



A Data Mining Approach to Clustering Cases of Violence Against Children in Indonesia

¹Agnes Irene Silitonga, Universitas Negeri Medan, Indonesia

²Mutiara Gracella Angel Lumban Gaol, Universitas Negeri Medan, Indonesia

³Victoria Ezra Stephani Mendrofa, Universitas Negeri Medan, Indonesia

⁴Daniel Junior Panggabean, Universitas Negeri Medan, Indonesia

⁵Dewi Pika Lbn Batu, Universitas Negeri Medan, Indonesia

Correspondence: E-mail: agnesirenesilitonga@unimed.ac.id

Article Info

Article history:

Received October 05, 2025

Revised November 10, 2025

Accepted December 12, 2025

Keywords:

*Child Violence,
Data Mining,
Clustering,
K-Means Clustering*

ABSTRACT

Violence against children in Indonesia remains a crucial issue that requires a data-driven approach for more targeted interventions. This study aims to cluster provinces in Indonesia based on the number of cases of violence and the types of violence committed, namely physical, psychological, and sexual violence. The method used in this study is the K-Means Clustering algorithm, an unsupervised learning technique in data mining that is able to find hidden patterns in large data sets. Data was obtained from the Ministry of Women's Empowerment and Child Protection's Gender and Child Information System (SIGA), which covers 38 provinces. The clustering results produced three main groups: a cluster with high levels of violence consisting of the provinces of North Sumatra, DKI Jakarta, West Java, Central Java, and East Java; a cluster with moderate levels of violence consisting of 16 provinces; and a cluster with low levels of violence covering 17 provinces. These findings are expected to form the basis for the development of evidence-based child protection policies geographically and thematically.

1. INTRODUCTION

The issue of violence against children remains a serious issue requiring cross-sectoral attention in Indonesia. These forms of violence include physical, psychological, sexual, neglect, and exploitation, occurring in various settings, including homes, schools, and the wider community. These conditions

not only damage children's physical and emotional well-being in the short term but also have long-term impacts on their overall growth and development. According to the latest data compiled from the Gender and Child Information System (SIGA) of the Ministry of Women's Empowerment and Child Protection (KemenPPPA), violence

against children in Indonesia shows a concerning trend [1]. Although public awareness of reporting violence is increasing, the fact that cases of violence continue to occur in large numbers demonstrates the weakness of the existing protection system. A 2023 report by the National Commission on Child Protection revealed that thousands of children experienced various forms of violence, documented through various reporting channels, from hotlines to online applications [2].

UNICEF Indonesia highlighted that one contributing factor to the low level of child protection is inadequate budget allocation. The government's child protection budget is less than 0.1% of total state spending [3]. This lack of funding impacts the quality of rehabilitation services, victim outreach, and preventative education within the community. Furthermore, many cases of violence are also found in the educational setting. School environments, which should be safe havens for children, are often the site of verbal abuse, bullying, and even sexual harassment. The 2024 Indonesian Child Protection Commission (KPAI) report stated that violence occurring in schools is often not officially recorded because it is covered up by educational institutions to maintain their image [4]. This indicates the lack of a transparent and structured reporting and handling system within educational institutions. Given the complexity and diversity of cases of violence against children, analytical methods are needed that can identify hidden patterns in the data. One approach that can be used is unsupervised learning techniques in data mining, namely clustering. This method is useful for grouping data based on certain shared characteristics, thereby revealing previously unseen relationships between variables.

K-Means Clustering is the most commonly used clustering method in social data processing. This algorithm works by dividing a data set into several clusters based on the distance between data points and the cluster center (centroid) [5]. In the context of violence against children, this method can help group types of violence based on region,

age, or frequency of occurrence, resulting in specific clusters that can be used for more targeted policy formulation. By identifying clusters of violence types, policymakers can tailor intervention approaches based on victim characteristics and location of incidents. For example, if a region experiences a high prevalence of sexual violence against school-age children, prevention and protection campaigns can be focused on educational settings in that area. This evidence-based approach has the potential to increase the effectiveness of national child protection programs.

This study aims to apply the K-Means Clustering method to child violence data obtained from the Ministry of Women's Empowerment and Child Protection's (KemenPPPA) SIGA (Standardized Data Collection), focusing on the types of violence experienced. The data used is secondary data collected nationally and is open to public access. Through this analysis, it is hoped that clusters of violence will be formed that describe thematic case patterns, which can then be utilized for data-driven decision-making and the development of child protection policies.

2. LITERATURE REVIEW

This study discusses the clustering of cases of violence against children in Indonesia based on the type of violence. Several key theories will be discussed, including violence against children, the concept of child protection, data mining, and the K-Means Clustering algorithm.

2.1. Violence Against Children

Violence against children is a serious human rights violation and can have long-term impacts on a child's physical, mental, social, and emotional well-being. According to UNICEF, violence against children encompasses all forms of abuse or harm, whether physical, sexual, emotional, or neglect, that can negatively impact a child's well-being [6]. Widodo (2011) emphasized that violence against children not only has immediate effects but can also hinder a child's ongoing psychological development [7].

Therefore, adequate data is crucial for effective intervention.

In Indonesia, according to data from the Indonesian Child Protection Commission (KPAI), thousands of cases of violence against children are reported annually, although this figure is believed to be significantly lower than the actual number of cases due to stigma and public ignorance about the reporting system [8]. Therefore, mapping patterns of violence is crucial for developing evidence-based policies.

2.2. Child Protection and Policy

Child protection is all forms of activities that guarantee and safeguard children so they can live, grow, develop, and participate optimally in accordance with human dignity. This protection encompasses legal, social, and psychological aspects. The Indonesian government has established institutions such as the UPTD PPA and developed online reporting systems such as SIGA (Gender and Child Information System) as concrete measures to strengthen the child protection system.

2.3.Data Mining in Social Analysis

Hanif and Setiawan (2019) explain that data mining is the process of exploring large amounts of data to discover previously hidden patterns and useful information [9]. In a social context, data mining is particularly helpful in detecting trends and behavioral patterns, including in child protection issues. The application of data mining in the social sector enables policymakers to not only analyze data statistically but also understand the relationships between variables on a large scale, which is difficult to achieve with conventional methods [10].

2.4.K-Means Clustering

Clustering is an unsupervised learning method in data mining that aims to group data into clusters with a high degree of similarity between its members [11]. Jain et al. (1999) defined clustering as the process of grouping objects based on certain characteristics, so that objects within one group are more similar to each other than to objects in other groups

[12]. One of the most widely used clustering algorithms is K-Means Clustering. This algorithm groups data into k groups based on the Euclidean distance from the cluster center (centroid), which is updated iteratively until convergence is achieved. Witten and Frank (2005) described K-Means as an efficient and easy-to-implement method, although it has weaknesses such as sensitivity to the initial centroid selection [13].

3. METHODS

This study aims to clusters provinces based on cases and type of violence against children. The main technique used in this study is data clustering using the K-Means Clustering algorithm, which is one of the unsupervised learning methods in data mining.

3.1.Data

The data used in this study are secondary data obtained from <https://siga.kemenpppa.go.id/dataset?ids=MTES5Mw==&entity=cHJvdmluY2U=>. This dataset contains detailed information on cases of violence against children in Indonesia based on reports, which have been compiled by the Regional Technical Implementation Units for the Protection of Women and Children (UPTD PPA) throughout Indonesia [1]. Before being analyzed, the data underwent a preprocessing process, which included cleaning data from duplication, handling missing values, numeric normalization, and transforming categorical attributes into numeric form (e.g., with one-hot encoding).

3.2.Data Analysis – Clustering Method

This study used the K-Means Clustering algorithm with the following steps:

1. Initialization: Determine the initial number of clusters (k value).
2. Initial Centroid Determination: Randomly select the initial centroid from the normalized data.
3. Iterative Process:
 - a. Calculate the Euclidean distance from each data point to the cluster centroid using Equation 1.

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

- b. Assign data to the cluster with the shortest distance.
- c. Recalculate the cluster centroid based on the average of the points within the cluster.
- d. The process is repeated until the centroid does not experience significant changes (convergence).

The K-Means Clustering algorithm was chosen because it is highly efficient in processing large amounts of data and is capable of identifying hidden structures in the data. However, the quality of the final results is highly dependent on the choice of the initial k value and centroid.

3.3. Interpretation of Results

The results of cluster analysis obtained using the K-Means Clustering algorithm provide important insights into grouping regions based on shared characteristics in cases of violence against children. In this context, the resulting clusters represent the distribution patterns of the dominant types of violence in each region or province in Indonesia.

Using the number of cases for each type of violence as the main feature (physical violence, psychological violence, sexual violence), each cluster exhibits specific patterns. For example, one cluster may be dominated by areas with high rates of sexual violence, while another cluster may exhibit a tendency toward emotional violence or neglect.

4. RESULTS AND DISCUSSIONS

The variables used in this study are unit, physical, and psychological variables. These three variables form the basis for the data grouping process using the K-Means Clustering algorithm. The initial step in this method is determining the desired number of clusters, followed by determining the initial cluster center points.

4.1. First Iteration

In the K-Means Clustering algorithm, initial cluster center points are randomly selected from the existing data. This selection serves as the initial reference for the process of grouping the data into clusters. The initial center points used in this study were randomly determined and are shown in Table 1.

Table 1. Number of children experiencing violence

No	Province	Physique	Psychic	Sexual
1	Aceh	201	233	308
2	Sumatera Utara	255	200	651
3	Sumatera Barat	142	228	492
4	Riau	142	193	615
5	Jambi	88	87	227
6	Sumatera Selatan	129	106	179
7	Bengkulu	45	17	145
8	Lampung	99	68	511
9	Kepulauan Bangka Belitung	48	41	112
10	Kepulauan Riau	66	86	303
11	Dki Jakarta	302	312	636
12	Jawa Barat	460	378	1231
13	Jawa Tengah	287	341	775
14	Di Yogyakarta	145	178	280
15	Jawa Timur	449	422	742
16	Banten	145	260	493
17	Bali	40	103	70
18	Nusa Tenggara Barat	117	74	296
19	Nusa Tenggara Timur	159	266	363
20	Kalimantan Barat	68	30	362
21	Kalimantan Tengah	55	61	203
22	Kalimantan Selatan	120	186	250
23	Kalimantan Timur	142	130	458
24	Kalimantan Utara	47	33	131
25	Sulawesi Utara	168	112	276
26	Sulawesi Tengah	194	177	270
27	Sulawesi Selatan	359	244	459
28	Sulawesi Tenggara	129	37	214

29	Gorontalo	48	5	152
30	Sulawesi Barat	34	18	43
31	Maluku	74	38	151
32	Maluku Utara	54	27	181
33	Papua Barat	7	54	57
34	Papua	32	41	53
35	Papua Pegunungan	2	1	2
36	Papua Selatan	11	27	22
37	Papua Tengah	6	4	26
38	Papua Barat Daya	21	20	32
Total		4890	4838	11771

Based on the Table 1, three initial cluster center points were determined, each representing a category of child violence level: high, moderately high, and low. These initial center points are:

1. Cluster center-1 (high level of violence): The 15th data point from East Java with a physical score of 449, psychological score of 422, and sexual score of 742.
2. Cluster center-2 (moderate level of violence): The 27th data point from South Sulawesi with a physical score of 359, psychological score of 244, and sexual score of 459.
3. Cluster center-3 (low level of violence): The 15th data point from East Java with a

physical score of 2, psychological score of 1, and sexual score of 2.

In the K-Means Clustering algorithm, the distance between each data point and the cluster center is calculated to determine the position of the new cluster center. This calculation uses the Euclidean Distance.

This process is repeated until the cluster center point remains unchanged from the previous iteration. When this condition is met, the iteration process is stopped because the results are considered converged. The distance between the data and the cluster center point is calculated at each iteration, as shown in Table 2.

Table 2. Results of calculation of data distance to cluster center point

Calculation of Data Distance to Cluster Center Point-1		Calculation of Data Distance to Cluster Center Point-2		Calculation of Data Distance to Cluster Center Point-3	
Province	Data Distance to Cluster Center Point-1	Province	Data Distance to Cluster Center Point-2	Province	Data Distance to Cluster Center Point-3
Aceh	534,3977919	Aceh	218,8287001	Aceh	432,5054913
Sumatera Utara	308,5465929	Sumatera Utara	222,7464927	Sumatera Utara	724,4384032
Sumatera Barat	440,891143	Sumatera Barat	220,0772592	Sumatera Barat	557,8790191
Riau	403,5083642	Riau	272,0771949	Riau	657,4442942
Jambi	712,5805218	Jambi	389,7614655	Jambi	255,7674725
Sumatera Selatan	720,5726889	Sumatera Selatan	387,7421824	Sumatera Selatan	241,8325867
Bengkulu	826,8313008	Bengkulu	498,71936	Bengkulu	150,1798921
Lampung	548,7959548	Lampung	318,2451885	Lampung	522,4739228

Kepulauan Bangka Belitung	838,368654	Kepulauan Bangka Belitung	508,2705972	Kepulauan Bangka Belitung	125,7616794
Kepulauan Riau	672,5369878	Kepulauan Riau	367,6261688	Kepulauan Riau	319,2522514
Dki Jakarta	212,0023585	Dki Jakarta	197,9949494	Dki Jakarta	767,2528918
Jawa Barat	491,0987681	Jawa Barat	790,0259489	Jawa Barat	1364,673587
Jawa Tengah	184,1032319	Jawa Tengah	338,3031185	Jawa Tengah	891,2653926
Di Yogyakarta	604,4799418	Di Yogyakarta	286,6932158	Di Yogyakarta	359,2520007
Jawa Timur	0	Jawa Timur	346,2268043	Jawa Timur	961,5872295
Banten	425,0423508	Banten	217,2740205	Banten	573,2460205
Bali	848,8969313	Bali	522,4586108	Bali	128,3432897
Nusa Tenggara Barat	655,9298743	Nusa Tenggara Barat	337,6877256	Nusa Tenggara Barat	324,0216042
Nusa Tenggara Timur	502,0727039	Nusa Tenggara Timur	222,9349681	Nusa Tenggara Timur	474,5471526
Kalimantan Barat	665,7514551	Kalimantan Barat	374,0133687	Kalimantan Barat	367,1471095
Kalimantan Tengah	758,9980237	Kalimantan Tengah	437,5397125	Kalimantan Tengah	216,3561878
Kalimantan Selatan	637,1820776	Kalimantan Selatan	322,7475794	Kalimantan Selatan	331,1389436
Kalimantan Timur	510,0676426	Kalimantan Timur	245,1244582	Kalimantan Timur	494,1426919
Kalimantan Utara	828,399662	Kalimantan Utara	499,4486961	Kalimantan Utara	140,3210604
Sulawesi Utara	626,272305	Sulawesi Utara	295,6247622	Sulawesi Utara	339,0471944
Sulawesi Tengah	589,7745332	Sulawesi Tengah	259,6824984	Sulawesi Tengah	373,7164701
Sulawesi Selatan	346,2268043	Sulawesi Selatan	0	Sulawesi Selatan	628,7662523
Sulawesi Tenggara	727,6049752	Sulawesi Tenggara	394,6821506	Sulawesi Tenggara	249,7378626
Gorontalo	826,3110794	Gorontalo	498,0873417	Gorontalo	156,9458505
Sulawesi Barat	907,7675914	Sulawesi Barat	574,2447214	Sulawesi Barat	54,71745608

Maluku	798,3495475	Maluku	467,4665763	Maluku	169,5700445
Maluku Utara	791,6887014	Maluku Utara	466,2595844	Maluku Utara	188,2046758
Papua Barat	894,4344582	Papua Barat	567,1049286	Papua Barat	76,54410493
Papua	890,9382695	Papua	559,4407922	Papua	71,42128534
Papua Pegunungan	961,5872295	Papua Pegunungan	628,7662523	Papua Pegunungan	0
Papua Selatan	930,7357305	Papua Selatan	599,3012598	Papua Selatan	34,0147027
Papua Tengah	940,0154254	Papua Tengah	608,0279599	Papua Tengah	24,51530134
Papua Barat Daya	921,3511817	Papua Barat Daya	588,8539717	Papua Barat Daya	40,27406113

The next step in the K-Means Clustering process is to determine the closest distance of each data point to the initial cluster center that was previously randomly selected. This closest distance determination aims to group each data point into the most appropriate cluster based on its proximity. To determine the closest distance, the MIN ($C_{ij} : C_{kj} : C_{lj}$) formula can be used, which compares the distance of each data point to each cluster center and then selects the smallest value among them. Based on the minimum value calculated using the previous formula, the results of grouping the data into each cluster can be determined. This process is carried out by grouping each data point to the cluster center that has the shortest distance. The final results of the clustering process can be seen in Table 3.

Table 3. First Iteration Clustering Results

Cluster	Data	Total
Cluster-1 (High)	{12,13,15}	3
Cluster -2 (Middle)	{1,2,3,4,8,10,11,14,16,18,20,22,23,25,26,27,30,31,32,33,34,36,37,38}	24
Cluster -3 (Low)	{5,6,7,9,17,19,21,24,28,29,35}	11

The next step in the K-Means Clustering process is to determine the new cluster centroid. This new centroid is obtained by

calculating the average value of all members in each cluster. From the previous calculation, a new centroid is obtained, which will be used as a reference in the next iteration. The new centroid values are presented in Table 4.

Table 4. New Centroid First Iteration

Cluster-1 (High)	X_1	398,6666667
	X_2	380,3333333
	X_3	916
Cluster-2 (Middle)	X_1	113,8333333
	X_2	111,0833333
	X_3	254,5416667
Cluster-3 (Low)	X_1	71,81818182
	X_2	68,81818182
	X_3	163,4545455

4.2. Second Iteration

After performing calculations in the second iteration, the final results of the clustering process can be seen in Table 5.

Table 5. Second Iteration Clustering Results

Cluster	Data	Total
Cluster-1 (High)	{11, 12, 13, 15}	4
Cluster-2 (Middle)	{1, 2, 3, 4, 5, 8, 10, 14, 16, 18, 19, 20, 22, 23, 25, 26, 27}	17
Cluster-3 (Low)	{6, 7, 9, 17, 21, 24, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38}	17

The new cluster center points, along with the mean values of each variable, are presented in Table 6.

Table 6. New Centroid Second Iteration

Cluster-1 (High)	X ₁	374,5
	X ₂	363,25
	X ₃	665
Cluster-2 (Middle)	X ₁	153,5294118
	X ₂	161,8823529
	X ₃	389,0588235
Cluster-3 (Low)	X ₁	38,41176471
	X ₂	35,05882353
	X ₃	91,70588235

4.3.Third Iteration

After performing calculations in the third iteration, the final results of the clustering process can be seen in Table 7.

Table 7. Third Iteration Clustering Results

Cluster	Data	Total
Cluster-1 (High)	{2, 11, 12, 13, 15}	5
Cluster-2 (Middle)	{1, 3, 4, 5, 8, 10, 14, 16, 18, 19, 20, 22, 23, 25, 26, 27}	18
Cluster-3 (Low)	{6, 7, 9, 17, 21, 24, 28, 29, 30, 32, 33, 34, 35, 36, 37, 38}	15

The new cluster center points, along with the average values of each variable, are shown in Table 8.

Tabel 8. New Centroid Third Iteration

Cluster-1 (High)	X ₁	350,6
	X ₂	330,6
	X ₃	807
Cluster-2 (Middle)	X ₁	155,7777778
	X ₂	165,2222222
	X ₃	372,5
Cluster-3 (Low)	X ₁	47,06666667
	X ₂	39,6
	X ₃	108

4.4.Fourth Iteration

After performing calculations on the fourth iteration, the final results of the clustering process can be seen in Table 9.

Table 9. Fourth Iteration Clustering Results

Cluster	Data	Total
Cluster-1 (High)	{2, 11, 12, 13, 15}	5
Cluster-2 (Middle)	{1, 3, 4, 8, 10, 14, 16, 18, 19, 20, 22, 23, 25, 26, 27, 28}	16
Cluster-3 (Low)	{5, 6, 7, 9, 17, 21, 24, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38}	17

4.5.Fifth Iteration

After performing calculations on the fifth iteration, the final results of the clustering process can be seen in Table 10.

Table 10. Fifth Iteration Clustering Results

Cluster	Data	Total
---------	------	-------

Cluster-1 (High)	{2, 11, 12, 13, 15}	5
Cluster-2 (Middle)	{1, 3, 4, 8, 10, 14, 16, 18, 19, 20, 22, 23, 25, 26, 27, 28}	16
Cluster-3 (Low)	{5, 6, 7, 9, 17, 21, 24, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38}	17

After comparing the results of the fifth iteration with the results of the fourth iteration, it was found that the cluster members remained unchanged. Therefore, the iteration process was stopped due to convergence. The results of this clustering process include:

1. The first cluster (high level of violence) consists of five provinces: North Sumatra, Jakarta, West Java, Central Java, and East Java.
2. The second cluster (middle level of violence) consists of sixteen provinces: Aceh, West Sumatra, Riau, Lampung, Riau Islands, Yogyakarta, Banten, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, and Southeast Sulawesi. Violence patterns in these regions tend to vary and indicate the need for more adaptive and regional interventions.
3. The third cluster (low level of violence) consists of seventeen provinces, including Jambi, South Sumatra, Bengkulu, Bangka Belitung Islands, Bali, Central Kalimantan, North Kalimantan, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Papua, Papua Mountains, South Papua, Central Papua, and Southwest Papua.

5. CONCLUSION

The clustering results show a pattern of child violence distribution that varies across regions, with a total of 21,499 cases across Indonesia. The clustering results yielded three main groups. The first cluster includes five provinces with high levels of violence: West Java, Central Java, East Java, Jakarta, and North Sumatra. These provinces generally have large populations and high levels of urbanization, which contribute to the high rate of violence against children. The second cluster consists of 17 provinces with moderate

levels of violence, while the third cluster includes 17 other provinces with relatively low numbers of violence cases. However, the low number of cases in the third cluster does not necessarily reflect a safe environment, as weak reporting systems and a culture of silence can contribute to low case documentation.

The most dominant form of violence found in the data is sexual violence, followed by physical and psychological violence. These findings indicate that sexual violence is a significant challenge to Indonesia's child protection system. Through the clustering approach, the government and stakeholders can identify priority areas requiring specific interventions, both in terms of prevention and response. The practical implications of this research lie in utilizing the clustering results as a basis for formulating evidence-based child protection policies. Policy interventions can be tailored to the characteristics of each cluster. Areas with high levels of violence require special attention in the form of increased budgets, integrated reporting systems, and strengthening child protection institutions. Areas with moderate levels of

violence should focus on strengthening rehabilitation services and community education, while areas with low levels of violence should strengthen early detection and reporting mechanisms to prevent potential cases from being hidden.

However, this research has limitations, such as using only one year of data, not including additional sociodemographic variables, and the limited flexibility of the K-Means method in determining the optimal number of clusters. For further research, multivariate analysis and the use of alternative clustering methods are recommended for more robust results.

Overall, this study confirms that the clustering method can be a strategic approach to understanding the distribution of violence against children in Indonesia. This approach enables the government to formulate more contextual, efficient, and targeted policies to prevent and address violence against children, thereby creating a safe and appropriate environment for the growth and development of the nation's next generation.

6. REFERENCES

- [1] Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KemenPPPA), "Jumlah Kekerasan Terhadap Anak menurut Jenis Kekerasan yang Dialami," *Sistem Informasi Gender dan Anak*, 2024. [Online]. Available: <https://siga.kemenpppa.go.id/dataset?entity=cHJvdmluY2U%3D&ids=MTE5Mw%3D%3D>
- [2] Komisi Nasional Perlindungan Anak (Komnas PA), "Komnas PA: Ada 3.547 Kasus Kekerasan terhadap Anak Sepanjang 2023," *Kompas.com*, 28 Desember 2023. [Online]. Available: <https://megapolitan.kompas.com/read/2023/12/28/16152591/komnas-pa-ada-3547-kasus-kekerasan-terhadap-anak-sepanjang-2023>
- [3] UNICEF Indonesia, "Perlindungan Anak," *UNICEF Indonesia*, 2023. [Online]. Available: <https://www.unicef.org/indonesia/id/child-protection>
- [4] Komisi Perlindungan Anak Indonesia (KPAI), "Kekerasan Pada Anak di Satuan Pendidikan," *Puslit DPR RI*, Feb. 2024. [Online]. Available: https://berkas.dpr.go.id/pusaka/files/isu_sepekan/Isu%20Sepekan---I-PUSLIT-Februari-2024-190.pdf
- [5] M. Fahmi, S. Widiyari, and E. C. Puspaningtyas, "Implementasi K-Means Clustering untuk Pengelompokan Data Sosial," *Jurnal Teknik Informatika*, vol. 14, no. 2, pp. 85–92, 2023.
- [6] UNICEF, *Hidden in Plain Sight: A statistical analysis of violence against children*, New York: UNICEF, 2020.

- [7] Widodo, S., “Perlindungan Anak dalam Perspektif Sosial dan Psikologis,” *Jurnal Ilmu Sosial dan Politik*, vol. 15, no. 2, pp. 120-135, 2011.
- [8] Komisi Perlindungan Anak Indonesia (KPAI), *Data Kekerasan terhadap Anak 2023*, Jakarta: KPAI, 2023. [Online]. Tersedia: <https://siga.kemenpppa.go.id/>
- [9] Hanif, M. dan Setiawan, B., “Penerapan Data Mining dalam Dunia Sosial,” *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 1, pp. 55–60, 2019.
- [10] M. Fahmi, S. Widiyari, dan E. C. Puspaningtyas, “Implementasi K-Means Clustering untuk Pengelompokan Data Sosial,” *Jurnal Teknik Informatika*, vol. 14, no. 2, hlm. 85–92,
- [11] A. Kumar dan M. V. Kumar, “An Enhanced K-Means Algorithm for Clustering Large Datasets,” *Procedia Computer Science*, vol. 167, hlm. 373–382, 2023.
- [12] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [13] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005