The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00043 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Application of the K-Medoids Clustering Method for Grouping High-Risk Areas of Violence Against Women and Children

Annisa Afrilia Zahra[1] **Rozzi Kesuma Dinata**[2] **Maryana**[3]
[1]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia
annisa.200170087@mhs.unimal.ac.id
[2]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia
rozzi@unimal.ac.id
[3]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia
maryana@unimal.ac.id
Corresponding Author : **annisa.200170087@mhs.unimal.ac.id** | **Phone : +62853-7232-2553**

Abstract
Violence against women and children has been increasing in both quantity and variety, necessitating special attention. This study aims to cluster areas prone to violence against women and children in North Aceh using the K-Medoids Clustering method. The data used includes physical, sexual, exploitation, and neglect violence, obtained from 542 villages sourced from Unit II PPA Polres North Aceh for the period of 2021-2023. The clustering is categorized into three clusters: very prone, prone, and not prone. The results show that in 2021, there were 16 very prone villages, 22 prone villages, and 506 not prone villages, with the smallest DBI value of 0.12263 from 8 trials. In 2022, there were 22 very prone villages, 18 prone villages, and 502 not prone villages, with a DBI value of 0.10517 from 10 trials. In 2023, there were 15 very prone villages, 11 prone villages, and 516 not prone villages, with a DBI value of 0.21408 from 6 trials. The developed web-based system, using PHP and UML, is expected to assist authorities in preventing and addressing violence in prone areas, thereby reducing the incidence of violence in North Aceh.
Keywords: Clustering, K-Medoids, Davies-Bouldin Index, Violence

Introduction
Violence against women and children is a longstanding social issue that has been increasingly prevalent in recent times. Both the quantity and variety of forms and actions of violence against women and children continue to rise. Acts such as rape, sexual harassment, trafficking of women and children, domestic violence, sexual exploitation, violence against domestic workers, exploitation of migrant workers, and neglect seem to persist in daily life, both within households and in the broader community[1].

North Aceh, a region in the Aceh Province of Indonesia, faces serious issues related to violence against women and children, which has a wide-ranging impact on the community. Gender inequality, low levels of education, and economic factors are potential causes of this violence.

At the City or Regency level, reporting of services related to violence cases is handled by the Department of Social Services for the Empowerment of Women and Children, as well as the local Police Unit II for Women and Children Protection (PPA). These agencies are responsible for receiving and addressing all reports of violence against women and children.

Based on the data on the recapitulation of cases of violence handled by service agencies, there are 3 service agencies with the highest cases of violence compared to other service agencies. For the highest number of cases of violence against women and children, the service agencies from Banda Aceh City recorded 135 cases. Bener Meriah District Service Agency recorded 81 cases. North Aceh District Service Institution recorded 74 cases. From the results of the recapitulation of cases of violence against women and children in January-December 2023 by the Aceh Social Service Office for Women's Empowerment and Child Protection (DPPPA), the North Aceh District Service Institution was the third highest case of violence against women and children [2].

One initial step to address the issue of violence against women and children in North Aceh is to identify and cluster high-risk areas using the K-Medoids Clustering algorithm. This approach allows the Department of Social Services for the Empowerment of Women and Children (DPPPA) in North Aceh and the Unit II PPA of the North Aceh Police to focus on areas with elevated risk levels, enabling them to reduce the likelihood of future violence and protect potential victims. This clustering will support more focused and targeted protection and prevention efforts.

Literature Review

Clustering

Data mining techniques commonly used for regional grouping include clustering. Clustering is a method of grouping data into clusters based on data similarities, placing related data within the same cluster [3]. According to [4], there are two main approaches used in developing clustering methods:

1. **Hierarchical Clustering**: This technique aims to identify similarities between data points. Various hierarchical clustering methods include Average Group Linkage, Complete Linkage, Single Linkage, and Average Linkage.
2. **Non-Hierarchical Clustering**: This approach differs from hierarchical clustering. It begins by determining the number of clusters, where a cluster can consist of multiple clusters, usually at least two or more. Once the number of clusters is set, the clustering process follows a different approach than hierarchy-based methods. This category includes methods like K-Means, Fuzzy K-Means, K-Medoids, and Mixture Modeling.

K-Medoids Clustering

K-Medoids is a data mining algorithm that can be used to cluster data with similarities among the data points [5]. The K-Medoids algorithm utilizes representative data points, known as medoids, which serve as the center for each cluster. The goal of K-Medoids is to minimize the overall dissimilarity between data points within each cluster and their respective medoid [4].

According to [6], several steps are involved in calculating the K-Medoids Clustering Algorithm, as follows:

1. Initialize the cluster centers by selecting k (the number of clusters). In this study, the author uses 3 clusters: Highly Vulnerable (C1), Vulnerable (C2), and Not Vulnerable (C3).
2. Allocate each data point (object) to the nearest cluster using the Euclidean Distance formula:
   $d\ (x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2}$ ..... (1)
   where:
   d (x,y) = Distance of data to the cluster center
   xi = Attribute value of the original data
   yi = Attribute value of the medoid data
   n = Number of attributes used.
3. Randomly select an object within each cluster as a candidate for the new medoids.
4. Calculate the distance of each object within each cluster to the candidate medoids.
5. Calculate the total deviation (S) by finding the difference between the new total distance and the old total distance. If S<0, swap the object with the cluster data to form a new set of k objects as medoids.
6. Repeat steps 3 to 5 until no change in the medoids occurs, resulting in clusters with their respective members.

Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is a cluster validation measure introduced by D.L. Davies and Donald W. Bouldin, and it is commonly used to evaluate the quality of clusters formed during clustering. The DBI value is a function that represents the ratio of intra-cluster distances to inter-cluster distances [7]. There are four main steps to calculate the DBI:

1. Calculate the Sum of Squares Within Cluster (SSW)
   The formula for calculating SSW is: $SSW_i = \frac{1}{mi}\sum_{j=i}^{mi} d(x_j, c_i)$ ..... (2)
   Where :
   mi = Number of data points in the i-th cluster
   x = Data Points in the cluster
   d(x,c) = Distance of the data point to the medoid
   x_j = Data point in the j-th cluster
   c_i = Medoid of the i-th cluster
2. Calculate the Sum of Squares Between Cluster (SSB)
   The formula for calculating SSB is: $SSB_{i,j} = d(c_i, c_j)$ ..... (3)
   Where :
   C_i = i-th cluster
   C_j = j-th cluster
   $d(c_i,c_j)$ = Distance between the i-th and j-th cluster centers
3. Calculate the Ratio
   The formula for calculating the ratio is : $R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{ij}}$ ..... (4)
   Where :
   $R_{i,j}$ = Ratio between clusters
   SSW_i = Sum of squares within the i-th cluster
   SSW_j = Sum of squares within the j-th cluster
   SSB_{ij} = Separation between the i-th and j-th clusters
4. Calculate the Davies-Bouldin Index
   The formula for calculating the DBI is : $DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i=j}\ (R_{i,j})$ ..... (5)
   Where :
   k = Total number of cluster

$R_{i,j}$ = Ratio between the i-th and j-th clusters
Max = The maximum ratio found among the clusters

Materials & Methods

The steps of this research are conducted actively and efficiently. To facilitate the system in identifying areas prone to violence against women and children using the K-Medoids method, several stages are carried out, which will be outlined as follows:

1. Data Collecyion
   Data collection in this research is conducted using interview techniques with the Unit II PPA POLRES North Aceh. In addition to interviews, the author also gathers data from national journals, the internet, and other relevant sources as references for this research.
2. Data Processing
   Data processing employs various basic statistical methods using different software, such as Excel.
3. System Design
   System design plays a crucial role as it involves the representation of the system to be developed into a workflow or system scheme. With proper system design, the development process of the system will be more directed.
4. System Implementation
   System implementation is the process of developing applications using programming languages. In this case, the programming languages used will be PHP, HTML, and CSS.
5. System Testing
   The testing phase is conducted once the system development is complete. System testing aims to ensure that the developed system functions properly and meets the requirements or expectations, and to identify any discrepancies or areas that do not align with the designed system schema, necessitating further improvements.
6. System Evaluation
   After the system software is complete, testing will be conducted on the developed system. The purpose of testing is to identify any errors within the system, allowing for corrections to be made, with the goal of aligning the system with the analysis that has been conducted.

The system schema is a diagram that illustrates the operational processes of the system from start to finish. Below is the system schema for clustering areas prone to violence against women and children.

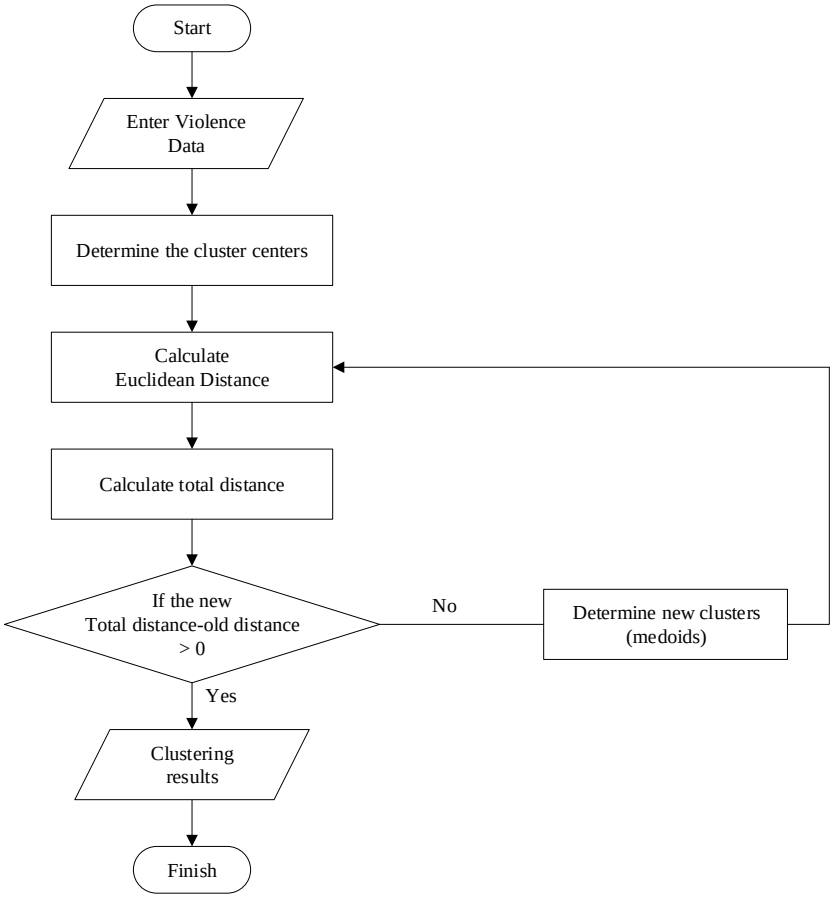


Figure 1. System Architecture of the K-Medoids Clustering System

Here is the explanation of each process in the system schematic for clustering areas vulnerable to violence against women and children:

1. Start
   This is the initial process to begin the system.
2. Input Violence Data
   In this stage, the admin inputs (enters) violence data containing attributes such as physical violence, sexual violence, exploitation, and neglect.
3. Determine Cluster Centers
   expected by the editor. To use this template, please *Save As* to your document, then copy and paste your document here. The work should not have been published or submitted for publication elsewhere.
4. K-Medoids Method Process
   Once the cluster center determination process is complete, the next step is to calculate the violence data using the K-Medoids method by computing distances with the Euclidean Distance formula. At this stage, the system will classify the violence data into three groups according to the value of K: highly vulnerable, vulnerable, and not vulnerable.
5. Calculate Total Distance
   If the new total distance minus the old total distance is greater than 0, the calculation process is complete. However, if the total distance is less than 0, it will recalculate the total distance by determining new medoids until the difference is 0 or repeat steps 3-5 until no changes in medoids occur.
6. Clustering Results
   The clustering result of the violence data is displayed if the total distance is greater than 0. The clustering result is the output that will be shown by the system, indicating the levels of areas vulnerable to violence against women and children. The mentioned levels are very vulnerable, vulnerable, and not vulnerable.
7. End
   End marks the completion of all processes carried out by the system

Results and Discussion

The data on violence against women and children was obtained from Unit II PPA POLRES North Aceh over a period of three years, from 2021 to 2023. The collected data includes physical violence, sexual violence, exploitation, and neglect. This data will be grouped into three categories: highly vulnerable, vulnerable, and not vulnerable. Here are the weight values that will be assigned to cases of violence:

Table 1. Violence weight value

Violence value	Weight
0	0,3
1	5
2	10
3	15

1. Datasets

Table 2. Violence data for 2021

No	Village	Physical violence	Sexual Violence	Exploitation	Neglect
1	Alue Anou Barat	0,3	0,3	0,3	0,3
2	Alue Anou Timur	0,3	0,3	0,3	0,3
3	Alue Bili Geulumpang	0,3	5	0,3	0,3
4	Alue Bili Rayeuk	0,3	0,3	0,3	0,3
5	Alue Buya	0,3	0,3	0,3	0,3
…	…	…	…	…	…
542	Ulee Tanoh (TP)	0,3	0,3	0,3	0,3

2. Determining Cluster
   Initialize the cluster centers with 3 clusters from the dataset, where each medoid is selected randomly.

Table 3. Initial Medoid Values for the Year 2021

		Initial Medoid			
No	**Village**	**Physical**	**Sexual**	**Exploitation**	**Neglect**
154	Bintang Hu	0,3	5	0,3	0,3
535	Matang Ranup Laseh	5	0,3	0,3	0,3
33	Matang Baro	0,3	0,3	0,3	0,3

3. Calculate Euclidean Distance

Calculate the nearest distance using the Euclidean Distance formula. To perform clustering on each piece of data obtained, the next step is to compute the Euclidean Distance.

Table 4. Iteration 1 Calculation Results for the Year 2021

Iterasion 1					
Village	Distance to Medoid			Proximity	Nearest Cluster
	C1	C2	C3		
Alue Anou Barat	4,7	4,7	0	0	3
Alue Anou Timur	4,7	4,7	0	0	3
Alue Bili Geulumpang	0	6,6468	4,7	0	1
Alue Bili Rayeuk	4,7	4,7	0	0	3
Alue Buya	4,7	4,7	0	0	3
…	…	…	…	…	…
Ulee Tanoh (TP)	4,7	4,7	0	0	3
Total Proximity				**19,39**	

4. New Medoids
   Determine the new medoid values by selecting new medoids randomly, ensuring that each previously selected medoid cannot be chosen again as a new medoid.

Table 5. New Medoids

New Medoids					
No	Village	Physical	Sexual	Exploitation	Neglect
523	Teupin Gajah	0,3	5	0,3	0,3
236	Glumpang VII	5	0,3	0,3	0,3
1	Alue Anou Barat	0,3	0,3	0,3	0,3

5. Calculate Euclidean Distance for Iteration 2

Table 6. Iteration 2 Calculation Results for the Year 2021

Iterasion 1					
Village	Distance to Medoid			Proximity	Nearest Cluster
	C1	C2	C3		
Alue Anou Barat	4,7	4,7	0	0	3
Alue Anou Timur	4,7	4,7	0	0	3
Alue Bili Geulumpang	0	6,6468	4,7	0	1
Alue Bili Rayeuk	4,7	4,7	0	0	3
Alue Buya	4,7	4,7	0	0	3
…	…	…	…	…	…
Ulee Tanoh (TP)	4,7	4,7	0	0	3
Total Proximity				**19,4**	

6. Total Deviation (Standard Deviation)
   fter obtaining the distance values between iteration 1 and iteration 2, calculate the total deviation (S) by finding the difference between the new total cost (sum of proximity) and the old total cost (sum of proximity). The criteria are as follows: if S < 0, then swap the values of the objects by determining new medoids; conversely, if S > 0, the calculation is halted or completed.
   S = New total cost−Old total cost
   S = 19,4−19,39
   S = 0,01
   Since the total deviation is greater than zero (S > 0), this test is halted at iteration 2.
7. Final Results

Table 7. Final Results 2021

No	Village	*Cluster*
1	Alue Anou Barat	Not Vulnerable
2	Alue Anou Timur	Not Vulnerable
3	Alue Bili Geulumpang	Very Vulnerable
4	Alue Bili Rayeuk	Not Vulnerable
5	Alue Buya	Not Vulnerable
6	Alue Dama	Very Vulnerable
7	Alue Geudong	Not Vulnerable
8	Alue Ie Tarek	Not Vulnerable

9	Alue Jamok	Not Vulnerable
10	Alue Keutapang	Not Vulnerable
…	…	…
535	Matang Ranup Laseh	Vulnerable
536	Me Matang Panyang	Not Vulnerable
537	Meunasah Alue	Not Vulnerable
538	Meunasah Cangguek	Not Vulnerable
539	Meunasah Cibrek	Not Vulnerable
540	Meunasah Prey	Not Vulnerable
541	Teupin Gapeuh	Not Vulnerable
542	Ulee Tanoh (TP)	Not Vulnerable

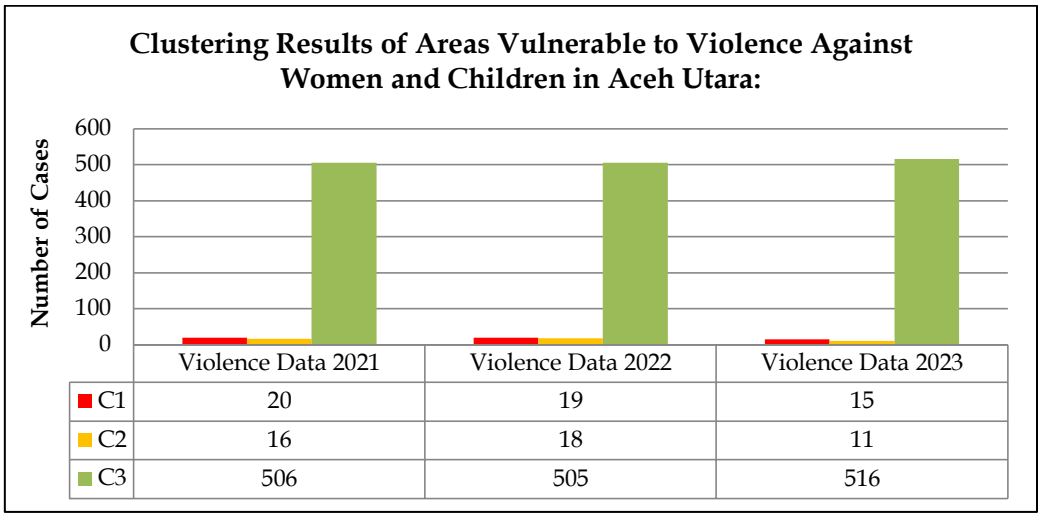


Figure 2. Clustering Results for 2021-2023

Table 8. DBI Evaluation Results for 2021-2023

Year	Test Number	Lowest Davies-Bouldin Index (DBI):
2021	1,3,4,6,7,8	0,12263
2022	1,9,10	0,10517
2023	1,2,3,4,5	0,21408

System Implementation

Below are several interface displays from the system implementing the K-Medoids Clustering Method to Classify High-Risk Areas for Violence Against Women and Children in North Aceh :

1. Login Page

   This page is designed to grant the admin access to the dashboard. To proceed, the admin must enter a username and password. After entering these credentials, the admin must click the login button to access the next page.

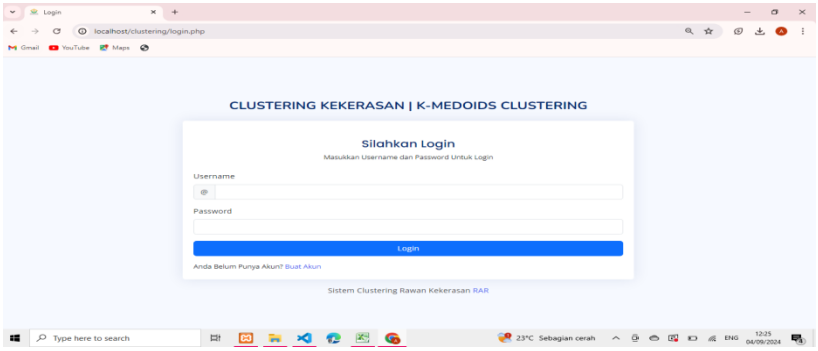
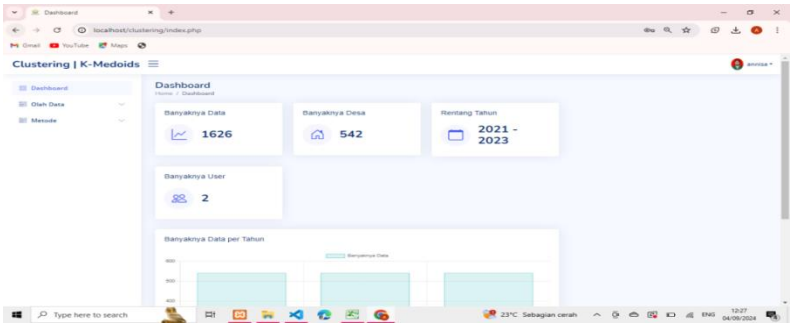


Figure 3. Login

2. Dashboard Page

   The dashboard page serves as the main interface that the admin can access after logging into the system. The

dashboard includes several menus, such as the data processing menu with submenus to manage data, manage villages, and manage years.



Figutre 4. Dashboard Page

3. Final K-Medoids Calculation Results Page
   This page displays the final clustering results. The admin can view the clustering results by individual cluster or as an overall summary.

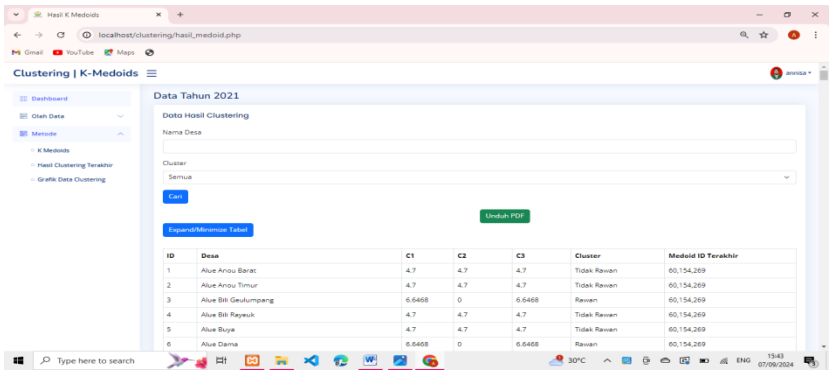


Figure 5. Final K-Medoids Calculation Results Page

4. Clustering Graph Page
   This page displays a graph of violence data per year, based on the year selected during the K-Medoids calculation process. It also shows the names of villages along with their respective cluster classifications.

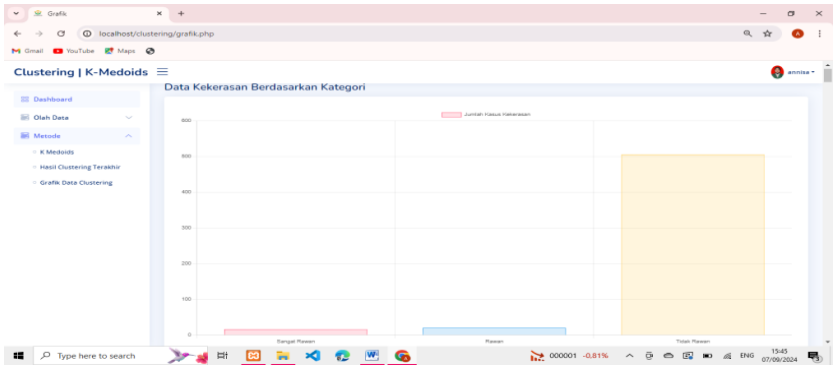


Figure 6. Clustering Graph Page

Conclusions

The conclusions derived from this research are as follows:

1. This classification utilized data obtained from Unit II PPA POLRES North Aceh, encompassing a total of 542 villages. The types of violence recorded include physical violence, sexual violence, exploitation, and neglect, covering the years 2021–2023.
2. The K-Medoids method successfully classified the villages into three clusters: high risk, medium risk, and low risk. Additionally, the method's performance was evaluated using the Davies-Bouldin Index (DBI).
3. 2021 Clustering Results: 20 villages were classified as C1 (Very Vulnerable), 16 as C2 (Vulnerable), and 506 as C3 (Not Vulnerable). With eight test iterations, the average number of iterations per test was two, yielding a minimum DBI of 0.12263 and a maximum of 0.79491. 2022 Clustering Results: 19 villages fell under C1 (Very Vulnerable), 18 under C2 (Vulnerable), and 505 under C3 (Not Vulnerable). Conducted over ten test iterations, the average number

of iterations per test was two, with a minimum DBI of 0.10517 and a maximum of 0.73378. 2023 Clustering Results: 15 villages were categorized as C1 (Very Vulnerable), 11 as C2 (Vulnerable), and 516 as C3 (Not Vulnerable). After six test iterations, the average number of iterations per test was two, with a minimum DBI of 0.21408 and a maximum of 0.25647.

Acknowledgments

I would like to express my deepest gratitude to all parties for their support during the research and until now. I would like to express my deepest gratitude to my supervisors, Mr. Rozzi Kesuma Dinata, S.T.,M.Eng and Mrs. Maryana, S.Si., M.Si, for their invaluable guidance and input, as well as their continuous encouragement. Their dedication was critical in shaping all to this. I am also very gratefu to Unit II PPA POLRES North Aceh for helping me in collecting data on violence in North Aceh. Finally, i would like to thank myself, family and friends for their support and encouragement during this time.

References

[1] H. Haryati and Sekar Ayu Aryani, "Konseling Multikultural Dengan Terapi Feminis Dalam KDRT Pada Perempuan," *J-CEKI J. Cendekia Ilm.*, vol. 1, no. 6, pp. 809–816, 2022, doi: 10.56799/jceki.v1i6.1009.

[2] Asmulyadi, "Rekap Kasus Kekerasan Yang Ditangani Oleh Lembaga Layanan (Januari - Desember 2023)," *DPPPA Aceh*, 2024. https://dinaspppa.acehprov.go.id/berita/kategori/rekap-kekerasan/rekap-kasus-kekerasan-yang-ditangani-oleh-lembaga-layanan-januari-desember-2023

[3] N. Adawiyah, N. Sulistiyowati, and M. Jajuli, "Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means," *Gener. J.*, vol. 5, no. 2, pp. 69–80, 2021, doi: 10.29407/gj.v5i2.15995.

[4] N. Ulfauza and Z. Yunizar, "Clustering Status Pemberian Imunisasi Dasar Di Dinas Kesehatan Kabupaten Bireuen Menggunakan Metode K-Medoids," pp. 1–5, 2022.

[5] R. K. Dinata, Furqan, and S. Retno, "Pengelompokan daerah padat penduduk untuk penentuan kawasan perumahan di kota lhokseumawe menggunakan k-medoids clustering," *J. Elektron. dan Teknol. Inf.*, vol. 4, no. 1, pp. 2721–9380, 2023.

[6] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.

[7] I. W. Septiani, A. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 556, 2022, doi: 10.30865/json.v3i4.4055.