purity K-Medoids methods, as lower iterations signify higher computational efficiency.

Based on the comparative analysis, this study aimed to provide perspectives on the suitability and effectiveness of clustering methods for identifying accident-prone areas in North Aceh Regency. The results provided valuable insights for policymakers and stakeholders in implementing targeted interventions to improve road safety measures. Additionally, the study provided a methodological framework for evaluating clustering algorithms in similar geographical contexts, enhancing the understanding of their applicability and performance in real-world scenarios. By exploring the strengths and limitations of each method, policymakers and stakeholders can gain valuable insights to guide their decisions on road safety interventions. This informed method has the potential to enhance public safety and mitigate the occurrence of road accidents in the region.

METHODS

Several methods were used for data analysis, including K-Medoids and Purity K-Medoids, with the study framework presented in Figure 1.

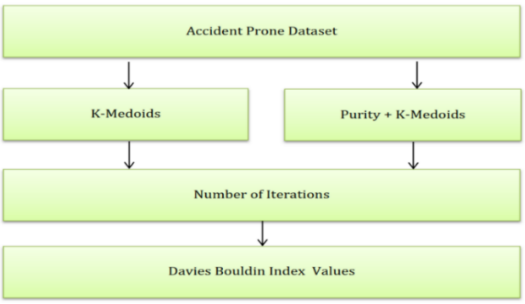

Figure 1. Research scheme

Figure 1 shows the stages in this study, which include gathering of accident-prone data within the scope of North Aceh. This is followed by clustering of the data using k-medoids, analyzing the number of iterations, and evaluating the DBI value. Subsequently, data purity is calculated using the purity method, where the minimum, medium, and maximum results are used as clusters in calculating K-Medoids. These results are further compared with the number of iterations and analyzed for DBI value.

K-Medoids

In K-Medoids, the cluster representatives are known as "medoids," which are positioned centrally within their respective clusters and are less influenced by outliers. The clustering process includes assessing the distance between medoids and other objects, with Figure 2 showing the computational steps of the K-Medoids method [16].

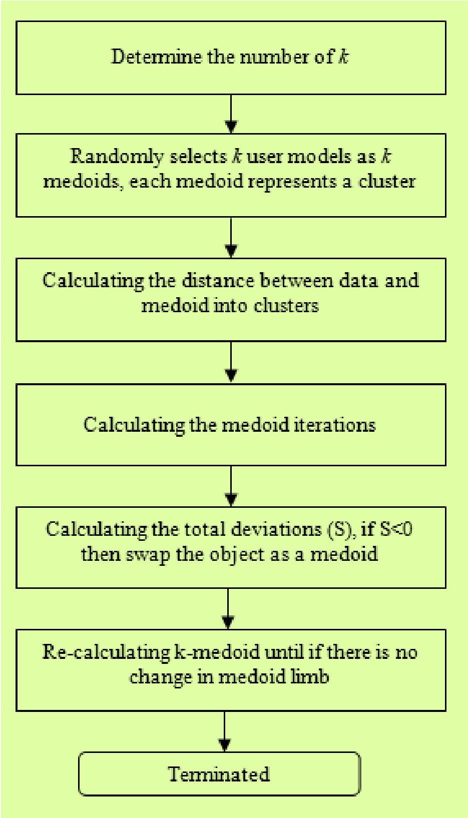

Figure 2. Framework of k-medoids clustering

Figure 2 shows the computational process of K-Medoids method, which comprises several stages. Initially, the algorithm determines the number of clusters, labeled as k, followed by a random selection of k initial medoids from the dataset. Subsequently, the Euclidean distance between each object and the medoids is computed to gauge their closeness. Objects are grouped based on their proximity to the medoids, with each cluster represented by medoids. The Euclidean distance formula, which measures the distance between two points in Euclidean space, is used for this computation. The iterative process continues until convergence is reached, resulting in clearly defined clusters represented by their respective medoids [17]:

$$d(xi,\mu j)=\sum_{i=1}^{n}(xi-\ \mu j)^2 \qquad (1)$$

The next step in the algorithm is iteration through the medoids, followed by the computation of total deviation (S) using the specified formula:

$$S = b - a \quad (2)$$

In this context, 'a' represents the summation of the closest distances between the object and the initial medoid, while 'b' denotes the summation of the closest distances between the object and the new medoid. When the total deviation (S) is less than zero (S < 0), the algorithm exchanges the object with other data to designate a new k as a medoid. This iterative process persists through steps 3 to 5, and the algorithm stops when there is no alteration in the medoid.

Purity algorithm

Purity is used to determine the purity value of a cluster, showing the most appropriate cluster member within a class. The formula used to calculate purity is depicted in Equation 3 [18].

$$\text{Purity}(y)=\frac{1}{N_y}\max(n_{xy}) \quad (3)$$

Purity (y) represents the level of purity for the y-variable, where N_y indicates the amount of data within the y-cluster, and y denotes the cluster index.

Davies-Bouldin Index (DBI)

When computing the DBI, the first step includes calculating the Sum of Square Within Cluster (SSW), which represents the cohesion value. The computation of DBI is achieved using the following formula expressed below [19]:

$$SSWi=\frac{1}{mi}\sum_{j=i}^{mi} d(xj,ci) \quad (4)$$

After computing SSW, the next step is calculating the Sum of Square Between Clusters (SSB), which indicates the separation value among clusters [20]. This is accomplished using the following formula [21]:

$$SSBi,j=d(ci,cj) \quad (5)$$

The next step is calculating the Ratio to compare the i-cluster with the j-cluster [22] using the formula stated below [23]:

$$Rij=\frac{SSWi+SSWj}{SSBij} \quad (6)$$

After obtaining the ratio value, the final step is to calculate the DBI value using the formula below [24]:

$$DBI=\frac{1}{k}\sum_{i=1}^{k} \max_{i\neq j}(R_{i,j}) \quad (7)$$

In DBI, a smaller value indicates superior clustering outcomes, suggesting that clusters are more internally cohesive and more distinct from each other, which is desirable in clustering tasks [25].

RESULTS AND DISCUSSIONS

Study dataset

The study dataset comprises accident-prone areas in North Aceh Regency based on the number of incidents, vehicles, and casualties. This classification is carried out to facilitate the community in identifying accident-prone areas, with three vulnerability levels, namely Not Vulnerable, Vulnerable, and Highly Vulnerable. The data used in this study are obtained from accidents that occurred in 2022, where the dataset comprises cases from 7 sub-districts along the National Highway Medan-Banda Aceh, under the jurisdiction of North Aceh Police Resort (Polres Aceh Utara). These sub-districts include Tanah Jambo Aye, Baktiya, Baktiya Barat, Lhoksukon, Tanah Luas, Tanah Pasir, and Syamtalira Aron.

Table 1. Study dataset

No	Village	Sub-district	Number of Incidents	Number of Vehicles Involved	Number of Casualties	Σ
1	Rawa Iteuk	0,000	0,833	1,670	0,769	3,273
2	Ceumpeudak	0,167	0,000	0,100	0,000	0,100
3	Matang Kumbang	0,167	0,800	0,900	1,000	2,700
4	Lhok Seutui	0,167	0,400	0,500	0,385	1,285
5	Alue Ie Puteh	0,333	0,400	0,500	0,385	1,285
6	Alue Bili Rayeuk	0,333	0,000	0,100	0,000	0,100
7	Matang Panyang	0,333	0,000	0,100	0,231	0,331
8	Ceubrek	0,333	0,400	0,500	0,385	1,285
9	Meunasah Dayah	0,500	0,400	0,500	0,385	1,285
10	Alue Drien	0,500	0,000	0,100	0,077	0,177
11	Meunasah Ranto	0,500	0,000	0,300	0,154	0,454
12	Alue Mudem	0,500	0,400	0,500	0,231	1,131
13	Kuta	0,500	0,000	0,500	0,462	0,962
14	Bintang Hu	0,500	0,200	0,300	0,077	0,577
15	Meunasah Geumata	0,500	0,000	0,100	0,077	0,177
16	Meunasah Trieng	0,500	0,200	0,300	0,154	0,654
17	Meunasah Nga	0,500	0,200	0,200	0,077	0,477
18	Paya Beurandang	0,500	0,200	0,200	0,077	0,477
19	Alue Keujruen	0,667	0,600	0,500	0,308	1,408
20	Pulo U	0,667	1,000	0,000	0,460	1,460
21	Ulee Tanoh	0,833	1,000	0,100	0,260	1,360
22	Teupin Punti	1,000	0,200	0,300	0,154	0,654
23	Dayah Teungku	1,000	0,400	0,500	0,231	1,131
24	Cibrek Tunong	1,000	0,000	0,100	0,077	0,177

Results of the purity calculations

The results of the purity calculations are as follows:

$$\text{Purity (d1)} = \frac{1}{(3{,}273)}(1{,}670)$$
$$= 0{,}510$$
$$\text{Purity (d2)} = \frac{1}{(0{,}100)}(0{,}100)$$
$$= 1{,}677$$
$$\text{Purity (d3)} = \frac{1}{(2{,}700)}(1{,}000)$$
$$= 0.370$$

Table 2. The results of the purity calculations

Village	Purity Value
Rawa Iteuk	0,510
Ceumpeudak	1,667
MatangKumbang	0,370
LhokSeutui	0,389
Alue Ie Puteh	0,389
Alue Bili Rayeuk	3,333
MatangPanyang	1,008
Ceubrek	0,389
Meunasah Dayah	0,389
Alue Drien	2,826
Meunasah Ranto	1,102
Alue Mudem	0,442
Kuta	0,520
Bintang Hu	0,867
MeunasahGeumata	2,826
MeunasahTrieng	0,765

Meunasah Nga	1,048
Paya Beurandang	1,048
Alue Keujruen	0,474
Pulo U	0,685
Ulee Tanoh	0,735
TeupinPunti	1,529
Dayah Teungku	0,884
CibrekTunong	5,652

The minimum and purity values observed in the accident-prone dataset are found to be 0.370 and 5.652, occurring at the 3rd and 24th data point. In this study, the data point with the minimum purity value is designated as the initial Medoid, while the value serves as the new Medoid.

Clustering process using K-Medoids

The initialization of centers for 3 clusters from the sample data includes randomly selecting Medoids, with the initial cluster center values for the first test presented in Table 3.

Table 3. Data testing

Test Number	Cluster Center		
1	5	10	12
2	1	2	3

Table 4. Initial cluster center values

No	Village	Sub-district (X1)	Number of Incidents (X2)	Number of Vehicles (X3)	Number of Casualties (X4)
5	Alue Ie Puteh	0.16667	0.80000	0.88889	0.63636
10	Alue Drien	0.50000	0.20000	0.44444	0.36364
12	Alue Mudem	0.50000	0.20000	0.33333	0.45455

Calculating the nearest distance using the euclidean distance equation

The distance equation was calculated using the Euclidean Distance method for each cluster as shown in Table 5, to perform clustering on each data point.

Table 5. Results of k-medoids algorithm calculation iteration 1

No	Cost 1	Cost 2	Cost 3	Proximity
1	0.91030	0.52020	0.53203	0.52020
2	1.29525	0.73030	0.73891	0.73030
3	0.27029	1.07483	1.09178	0.27029
4	1.28448	0.62806	0.63805	0.62806
5	0.00000	0.86199	0.90156	0.00000
6	0.38023	1.08844	1.10140	0.38023
7	0.93020	0.29228	0.20031	0.20031
8	1.32703	0.53230	0.54405	0.53230
9	1.39506	0.60808	0.59810	0.59810
10	0.86199	0.00000	0.14356	0.00000
11	1.35601	0.55849	0.53230	0.53230
12	0.90156	0.14356	0.00000	0.00000
13	0.69058	0.27029	0.24619	0.24619
14	0.34551	0.76850	0.82274	0.34551
15	0.77516	0.21969	0.29224	0.21969
16	1.25006	0.42915	0.40468	0.40468
17	1.39506	0.60808	0.59810	0.59810
18	1.33883	0.50326	0.49939	0.49939
19	1.33883	0.50326	0.49939	0.49939
20	1.33883	0.50326	0.49939	0.49939
21	1.37695	0.54340	0.52428	0.52428
22	1.25288	0.54351	0.56954	0.54351
23	1.49563	0.68956	0.68674	0.68674
24	1.06130	0.53852	0.55732	0.53852

Grouping data based on nearest distance Iteration 1

After calculating the distance for each data point, grouping is carried out according to their clusters. The cluster group of a data point is determined by its nearest distance to a cluster.

Determining new medoids values

As shown in Table 6, new Medoids values are determined through random selection, ensuring that each previously selected medoid cannot reselected again as Non-Medoids.

Table 6. New medoids

No	Village	Sub-district (X1)	Number of Incidents (X2)	Number of Vehicles (X3)	Number of Casualties (X4)
1	Rawa Iteuk	0.16667	0.80000	0.88889	0.63636
2	Ceumpeudak	0.50000	0.20000	0.44444	0.36364
3	Matang Kumbang	0.50000	0.20000	0.33333	0.45455

Recalculate the distance of each object in Iteration 2 using the new medoids, followed by data point clustering through the computation of the Euclidean Distance method.

Grouping data based on nearest distance in iteration 2

After calculating the distance for each data point, grouping is carried out according to data clusters. Moreover, the cluster group of a data point is determined by its nearest distance to a cluster.

Calculating the total deviation (S)

After obtaining the distance values between Iterations 1 and 2, total deviation (S) is calculated by determining the difference between the new and the old total cost value. When $S < 0$, the object values are swapped by determining new medoids, leading to repetition of the process from steps 4 to 6.

$$
\begin{aligned}
S &= \text{new total cost} - \text{old total cost} \\
&= 12.30 - 10.00 \\
&= 2.30
\end{aligned}
$$

The total deviation result is greater than zero ($S > 0$), leading to the termination of the test at Iteration 2. The results from the last iteration serve as clustering parameters and the cluster members are classified as Not Vulnerable, Vulnerable, and Highly Vulnerable based on their centroids, namely C1, C2, and C3.

Table 7. Final results

No	Village	Prediction
1	Rawa Iteuk	Vulnerable
2	Ceumpeudak	Vulnerable
3	Matang Kumbang	Not Vulnerable
4	Lhok Seutui	Vulnerable
5	Alue Ie Puteh	Not Vulnerable
6	Alue Bili Rayeuk	Not Vulnerable
7	Matang Panyang	Highly Vulnerable
8	Ceubrek	Vulnerable
9	Meunasah Dayah	Highly Vulnerable
10	Alue Drien	Vulnerable
11	Meunasah Ranto	Highly Vulnerable
12	Alue Mudem	Highly Vulnerable
13	Kuta	Highly Vulnerable
14	Bintang Hu	Not Vulnerable
15	Meunasah Geumata	Vulnerable
16	Meunasah Trieng	Highly Vulnerable
17	Meunasah Nga	Highly Vulnerable
18	Paya Beurandang	Highly Vulnerable
19	Alue Keujruen	Highly Vulnerable
20	Pulo U	Highly Vulnerable
21	Ulee Tanoh	Highly Vulnerable

22	Teupin Punti	Vulnerable
23	Dayah Teungku	Highly Vulnerable
24	Cibrek Tunong	Vulnerable

Iteration results

In each test conducted on the annual data, all iterations were concluded or completed by the second iteration.

Table 8. Comparison of iteration results

Test No	Number of Iterations	
	Conventional K-Medoids	Purity K-Medoids
1	13	2
2	17	2
3	10	2
4	13	2
5	16	2
6	9	2
7	17	2
8	10	2
9	16	2
10	13	2
Average	13.4	2

Result of the DBI value

The validation results of DBI obtained during the clustering process using traditional K-Medoids and Purity K-Medoids are presented in Table 9.

Table 9. Validation results for DBI

No	Dataset	Number of K	DBI Value	
			Conventional K-Medoids	Purity K-Medoids
		3	1.5128	0.7847

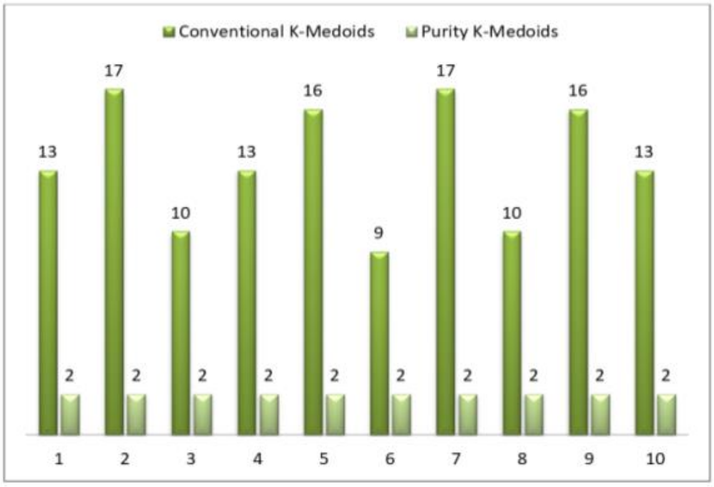

Figure 3. Graph displaying the iteration count of k-medoids clustering with k = 3.

CONCLUSION

In conclusion, this study conducted a comparative analysis between K-Medoids and Purity K-Medoids clustering methods for identifying accident-prone areas in North Aceh Regency. The Purity K-Medoids showed superior performance compared to the traditional K-Medoids, with an average of only 2 iterations and a DBI value of 0.7847 across 10 testing runs. This showed the effectiveness and efficiency of the method in accurately identifying accident-prone areas. Meanwhile, K-Medoids method required an average of 13.4 iterations and DBI value of 1.5128 under similar testing conditions. The results provided valuable guidance for policymakers and stakeholders in implementing targeted interventions to improve road safety measures in North Aceh Regency. By using clustering methods, such as Purity K-Medoids, authorities could identify accident-prone areas more accurately, allowing for the deployment of preventive measures, increased surveillance, and infrastructure improvements. Furthermore, this study contributed methodological insights into evaluating clustering algorithms in similar geographical contexts.

REFERENCES

[1] M. Muliari, Mahdaliana, I. Irfannur, Y. Akmal, and A. S. Batubara, "Pollutant levels in the waters of the industrial area of North Aceh and Lhokseumawe Regency, Indonesia," *Mar. Pollut. Bull.*, vol. 201, no. February, p. 116170, 2024, doi: 10.1016/j.marpolbul.2024.116170.

[2] A. Chand, S. Jayesh, and A. B. Bhasi, "Road traffic accidents: An overview of data sources, analysis techniques and contributing factors," *Mater. Today Proc.*, vol. 47, pp. 5135–5141, 2021, doi: 10.1016/j.matpr.2021.05.415.

[3] D. Petrovic, R. Mijailović, and D. Pešić, "Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers," *Transp. Res. Procedia*, vol. 45, no. 2019, pp. 161–168, 2020, doi: 10.1016/j.trpro.2020.03.003.

[4] A. Alamsyah *et al.*, "Customer Segmentation Using the Integration of the Recency Frequency Monetary Model and the K-Means Cluster Algorithm," *Sci. J. Informatics*, vol. 9, no. 2, pp. 189–196, 2022, doi: 10.15294/sji.v9i2.39437.

[5] F. R. Devi, E. Sugiharti, and R. Arifudin, "The Comparison Combination of Naïve Bayes Classification Algorithm with Fuzzy C-Means and K-Means for Determining Beef Cattle Quality in Semarang Regency," *Sci. J. Informatics*, vol. 5, no. 2, pp. 194–204, 2018, doi: 10.15294/sji.v5i2.15452.

[6] N. Hazimah, S. Harahap, A. Amirullah, M. B. Saputro, and I. A. Tamaroh, "Classification of potential customers using C4.5 and k-means algorithms to determine customer service priorities to maintain loyalty," *J. Soft Comput. Explor.*, vol. 3, no. 2, pp. 123–130, 2022, doi: 10.52465/joscex.v3i2.89.

[7] R. K. Dinata, S. Retno, and N. Hasdyna, "Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm," *Rev. d'Intelligence Artif.*, vol. 35, no. 3, pp. 193–199, 2021, doi: 10.18280/ria.350302.

[8] N. Sureja, B. Chawda, and A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," *J. Comput. Math. Data Sci.*, vol. 3, no. April, p. 100034, 2022, doi: 10.1016/j.jcmds.2022.100034.

[9] A. Sobrinho Campolina Martins, L. Ramos de Araujo, and D. Rosana Ribeiro Penido, "K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow," *Int. J. Electr. Power Energy Syst.*, vol. 157, no. August 2023, 2024, doi: 10.1016/j.ijepes.2024.109861.

[10] Y. Shen, D. Zhang, R. Wang, J. Li, and Z. Huang, "SBD-K-medoids-based long-term settlement analysis of shield tunnel," *Transp. Geotech.*, vol. 42, no. May, p. 101053, 2023, doi: 10.1016/j.trgeo.2023.101053.

[11] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.

[12] S. Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic," *SinkrOn*, vol. 5, no. 1, p. 116, 2020, doi: 10.33395/sinkron.v5i1.10649.

[13] C. Oktarina, K. A. Notodiputro, and I. Indahwati, "Comparison of K-Means Clustering Method and K-Medoids on Twitter Data," *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, pp. 189–202, 2020, doi: 10.29244/ijsa.v4i1.599.

[14] A. V. Ushakov and I. Vasilyev, "Near-optimal large-scale k-medoids clustering," *Inf. Sci. (Ny).*, vol. 545, pp. 344–362, 2021, doi: 10.1016/j.ins.2020.08.121.

[15] B. Lund and J. Ma, "A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering," *Perform. Meas. Metrics*, vol. 22, no. 3, pp. 161–173, 2021, doi: 10.1108/PMM-05-2021-0026.

[16] D. Dwi Aulia and N. Nurahman, "Comparison Performance of K-Medoids and K-Means Algorithms In Clustering Community Education Levels," *J. Nas. Pendidik. Tek. Inform.*, vol. 12,

no. 2, pp. 273–282, 2023, doi: 10.23887/janapati.v12i2.59789.
[17] M. Tiwari *et al.*, “BanditPAM++: Faster k-medoids Clustering,” *Adv. Neural Inf. Process. Syst.*, vol. 36, no. NeurIPS, 2023.
[18] R. K. Dinata, R. T. Adek, N. Hasdyna, and S. Retno, “K-nearest neighbor classifier optimization using purity,” *AIP Conf. Proc.*, vol. 2431, no. 1, 2023.
[19] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, 2023, doi: 10.1016/j.neucom.2023.01.043.
[20] E. Muningsih, C. Kesuma, Sunanto, Suripah, and A. Widayanto, “Combination of K-Means method with Davies Bouldin index and decision tree method with parameter optimization for best performance,” *AIP Conf. Proc.*, vol. 2714, no. 1, 2023.
[21] J. Chen, Y. Gu, S. Cai, H. Chen, and J. Chen, “A novel test case prioritization approach for black-box testing based on K-medoids clustering,” *J. Softw. Evol. Process*, vol. 36, no. 4, 2023.
[22] M. Mughnyanti, S. Efendi, and M. Zarlis, “Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
[23] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, “Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities,” *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.
[24] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, “Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies,” *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 07–15, 2022, doi: 10.30871/jaic.v6i1.3485.
[25] S. Ramadhani, D. Azzahra, and T. Z, “Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student’s Thesis,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.