Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 3, No. 1, Januari 2025, Hal 1 - 11
ISSN: **2963-0169** (Online - Elektronik)
https://journal.grahamitra.id/index.php/jurikti

Classification of Types of Crimes Against Human Physique Using the K-Means Clustering Method

Rizkah Fadillah[1,*], Fiqri Hidayat Rangkuti[2], Anggi Jaya Maulana Siregar[2], Rizky Ananda[2], Muhammad Syahrizal[3,*]

[1]Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia
[2]Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia
[3]Prodi Teknologi Rekayasa Komputer Grafis, Politeknik Cendana, Medan, Indonesia
Jl. Sisingamangaraja No.338, Siti Rejo I, Kec. Medan Kota, Kota Medan, Sumatera Utara, Indonesia
Jl. Williem Iskandar No.261, Sidorejo Hilir, Kec. Medan Tembung, Kota Medan, Sumatera Utara, Indonesia
Email: [1]rizkahfadillah52@gmail.com [2]panditorangkuti@email.com, [3]Anggijaya003@email.com, [4]dhitisyhla@email.com,
Email Penulis Korespondensi: [5,*]m.syahrizal156@gmail.com

Abstract−*Human physical crimes are unlawful acts and prohibited by the rule of law, which can harm or damage the body of others. This study aims to examine the number of groupings of types of crimes against human physique in 2019 to 2020 in all areas of East Nusa Tenggara province. To do this, we use the K-Means Clustering method to group the types of physical crimes against humans. The data used came from the Central Statistics Agency of East Nusa Tenggara province. The K-Means method is one of the non-hierarchical data clustering methods that seeks to partition data into the form of one or more clusters/groups. After the application of the K-Means algorithm in the grouping of types of crimes against human bodies in 2019 to 2020 in the East Nusa Tenggara region, there are 3 centroids, C1 for areas with low crimes, C2 for areas with moderate crimes and C3 for areas with high crimes. The initial centroid value is determined randomly and then for the next centroid is adjusted to the result of the calculation of the closest distance (minimum). The final results obtained are areas with low crime totaling 13 regions, namely East Sumba, Lembata, Sikka, Ende, Ngada, Manggarai, Rote Ndao, West Manggarai, Central Sumba, Southwest Sumba, Nagekeo, East Manggarai, and Sabu Raijua. There are 7 areas with moderate crimes, namely West Sumba, Kupang, South Central Timor, North Central Timor, Belu, Alor, and East Flores. As for the area with high crime, there is 1 area, namely Kupang City.*

Keywords: *East Nusa Tenggara; Crime; Data Mining; Clustering; K-Means.*

1. INTRODUCTION

This study aims to examine the number of groupings of types of crimes against human physique in 2019 to 2020 in all areas of East Nusa Tenggara province. Human physical crimes are unlawful acts and are prohibited by the rule of law, which can harm or damage the body of others. The definition of a human physical crime can be known through the term *"strafbaar feit"* which comes from the Dutch language, which means "an unlawful act committed intentionally or unintentionally by a person whose actions can be held accountable and which are made by a person who can be held accountable.

Crimes against human bodies in East Nusa Tenggara province are complex problems and require serious attention. Various forms of violence, such as violence against children, human trafficking, and other crimes, have become part of people's lives in the province. The factors that cause physical violence against children in Kupang City, such as lack of legal protection and lack of community participation, have become a major concern in research and countermeasures. This research will help those responsible for managing human physical crimes in the province of East Nusa Tenggara. In addition, this research will also help the responsible parties in developing strategies for the prevention and control of crimes against humans in East Nusa Tenggara province. By knowing the number of groupings of types of physical crimes against humans, we can improve the system of error management and prevention of physical crimes against humans in the province of East Nusa Tenggara.

In this study, we will conduct statistical analysis and data collection from the Central Statistics Agency of East Nusa Tenggara province. We will also conduct a comparison and comparison of data from year to year. By conducting an in-depth analysis, we will find trends and similarities in the number of types of crimes against physical humans in the province of East Nusa Tenggara. To do this, a technique called data mining is needed. Data mining is a process that uses *machine learning*, statistical engineering, mathematics, and artificial intelligence to identify valuable information from various large databases. Data mining is a systematic process of extracting useful information from a set of data consisting of unknown knowledge by manual means. [1] Data mining allows us to identify hidden patterns in data, so that it can provide deeper insight and understanding of a phenomenon or problem. In simple terms, data mining can be said to be the process of filtering or "mining" knowledge from a large amount of data. Another term for data mining is Knowledge Discovery in Database (KDD). Although data mining itself is part of the KDD process stages. Data mining is an analytical process designed to explore large amounts of data for valuable, consistent, and hidden knowledge. Data mining is a series of processes that extract previously unknown knowledge from a set of data. Data mining is used to improve previous techniques so that it can handle various problems that are often encountered. [2][3][4][5][6] The methods used in data mining include several main classifications, namely *Classification, Clustering, Association, Regression, and Forecasting.* Based on several groupings in data mining, this research uses the K-Means clustering method in classifying crimes that occur in NTT.

There are several previous studies or previous studies that have been carried out so that they can be a reference in completing this research. The research conducted in 2020 by Wahyu Saputro, and several other research colleagues,

where they conducted research on corruption crimes by utilizing the K-Means algorithm for classification, they conducted research in jurisdictions in Indonesia. The final result obtained from the research they have conducted is that they formed 3 clusters which show that the DBI value obtained is good, which is 0.113 so that it can be concluded that as many as 5 areas are prone to corruption on the island of Samatera, 3 areas are prone to corruption on the island of Java and Sulawesi and one area is prone to corruption on the island of Kalimantan. [7]

The research conducted in 2023 by Ikhlasul Amal and Raissa Amanda Putri, where they conducted research on drug addicts using the K-Means algorithm for classification, they conducted research in the North Sumatra region, precisely in the city of Medan. The final result obtained from the study is that they formed 4 clusters, so it can be concluded that there are 2 sub-districts that are included in the area with the highest level of drug addicts, 7 sub-districts that are included in the area with a high level of addicts, 7 sub-districts that are included in the area with a low level of drug addicts, and 5 sub-districts that are included in the area with a very low level of addicts. [8]

Based on research conducted in 2022 by Nisriina Nuur Hasanah and Agus Sidiq Purnomo, they conducted research on the implementation of data mining for library book grouping using the K-Means algorithm method, the research was conducted at the LPP Yogyakarta Polytechnic Library. The final result obtained from the study is to form 3 clusters, so that the results obtained are 3 books that are most in demand, 9 books that are quite in demand, and 18 books that are in little interest. [9]

In 2021, a study was conducted by Lalu Ganda Rady Putra and Anthony Anggrawan, they conducted a study on social assistance recipients who deserve to receive assistance and those who are less eligible to receive assistance using the K-Means algorithm method for grouping, the research was conducted in Ampenan district, West Nusa Tenggara province. The final result obtained from the study is that they form 2 clusters, so it can be concluded that out of 257 data, there are 196 data with the status of social assistance recipients on target and 61 data with the status of social assistance recipients who are not on target. [10]

2. RESEARCH METHODOLOGY

2.1 Research Stages

At this stage, it explains the overall description of the research method, the steps of this stage can be seen in figure 1.

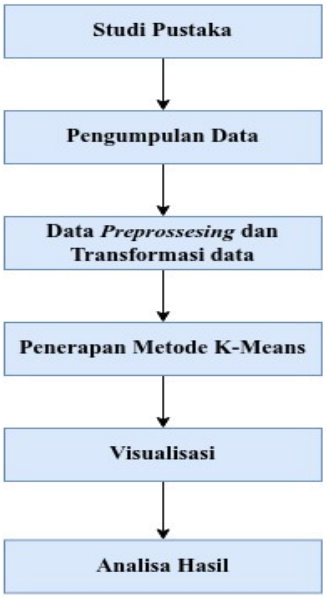


Figure 1. Research Stages

a. Study book
   Literature study is a method of collecting information and data with the help of various kinds of literature in the library such as documents, books, and research journals.[11]
b. Data collection
   At this stage, data was collected from human physical crime data in NTT in 2019-2020 obtained from the Central Statistics Agency (BPS) and the National Crime Information Center (Pusiknas) of the National Police Criminal Investigation Department.
c. Data *Preprocessing* and Data Transformation
   At this stage, value transformation or weighting of the original data is carried out, weighting the data from nominal data to numerical data to make it easier to calculate the K-Means algorithm.
d. Application of the K-Means Method

Applying the K-Means algorithm to classify the types of crimes against human physicality. With the aim of increasing efficiency and accuracy in the analysis of data on physical crimes against humans.

e. Visualization
The next stage of this research is the visualization of data that has been weighted by the application of the K-Means algorithm method. The data visualized is data on the Types of Crimes Against Human Physique in 2019 and 2020.

f. Analyze the results
In this study, the last stage is the analysis of results. The analysis of the results was carried out to determine the success and conclusions of the visualization that had been implemented.

2.2 Clustering

Clustering is a process flow where a set of information subjects is put into a set of sections called clusters. Icons or subjects in clusters have similar characters to each other and are different from other clusters. [12] Clustering is a technique that is often used to group data into separate groups. The main goal is to separate and group data that have similarities or similarities in nature or characteristics among the data present in the dataset. The clustering process is used to divide data into classes or clusters based on their degree of similarity. Clustering in data mining is grouping a set of data or objects into clusters (groups) so that all data in a cluster contains data that is as similar as possible to objects in other clusters. Two methods of grouping are known: hierarchical grouping and partitioning. Clustering is also defined as the process of grouping the same data into different groups, or rather partitions of a set into subsets, so that the data in each subset has a useful meaning. The way clustering works has several processes of grouping data into several clusters, so that the data in the cluster has the greatest similarity which also makes it possible to determine and retrieve data between different clusters that show minimal similarities, can also be used to determine in identifying cluster data groups that result from grouping small items based on their similarities. [13] [14] [15][16].

2.3 Algoritma K-Means

The k-means method is a clustering method where the cluster center uses the average value of the data in each cluster formed. K-Means is a non-hierarchical data clustering method that seeks to partition existing data into one or more clusters/groups. This method partitions data into clusters or groups so that data that have the same characteristics are grouped into the same cluster. K-Means is a data analysis technique or known as the data mining technique to apply the unsupervised data modeling process and is one of the techniques to group each data in the form of partitions. In the K-Means method, the data is grouped into several groups where each group has similar or same characteristics as the others but with other groups have different characteristics. K-Means is a non-hierarchical method, in the K-Means process stage the cluster center is randomly selected from a set of component data in the data population and marks the component to one of the cluster centers that has been defined depending on the minimum distance between the components and each cluster. Some of the advantages of the K-Means algorithm include being easy to implement, having a high level of convergence, and producing denser clusters when compared to hierarchical methods. The calculation process in the K-Means algorithm has a flowchart flow shown in Figure 2. [17][18][19][20][21][22].

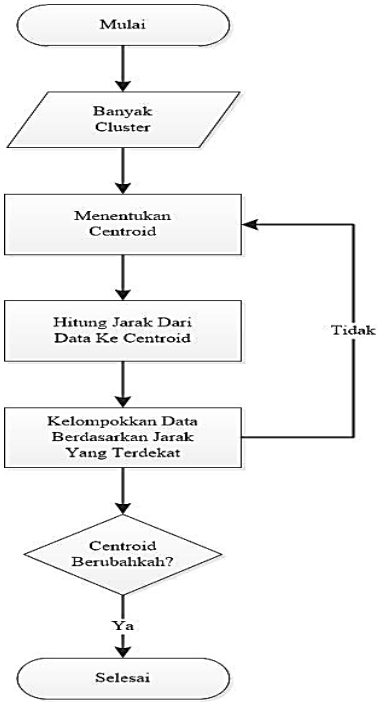


Figure 2. K-Means Algorithm Flowchart

The following are the research steps carried out for sampling so that the data used can be processed, including: [23]

a. Specify K as the number of clusters to be formed
b. Determine the initial centroid.
   Formula of the equation (1):

$$Ci = \frac{1}{M}\sum_{j=1}^{M} X_j \quad (1)$$

c. Calculate the distance of each input data for each centroid using the Euclidean Distance formula until the closest distance from each data to the centroid is found.
   Formula equation (2)

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2} \quad (2)$$

   Information:
   D = Distance
   X= Data
   Y= Centroid
d. Classify data based on their proximity to centroids.
e. Recount the cluster center with the current cluster members. The cluster center is the average value of all the object data in a given cluster.
f. Count each object again using the new cluster center. If the cluster center does not change anymore then the clustering process is complete. Alternatively, go back to step number 3 until the center of the cluster does not change again.

3. RESULTS AND DISCUSSION

3.1 Data collection

In this study, we collected data from the Central Statistics Agency of East Nusa Tenggara province. The data we used in this study is the number of groupings of types of crimes against human physicality, in 2019-2020.

Table 1. Number of Classifications of Types of Crimes Against Human Physicality, 2019-2020

Region	Number of Classifications of Types of Crimes Against Human Physique									
	Murder		Rape		Normal/Mild Persecution		Severe Persecution		Abduction	
	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
West Sumba	7	3	2	7	116	133	2	4	1	2
East Sumba	2	2	1	6	49	43	0	0	1	1
Kupang	2	4		5	168	165	0	0	0	0
South Central Timor	6	6	6	3	154	156	3	2	0	0
North Central Timor	2	3	0	1	84	177	0	2	0	0
Belu	4	3	1	4	143	98	0	1	0	0
Alor	3	2	8	7	161	220	0	0	0	2
Lembata	0	2	0	0	10	42	2	0	0	0
East Flores	1	3	3	1	99	135	2	0	0	0
Sikka	2	1	1	5	21	38	0	1	0	0
End	1	1	1	2	36	82	0	1	0	0
Ngada	0	0	4	1	53	80	1	1	0	2
Manggarai	1	3	0	0	81	83	0	0	0	2
Rote Ndao	0	0	3	6	19	31	0	1	0	0
West Manggarai	0	0	0	7	16	36	0	0	0	0
Central Sumba	0	3	0	7	0	133	0	4	0	2
Southwest Sumba	0	8	0	7	0	12	0	13	0	1
Nagekeo	0	1	0	0	0	25	0	0	0	0
East Manggarai	0	1	0	2	0	38	0	0	0	0
Sabu Raijua	0	1	0	0	0	7	0	0	0	0
Kupang City	32	2	5	7	409	376	0	0	1	0
East Nusa Tenggara	63	49	37	78	1619	2110	10	30	3	12

3.2 Application of K-Means Clustering Method

Iteration 1

a. Number of clusters C=3 (C1, C2, and C3)
b. Early cluster centroid
   The initial centroid center was used for the grouping or cluster used in this study. The initial centroid center is determined randomly, so the data used is free to any data. For the early centroid center, it can be seen in the following table 2.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 3, No. 1, Januari 2025, Hal 1 - 11
ISSN: **2963-0169** (Online - Elektronik)
https://journal.grahamitra.id/index.php/jurikti

Table 2. Early Centroid Center(Early Centroid)

Region	Number of Classifications of Types of Crimes Against Human Physique										CN
	Murder		Rape		Normal/Mild Persecution		Severe Persecution		Abduction		
	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	
East Sumba	2	2	1	6	49	43	0	0	1	1	C1
Alor	3	2	8	7	161	220	0	0	0	2	C2
Kupang City	32	2	5	7	409	376	0	0	1	0	C3

$Ci = \frac{1}{M}\sum_{j=1}^{M} X_j$

c. Calculate the distance of each centroid's input data using the Euclidean Distance formula $D(x,y) = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2}$

Data 1:

$$C1 = \sqrt{\begin{matrix}(7-2)^2 + (3-2)^2 + (2-1)^2 + (7-6)^2 + (116-49)^2 + \\ (133-43)^2 + (2-0)^2 + (4-0)^2 + (1-1)^2 + (2-1)^2 = 112{,}4188596\end{matrix}}$$

$$C2 = \sqrt{\begin{matrix}(7-3)^2 + (3-2)^2 + (2-8)^2 + (7-7)^2 + (116-161)^2 + (133-220)^2 + \\ (2-0)^2 + (4-0)^2 + (1-0)^2 + (2-2)^2 = 98{,}32598843\end{matrix}}$$

$$C3 = \sqrt{\begin{matrix}(7-32)^2 + (3-2)^2 + (2-5)^2 + (7-7)^2 + (116-409)^2 + (133-376)^2 + \\ (2-0)^2 + (4-0)^2 + (1-1)^2 + (2-0)^2 = 381{,}5193311\end{matrix}}$$

Data 2:

$$C1 = \sqrt{\begin{matrix}(2-2)^2 + (2-2)^2 + (1-1)^2 + (6-6)^2 + (49-49)^2 + \\ (43-43)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2 + (1-1)^2 = 0\end{matrix}}$$

$$C2 = \sqrt{\begin{matrix}(2-3)^2 + (2-2)^2 + (1-8)^2 + (6-7)^2 + (49-161)^2 + (43-220)^2 + \\ (0-0)^2 + (0-0)^2 + (1-0)^2 + (1-2)^2 = 209{,}5853048\end{matrix}}$$

$$C3 = \sqrt{\begin{matrix}(2-32)^2 + (2-2)^2 + (1-5)^2 + (6-7)^2 + (49-409)^2 + (43-376)^2 + \\ (0-0)^2 + (0-0)^2 + (1-1)^2 + (1-0)^2 = 491{,}3318634\end{matrix}}$$

Data 3:

$$C1 = \sqrt{\begin{matrix}(2-2)^2 + (4-2)^2 + (2-1)^2 + (5-6)^2 + (168-49)^2 + \\ (165-43)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 = 170{,}449406\end{matrix}}$$

$$C2 = \sqrt{\begin{matrix}(2-3)^2 + (4-2)^2 + (2-8)^2 + (5-7)^2 + (168-161)^2 + (165-220)^2 + \\ (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-2)^2 = 55{,}88380803\end{matrix}}$$

$$C3 = \sqrt{\begin{matrix}(2-32)^2 + (4-2)^2 + (2-5)^2 + (5-7)^2 + (168-409)^2 + (165-376)^2 + \\ (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 = 321{,}7452408\end{matrix}}$$

Perform the above calculation steps until the 21st data. The following is the closest distance based on the initial centroid can be seen in the following table 3.

Table 3. Minimum Distance of Iteration 1

Region	C1	C2	C3	Nearby	Cluster
West Sumba	112,4188596	98,32598843	381,5193311	98,32598843	C2
East Sumba	0	209,5853048	491,3318634	0	C1
Kupang	170,449406	55,88380803	321,7452408	55,88380803	C2
South Central Timor	154,5153714	64,86139067	337,8579583	64,86139067	C2
North Central Timor	138,6145735	88,81441324	382,3519321	88,81441324	C2
Belu	108,9632966	123,5839795	385,8134264	108,9632966	C1
Alor	209,5853048	0	294,4401467	0	C2
Lembata	39,61060464	233,6985237	521,4019563	39,61060464	C1
East Flores	104,885652	105,5604092	393,9390816	104,885652	C1

Region	C1	C2	C3	Nearby	Cluster
Sikka	28,53068524	229,7476877	515,4716287	28,53068524	C1
End	41,36423576	186,4215653	475,9936974	41,36423576	C1
Ngada	37,82856064	177,0056496	464,1379105	37,82856064	C1
Manggarai	51,623638	159,0188668	440,9920634	51,623638	C1
Rote Ndao	32,54228019	236,4931289	521,6895629	32,54228019	C1
West Manggarai	33,91164992	234,4397577	520,6755228	33,91164992	C1
Central Sumba	102,5962962	183,2484652	476,8658092	102,5962962	C1
Southwest Sumba	59,78294071	263,5602398	548,6647428	59,78294071	C1
Nagekeo	52,62128847	253,1264506	539,9833331	52,62128847	C1
East Manggarai	49,49747468	243,2036184	531,602295	49,49747468	C1
Sabu Raijua	61,16371473	267,2395929	551,8532414	61,16371473	C1
Kupang City	491,3318634	294,4401467	0	0	C3

d. The data has been classified and grouped into centroids based on the closest distance (minimum distance) of the clusters that have been calculated.
e. After clustering clusters based on the closest distance, then perform iteration calculations using the new centroid value. The new centroid value is adjusted to the location of the nearest cluster and then do the equation one.

$$Ci = \frac{1}{M}\sum_{j=1}^{M} X_j$$

Murder:

$$C1(2019) = \frac{1}{15}(2+4+0+1+2+1+0+1+0+0+0+0+0+0+0) = 0{,}733333333$$

$$C1(2020) = \frac{1}{15}(2+3+2+3+1+1+0+3+0+0+3+8+1+1+1) = 1{,}933333333$$

Rape:

$$C1(2019) = \frac{1}{15}(1+1+0+3+1+1+4+0+3+0+0+0+0+0+0) = 0{,}933333333$$

$$C1(2020) = \frac{1}{15}(6+4+0+1+5+2+1+0+6+7+7+7+0+2+0) = 3{,}2$$

Normal/Mild Persecution:

$$C1(2019) = \frac{1}{15}(49+143+10+99+21+36+53+81+19+16+0+0+0+0+0) = 35{,}13333333$$

$$C1(2020) = \frac{1}{15}(43+98+42+135+38+82+80+83+31+36+133+12+25+38+7) = 58{,}86666667$$

Severe Persecution:

$$C1(2019) = \frac{1}{15}(0+0+2+2+0+0+1+0+0+0+0+0+0+0+0) = 0{,}333333333$$

$$C1(2020) = \frac{1}{15}(0+1+0+0+1+1+1+0+1+0+4+13+0+0+0) = 1{,}466666667$$

Abduction:

$$C1(2019) = \frac{1}{15}(1+0+0+0+0+0+0+0+0+0+0+0+0+0+0) = 0{,}066666667$$

$$C1(2020) = \frac{1}{15}(1+0+0+0+0+0+2+2+0+0+2+1+0+0+0) = 0{,}533333333$$

Perform the above steps to find the C2 and C3 centroids. The following is a table of 4 initial Centroids that have been calculated based on equation (1).

Table 4. Centroid Early Iteration 2

CN	Number of Classifications of Types of Crimes Against Human Physique									
	Murder		Rape		Normal/Mild Persecution		Severe Persecution		Abduction	
	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
C1	0,7333333 33	1,93333333 3	0,93333333 3	3,2	35,13333 333	58,8666666 7	0,333333 333	1,466666 667	0,06666666 7	0,53333333 3
C2	4	3,6	3,6	4,6	136,6	170,2	1	1,6	0,2	0,8

CN	Number of Classifications of Types of Crimes Against Human Physique									
	Murder		Rape		Normal/Mild Persecution		Severe Persecution		Abduction	
	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020
C3	32	2	5	7	409	376	0	0	1	0

f. Recalculate each object using the new cluster center. If the cluster center does not change anymore then the clustering process is complete. Alternatively, go back to step number 3 until the center of the cluster does not change again.

After the calculation is carried out, the iteration calculation process stops at the calculation of the 5th iteration, where the results of the 4th and 5th iterations are the same (the location of the centroid or the nearest distance does not change or is fixed). The following can be seen in table 5, the closest distance of the 4th iteration and the closest distance of the 5th iteration can be seen in table 6.

Table 5. Closest Distance of Iteration 4

Region	C1	C2	C3	Nearby	Cluster
West Sumba	131,2380682	17,31193591	381,5193311	17,31193591	C2
East Sumba	32,4771287	128,4112656	491,3318634	32,4771287	C1
Kupang	191,5222282	46,45646484	321,7452408	46,45646484	C2
South Central Timor	175,1583014	30,52381243	337,8579583	30,52381243	C2
North Central Timor	146,0688099	52,24416833	382,3519321	52,24416833	C2
Belu	135,8943115	50,94068242	385,8134264	50,94068242	C2
Alor	225,0639107	82,35565023	294,4401467	82,35565023	C2
Lembata	9,836863773	155,6388548	521,4019563	9,836863773	C1
East Flores	120,1829046	29,12821871	393,9390816	29,12821871	C2
Sikka	10,36487123	150,4109475	515,4716287	10,36487123	C1
End	39,65808731	110,2381655	475,9936974	39,65808731	C1
Ngada	48,90395917	98,30286428	464,1379105	48,90395917	C1
Manggarai	73,45132553	77,35924718	440,9920634	73,45132553	C1
Rote Ndao	16,80567034	156,9122466	521,6895629	16,80567034	C1
West Manggarai	12,09396084	155,4411565	520,6755228	12,09396084	C1
Central Sumba	87,54292598	126,5699535	476,8658092	87,54292598	C1
Southwest Sumba	41,36541094	184,1988141	548,6647428	41,36541094	C1
Nagekeo	28,31249234	174,5349625	539,9833331	28,31249234	C1
East Manggarai	19,5387791	165,7670146	531,602295	19,5387791	C1
Sabu Raijua	43,8930202	187,451202	551,8532414	43,8930202	C1
Kupang City	512,612359	366,1109164	0	0	C3

Table 6. Closest Distance of Iteration 5

Region	C1	C2	C3	Nearby	Cluster
West Sumba	125,7396813	27,77974596	381,5193311	27,77974596	C2
East Sumba	28,20539627	139,4325234	491,3318634	28,20539627	C1
Kupang	185,952243	37,37264087	321,7452408	37,37264087	C2
South Central Timor	169,6011332	22,48491558	337,8579583	22,48491558	C2
North Central Timor	141,3963542	53,21116961	382,3519321	53,21116961	C2
Belu	130,3035146	57,93962375	385,8134264	57,93962375	C2
Alor	219,8167203	71,52621897	294,4401467	71,52621897	C2
Lembata	14,96939679	166,4310239	521,4019563	14,96939679	C1
East Flores	114,8219615	38,87893297	393,9390816	38,87893297	C2
Sikka	12,31033625	161,3186731	515,4716287	12,31033625	C1
End	35,00887461	120,7234147	475,9936974	35,00887461	C1
Ngada	43,44853447	109,1034895	464,1379105	43,44853447	C1
Manggarai	67,80122357	88,41056821	440,9920634	67,80122357	C1
Rote Ndao	19,64507557	167,8447412	521,6895629	19,64507557	C1
West Manggarai	15,865	166,3206885	520,6755228	15,865	C1
Central Sumba	85,98393272	134,0959145	476,8658092	85,98393272	C1
Southwest Sumba	45,90967463	195,0919197	548,6647428	45,90967463	C1
Nagekeo	33,48142645	185,3972569	539,9833331	33,48142645	C1
East Manggarai	25,11523737	176,5017199	531,602295	25,11523737	C1
Sabu Raijua	48,42526115	198,421269	551,8532414	48,42526115	C1
Kupang City	507,084508	355,5002512	0	0	C3

After the calculation process is completed, it can be grouped into areas with low crime (C1), areas with medium crime (C2) and areas with high crime (C3) based on the application of the K-Means algorithm. The following can be seen in table 7, table 8, and table 9 for the grouping of types of crimes against human physicality.

Jurnal Ilmu Komputer, Teknologi Dan Informasi
Vol. 3, No. 1, Januari 2025, Hal 1 - 11
ISSN: **2963-0169** (Online - Elektronik)
https://journal.grahamitra.id/index.php/jurikti

Table 7. Areas with Low Crime (C1)

Region
East Sumba
Lembata
Sikka
Ende
Ngada
Manggarai
Rote Ndao
West Manggarai
Central Sumba
Southwest Sumba
Nagekeo
East Manggarai
Sabu Raijua

Table 8. Regions with Moderate Crimes (C2)

Region
West Sumba
Kupang
South Central Timor
North Central Timor
Belu
Alor
East Flores

Table 9. Areas with High Crime (C3)

Region
Kupang City

Based on the data in table 7, it can be seen that the group of areas with low crime (C2) totals 13 regions, the group of areas with medium crime (C2) amounts to 7 regions, and the group of areas with high crime (C3) amounts to 1 region. So that from this grouping, the government or the authorities must pay more attention to areas based on their crime groups to reduce cases of crimes against human beings.

3.3 Visualization

In this study, we used rapidminer to visualize the K-Means clustering method. RapidMiner is a data science platform that allows users to perform data mining, text mining, and predictive analysis with more than 500 data mining operators. RapidMiner allows users to design and run analytical pipelines automatically, has powerful data processing capabilities, is available in a free version for students and early users, is compatible with various operating systems, and can be used for prototyping, application development, and research. The following can be seen in figure 3 of the operator input process used in determining the number of clusters to be formed.

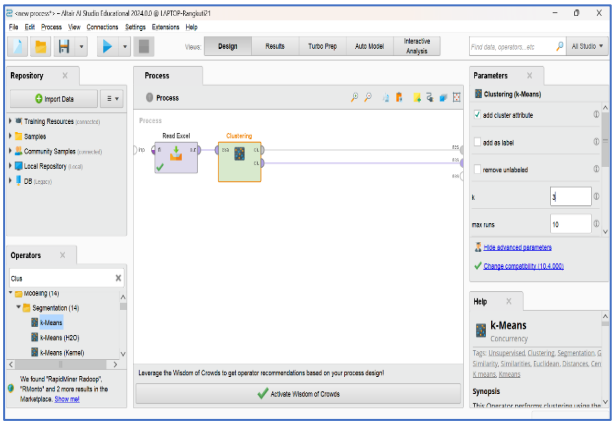


Figure 3. RapidMiner Operator Input Proser

Based on figure 3, it can be seen in the process using the Read Excel and K-Means operators. Read Excel aims to read the data to be processed and K-Means aims to classify the data, in the K-Means operator determines the K parameters, in this study the K parameters are determined as much as K=3. Once all the operators are well connected, then run the process and the results can be seen in table 10.

Table 10. K-Means Classification Results with RapidMiner

Region	Number of Classifications of Types of Crimes Against Human Physique										Cluster
	Murder		Rape		Normal/Mild Persecution		Severe Persecution		Abduction		
	2019	2020	2019	2020	2019	2020	2019	2020	2019	2020	
West Sumba	7	3	2	7	116	133	2	4	1	2	cluster_0
East Sumba	2	2	1	6	49	43	0	0	1	1	cluster_2
Kupang	2	4	2	5	168	165	0	0	0	0	cluster_0
South Central Timor	6	6	6	3	154	156	3	2	0	0	cluster_0
North Central Timor	2	3	0	1	84	177	0	2	0	0	cluster_0
Belu	4	3	1	4	143	98	0	1	0	0	cluster_0
Alor	3	2	8	7	161	220	0	0	0	2	cluster_0
Lembata	0	2	0	0	10	42	2	0	0	0	cluster_2
East Flores	1	3	3	1	99	135	2	0	0	0	cluster_0
Sikka	2	1	1	5	21	38	0	1	0	0	cluster_2
End	1	1	1	2	36	82	0	1	0	0	cluster_2
Ngada	0	0	4	1	53	80	1	1	0	2	cluster_2
Manggarai	1	3	0	0	81	83	0	0	0	2	cluster_2
Rote Ndao	0	0	3	6	19	31	0	1	0	0	cluster_2
West Manggarai	0	0	0	7	16	36	0	0	0	0	cluster_2
Central Sumba	0	3	0	7	0	133	0	4	0	2	cluster_2
Southwest Sumba	0	8	0	7	0	12	0	13	0	1	cluster_2
Nagekeo	0	1	0	0	0	25	0	0	0	0	cluster_2
East Manggarai	0	1	0	2	0	38	0	0	0	0	cluster_2
Sabu Raijua	0	1	0	0	0	7	0	0	0	0	cluster_2
Kupang City	32	2	5	7	409	376	0	0	1	0	cluster_1

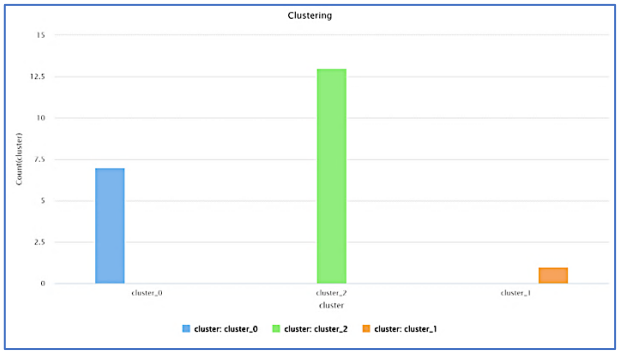


Figure 4. Results of Bar Chart Visualization

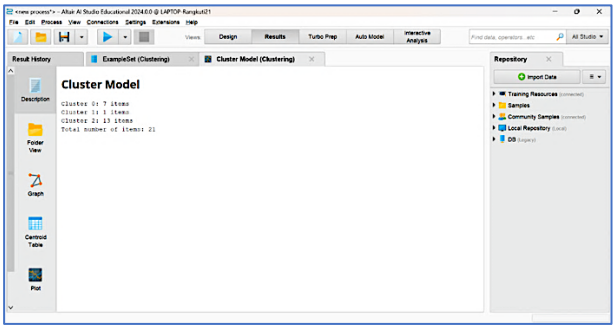


Figure 5. Results of Visualization of K-Means with RapidMiner

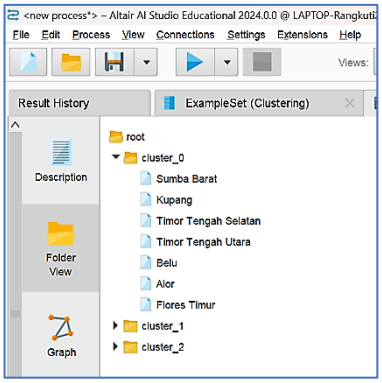
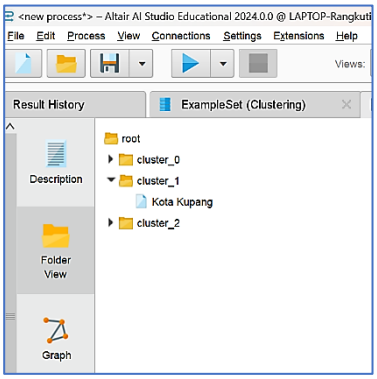
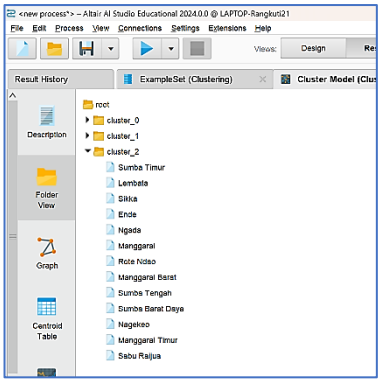


Figure 6. Results Cluster_0



Gambar 7. Hasil Cluster_1



Gambar 8. Hasil Cluster_2

3.4 Result Analysis

The last stage is the analysis of the results, where in this study we use the K-Means method for the grouping of types of crimes against the human body in the East Nusa Tenggara region by performing manual calculations using the Euclidean equation formula (Euclidean Distance) to determine the closest distance from each data to the centroid. This study also carried out visualization using rapidminer applications, from manual calculations and rapidminer the results were different cluster arrangements but the areas grouped were the same, namely, the results of manual calculations C1 there were 13 regions, C2 there were 7 regions and C3 there was 1 region. The results of the rapidminer Cluster_0 there are 7 regions, Cluster_1 there are 1 regions and Cluster_2 there are 13 regions. The regions that are divided into groupings using rapidminer or manual calculation have the same results and there is no difference in region.

4. CONCLUSION

The conclusion that can be reached from this study after applying the K-Means algorithm in classifying the types of crimes against human bodies in 2019 to 2020 in the East Nusa Tenggara region is that there are 3 centroids, C1 for areas with low crime, C2 for areas with moderate crime and C3 for areas with high crime. The initial centroid value is determined randomly and then for the next centroid is adjusted to the result of the calculation of the closest distance (minimum). The final results obtained are areas with low crime totaling 13 regions, namely East Sumba, Lembata, Sikka, Ende, Ngada, Manggarai, Rote Ndao, West Manggarai, Central Sumba, Southwest Sumba, Nagekeo, East Manggarai, and Sabu Raijua. There are 7 areas with moderate crimes, namely West Sumba, Kupang, South Central Timor, North Central Timor, Belu, Alor, and East Flores. As for the area with high crime, there is 1 area, namely Kupang City.

REFERENCES

[1] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi”.
[2] “FullBook Pengenalan Data Mining”.
[3] “ARTIKEL PENERAPAN DATA MINING UNTUK MENGANALISIS DATA.”
[4] B. I. Nugroho, Z. Ma’arif, and Z. Arif, “Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media,” 2022. [Online]. Available: www.journal.peradaban.ac.id
[5] “IMPLEMENTASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN DI ARMADA COMPUTER MENGGUNAKAN ALGORITMA APRIORI”, doi: 10.55123.
[6] N. A. Sudibyo, A. Iswardani, K. Sari, S. Suprihatiningsih, U. Duta, and B. Surakarta, “PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI INDONESIA,” vol. 1, no. 3, p. 2020, doi: 10.46306/lb.v1i3.

[7] W. Saputro, M. Reza Pahlevi, A. Wibowo, M. Ilmu Komputer, and F. Ilmu Komputer, "ANALISIS ALGORITMA K-MEANS UNTUK KLASTERISASI TINDAK PIDANA KORUPSI DI WILAYAH HUKUM INDONESIA," Dec. 2020, doi: 10.33387/jiko.
[8] I. Amal and R. Amanda Putri, "Clustering Pecandu Narkoba Menggunakan Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON) Hal: 434−*, vol. 443, no. 2, 2023, doi: 10.30865/json.v5i2.7009.
[9] N. N. Hasanah and A. S. Purnomo, "Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta)," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 4, no. 2, pp. 300–311, Jul. 2022, doi: 10.47233/jteksis.v4i2.499.
[10] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 205–214, Nov. 2021, doi: 10.30812/matrik.v21i1.1554.
[11] J. Homepage, K. Pratama Simanjuntak, and U. Khaira, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm Pengelompokkan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering," vol. 1, pp. 7–16, 2021.
[12] Y. Anggiat, J. Rizal Hermawan, T. C. Widianto, S. Anggraeni, and S. D. Saraswati, "DATA MINING CLUSTERING PENYEBARAN VIRUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA METODE K-MEANS," *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 6, no. 2, 2022.
[13] P. sari and R. Syahri, "ALGORITMA K-MEANS CLUSTERING: SEBUAH STUDI LITERATUR," *Jurnal Informatika (JURI)*, doi: 10.12345/juri.
[14] R. L. Pattipeilohy and M. A. I. Pakereng, "Penerapan K-Means Clustering Pada Data Mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata Untuk Menentukan Strategi Promosi," 2023.
[15] N. Aswan, Y. Fadhillah, M. Noor, and H. Siregar, "Clusterirasi Kabupaten/Kota di Sumatera Utara Berdasarkan Luas Tanam dan Hasil Produksi Kelapa Sawit," *EDUSAINS : Journal of Education and Science*, vol. 01, no. 01, pp. 25–32, 2023, doi: 10.57255/edusain.v1i1.248.
[16] C. Naya, A. Siswandi, H. Rahendra Herlianto, and A. Fauzi, "SIGMA-Jurnal Teknologi Pelita Bangsa IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKAN PENGANGGURAN TERBUKA DI INDONESIA DENGAN METODE CLUSTERING," vol. 14, 2023.
[17] A. Septianingsih, S. A. Pertama, D. Kependudukan, P. Sipil, and K. Tangerang, "ANALISIS K-MEANS CLUSTERING PADA PEMETAAN PROVINSI INDONESIA BERDASARKAN INDIKATOR RUMAH LAYAK HUNI," vol. 3, no. 1, 2022, doi: 10.46306/lb.v3i1.
[18] F. Kurnia Alvisan, "Clustering Minimarket Untuk Menentukan Jumlah Kebutuhan Pembelian Menggunakan Metode K-Means," *Jurnal NOE*, vol. 4, no. 2, 2021, [Online]. Available: https://ojs.unpkediri.ac.id/index.php/noe
[19] J. Penerapan, T. Informasi, D. Komunikasi, G. B. Kaligis, and S. Yulianto, "IT-EXPLORE ANALISA PERBANDINGAN ALGORITMA K-MEANS, K-MEDOIDS, DAN X-MEANS UNTUK PENGELOMPOKKAN KINERJA PEGAWAI (Studi Kasus: Sekretariat DPRD Provinsi Sulawesi Utara)," 2022.
[20] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, Mar. 2023, doi: 10.56211/helloworld.v2i1.230.
[21] A. Febrianto, S. Achmadi, and A. P. Sasmito, "PENERAPAN METODE K-MEANS UNTUK CLUSTERING PENGUNJUNG PERPUSTAKAAN ITN MALANG," 2021.
[22] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. PRISMA," *Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161–170, 2019, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/
[23] S. Nurajizah and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, 2021, doi: 10.31294/jtk.v4i2.