Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)
Vol. 6, No. 1, March 2025, pp. 737~740
ISSN: 2721-3838, DOI: 10.30596/jcositte.v6i1.21671

Clustering Women Violence Cases Based on Number in Central Java Province Using K-Means Algorithm

Fanika Putri Tantisnio[1], Nur Wakhidah[2]
[1,2]Study Program of Informatics Engineering, University of Semarang, Indonesia

ABSTRACT

Cases of violence against women are one of the cases that require special attention, especially in the Central Java Province area. This study aims to classify the number of cases of violence against women (age 18+) based on data from 2022 and 2023 using the K-Means algorithm. Through the K-Means método, the regions are classified into 2 clusters based on the pattern of the number of cases in the 2 years. The results of clustering show two main groups: the first cluster which is from the majority of regions with a fairly low number of violent cases, and the second cluster which includes Semarang City as a very high number of violent cases. This study provides an overview of cases of violence against women in Central Java Province and identifies areas with higher case rates. This research is expected to be used as a basis for more appropriate policy-making to reduce the rate of violence against women in the most affected areas.

Keyword : K-Means, Clustering, Violence Against Women

Corresponding Author:
Nur Wakhidah,
Universitas Semarang
Jl. Soekarno Hatta, RT.7/RW.7, Tlogosari Kulon, Kec. Pedurungan, Kota Semarang, Jawa Tengah 50196.
Email : ida@usm.ac.id

Article history:
Received Nov 04, 2024
Revised Nov 06, 2024
Accepted Mar 10, 2025

1. INTRODUCTION

Violence against women is one of several cases that require special scrutiny Violence against women is one of several cases that require special attention(Hardianti et al. 2021), especially in the Central Java region. Violence against women still occurs in several forms, including human trafficking, sexual harassment, and domestic violence(Fatma 2024). In Central Java, the high number of cases of violence against women is of particular concern, as it demonstrates the need to understand the patterns and characteristics of these cases to support effective policy-making(Okta Windya Ningrum and Yana S. Hijri 2022).

Based on existing data, there are differences in the number of cases of violence in each district and city in this region(Sa'diah, Enri, and Nur Padilah 2023). This indicates the existence of a wide variety of risk factors that have the potential to influence the rate of violence in each region(Muhammad Jadi 2021). However, the treatment carried out is often generalized and does not consider differences in characteristics between regions, so that the policies implemented are less precise and have not been effective in significantly reducing the rate of violence(Lalu Muhamad Rofiâ€™I and Mawardi 2022).

Cases of violence have been a concern, especially for research from year to year. In previous research(Stepanus Ginting et al. 2022), raised the case of Domestic Violence in Samarinda. The reviewer conducted a field study to obtain data from the Samarinda Police. The data that has been obtained is normalized in previous research. The results obtained are months with the highest domestic violence cases.

This study utilizes the K-Means algorithm to cluster regions in Central Java Province based on the number of cases of violence against women(Annurfariz, Irma Purnamasari, and Ali 2024). The K-Means algorithm was chosen because it can classify data into several clusters according to the existing pattern(Adelina Bui and Bahtiar 2024). With this approach, two main clusters are expected to be formed, namely areas with low levels of violence cases and areas with high levels of violence cases.

This approach allows for data-driven clustering by considering the number of violence cases in each region directly(Hoerunnisa et al. 2024). This study provides a new solution in understanding the patterns of violence against women through regional clustering(Sundari, Pane, and Rohani 2023), and overcomes the weaknesses of previous studies that do not consider differences between regions.

2. RESEARCH STAGES

This study used a clustering approach using the K-Means algorithm to analyze cases of violence against women in Central Java in 2023(Hutagalung and Sonata 2021). The data used includes the number of female victims of violence (age 18+) in all districts and cities in Central Java. The data was then analyzed using the K-Means algorithm to classify regions according to the number of violence cases. The following are the stages of this study method:

1. Data Collection: This study uses data on the number of cases of violence against women in Central Java in 2022 and 2023. This data was obtained by conducting literature searches on the Central Java Provincial Statistics Agency website. It includes the number of victims of violence against women for all cities and districts in Central Java.
2. Data Processing: Before conducting the analysis, the data that has been collected is first processed to ensure consistency and completeness. This process includes data cleaning, where incomplete or irrelevant data will be removed(Siswa and Pranoto 2023).
3. *K-Means* Algorithm: In this study, the number of clusters set is two, to group areas that have a high and low number of violence cases. The K-Means algorithm is applied with Google Collaboratory, where clusters will be determined and each data will be grouped in the cluster according to the distance to the centroid(Akbar 2023).
4. Clustering Evaluation: Clustering results are evaluated by interpreting the centroid value of each cluster. The centroid is the average value of the data in the cluster(Sulistiyawati and Supriyanto 2021), it shows the main characteristics of each group of regions. Areas that belong to one cluster are similar in terms of the number of violence cases.

3. RESULTS AND DISCUSSION

```
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(X)
```

Fig 1. Application of K-Means Algorithm

The command n_clusters=2 specifies the number of clusters to be formed in this case, while random_state=42 specifies consistent clustering results each time the code runs. The value 42 was chosen as a random seed, but it can be any number. The use of random_state is useful for reproducibility, meaning the results will remain the same on each execution if this parameter is set. The kmeans.fit(X) command is used to train the K-Means model on the data stored in the X variable. The K-Means command produces the following graphical visualization and data table of clustering results:

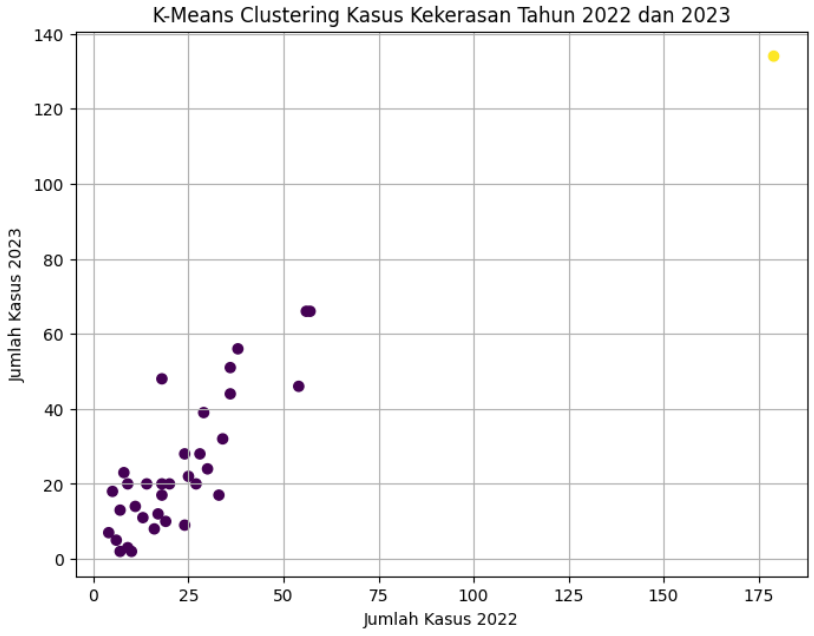

Fig 2. Visualization of K-Means Algorithm Results

Table 1. K-Means Clustering Result Table

	Distric/City	Number of Victims 2022	Number of Victims 2023	Cluster
1	Cilacap Regency	8	23	0
2	Banyumas Regency	36	51	0
3	Purbalingga Regency	17	12	0
4.	Banjarnegara Regency	16	8	0
5	Kebumen Regency	54	46	0
6	Purworejo Regency	24	28	0
7	Wonosobo Regency	56	66	0
8	Magelang Regency	27	20	0
9	Boyolali Regency	25	22	0
10	Klaten Regency	24	9	0
11	Sukoharjo Regency	34	32	0
12	Wonogiri Regency	10	2	0
13	Karanganyar Regency	18	20	0
14	Sragen Regency	5	18	0
15	Grobogan Regency	18	17	0
16	Blora Regency	4	7	0
17	Rembang Regency	7	2	0
18	Pati Regency	19	10	0
19	Kudus Regency	6	5	0
20	Jepara Regency	14	20	0
21	Demak Regency	29	39	0
22	Semarang Regency	57	66	0
23	Temanggung Regency	9	3	0
24	Kendal Regency	28	28	0
25	Batang Regency	11	14	0
26	Pekalongan Regency	38	56	0
27	Pemalang Regency	36	44	0
28	Tegal Regency	33	17	0
29	Brebes Regency	30	24	0
30	Magelang Distric	7	13	0
31	Surakarta Distric	18	48	0
32	Salatiga Distric	20	20	0
33	Semarang Distric	179	134	1
34	Pekalongan Distric	13	11	0
35	Tegal Distric	19	20	0

The summary results:

- *Cluster 0*: Areas with a low number of violence cases. This cluster includes most of the districts/cities in Central Java Province.
- *Cluster 1*: Areas with high total cases of violence.

4. CONCLUSION

Based on a clustering analysis using the K-Means algorithm of data on cases of violence against women in Central Java Province, two main clusters were obtained that showed differences in the number of cases of violence. The clustering results show that:

1. The regions in Central Java Province can be grouped into two clusters, namely clusters with a low number of violence cases and clusters with a high number of violence cases.
2. The clustering table shows that most areas have a relatively low number of cases of violence, while certain areas have much higher levels of violence and require more attention from the Government.
3. This analysis shows the different distribution patterns of violence cases in each region, which can be used as a basis for authorities to formulate more appropriate policies to prevent and address violence against women.

The results of this clustering are expected to assist in identifying priority areas for more intensive policies, as well as provide insights into understanding patterns of violence against women in Central Java. Thus, this study contributes to supporting efforts to reduce the number of violence against women through a data-driven approach.

REFERENCES

Adelina Bui, Maria, and Agus Bahtiar. 2024. "Implementasi Metode Algoritma K-Means Clustering Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8 (2): 1451–56. https://doi.org/10.36040/jati.v8i2.8975.

Akbar, Emir. 2023. "Perbandingan Algoritma Dbscan-K Means Dan K Means Untuk Pengelompokan Madrasah Aliyah Provinsi Jawa Timur." *Repository.Uinjkt.Ac.Id*, 32–51.

Annurfariz, Aditya, Ade Irma Purnamasari, and Irfan Ali. 2024. "Implementasi Algoritma K-Means Pada Kasus Kekerasan Dalam Rumah Tangga Di Jawa Barat." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8 (2): 1904–10. https://doi.org/10.36040/jati.v8i2.8348.

Fatma, Mawarni. 2024. "Perlindungan Hukum Terhadap Perempuan Korban Tindak Pidana Kekerasan Dalam Rumah Tangga." *Biram Samtani Sains* 2 (1): 1–17. https://doi.org/10.55542/jbss.v2i1.27.

Hardianti, Firda Yanis, Reno Efendi, Putri Diah Lestari, and Elisabeth Septin Puspoayu. 2021. "Urgensi Percepatan Pengesahan Rancangan Undang-Undang Penghapusan Kekerasan Seksual." *Jurnal Suara Hukum* 3 (1): 26. https://doi.org/10.26740/jsh.v3n1.p26-52.

Hoerunnisa, Anis, Gifthera Dwilestari, Fatihanursari Dikananda, Heliyanti Sunana, and Denni Pratama. 2024. "Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8 (1): 103–10. https://doi.org/10.36040/jati.v8i1.8249.

Hutagalung, Juniar, and Fifin Sonata. 2021. "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah." *Jurnal Media Informatika Budidarma* 5 (3): 1187. https://doi.org/10.30865/mib.v5i3.3113.

Lalu Muhamad Rofiâ€™I, and Mawardi. 2022. "Analisis Kebijakan Hukum Pidana Dalam Penjatuhan Sanksi Kebiri Pada Pelaku Kejahatan Seksual Terhadap Anak." *Jurnal Kolaboratif Sains* 5 (10): 706–19. https://doi.org/10.56338/jks.v5i10.2840.

Muhammad Jadi. 2021. "Kekerasan Terhadap Perempuan: Pemicu Dan Alternatif Penanganan." *Afiasi : Jurnal Kesehatan Masyarakat* 6 (2): 110–26. https://doi.org/10.31943/afiasi.v6i2.161.

Okta Windya Ningrum, and Yana S. Hijri. 2022. "Implementasi Kebijakan Penanganan Kasus Kekerasan Terhadap Perempuan." *Jurnal Inovasi Dan Kreativitas (JIKa)* 1 (2): 109–25. https://doi.org/10.30656/jika.v1i2.4180.

Sa'diah, Halimatu, Ultach Enri, and Tesa Nur Padilah. 2023. "Penerapan Algoritme K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat." *JATI (Jurnal Mahasiswa Teknik Informatika)* 7 (2): 1351–57. https://doi.org/10.36040/jati.v7i2.6838.

Siswa, Yoga Azhima Taghfirul, and Wawan Joko Pranoto. 2023. "Implementasi Seleksi Fitur Information Gain Ratio Pada Algoritma Random Forest Untuk Model Data Klasifikasi Pembayaran Kuliah." *Dinamika Informatika* 15 (1): 41–49.

Stepanus Ginting, Roni, Hamdani Hamdani, Anindita Septiariani, and Faza Alameka. 2022. "The Clustering Tindak Kekerasan Dalam Rumah Tangga Di Kota Samarinda Menggunakan Algoritma K-Means." *Metik Jurnal* 6 (2): 172–77. https://doi.org/10.47002/metik.v6i2.378.

Sulistiyawati, Ari, and Eko Supriyanto. 2021. "Implementasi Algoritma K-Means Clustring Dalam Penetuan Siswa Kelas Unggulan." *Jurnal Tekno Kompak* 15 (2): 25. https://doi.org/10.33365/jtk.v15i2.1162.

Sundari, Mitha Amelia, Rahmadhani Pane, and Rohani Rohani. 2023. "Data Mining Clustering Korban Kejahatan Pelecehan Seksual Dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC." *Building of Informatics, Technology and Science (BITS)* 5 (1): 364–75.