Application of Clustering Methods On Sexual Harassment Cases

by Nurhafieza Nurhafieza

Application of Clustering Methods On Sexual Harassment Cases

Nurhafieza[1], Novriyenni[2], Milli Alfhi Syari[3]

[1,2,3]Information System, STMIK Kaputama, Kota Binjai

Email address:[1]nurhafiezaa94@gmail.com , [2]novriyenni.sikumbang@gmail.com , [3]milli.fhisya@gmail.com

Abstract: Sexual harassment is one of the most common crimes in Indonesia, this act of sexual harassment can occur in daily life regardless of time, whether at work, on the street, or at home. Sexual abuse can come from unknown people, people who have hate, even people we care about. To solve problems that often occur in some cases including cases of sexual harassment that often occur in women based on certain factors, resulting in trauma to the victims who suffer physically, sexually, and psychologically, are required quick action to reduce the number in cases of sexual abuse in the area that often occur using clustering methods so that later it is expected to help the agency in socializing so that the community is more alert while in the place. From the testing conducted using 20 sexual harassment cases data there are 3 groups, namely group 1 there are 9 data and 2 groups there are 5 data and group 3 there are 6 data and it can be known that in cluster 1 is a group in the case data on sexual harassment based on the factors that are many causes with a total of 9 data and located in the age group (X) is 12-16 years, and for the group Sexual Harassment (Y) namely Physical Harassment and causing factor (Z) that is a lot due to individual factors.

Keywords : Data Mining, Clustering, Sexual Harassment

1. INTRODUCTION

Sexual harassment is one of the most common crimes in Indonesia, this act of sexual harassment can occur in daily life regardless of time, whether at work, on the street, or at home. Sexual abuse can come from unknown people, people who have hate, even people we care about. (Sundari et al., 2023)

The Office of Women Empowerment, Child Protection, Population Control and Family Planning Stabat has the task of solving problems that often occur in some cases including cases of sexual abuse that often occur in women based on certain factors that can result in victims experiencing trauma to physical, sexual, and psychological suffering.

To solve this problem, quick action is needed to reduce the number in cases of sexual harassment in the region that often occurs using clustering methods so that it will be expected to help the agency in socializing so that people are more alert while in the place. Additionally, it is necessary to improve security in the region and provide consultation venues such as psychologists. Based on the background above, the author attempts to create the research title, "Application of Clustering Methods On the Case of Sexual Harassment".

2. METHODS

Data Mining is a series of processes to extract added value from a data set in the form of knowledge that has not been known manually from a data set. (Relita Buaton et al., 2019) Data mining is the process of extracting useful information and patterns from very large data. Data mining includes data collection, data extraction, data analysis, and data statistics. Data mining is also known as Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, and others. (Arhami et al., 2020).

The advantages of using data mining are as follows:

1. Better customer relationship management
2. Market trend forecasting
3. Assist in Competition
4. Attracting and retaining customers

While the shortcomings of data mining are as follows:

1. Data security is at risk of being hacked by unwanted parties.
2. Data mining violates user privacy. We must be aware of this reminder that user profiles can be sold, traded, or can be used as a tool to seek profit that harms people. and others. (Prasetyo, 2012)

Data mining techniques are methods that can be applied to various fields. Therefore, this method needs to be adjusted to the problem or needs of its users. There are seven classifications of data mining that are distinguished based on how it works, as follows.

1. Tracking Patterns/Sequencing

    The first data mining technique is to track patterns or sequences of events. This technique serves to find a pattern on a series of sequential events. Tracking Patterns can detect something at certain intervals, such as a surge in product demand when the weekend or the number of people visiting your site during certain weather.

2. Classification

    Classification techniques require more complex data mining techniques because it demands you to collect all data from a specific class or category. This technique is the most commonly used. You can apply it to group data based on the label you want. For example, based on financial and transaction information, you can group customers into low, medium, or high credit risk.

3. Association

The next classification of data mining is association, market basketball analysis related to product marketing. Basket analysis aims to know or identify products that are often purchased simultaneously by customers. For example, when buying a certain potato snack, customers also buy a packaged soda drink. By knowing the habits of customers like this, the company can also label certain products as "people too bought this" on the marketplace.

4. Outlier Detection

This technique aims to identify when anomalies occur in data patterns. For example, when your product is usually always bought by male sex customers, but one week in February, there is a sudden surge in purchases made by female customers. Outlier Detection techniques play a role in analyzing the surge and the cause, so you can decide the next sales move.

5. Clustering

Clustering techniques are almost similar to classification, but require more labels or data groups based on similar patterns. For example, you want to group different audience demographics into groups based on background, financial, or purchasing amounts when shopping in your store.

6. Regression

The next classification of data mining, regression, aims to find its numerical value pattern instead of its class. The result of this technique is a function as a determinant based on the value of the input. For example, you can use it to determine product prices based on other factors such as availability, customer demand, and competitors.

7. Forecasting/Prediction

The latter data mining technique can be said to be most valuable because it aims to predict the value that will be achieved in a given period. With predicition techniques, data noise and value in the previous period are used as reference or basis of predictions. For example, based on the purchase data last month, you can predict about what kind of purchases will be made in the next month.(Prasetyo, 2012)

2.1 K-Means Algorithm (Clustering)

K-Means algorithm is one of the techniques in data mining to group data into multiple groups based on distance, criteria, conditions or characteristics. Data in one group must have the shortest distance, criteria, conditions or characteristics of the same or almost the same or

one another. K-Means algorithms can group objects that have similarities. (Buulolo Efori, 2020)

The steps in grouping data with K-Means algorithm are as follows:

1. Determine the number of cluster(k) in the data set.

2. Determining Center value (centroid)

3. Calculate close distance with centroid

4. The centroid distance used is Euclidean Distance, with the formula as below

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \ldots + (x_{ki} - k_j)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Description:

d_{ij} = distance from ith data to cluster center j

x_{kj} = data from i-th on kth data attribute

x_{kj} = data from the jth attribute in the kth data attribute

2.2 Definition of Sexual Hardness

Sexual violence is any act that appears in the form of coercion or threatens to have sexual intercource, torture or act sadistically and leaves the victim after sexual acts. Sexual violence can be the treatment of sexual precontact between children and larger people (through words, touch, visual images, exhibitionism), as well as the treatment of direct sexual contact between children and adults (incest, rape, sexual exploitation). Sari et al., 2022)

As for factors that affect sexual violence are divided over; (Hasan et al., 2023)

1. Individual Factors are conditions in which the child is too innocent, has weak temperament, the child's ignorance of his rights, the child is too dependent on adults and the most important factor is the habit of children lying. Perhaps from the child who initially tried to do something outside the limit that could provoke the other party to act abuse him. For example, a daughter permits her parents to go to play the house of her friend, but it turns out that the child is going to koskan her boyfriend or other places that can cause abuse. Religion is the foundation of human life. However, the social cultural changes occur, the religious education should be preceded. For, from him it contains moral values, ethics, and healthy living guidelines that are universal and lasting. Parents have a great responsibility for the child's development so that if adults are knowledgeable and faithful. Merging is the process of interaction performed by individuals with individuals, can also by individuals with groups.

2. Family Factor is the child of the victim of divorce, or comes from an intact family, thus forming a pattern of child behavior that is not good to others that increase the risk of sexual violence in children.

3. Environmental Factors where factors of unhealthy or vulnerable social environmental conditions, such as unnatural child/teen behavior, vulnerable areas (security and public order disorders). Poor social environment, slum settlements, displaced child playground, indifferent attitudes to exploitation, views on child values are too low, wage economics, weak legal devices, absence of stable social control mechanisms trigger increased incidence of sexual violence in children.

As for the impact of sexual violence, namely;

1. Sexual impact, occurring on the life of the victim in the next life, such problems include the fear of having sex with her husband because her virginity has been destroyed, frigid, pain when having sex and so on.

2. Psychological impacts, regrets within yourself, depression, panic, feeling guilty and losing self-esteem. Victims can also experience phobias in crowded places, fear at home, phobias against sex and other types of phobias.

3. Social impacts, such as victims, have difficulty living their social life as they were before they had been given the incident. Not getting more confident, suspicious of others and attractive of social life. Based on the above description, it can be concluded that sexual violence not only affects physical, psychological, and social, but also accompanies the life of the victim later in life when the victim has a partner. (Sari et al., 2022)

In general, sexual harassment has 5 forms, namely:

1. Physical abuse, namely: Unwanted touch leads to sexual acts such as kissing, patting, hugging, pinching, stroking, massaging the nape, attaching the body or other physical touch.

2. Oral abuse, namely: Unwanted verbal remarks about personal life or body parts or appearance of a person, including jokes and sexual charged comments.

3. Non-verbal/signal abuse, namely: Body language and/or sexual pitched movement, rhythm performed over and over again, staring at the body full of lust, signal with fingers, licking lips, or others.

4. Visual harassment, namely: Showing pornographic material in the form of photos, posters, cartoon images, screensavers or other, or harassment via email, SMS and other media.

5. Psychological/emotional abuse, namely: Continuous and unwanted requests and calls, unexpected dating calls, sexual humiliation or reproach. Sexual harassment facing both men and women in its various forms, ranging from the light comments made and sexual assault. Connotate sexual and physical contact in secret (holding, touching to certain parts of the body) to a blatant call and sexual assault. Sari et al., 2022)

3. RESULTS AND DISCUSSION

3.1. Research methodology

In the application of this method there are several initial processes performed using the K-Means algorithm method.

1) Determine the number of clusters to be used in data sharing,

2) Determine the initial centroid obtained randomly and the number of centroid as many clusters to be made,

3) Calculate the distance on each data input to the cluster center until the closest distance is found from each data to the centroid.

Distance calculation is done using the Euclidean Distance equation, can etc. see below (3.1);

$$d\text{-}euclidean \qquad (X,Y) \qquad = \sqrt{(x_i - s_i)^2 + (y_i + t_i)^2}$$
....(3.1)

Table 3. 1 Data Case Sexual Harassment

No.	Age	Type of Sexual Harassment	Causing Factor
1	26 Year	Physical Harassment	Environmental Factors
2	8 Year	Physical Harassment	Individual Factor
3	18 Year	Physical Harassment	Individual Factor
4	33 Year	Physical Harassment	Environmental Factors
5	38 Year	Physical Harassment	Family Factor
6	28 Year	Visual harassment	Environmental Factors
7	30 Year	Visual harassment	Individual Factor
8	40 Year	Physical Harassment	Family Factor
9	10 Year	Physical Harassment	Family Factor

No.	Age	Type of Sexual Harassment	Causing Factor
10	48 Year	Physical Harassment	Family Factor
11	16 Year	Physical Harassment	Individual Factor
12	20 Year	Oral Harassment	Environmental Factors
13	11 Year	Physical Harassment	Family Factor
14	46 Year	Physical Harassment	Family Factor
15	50 Year	Physical Harassment	Family Factor
16	14 Year	Physical Harassment	Individual Factor
17	13 Year	Physical Harassment	Individual Factor
18	36 Year	Non-verbal abuse/requisite	Individual Factor
19	23 Year	Psychological/emotional abuse	Environmental Factors
20	28 Year	Physical Harassment	Individual Factor

The following below are the initialization tables of Age criteria, Type of Sexual Harassment criteria and Causing Factor criteria that can be seen in table III.3, table III.4, and table III.5 as follows:

Tabel 3. *2 Inisialisasi Kriteria Age*

Code	Age
1	≤ 11 Year
2	12 - 16 Year
3	17 - 25 Year
4	26 - 35 Year
5	36 – 45 Year
6	≥ 46 Year

Tabel 3.3 *Type of Sexual Harassment*

Code	Type of Sexual Harassment
1	Physical Harassment
2	Visual harassment

Code	Type of Sexual Harassment
3	Oral Harassment
4	Non-verbal abuse/requisite
5	Psychological/emotional abuse

Tabel 3.4 Causing Factor

Code	Causing Factor
1	Individual Factor
2	Family Factor
3	Environmental Factors

Next do the data in the form of numbers, then this data can be expressed in an independent variables namely Age (X), Type of Sexual Harassment (Y), and Causing Factor (Z). The table below is the data that has been transformed.

Tabel 3.5 Tranformasi Data

No.	Age (X)	Type of Sexual Harassment (Y)	Causing Factor (Z)
1	4	1	3
2	1	1	1
3	3	1	1
4	4	1	3
5	5	1	2
6	4	2	3
7	4	2	1
8	5	1	2
9	1	1	2
10	6	1	2
11	2	1	1
12	3	3	3
13	1	1	2
14	6	1	2
15	6	1	2

No.	Age (X)	Type of Sexual Harassment (Y)	Causing Factor (Z)
16	2	1	1
17	2	1	1
18	5	4	1
19	3	5	3
20	4	1	1

The next step is calculated data based on the algorithm of k-means clustering

Iteration 1

Centroid 1 = (3, 1, 1) is taken randomly from data 3

Centroid 2 = (4, 1, 3) is taken randomly from data 4

Centroid 3 = (5, 1, 2) is taken randomly from data 5

Description:

The centroid value is randomly taken from the data that has been transformed.

Part A (4, 1, 3)

K=3, centroid 1 = (3, 1, 1), centroid 2 = (4, 1, 3), centroid 3 = (5, 1, 2)

Distance from C1 (X) = $\sqrt{(4-3)^2 + (1-1)^2 + (3-1)^2}$ = 2.24

Distance from C2 (Y) = $\sqrt{(4-4)^2 + (1-1)^2 + (3-3)^2}$ = 0

Distance from C3 (Z) = $\sqrt{(4-5)^2 + (1-1)^2 + (3-2)^2}$ = 1.41

Part B (1, 1, 1)

K=3, centroid 1 = (3, 1, 1), centroid 2 = (4, 1, 3), centroid 3 = (5, 1, 2)

Distance from C1 (X) = $\sqrt{(1-3)^2 + (1-1)^2 + (1-1)^2}$ = 2

Distance from C2 (Y) = $\sqrt{(1-4)^2 + (1-1)^2 + (1-3)^2}$ = 3.61

Distance from C3 (Z) = $\sqrt{(1-5)^2 + (1-1)^2 + (1-2)^2}$ = 4.12

Part C (3, 1, 1)

K=3, centroid 1 = (3, 1, 1), centroid 2 = (4, 1, 3), centroid 3 = (5, 1, 2)

Distance from C1 (X) = $\sqrt{(3-3)^2 + (1-1)^2 + (1-1)^2}$ = 0

Distance from C2 (Y) = $\sqrt{(3-4)^2 + (1-1)^2 + (1-3)^2}$ = 2.24

Distance from C3 (Z) = $\sqrt{(3-5)^2 + (1-1)^2 + (1-2)^2}$ = 2.24

Part D (4, 1, 3)

K=3, centroid 1 = (3, 1, 1), centroid 2 = (4, 1, 3), centroid 3 = (5, 1, 2)

Distance from C1 (X) = $\sqrt{(4-3)^2 + (1-1)^2 + (3-1)^2}$ = 2.24

Distance from C2 (Y) = $\sqrt{(4-4)^2 + (1-1)^2 + (3-3)^2}$ = 0

Distance from C3 (Z) = $\sqrt{(4-5)^2 + (1-1)^2 + (3-2)^2}$ = 1.41

Table 3.6 Results of Iteration Calculation

No	Age (X)	Type of Sexual Harassment (Y)	Causing Factor (Z)	Distance from C1	Distance from C2	Distance from C3	Group
1	4	1	3	2.24	0	1.41	2
2	1	1	1	2	3.61	4.12	1
3	3	1	1	0	2.24	2.24	1
4	4	1	3	2	0	1.41	2
5	5	1	2	2.24	1.41	0	3
6	4	2	3	2	1	1.73	2
7	4	2	1	1.41	2.24	1.73	1
8	5	1	2	2.24	1.41	0	3
9	1	1	2	2.24	3.16	4	1
10	6	1	2	3.16	2.24	1	3
11	2	1	1	1	2.83	3.16	1
12	3	3	3	2.83	2.24	3	2
13	1	1	2	2.24	3.16	4	1
14	6	1	2	3.16	2.24	1	3
15	6	1	2	3.16	2.24	1	3
16	2	1	1	1	2.83	3.16	1
17	2	1	1	1	2.83	3.16	1
18	5	4	1	3.61	3.74	3.16	3
19	3	5	3	4.47	4.12	4.58	2
20	4	1	1	1	2	1.41	1

Description:

1. If the centroid 1 is smaller then the cluster results are in group 1.

2. If the centroid 2 is smaller then the cluster results are in group 2.

3. If the centroid 3 is smaller then the cluster results are in group 3.

Old Group : {0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}

New Group : {2,1,1,2,3,2,1,3,1,3,1,2,1,3,3,3,1,1,3,3,2,1}

There is a change in the group, then proceed to the next iteration.

For Group 1 there are 9 data;

C_1 1 = (1+3+4+1+2+1+2+2+4)/9 = 2.22

C_1 2 = (1+1+2+1+1+1+1+1+1)/9 = 1.11

C_1 3 = (1+1+1+2+1+2+1+1+1)/9 = 1.22

For Group 2 there are 5 data;

$C_2\,1 = (4+4+4+3+3)/5 = 3.60$

$C_2\,2 = (1+1+2+3+5)/5 = 2.40$

$C_2\,3 = (3+3+3+3+3)/5 = 3$

For Group 3 there are 6 data;

$C_3\,1 = (5+5+6+6+6+5)/6 = 5.50$

$C_3\,2 = (1+1+1+1+1+4)/6 = 1.50$

$C_3\,3 = (2+2+2+2+2+1)/6 = 1.83$

Iterasi 2

Centroid 1 = (2.22, 1.11, 1.22)

Centroid 2 = (3.6, 2.4, 3)

Centroid 3 = (5.5, 1.5, 1.83)

Part A (4, 1, 3)

K=3, centroid 1 = (2.22, 1.11, 1.22), centroid 2 = (3.6, 2.4, 3), centroid 3 = (5.5, 1.5, 1.83)

Distance from C1 (X) = $\sqrt{(4-2.22)^2 + (1-1.11)^2 + (3-1.22)^2}$ = 2.52

Distance from C2 (Y) = $\sqrt{(4-3.6)^2 + (1-2.4)^2 + (3-3)^2}$ = 1.46

Distance from C3 (Z) = $\sqrt{(4-5.5)^2 + (1-1.5)^2 + (3-1.83)^2}$ = 1.97

Part B (1, 1, 1)

K=3, centroid 1 = (2.22, 1.11, 1.22), centroid 2 = (3.6, 2.4, 3), centroid 3 = (5.5, 1.5, 1.83)

Distance from C1 (X) = $\sqrt{(1-2.22)^2 + (1-1.11)^2 + (1-1.22)^2}$ = 1.24

Distance from C2 (Y) = $\sqrt{(1-3.6)^2 + (1-2.4)^2 + (1-3)^2}$ = 3.57

Distance from C3 (Z) = $\sqrt{(1-5.5)^2 + (1-1.5)^2 + (1-1.83)^2}$ = 4.60

Part C (3, 1, 1)

K=3, centroid 1 = (2.22, 1.11, 1.22), centroid 2 = (3.6, 2.4, 3), centroid 3 = (5.5, 1.5, 1.83)

Distance from C1 (X) = $\sqrt{(3-2.22)^2 + (1-1.11)^2 + (1-1.22)^2}$ = 0.82

Distance from C2 (Y) = $\sqrt{(3-3.6)^2 + (1-2.4)^2 + (1-3)^2}$ = 2.51

Distance from C3 (Z) = $\sqrt{(3-5.5)^2 + (1-1.5)^2 + (1-1.83)^2}$ = 2.6

Part D (4, 1, 3)

K=3, centroid 1 = (2.22, 1.11, 1.22), centroid 2 = (3.6, 2.4, 3), centroid 3 = (5.5, 1.5, 1.83)

Distance from C1 (X) = $\sqrt{(4-2.22)^2 + (1-1.11)^2 + (3-1.22)^2}$ = 3

Distance from C2 (Y) = $\sqrt{(4-3.6)^2 + (1-2.4)^2 + (3-3)^2}$ = 1.46

Distance from C3 (Z) = $\sqrt{(4-5.5)^2 + (1-1.5)^2 + (3-1.83)^2}$ = 1.97

Table 3. 7 Results of Iteration II

No	Age (X)	Type of Sexual Harassment (Y)	Causing Factor (Z)	Distance from C1	Distance from C2	Distance from C3	Group
1	4	1	3	2.52	1.46	1.97	2
2	1	1	1	1.24	3.57	4.60	1
3	3	1	1	0.82	2.51	2.68	1
4	4	1	3	3	1.46	1.97	2
5	5	1	2	2.89	2.22	0.73	3
6	4	2	3	3	0.57	1.97	2
7	4	2	1	2	2.08	1.79	3
8	5	1	2	2.89	2.22	0.73	3
9	1	1	2	1.45	3.12	4.53	1
10	6	1	2	3.86	2.95	0.73	3
11	2	1	1	0.33	2.92	3.63	1
12	3	3	3	2.71	0.85	3.14	2
13	1	1	2	1.45	3.12	4.53	1
14	6	1	2	3.86	2.95	0.73	3
15	6	1	2	3.86	2.95	0.73	3
16	2	1	1	0.33	2.92	3.63	1
17	2	1	1	0.33	2.92	3.63	1
18	5	4	1	4.02	2.92	2.68	3
19	3	5	3	4.35	2.67	4.46	2
20	4	1	1	1.80	2.47	1.79	1

From the I and II iterations results in the table above, there is a value in the old group as follows:

Old Group : {2,1,1,2,3,2,1,3,1,3,1,2,1,3,3,3,1,1,3,3,2,1}

New Group : {2,1,1,2,3,2,1,3,1,3,1,2,1,3,3,3,1,1,3,3,2,1}

Because in the 1st and 2nd iterations the cluster position does not change or there is equation, the calculation of iteration is stopped and obtain the following results:
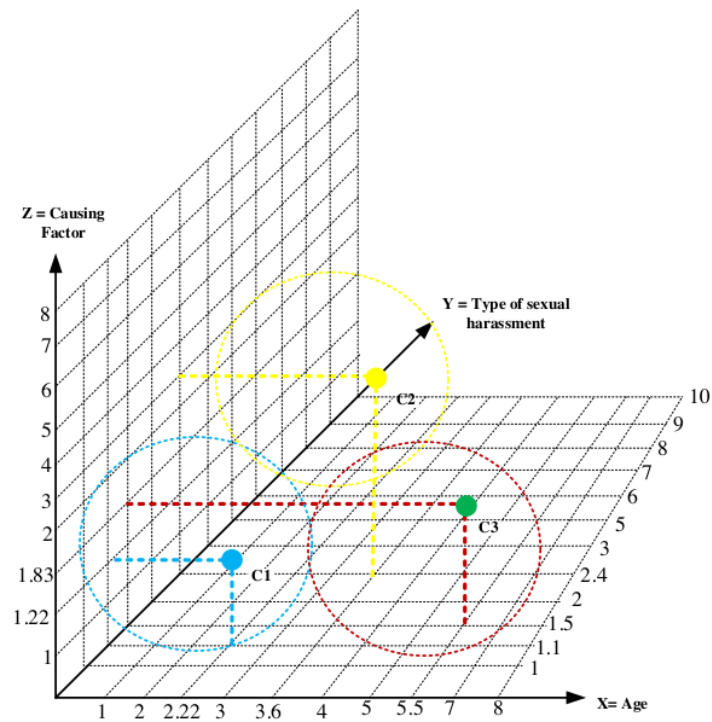
Figure 3. 1 Chart Cluster

Of the 20 data there are 3 groups, namely group 1 there are 9 data and 2 groups there are 5 data and group 3 there are 6 data. The following explanation is that of the three groups:

1. Cluster 1 There are 9 data

   2(2.22); 1(1.11); 1(1.22);

   Based on the calculations above, it is known that cluster 1 is a group on sexual harassment case data based on the Causing Factor which has a total of 9 data and is located in the group Age (X) is Age 12-16 Year, and for the Type of Sexual Harassment (Y) group Physical Harassment and Causing Factor (Z) that is widely occurring due to the Individual Factor.

2. Cluster 2 There are 5 Data

   4(3.60); 2(2.4); 3(3);

   Based on the calculations above, it is known that cluster 2 is a group on sexual harassment case data based on the Causing Factor which has a total of 5 data and is located in the group Age (X) is Age 26-35 Year, and for the Type of Sexual Harassment (Y) group Visual Harassment and Causing Factor (Z) which is widely occurring due to Environmental Factors.

3. Cluster 3 There are 6 data

6 (5.50); 2(1.50); 2(1.83);

Based on the calculations above, it is known that cluster 3 is a group on sexual harassment case data based on the Causing Factor which has a total of 6 data and is located in the group Age ($\geq$ 46 Year, and for the Type of Sexual Harassment (Y) group Visual Harassment and Causing Factor (Z) which is widely occurring due to Family Factor.

3.2. Flowchart Design

The design of the process that will be built in the grouping of sexual harassment cases data based on the causes using clustering methods will have the following picture of flowchart:

```
                    ┌──────────┐
                    │  Start   │
                    └──────────┘
                         │
                    ┌─────────────┐
                   ╱ Data Input    ╱
                  ╱(which has been ╱
                 ╱  transformed)  ╱
                 └───────────────┘
                         │
                  ┌─────────────┐
                 ╱ Determine the ╱
                ╱  number of    ╱
               ╱   centroid    ╱
               └──────────────┘
                       │
              ┌─────────────────┐
              │  Count Cluster  │◄───────┐
              └─────────────────┘        │
                       │                 │
          ┌──────────────────────┐       │
          │ Calculate the distance│       │
          │ of objects based on   │       │
          │      centroid         │       │
          └──────────────────────┘       │
                       │                 │
                ┌──────────────┐         │
               ╱ Group based on ╱         │
              ╱  the nearest   ╱          │
             ╱    distance    ╱           │
             └───────────────┘            │
                     │                    │
                 ◇──────────◇      Yes     │
                ╱ Is there any ╲───────────┘
               ╱ object moving ╲
                ╲    group?    ╱
                 ◇──────────◇
                     │ No
               ┌──────────────┐
              ╱ Analysis Results╱
              └────────────────┘
                     │
               ┌──────────┐
               │ Finished │
               └──────────┘
```
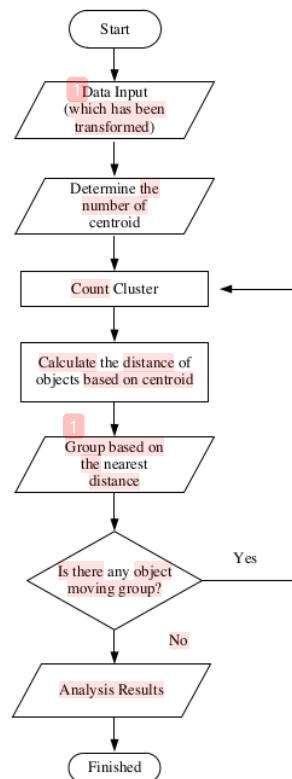
Figure 3. 2 Flowchart for the system to be designed

The description of the above picture is as follows:

1. Starting the system,

2. Performing data mining process.

3. Cluster objects into clusters based on the nearest distance, then group/group.

4. Ensure if there are objects that move if there is no further process and if there is an object that moves then do the data mining process again.

5. Displaying results from k-means cluster. Once the result of the k-means cluster appears then the process is completed.

4. CONCLUSION

From the testing conducted using 20 sexual harassment cases data there are 3 groups, namely group 1 there are 9 data and 2 groups there are 5 data and group 3 there are 6 data and it can be known that in cluster 1 is a group in the case data on sexual harassment based on the factors that are many causes with a total of 9 data and located in the age group (X) is 12-16 years, and for the group Sexual Harassment (Y) namely Physical Harassment and causing factor (Z) that is a lot due to individual factors.

REFERENCES

Adawiah, N., Suliatiyowati, N., & Jajuli, M. (2021). Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma K-Means. Generation Journal, 5(2), 69–80.

Dewi, ratna. (2023). Aplikasi Matlab untuk Simulasi Pengolahan Sinyal (agus Prijono, Ed.; 1st ed.). Zahir Publishing.

Hasan, Z., Novriyanti, F., Ramadhani Putri, A. T., & Al Munawwaroh, R. (2023). FAKTOR PENYEBAB TERJADINYA PELECEHAN DAN KEKERASAN SEKSUALTERHADAP ANAK DI BAWAH UMUR DI KOTA BANDAR LAMPUNG. JHM, 4(2), 84–91.

Laksana, F., Hidayat, R., & Dan Dewi, Y. (2023). Pengelompokan jumlah kekerasan pada anak menggunakan K-Means 123. Computational Intelligent Journal, 5(2), 123–135.

Laksana, F., Hidayat, R., & Dewi, Y. (2023). Pengelompokan Jumlah Kekerasan Terhadap Anak Berdasarkan Jenis Pelecehan Seksual Di Kabupaten Banyumas Menggunakan Penerapan Algoritma K-Means. INDEXIA : Informatic and Computational Intelligent Journal, 5(2), 123–135.

Matheos Sariole, F., & Hakim, L. (2024). Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang. Jurnal Sains Dan Teknologi, 5(3), 846–854. https://doi.org/10.55338/saintek.v5i1.2709

Prasetyo. (2012). Data Mining Konsep dan Aplikasi menggunakan MATLAB (Nikodemus WK, Ed.; 1st ed., Vol. 1). ANDI OFFSET.

Relita, B., Sundari, Y., & Maulita Yani. (2016). Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritm Dengan Perbandingan Jarak Kedekatan Manhattan City Dan

Eu oritma K-Means an Euclidean. MEANS (Media Informasi Analisa Dan Sistem), 1(2548–6985), 47–53.

Relita Buaton, Zarlis, M., Efendi, S., & Yasin, V. (2019). DATA MINING TIME SERIES (1st ed., Vol. 1). Wade Group.

Sari, K. indriyanti purnama, Farida, L. nur, Khayati, N., Maidaliza, Asmaret, D., Pramana, I., Meinarisa, Melvia Girsang, B., Alfianto, A. G., & Suminah. (2022). KEKERASAN SEKSUAL (Agustiawan, Ed.; 1st ed., Vol. 1). CV. Media Sains Indonesia. www.medsan.co.id

Sundari, M. A., Pane, R., & Rohani, R. (2023). Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC. Building of Informatics, Technology and Science (BITS), 5(1). https://doi.org/10.47065/bits.v5i1.3499

Wandana, J., Defit, S., & Sumijan, S. (2020). Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means. Jurnal Informasi Dan Teknologi. https://doi.org/10.37034/jidt.v2i4.73

Application of Clustering Methods On Sexual Harassment Cases

| 8 | seaninstitute.org
Internet Source | 1% |

| 9 | Tetti Solehati, Azzahra Salsabila, Ratu Nurafni, Valencia Trie Millenika et al. "Local Culture in Indonesia As Risk of Child Sexual Abuse: A Systematic Review", Jurnal Aisyah : Jurnal Ilmu Kesehatan, 2021
Publication | 1% |

| 10 | ejournal.sisfokomtek.org
Internet Source | 1% |

| 11 | ejournal.umm.ac.id
Internet Source | 1% |

| 12 | journal.stmikjayakarta.ac.id
Internet Source | 1% |

| 13 | documents1.worldbank.org
Internet Source | 1% |

Application of Clustering Methods On Sexual Harassment Cases

FINAL GRADE

/0

GENERAL COMMENTS