

Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means

Noviya Adawiyah¹, Nina Sulistiyowati², Mohamad Jajuli³

¹Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

E-mail: *¹noviya.adawiyah17165@student.unsika.ac.id, ²nina.sulistio@unsika.ac.id,

³mohamad.jajuli@unsika.ac.id

Abstrak – Kekerasan adalah tindakan atau ancaman terhadap diri sendiri, sekelompok orang atau masyarakat atau individu yang menyebabkan kematian, kerugian psikolog, trauma, atau perampasan hak. Kabupaten Karawang merupakan salah satu kabupaten yang ada di wilayah Jawa Barat. Kekerasan yang menimpa anak dan perempuan di wilayah Karawang marak terjadi, demikian halnya dengan rendahnya kesadaran korban untuk menindaklanjuti kasus yang dialaminya. Tujuan dari mengetahui hasil *cluster* kasus kekerasan terhadap anak dan perempuan menjadi tiga *cluster* yang tersebar pada setiap kecamatan di Kabupaten Karawang dengan kategori tingkat kekerasan rendah, sedang atau tinggi agar pemerintah karawang dapat memberi penanganan yang berbeda dan lebih terarah dan fokus dari hasil analisis untuk masing-masing kecamatan. *Data mining* yaitu proses pengalihan data untuk memperoleh informasi baru. Dalam penelitian ini menggunakan metodologi CRIPS-DM. Penelitian ini melakukan perhitungan algoritma *k-means* pada data kasus kekerasan terhadap anak dan perempuan tahun 2016-2020. Hasil pengujian menggunakan *tools* WEKA 3.8 dengan menentukan *cluster* (K) terbaik menggunakan metode *elbow* yaitu dengan menghitung perbandingan selisih *Sum of Square Error* K=2 sampai K=10 dan K=3 sebagai K terbaik maka didapatkan tiga *cluster* atau tiga kategori tingkat kekerasan yaitu *cluster* 0 terdapat 4 anggota yang dikategorikan tingkat kekerasan tinggi, *cluster* 1 terdapat 2 anggota yang dikategorikan tingkat kekerasan sedang, dan *cluster* 2 terdapat 24 anggota yang dikategorikan tingkat kekerasan rendah, hasil *clustering* dievaluasi menggunakan persamaan pengujian *purity measure*, menghasilkan nilai *purity* 0.617, hal tersebut menunjukkan *cluster* cukup baik.

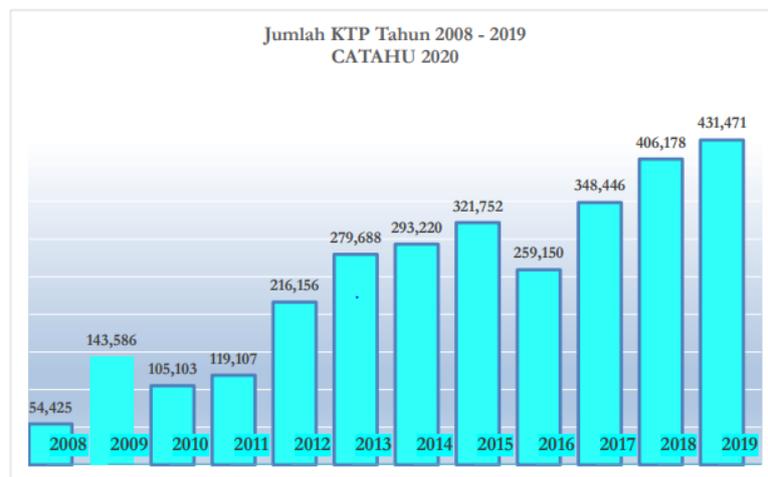
Kata Kunci — Algoritma K-Means, Clustering, Data Mining, dan Purity Measure.

Abstract– Violence is action or threats against themselves alone, a group of people or community a group of people or community, loss psychologist, trauma, or deprivation of rights. District Karawang is on of the district that exist in the province of Jawa Barat. Violence that befell children and women in the area of Karawang bloom occurs, such as the lacj awareness of the victim to follow up cases that happened. The purpose of knowing the results of the cluster of cases of violence against children and women into three clusters are statterd in every sub-district in the District Karawang with category level of hardness low, medium or high in order that the government Karawang can provide treatment that is defferent and more targeted and focused on the results of the analysis for each-each district. Data mining is the process of extracting data to obtain new information. In this study using CRIPS-DM methodology. Research is doing computation algorithm *k-means* clustering on the data of case of violence against children and women in 2016-2020. The results of the test using the WEKA 3.8 tools by determining the best K using the elbow method, namely by calculating the ratio of the difference between the Sum of Square Errors K = 2 to K = 10 and K = 3 as the best K, we get three clusters or the three categories of the level of violence that is cluster 0 there are 4 members who categorized the level of violence high, cluster 1 there are 2 members categorized the level of violence medium, and cluster 2 there are 24 members who categorized the level of violence low, the results of clustering is evaluated using equation testing *purity measure*, generate value *purity* 0,617, case that shows the cluster is quite good

Keywords — K-Means Algorithm, Clustering, Data Mining, and Purity Measure

1. PENDAHULUAN

Kekerasan dapat diartikan sebagai tindakan atau ancaman terhadap diri sendiri, sekelompok orang atau masyarakat atau individu yang menyebabkan kematian, kerugian psikologis, trauma atau perampasan hak[1]. Kasus kekerasan terhadap anak dan perempuan merupakan kejahatan serius yang saat ini dihadapi masyarakat Indonesia, sering kali anak-anak menjadi korban dari tindakan tidak manusiawi dan perempuan rentan menjadi korban kekerasan. Dari tahun ke tahun kasus kekerasan di Indonesia semakin meningkat dan menjadi perhatian dikalangan masyarakat. Kondisi ini tercatat di Komisi Nasional Anti Kekerasan Terhadap Perempuan (KOMNAS Perempuan), angka kekerasan terhadap perempuan setiap tahun mengalami peningkatan[2]. Pada gambar 1 menunjukkan grafik peningkatan kasus kekerasan terhadap perempuan dalam 12 tahun terakhir kasus kekerasan terhadap perempuan meningkat sebesar 792% (hampir 800%). Artinya, dalam 12 tahun terakhir jumlah kasus kekerasan terhadap perempuan di Indonesia meningkat 8 kali lipat.



Gambar 1. Jumlah Perempuan Korban Kekerasan Tahun 2019
(Sumber: komisi nasional anti kekerasan terhadap perempuan, 2020)

Salah satu kabupaten yang ada di wilayah Jawa Barat yaitu Kabupaten Karawang yang mana angka kasus kekerasan di wilayah karawang mengalami peningkatan dari tahun ke tahun[3]. Kasus kekerasan terhadap anak dan perempuan termasuk masalah yang sangat meresahkan masyarakat kabupaten karawang, dengan meningkatnya kasus kekerasan yang menimpa anak dan perempuan serta rendahnya kesadaran korban untuk menindaklanjuti kasus yang dialaminya[4]. Sehingga untuk memberikan perlindungan, pengawasan dan menekan angka kasus kekerasan terhadap anak dan perempuan maka pemerintah karawang dalam hal ini perlu melakukan upaya pengendalian terhadap kasus ini, untuk membantu upaya pengendalian kasus ini maka dibutuhkannya suatu metode untuk mengklasterisasi kasus kekerasan terhadap anak dan perempuan berdasarkan pemetaan wilayah dengan kategori tingkat kekerasan yang diharapkan dapat mengklasterisasi kasus kekerasan kedalam *cluster* yang tepat sehingga memudahkan pemerintah karawang untuk mengetahui daerah-daerah rawan kekerasan.

Algoritma *k-means* dari teknik *clustering* dapat membantu mengklasterisasi kasus kekerasan terhadap perempuan dan anak di kabupaten karawang. Klasterisasi bermanfaat untuk dinas pemberdayaan perempuan dan perlindungan anak (DP3A) kabupaten karawang salah satunya adalah untuk memberikan informasi kepada DP3A juga kepada pemerintah kabupaten karawang sebagai masukan untuk implementasi kota ramah anak. Selain itu, pihak DP3A dapat merumuskan kegiatan apa yang cocok dilakukan untuk masing-masing daerah rawan kekerasan seperti sosialisasi, pendampingan kepada korban kekerasan, *training of trainer* untuk orangtua dan satuan tugas kecamatan dan lain-lain. berdasarkan hasil klasterisasi, menjadikan metode penanganan yang berbeda dan lebih terarah dan fokus dari hasil analisis untuk masing-masing daerah.

Dampak kekerasan yang dialami tentunya berpengaruh besar pada kondisi kesehatan baik pada kondisi fisik maupun psikologis[5], maka penelitian ini berkontribusi untuk kasus kekerasan

terhadap anak dan perempuan di Kabupaten Karawang salah satu nya sebagai masukan untuk implementasi kota ramah anak disetiap daerah-daerah yang ada di Kabupaten Karawang.

Pada penelitian sebelumnya (Singh, Reddy, Kapoor, & Churi, 2020) tentang *K-means Clustering Analysis of Crimes on Indian Women* dapat disimpulkan bahwa *k-means clustering* merupakan metode kuantisasi vektor, bertujuan untuk mempartisi *cluster* dalam setiap pengamatan termasuk pada *cluster* dengan nilai rata-rata terdekat (pusat *cluster*) yang berfungsi sebagai prototipe *cluster*. Ini adalah pendekatan yang efektif dan digunakan di seluruh dunia untuk analisis pengelompokan[6].

Algoritma *k-means* memiliki kekurangan atau kelemahan dalam menganalisa dan menentukan jumlah *cluster* terbaik dalam mengklaster data pada suatu dataset[7]. Untuk mengatasi masalah tersebut dalam penelitian ini metode *elbow* digunakan untuk optimasi jumlah *cluster* serta mencari nilai *cluster* terbaik[8]. Metode *elbow* untuk menentukan jumlah *cluster* terbaik dapat digunakan sebagai *default cluster* berdasarkan studi kasus yang dilakukan dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik tertentu dengan menggunakan rumus *Sum of Square Error* dari masing-masing *cluster*[9].

Berdasarkan permasalahan diatas, artikel ini akan melakukan penelitian yang bertujuan untuk mengetahui pemetaan wilayah dengan kategori tingkat kekerasan di Kabupaten Karawang.

1.1 Tinjauan Pustaka

1.1.1 Data Mining

Data mining merupakan istilah yang digunakan untuk menggambarkan penemuan pengetahuan dari *database*. *Data mining* menggunakan teknik pengenalan pola (teknik matematika dan statistik) untuk memeriksa sejumlah besar data yang disimpan dalam memori untuk mendapatkan informasi[10].

1.1.2 Clustering

Clustering adalah algoritma yang mengelompokkan sejumlah besar data ke dalam kelompok data tersebut (*cluster*). *Clustering* merupakan metode yang digunakan untuk mengelompokkan data ke dalam *cluster* berdasarkan kemiripan data, sehingga data yang saling berhubungan ditempatkan pada *cluster* yang sama[11]. Terdapat dua bagian dalam algoritma *clustering* yaitu secara *partitional* dan *hirarkis*. Algoritma *partitional* menentukan semua *cluster* pada waktu tertentu sedangkan algoritma *hirarkis* secara berurutan mencari *cluster* dengan *cluster* yang telah ditentukan[12].

1.1.3 Algoritma K-means

Algoritma *k-means* merupakan suatu algoritma membagi data menjadi *cluster* maka data yang serupa berada di *cluster* sama, yang mempunyai ketidaksamaan berada di *cluster* berbeda[13]. Secara lebih spesifik, langkah-langkah algoritma *k-means* adalah sebagai berikut:

1. Tentukan *k* sebagai jumlah *cluster* yang akan dibuat.
2. Tentukan pusat *cluster* secara acak.
3. Menghitung jarak dari setiap input data ke setiap *centroid* untuk mencari jarak terdekat dari setiap data ke *centroid*, persamaan jarak *Euclidean Distance* dapat dilihat pada persamaan 1.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2} \quad (1)$$

Dimana :

x_i : data kriteria

μ_j : *centroid* pada *cluster* ke-*j*

4. Klasifikasikan setiap data menurut seberapa dekat dengan pusat dengan *centroid* (jarak minimum).
5. Perbaharui nilai *centroid*. Gunakan rumus pada persamaan 2 untuk mendapatkan nilai *centroid* baru dari nilai rata-rata setiap *cluster*.

(2)

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j$$

Dimana :

$\mu_j(t+1)$: *centroid* baru pada iterasi ke (t+1)

N_{sj} : banyak data pada *cluster* S_j .

6. Ulangi langkah 2 sampai 5 hingga anggota setiap *cluster* tidak berubah.

1.1.4 Metode Elbow

Metode *elbow* merupakan metode yang digunakan untuk menghasilkan informasi untuk menentukan jumlah *cluster* terbaik dengan cara memeriksa persentase perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik tertentu, Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Karena semakin besar jumlah cluster K maka nilai SSE akan semakin kecil[14]. Rumus SSE pada K-Means sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{X_1} \|X_1 - C_k\|^2 \quad (3)$$

Dimana :

K : Jumlah cluster

X_1 : Data ke - i

C_x : Centroid cluster

Persentase yang berbeda dari setiap nilai *cluster* dapat direpresentasikan dengan menggunakan grafik sebagai sumber informasi. Jika nilai *cluster* pertama bersama dengan nilai *cluster* kedua memberikan sudut pada grafik atau nilai mengalami penurunan paling besar, maka nilai *cluster* tersebut adalah yang terbaik[15].

1.1.5 Purity Measure

Purity measure merupakan salah satu pengukuran dalam evaluasi *clustering*. Setiap *cluster* menggunakan label kelas berdasarkan label yang paling sering muncul di data *cluster*, dan kemudian membagi jumlah data yang benar dengan jumlah data untuk menghitung akurasi *cluster*. Rentang *purity* dari 0 hingga 1. Semakin besar nilai *purity*, semakin baik *cluster* tersebut *Cluster* dianggap murni (*pure*) semua objek dari *class* yang sama berada dalam *cluster* yang sama[16]. Untuk mengukur tingkat akurasi *clustering* atau 'r', pengukuran nilai 'r' ini menggunakan persamaan 3 berikut ini:

$$r = \frac{1}{n} \sum_{i=1}^k a_i \quad (4)$$

Dimana :

r : tingkat akurasi *clustering*

k : jumlah *cluster*

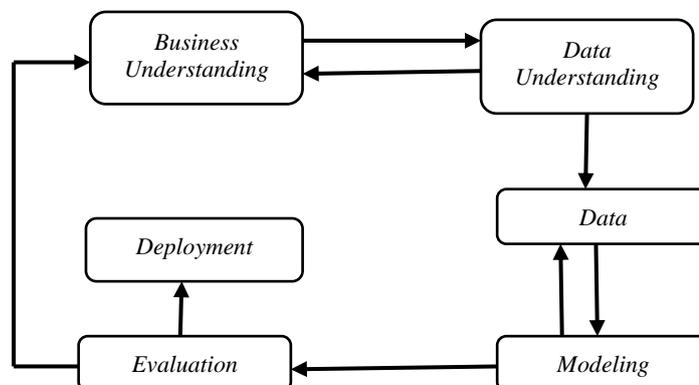
a_i : objek yang muncul di *cluster* C_i dan pada label terkait.

n : jumlah data

Semakin tinggi nilai r (mendekati 1), semakin baik kualitas *cluster* tersebut.

METODE PENELITIAN

Pada penelitian ini metode yang digunakan adalah metodologi datamining CRIPS-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri dari 6 tahap[17]. Gambar 2 merupakan metodologi penelitian ini.



Gambar 2. Model CRIPS-DM

2.1 Business Understanding

Pada tahap ini dilakukannya observasi ke tempat yang akan dijadikan objek penelitian yaitu Dinas Pemberdayaan Perempuan dan Perlindungan Anak (DP3A) Kabupaten Karawang dan mencari tau potensi data yang akan diolah maka akan didapatkannya informasi yang akan dibutuhkan untuk menyelesaikan penelitian ini.

2.2 Data Understanding

Pada tahap ini proses pengumpulan data, kemudian data tersebut akan dipelajari dan dipahami lebih lanjut.

2.3 Data Preparation

Pada tahap ini dilakukannya persiapan pada data yang diperoleh dari tahap sebelumnya serta perlu dilakukan prapemrosesan data terlebih dahulu. Pada tahap ini meliputi semua aktivitas untuk membuat *dataset final*.

2.4 Modeling

Pada tahap ini merupakan gambaran dari rangkaian kegiatan untuk melakukan pengujian dan pelatihan terhadap data yang sudah dikumpulkan dan diolah. Ditahap ini pemilihan teknik *clustering* dengan algoritma *k-means* yang akan digunakan.

2.5 Evaluation

Pada tahap ini akan dilakukan evaluasi keakuratan hasil yang dicapai oleh model yang sudah digunakan menggunakan *purity measure*.

2.6 Deployment

Pada tahap ini merupakan tahap akhir dari metodologi CRIPS-DM merupakan penyusunan laporan akhir penelitian dengan mempresentasikan hasil dalam bentuk tabel, grafik, atau deskripsi yang mudah dipahami.

2. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Tujuan bisnis berdasarkan pada fungsi analisis algoritma *k-means* yaitu agar data jumlah kasus kekerasan terhadap anak dan perempuan dapat dimanfaatkan dengan sebaik mungkin dan dapat memberikan informasi untuk mengklasterisasi kasus kekerasan terhadap perempuan dan anak

di Kabupaten Karawang berdasarkan pemetaan wilayah dengan kategori tingkat kekerasan yang diharapkan dapat mengklasterisasi kasus kekerasan kedalam *cluster* yang tepat sehingga memudahkan pemerintah karawang untuk mengetahui daerah-daerah rawan kekerasan serta dapat memberikan informasi pada DP3A mengenai metode *clustering* algoritma *k-means* untuk mengklaster kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang.

3.2 *Data Understanding*

Pemahaman data mengacu pada data-data yang diperoleh dari Dinas Pemberdayaan Perempuan dan Perlindungan Anak (DP3A) Kabupaten Karawang mengenai kasus kekerasan terhadap anak dan perempuan dan data yang didapat yaitu nama kecamatan, jenis kekerasan, kategori, dan total kasus. Dalam tahap ini akan dideskripsikan mengenai atribut, tipe data, dan keterangan data. Atribut dikatakan nominal jika memiliki nilai sebagai label, sedangkan tipe data numerik adalah tipe data yang digunakan untuk angka atau bilangan bulat yang menyimpan nilai dalam format numerik[18]. Deskripsi data dapat dilihat pada tabel 1 dibawah.

Tabel 1. Mendeskripsikan Data

Atribut	Tipe Data	Keterangan
No	Nominal	Merupakan atribut yang tidak diakumulasikan perkecamatan dalam satu tahun
Nama Kecamatan	Nominal	Merupakan data kecamatan di Kabupaten Karawang
Jenis Kekerasan (Fisik)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Jenis Kekerasan (Psikis)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Jenis Kekerasan (Seksual)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Jenis Kekerasan (Eksplotasi)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Jenis Kekerasan (Penelantaran)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Jenis Kekerasan (Lainnya)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Kategori (Anak)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Kategori (Perempuan)	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun
Total Kasus	Numerik	Merupakan atribut yang akan diakumulasikan perkecamatan dalam satu tahun

3.3 *Data Preparation*

Pada tahap ini dilakukan pemilihan data. Data yang digunakan yaitu data jumlah kasus kekerasan terhadap anak dan perempuan perkecamatan berdasarkan jenis kekerasan tahun 2016, 2017, 2018, 2019, dan 2020. Dikarenakan tidak semua data di gunakan. Atribut nomer dihilangkan karena tidak diperlukan pada saat perhitungan pada *tools* pemodelan data mining. Tahap berikutnya yaitu integrasi data merupakan tahap penggabungan data dari berbagai tabel ke dalam satu tabel baru. Dataset kasus kekerasan terhadap anak dan perempuan berasal dari satu sumber yaitu dari DP3A Kabupaten Karawang. Langkah berikutnya data harus melalui proses integrasi karena data akan digabungkan dari lima tahun sebelumnya. Tabel 2 merupakan tabel *dataset final* hasil integrasi data kasus kekerasan terhadap anak dan perempuan tahun 2016-2020 yang terdapat 314 korban kasus kekerasan.

Tabel 2. Integrasi Data

Kecamatan	Jenis Kekerasan Fisik	Jenis Kekerasan Psikis	Jenis Kekerasan Seksual	Jenis Kekerasan Eksplotasi	Jenis Kekerasan Penelantaran	Jenis Kekerasan Lainnya	Kategori Anak	Kategori Perempuan	Total Kasus
Karawang Barat	15	4	12	0	5	12	18	30	48
Pangkalan	1	0	0	0	0	0	1	0	1
Telukjambe Timur	16	7	9	0	7	16	14	41	55

Ciampel	1	0	1	0	1	0	2	1	3
Klari	6	1	6	2	6	8	14	15	29
Rengasdengklok	4	0	6	0	0	1	4	7	11
Cilamaya Kulon	0	0	3	1	0	1	2	3	5
Batujaya	0	0	5	0	0	0	5	0	5
Tirtajaya	0	0	1	1	0	0	1	1	2
Pedes	0	0	0	0	0	2	1	1	2
Cibuaya	0	0	0	1	0	1	0	2	2
Pakisjaya	0	0	0	0	0	0	0	0	0
Cikampek	7	1	4	0	3	5	9	11	20
Jatisari	1	0	0	1	0	2	2	2	4
Cilamaya Wetan	0	0	1	0	1	2	2	2	4
Tirtamulya	0	0	0	0	0	0	0	0	0
Telagasari	1	0	7	0	0	2	7	3	10
Rawamerta	0	0	0	0	0	0	0	0	0
Lemahabang	0	0	5	0	0	1	5	1	6
Tempuran	0	0	22	0	0	0	21	1	22
Majalaya	4	1	2	0	2	1	3	7	10
Jayakarta	0	0	2	1	1	0	2	2	4
Kutawaluya	0	0	1	0	0	0	0	1	1
Banyusari	0	0	1	0	0	0	1	0	1
Kota Baru	5	0	2	0	0	2	2	7	9
Karawang Timur	14	7	4	1	6	8	12	28	40
Telukjambe Barat	3	0	3	0	0	0	1	5	6
Tegalwaru	0	0	0	1	0	0	0	1	1
Purwasari	2	0	4	0	1	3	6	4	10
Cilebar	0	0	1	2	0	0	3	0	3

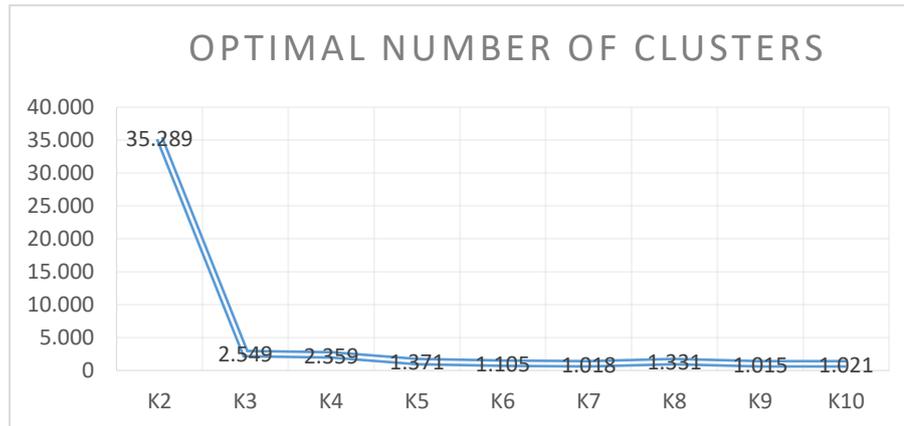
3.4 Modeling

Pemodelan data mining pada penelitian ini dibuat menggunakan perangkat lunak Weka 3.8. Pada *tools* ini tersedia algoritma *clustering* berupa algoritma *k-means* dengan pengukuran jarak yaitu *euclidean distance*[19]. Dalam melakukan penelitian ini dengan mencari nilai *cluster* (*k*) terbaik menggunakan metode *elbow*. Data yang akan diuji adalah jumlah data kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang. Jumlah *cluster* yang akan diuji adalah dari $K=2$ sampai dengan $K=10$. Dari hasil perhitungan *Sum of Square Error* terhadap 314 data kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang maka hasil yang mengalami penurunan yang paling besar adalah pada $K=3$, maka jumlah *cluster* yang ideal adalah $K=3$ yang akan dijadikan *default cluster*. Dapat dilihat pada tabel 3 dibawah.

Tabel 3. Hasil *Sum of Square Error* pada setiap *cluster*

Cluster	Hasil Sum of Square Error	Selisih
K2	35.289	35.289
K3	32.740	2.549
K4	30.381	2.359
K5	29.010	1.371
K6	27.905	1.105
K7	26.887	1.018
K8	25.556	1.331
K9	24.541	1.015
K10	23.520	1.021

Dari hasil perhitungan *Sum of Square Error* diatas dapat dilihat bahwa yang memiliki selisih atau penurunan yang sangat drastis dan membentuk siku pada titik $K=3$ dengan selisih 2.549. Dari hasil perhitungan *Sum of Square Error* diatas didapatkan grafik pada Gambar 3 dibawah.



Gambar 3. Grafik Sum of Square Error

Gambar 4 dibawah menunjukkan hasil pengujian *k-means clustering* dengan titik *cluster* yang telah ditentukan sebelumnya dengan menggunakan metode *elbow* pada *dataset* kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang.

```

Number of iterations: 4
Within cluster sum of squared errors: 32.74090126188894

Initial starting points (random):

Cluster 0: CIAMPEL,1,0,1,0,1,0,2,1,3
Cluster 1: BATUJAYA,0,0,5,0,0,0,5,0,5
Cluster 2: PEDES,0,0,0,0,0,2,1,1,2

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute                Full Data      Cluster#
                        (30.0)        (4.0)         1         2
                        (2.0)        (24.0)
-----
Nama Kecamatan           KARAWANG BARAT KARAWANG BARAT  CIKAMPEK  PANGKALAN
Jenis Kekerasan Fisik    2.6667         12.75        3.5       0.9167
Jenis Kekerasan Psikis   0.7            4.75         0.5       0.0417
Jenis Kekerasan Seksual  3.4            7.75         13        1.875
Jenis Kekerasan Eksploitasi 0.3667         0.75         0         0.3333
Jenis Kekerasan Penelentaraan 1.1            6            1.5       0.25
Jenis Kekerasan Lainnya  2.2333         11           2.5       0.75
Kategori Anak            4.6            14.5         15        2.0833
Kategori Perempuan       5.8667         28.5         6         2.0833
Total Kasus              10.4667        43           21        4.1667

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      4 ( 13%)
1      2 ( 7%)
2     24 ( 80%)

```

Gambar 4. Hasil perhitungan *k-means clustering* k=3

Pada *dataset* kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang tahun 2016-2020 *cluster* yang dihasilkan yaitu 3 *cluster* yaitu *cluster* 0, *cluster* 1, dan *cluster* 2. Adapun jumlah data atau persentase dari 3 *cluster* ini yaitu, *cluster* 0 sebanyak 13% (4 anggota), *cluster* 1 sebanyak 7% (2 anggota), *cluster* 2 sebanyak 80% (24 anggota). Adapun hasil *cluster* dari *dataset* kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang tahun 2016-2020 dapat dilihat pada tabel 4 dibawah.

Tabel 4. Anggota *Cluster*

Cluster 0	Cluster 1	Cluster 2
Karawang Barat	Cikampek	Pangkalan
Telukjambe Timur	Tempuran	Ciampel
Klari		Rengasdengklok
Karawang Timur		Cilamaya Kulon
		Tirtajaya

Batujaya
 Cibuaya
 Pedes
 Jatisari
 Pakisjaya
 Tirtamulya
 Cilamaya Wetan
 Telagasari
 Lemahabang
 Rawamerta
 Majalaya
 Jayakarta
 Kutawaluya
 Banyusari
 Kotabaru
 Telukjambe Barat
 Tegalwaru
 Purwasari
 Cilebar

Rata-rata total kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang dapat dilihat pada tabel 5 dibawah.

Tabel 5. Rata-rata total kasus

Cluster	Rata-rata total kasus kekerasan terhadap anak dan perempuan	Kategori
C0	43	Tinggi
C1	20	Sedang
C2	4.16	Rendah

Pada tabel 5 diketahui terdapat 3 cluster pada dataset kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang tahun 2016-2020, yaitu untuk cluster 0 dengan rata-rata total kasus kekerasan terhadap anak dan perempuan 43 dan memiliki 4 kecamatan yang dikategorikan sebagai tingkat kekerasan tinggi, untuk cluster 1 dengan rata-rata total kasus kekerasan terhadap anak dan perempuan 20 dan memiliki 2 kecamatan yang dikategorikan tingkat kekerasan sedang, untuk cluster 2 dengan rata-rata total kasus kekerasan terhadap anak dan perempuan 4.16 dan memiliki 24 kecamatan yang dikategorikan sebagai tingkat kekerasan rendah.

3.5 Evaluation

Hasil clustering dievaluasi dengan menggunakan persamaan pengujian purity measure, untuk algoritma k-means dengan tiga titik cluster. Kemiripan didalam cluster dihitung sebagai dasar penentu purity dari hasil pengelompokan[20]. Tabel 6 menunjukkan nilai max cluster dari setiap cluster Hasil perhitungan purity sebagai berikut:

Tabel 6. Nilai max cluster

Cluster	Fisik	Psikis	Seksual	Eksplorasi	Penelantaran	Lainnya	Anak	Perempuan	Nilai Max Cluster
C0	51	19	31	3	24	44	58	114	114
C1	7	1	26	0	3	5	30	12	30
C2	22	1	45	8	6	18	50	50	50

Dengan total sebanyak 314 jumlah korban kasus kekerasan terhadap anak dan perempuan, dapat dihitung nilai akurasi dari pemodelan cluster tersebut yaitu:

$$\begin{aligned}
 \text{Purity}(\Omega, \mathbb{C}) &= \frac{1}{n} \sum_{i=1}^k a_i \\
 &= \frac{1}{314} (114 + 30 + 50)
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{314} (194) \\
 &= 0.617 \\
 &= 61.7\%
 \end{aligned}$$

Berdasarkan perhitungan *purity measure* diatas didapatkan nilai *purity* 0.617 maka perhitungan *purity measure* yang telah dilakukan bernilai cukup baik karena mendekati 1.

3.6 Deployment

Setelah melalui proses *clustering* menggunakan algoritma *k-means* dari proses data mining dan telah melalui proses perhitungan *purity measure*, laporan disampaikan kepada pihak DP3A sebagai rekomendasi untuk membantu meminimalisir serta upaya pengendalian kasus kekerasan agar pemerintah Karawang dapat memberi perhatian serta penanganan yang efektif dalam mengatasi kasus kekerasan dan hasil penelitian sebagai masukan untuk implementasi kota ramah anak di Kabupaten Karawang.

3. SIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Penerapan algoritma *k-means* untuk mengklasterisasi kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang dilakukan dengan metode *clustering* untuk data kasus kekerasan terhadap anak dan perempuan pada tahun 2016 sampai 2020. Dari hasil *clustering* perhitungan $K=3$ terdapat tiga titik cluster yang tersebar pada setiap kecamatan di Kabupaten Karawang dengan kategori tingkat kekerasan rendah, sedang, atau tinggi. untuk *cluster* 0 dengan total 4 anggota yang dikategorikan tingkat kekerasan tinggi dan kecamatan yang termasuk kedalam kategori tingkat kekerasan tinggi yaitu Karawang Barat, Telukjambe Timur, Klari, dan Karawang Timur. Untuk *cluster* 1 dengan total 2 anggota yang dikategorikan tingkat kekerasan sedang dan kecamatan yang termasuk kedalam kategori tingkat kekerasan sedang yaitu Cikampek dan Tempuran. Untuk *cluster* 2 dengan total 24 anggota yang dikategorikan tingkat kekerasan rendah dan kecamatan yang termasuk kedalam kategori tingkat kekerasan rendah yaitu Pangkalan, Ciampel, Rengasdengklok, Cilamaya Kulon, Tirtajaya, Batujaya, Cibuaya, Pedes, Jatisari, Cilamaya Wetan, Pakisjaya, Rawamerta, Tirtamulya, Telagasari, Lemahabang, Rawamerta, Jayakarta, Majalaya, Kutawaluya, Banyusari, Kotabaru, Telukjambe Barat, Tegalwaru, Purwasari, dan Cilebar.
2. Penerapan metode *elbow* untuk menentukan *cluster* terbaik dengan cara melihat hasil perbandingan antara *cluster* yang membentuk siku pada titik tertentu menggunakan rumus *Sum of Square Error*. Dari hasil perhitungan *Sum of Square Error* $K=2$ sampai $K=10$ membentuk siku pada titik $K=3$ yang merupakan jumlah *cluster* yang ideals atau terbaik
3. Hasil evaluasi performa algoritma *k-means* untuk klasterisasi kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang dengan menggunakan persamaan pengujian *purity measure* menghasilkan nilai *purity measure* 0.617, hal tersebut menunjukkan *cluster* cukup baik karena mendekati angka 1.

4. SARAN

Beberapa saran untuk penelitian lebih lanjut yaitu:

1. Untuk metode *clustering* kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang yang dilakukan pada penelitian ini tahun 2016 – 2020 diharapkan untuk penelitian selanjutnya dapat ditambah lagi datanya.
2. Disarankan agar penelitian lebih lanjut menggunakan metode atau algoritma *clustering* lain dalam proses klasterisasi kasus kekerasan terhadap anak dan perempuan di Kabupaten Karawang.

3. Disarankan pada penelitian selanjutnya bisa menggunakan *tools* bantuan data mining lain yang memungkinkan mempunyai kemampuan lebih baik dalam melakukan analisis pengelompokan.

DAFTAR PUSTAKA

- [1] A. Anindya, Y. Indah, S. Dewi, and Z. D. Oentari, "Dampak Psikologis dan Upaya Penanggulangan Kekerasan Seksual Terhadap Perempuan," *Terap. Inform. Nusant.*, vol. 1, no. 3, pp. 137–140, 2020, [Online]. Available: <https://ejournal.seminar-id.com/index.php/tin/article/view/394>.
- [2] F. Nuraeni, N. N. Febriani Sm, L. Listiani, and E. Rahmawati, "Implementation of K-Means Algorithm with Distance of Euclidean Proximity in Clustering Cases of Violence Against Women and Children," *2019 1st Int. Conf. Cybern. Intell. Syst. ICORIS 2019*, vol. 1, no. August, pp. 162–167, 2019, doi: 10.1109/ICORIS.2019.8874883.
- [3] E. S. B. Ningsih and S. Hennyati, "Kekerasan Seksual Pada Anak Di Kabupaten Karawang," *Midwife J.*, vol. 4, no. 02, pp. 56–65, 2018, [Online]. Available: <http://jurnal.ibijabar.org/kekerasan-seksual-pada-anak-di-kabupaten-karawang/>.
- [4] O. Bariah and R. Marlina, "Tinjauan Psikologi dan Agama Terhadap Tindak Kekerasan pada Anak," *J. Stud. Insa.*, vol. 7, no. 2, p. 92, 2019, doi: 10.18592/jsi.v7i2.2646.
- [5] D. Fitriani, H. Fajriah, and Erika Shovianda, "Gender Equality: Internasional Journal of Child and Gender Studies ISSN: 2461-1468/E-ISSN: 2548-1959," *J. Child Gen. Stud. ISSN*, vol. 4, no. 1, pp. 39–54, 2018.
- [6] R. Singh, R. Reddy, V. Kapoor, and P. Churi, "K-means Clustering Analysis of Crimes on Indian Women K-means Clustering Analysis of Crimes on Indian Women," no. August, 2020, doi: 10.5281/zenodo.3909955.
- [7] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster," *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021, doi: 10.33633/tc.v20i2.4572.
- [8] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.
- [9] T. Santoso and F. Saftarina, "Klasterisasi Petani Padi Sawah di Kota Metro Provinsi Lampung Menggunakan Algoritma K-Means Cluster dan Elbow Method," *J. Agribus. Community Empower.*, vol. 3, no. 1, pp. 37–43, 2020.
- [10] Iskandar, P. Tarigan, and saidi ramadan Siregar, "Penerapan Data Mining Dalam Surat Jalan Transportasi Mobil Angkutan Menggunakan Metode Clustering," *J. Pelita Inform.*, vol. 17, pp. 460–464, 2018.
- [11] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020, doi: 10.35314/isi.v5i2.1457.
- [12] I. Ismawati, "Kajian Data Mining Profil Siswa Baru Dalam Penentuan Strategi Promosi Dengan Metode Two Step Clustering," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 3, pp. 47–59, 2019, doi: 10.33197/jitter.vol5.iss3.2019.302.
- [13] W. Supriyanti, Kusriani, and A. Amborowati, "Perbandingan Kinerja Algoritma c4.5 Dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa," *J. Inf. Politek. Indonusa*, vol. 1, no. 3, pp. 61–67, 2016.
- [14] E. Muningsih and A. B. S. I. Yogyakarta, "Optimasi jumlah cluster k-means dengan metode elbow untuk pemetaan pelanggan," *Pros. Semin. Nas. ELINVO*, no. September, pp. 105–114, 2017.
- [15] A. T. Rahman, Wiranto, and A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017, [Online]. Available: <https://jurnal.uns.ac.id/itsmart/article/download/11296/11108>.
- [16] mohamad jajuli nurul rohmawati, sofi defiyanti, "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa," *Jitter 2015*, vol. I, no. 2, pp. 62–68, 2015.
- [17] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.

- [18] S. E. D. A. Kusriani, "Algoritma K-Means untuk Diskretisasi Numerik Kontinyu Pada Klasifikasi Intrusion Detection System Menggunakan Naive Bayes," *Konf. Nas. Sist. Inform.*, pp. 61–66, 2017.
- [19] B. Juliartha, M. Putra, D. Ariani, and F. Yuniarti, "Analisis Hasil Belajar Mahasiswa Dengan Clustering Menggunakan Metode K-Means," vol. 12, no. 2, pp. 49–58, 2020.
- [20] N. Anastasya, A. W. Widodo, and M. A. Rahman, "Pengelompokan Musik berdasarkan Emosi menggunakan Metode Transformasi Haar Wavelet," *J. Pengemb. ...*, vol. 3, no. 11, pp. 10524–10530, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6706>.