



Application of the Clustering Algorithm for the Classification of General Criminal Cases at the Binjai District Attorney's Office

Ratih¹, Relita Buaton², Katen Lumbanbatu³

^{1,2,3}STMIK Kaputama

rantiratih@gmail.com^{1*}; bbcbuaton@gmail.com²; k.lumbanbatu@yahoo.com³;

Abstract

The Binjai District Attorney's Office in carrying out its duties and functions, one of which handles general crimes, where so far the SPDP (Warranty to Commence Investigation) from the police that has entered the Binjai District Prosecutor's Office amounted to approximately 50 (fifty) cases each month. This amount consists of several types of general criminal cases. It is known that the types of general criminal cases amount to approximately 215 (two hundred and fifteen) types of cases, from this data, a method of classifying/clustering is needed from the types of cases that exist each month so that the data can be processed so as to produce the highest, moderate and highest scores. the lowest value of a type of case. The Binjai District Attorney's Office often receives requests for data from other ministries or agencies such as the BPS (Central Statistics Agency), the National Commission on Women and the National Commission on Children in the form of data recapitulation of crimes against women and children as perpetrators of crimes. The Binjai District Attorney's Office has a case handling system where the recapitulation cannot be taken directly but instead collects data manually, because the existing case handling system does not have the recapitulation as requested. The application of clustering has been carried out by many previous researchers. Among them, the K-Means Clustering Algorithm Analysis Mapping the Number of Crimes. The research was carried out using a data mining model in classifying illegal fishing with the K-Means algorithm analysis by determining the shortest distance using the euclidean distance, more optimal than using the mahattan distance and chbchep distance in classifying student achievement, determining the centroid (central point) in the early stages of the algorithm K-Means is very influential on cluster results as the results of tests carried out using 267 records with different centroids produce different cluster results as well, a clustering model is obtained that can be used for illegal fishing in decision making for illegal fishing crimes high, medium, moderate.

Keywords: General, Criminal Cases, Clustering.

1. Introduction

In carrying out its duties and functions, the Attorney General's Office is required to be able to realize legal certainty, legal order, justice and truth based on law, and heed religious norms, decency and decency and must explore human values, law and justice that live in society. The Binjai District Attorney's Office in carrying out its duties and functions, one of which handles general crimes, where so far the SPDP (Warranty to Commence Investigation) from the police that has entered the Binjai District Prosecutor's Office amounted to approximately 50 (fifty) cases each month. This amount consists of several types of general criminal cases. It is known that the types of general criminal cases amount to approximately 215 (two hundred and fifteen) types of cases, from this data, a method of classifying/clustering is needed from the types of cases that exist each month so that the data can be processed so as to produce the highest, moderate and highest scores. the lowest value of a type of case.

The application of clustering has been carried out by many previous researchers. Among them, the K-Means Clustering Algorithm Analysis Mapping the Number of Crimes. The research was carried out using a data mining model in classifying illegal fishing with the K-Means algorithm analysis by determining the shortest distance using the euclidean distance, more optimal than using the mahattan distance and chbchep distance in classifying student achievement, determining the centroid (central point) in the early stages of the algorithm K-Means is very influential on cluster results as the results of tests carried

out using 267 records with different centroids produce different cluster results as well, a clustering model is obtained that can be used for illegal fishing in decision making for illegal fishing crimes high, medium, sufficient

2. Research Methods

2.1. Clustering Algorithm

Clustering is dividing data into groups that have objects with the same characteristics. Clustering plays an important role in data mining applications, such as data science exploration, information access and textmining, spatial database applications, and web analytics. Basically there are two clustering methods namely Hierarchical clustering method and Non Hierarchical clustering method. The hierarchical method is used if the number of groups is not known beforehand, while the non-hierarchical method is used if the number of groups is known from a number of objects. One of the non-hierarchical algorithms is the K-Means algorithm [2]. There is many approaches to create cluster, including creating rules that dictate are allowed within the same group by level among its members. Another approach is with create a function group measure some properties of grouping as a function of several parameters of a clustering [8].

2.2. Stages of Clustering

In the hierarchical method, the clustering stage begins with grouping two or more objects that have the closest similarity. Then it is passed on to other objects and so on until the cluster will form a kind of tree where there is a clear level (hierarchy) between objects, from the most similar to the least similar. Whereas the non-hierarchical method begins by first determining the desired number of clusters (two, three, or something else). After the number of clusters is determined, the cluster process is carried out without following the hierarchical process [3].

2.3. K-Means Algorithm

The K-Means algorithm is a popular clustering algorithm and is widely used in the industrial world. This algorithm is structured on the basis of a simple idea. Initially it is determined how many clusters will be formed. Any object or first element in the cluster can be selected to serve as the center point (Centroid Point) of the cluster [5]. The K-Means algorithm will then repeat the following steps until stability occurs (no objects can be moved). The concept of similarity is fundamental in cluster analysis. Similarity between objects is a measure of correspondence between objects. There are 3 methods that can be applied, namely the correlation measure, distance measure, and association measure [6]. K-Means is a method in data mining that can group data or Clustering a data into the form of one cluster or more clusters so that the data is owned The same characteristics are hidden in the same cluster and data with different characteristics vary among different groups [7].

K-Means Clustering is one of the simplest and most popular "unsupervised machine learning algorithms". K-Means Clustering is a data analysis method or Data Mining method that performs an unsupervised modeling process and is a method for grouping data with a partition system. K-means clustering is a non-hierarchical cluster analysis method that seeks to partition existing objects into one or more clusters or groups of objects based on their characteristics, so that objects with the same characteristics are grouped in the same cluster and objects with different characteristics. grouped into another cluster. In this learning algorithm, the computer groups its own data into its input without knowing the class target in advance. This learning is included in unsupervised learning. The inputs received are data or objects and the desired k clusters. This algorithm will group data or objects into these k groups. In each cluster there is a center point (centroid) that represents the cluster. The MOORA (Multi-Objective Optimization On The Basis Of Ratio Analysis) method is a multi-objective optimization technique that can be successfully applied to solve various types of complex decision-making problems in decision-making [1].

The steps for using the K-Means algorithm are as follows:

Determine the number of clusters by identifying data.

$$X_{ij} = (i=1, \dots, n; j=1, \dots, m) \quad (1)$$

where n is the amount of data to be clustered and m is the number of variables;

Determining the centroid (coordinate of the midpoint of each cluster), for the first iteration is taken randomly.

$$C_{kj} = (k=1, \dots; j=1, \dots, m) \quad (2)$$

Calculating the distance of the object to the centroid using the dEuclidean or dManhattan distance formula;

$$d_{\text{Euclidean}}: X, Y = \sqrt{\sum_i (X_i - Y_i)^2} \quad (3)$$

$$d_{\text{Manhattan}}: X, Y = \sqrt{\sum_i |X_i - Y_i|} \quad (4)$$

To calculate the new centroid, you can use the following equation:

$$C = \frac{\sum m}{n} \quad (5)$$

Where:

C : data centroids

m : data member that belongs to a certain centroid

n : the number of data that is a member of a certain centroid

Determine the distance of each object to the midpoint coordinates;
Grouping these objects based on the shortest distance.

Additional calculation steps, namely the ratio between BCV (Between Cluster Variation) and WCV (Within Cluster Variation) as follows:

The formula for calculating BCV:

$$BCV = d(c1,c2) + d(c1,c3) + d(c2,c3) \quad (6)$$

The formula for calculating WCV:

$$WCV = \sum(\text{jarak pusat tiap cluster paling minimum})^2 \quad (7)$$

To calculate the ratio, you can use the equation:

$$\text{Rasio} = \frac{BCV}{WCV} \quad (8)$$

2.4. Definition of General Crimes

There are 2 (two) types of criminal acts handled at the binjai state prosecutor's office, namely general crimes and special crimes. according to general crimes are all crimes contained in the criminal code as codification of material criminal law (Book II and Book III) 4. General Crimes have a legal source leading to the Criminal Code (KUHP) as a source of material law and the Criminal Procedure Code (KUHAP) as a source of formal law. Whereas Special Crimes are types of criminal cases whose legal arrangements are outside the Criminal Code (KUHP) which is a codified Code of Acts, has special and specific characteristics and handling of cases, both from the applicable legal rules, the procedural law. , law enforcement, and lawyers who handle it. Quite a lot of criminal acts are General Crimes, because they directly intersect with the life of society in general.[4]

General Crime Cases are divided into 4 (four) types, namely:

1. Person and Property (OHARDA)
2. State Security and Public Order and Other Common Crimes (KAMNEG TIBUM and TPUL)
3. Narcotics
4. Terrorism

Of the four types of general crime cases, each consists of several cases, and the total number of general criminal cases is approximately 215 (two hundred and fifteen) types of cases. With so many types of cases, grouping/clustering is needed to provide information on which types of cases are the most, moderate, and few in the Binjai State Prosecutor's Office, especially in the General Crimes Section.

3. Results And Discussion

3.1. Application of the Method

From research conducted on the general crimes section of the Binjai District Attorney's Office, it was obtained data on general criminal cases for 2022 as many as 349 cases and for this study only used 20 data samples. From the explanation of the observations that have been made, the following is the process of implementing the Clustering Algorithm for Grouping General Crime Cases at the Binjai District Attorney's Office based on the type of case

To classify general criminal cases in selecting data, the attributes used are the type of case/case, gender and age. Types of general criminal cases are as follows:

Narcotics crime:

1. Crime of theft;
2. Crime of embezzlement;
3. Crime of Persecution;
4. Criminal Acts of Extortion and Threats;

5. Criminal Acts Against the Origin of Marriage;
6. Crime of Destruction and Damage to Goods;
7. Gambling crime;
8. Crime of Fraud;
9. Crime of Obscenity;
10. Crimes Against Life;
11. Crime of theft with violence;
12. Criminal Acts of Collection, Publishing and Printing;
13. Crime of Child Protection;
14. Crime of Eliminating Domestic Violence;
15. Crimes Against State Security;
16. Crime of Forgery of Letters;
17. Crime of Violation of Public Order;
18. Traffic and Road Transportation Crimes;
19. Criminal Information and Electronic Transactions;
20. Criminal acts of committing violence together in public / beatings;
21. Crime of Sharp Weapons - Firearms and Ammunition and Explosives;
22. Oil and Gas crime;
23. Crimes that endanger public security for people/goods;
24. Crime of Rape;
25. Criminal Acts of Conservation of Living Natural Resources and Ecosystems;
26. Pornography Crime.

Table 1: 2022 Data Samples

Case ID	File No.	Suspect	Types of Things
P-001	BP/ 03 / I /2022 /Reskrim	Sahrial Efendi	Theft
P-002	BP/100/IX/2022/Drug Res	Elshadai Manalu	Narcotics Crime
P-003	BP/26/II/2022/Drug Res	Rahmat Riyadi	Narcotics Crime
P-004	BP/ 72 / VII /2022 /Reskrim	Wanda Fahroji Als Wanda	Embezzlement
P-005	BP/ 30 / X /2022	Chairul Hamdi Als Am	Theft
P-006	BP/ 31 / XII /2021 /Reskrim	Ayu Pratiwi	Embezzlement
P-007	BP/39/III/2022/Drug Res	Rilia Wati Dewi Wahyuni Syahputri	Narcotics Crime
P-008	BP/54/VI/2022/Drug Res	Doni Afrizal Harahap	Narcotics Crime
P-009	BP/69/VII/2022/Resba	Kelengi Tarigan	Narcotics Crime
P-010	BP/86/VIII/2022/Drug Res	Deni Mickael Manihuruk	Narcotics Crime
P-011	BP/ 03 / II /2022 /Reskrim	Budiman Als Akiong	Theft
P-012	BP/ 02 / I /2021 /Reskrim	Suheri Kapenia Sembiring Alias Blek	Theft
P-013	BP/ 12 / IV /2022 /Reskrim	Willi Gunawan Pane	Deceit
P-014	BP/38/XI/2022/Reskrim	Muhammad Nasri	Deceit
P-015	BP/42/III/2022/Drug Res	Akhirudin Sembiring	Narcotics Crime
P-016	BP/62/VII/2022/Drug Res	Kabul Evidence	Crimes against the origin of marriage
P-017	BP/65/VII/2022/Drug Res	Saiful Amri	Narcotics Crime
P-018	BP/75/VII/2022/Drug Res	Rudi Harjah Als Gogon	Narcotics Crime
P-019	BP/ 08 / IV /2022 /Reskrim	Lukman Hakim Als Kolok Jumadil Tsyani Hidayat Berry Prime	Theft
P-020	BP/ 32 / VIII /2022 /Reskrim	M. Azizi	Deceit

Table 2: Sample Criteria Data

No.	Criteria	Sub Criteria	Code
1.	K1 (Gender)	Man	2
		Woman	1
		> 45	4
2.	K2 (Age)	> 30 ≤ 45	3
		> 15 ≤ 30	2
		> 0 ≤ 15	1
3.	K3 (Criminal Type)	Drugs	4
		Oharda	3
		Kamnegtibus and TPUL	2
		Terrorism	1

3.2. Application of K-Means Clustering

The application of the K-Means Clustering algorithm and its calculation using several criteria including: Gender (K1), Age (K2) and Criminal Type (K3). The following table 3.3 of case data assessment can be seen below.

Table 3: Case Data Assessment

Case ID	Suspect	K1	K2	K3
P-001	Sahrial Effendi	2	3	3
P-002	Elshadai Manalu	2	2	4
P-003	Rahmat Riyadi	2	2	4
P-004	Wanda Fahroji As Wanda	2	3	3
P-005	Chairul Hamdi As Am	2	3	3
P-006	Ayu Pratiwi	1	2	3
P-007	Rilia Wati Dewi Wahyuni Syahputri	1	3	4
P-008	Doni Afrizal Harahap	2	3	4
P-009	Kelengi Tarigan	2	3	4
P-010	Deni Mickael Manihuruk	2	2	4
P-011	Budiman As Akiong	2	3	3
P-012	Suheri Kapenia Sembiring Alias Blek	2	3	3
P-013	Willi Gunawan Pane	2	2	3
P-014	Muhammad Nasri	2	2	3
P-015	Akhirudin Sembiring	2	3	4
P-016	Kabul Evidence	2	3	3
P-017	Saiful Amri	2	3	4
P-018	Rudi Harjah As Gogon	2	3	4
P-019	Lukman Hakim Als Kolok Jumadil Tsyani Hidayat Berry Prime	2	3	3
P-020	Mr. AZIZI	2	2	3

Based on the data above, 3 clusters will be formed which will be described in table 3.4 below.

Table 4: Cluster Data

No.	Cluster	Information
1	C3	High clusters
2	C2	Medium cluster
3	C1	Low cluster

The data in table 3.3 of the case data assessment above are normalized or cleaned up unused data and initialized the data so that it can be processed using the K-Means algorithm. Here are the steps on the K-Means algorithm:

- Determine the number of clusters i.e. as much as $k = 3$
- Determine the centroid (arithmetic mean of an object shape from all points in the object) of each cluster taken from the source data based on random selections (from table 3.3 in gray).

Table 5: Initial Centroid Data (Iteration 1)

Centroid	Case ID	K1	K2	K3
Centroid 1	P-007	1	3	4
Centroid 2	P-010	2	2	4
Centroid 3	P-015	2	3	4

- Calculate the distance of the data to the centroid using the Euclidean formula, the data is from the cluster closest to it.

- Distance between Case number one and point m1

$$\begin{aligned}
 &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \\
 &= \sqrt{(2 - 1)^2 + (3 - 3)^2 + (3 - 4)^2} \\
 &= 1,414
 \end{aligned}$$

- Distance between Case number one and point m2

$$\begin{aligned}
 &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \\
 &= \sqrt{(2 - 2)^2 + (3 - 2)^2 + (3 - 4)^2} \\
 &= 1,414
 \end{aligned}$$

- Distance between Case number one and point m3

$$\begin{aligned}
 &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \\
 &= \sqrt{(2 - 2)^2 + (3 - 3)^2 + (3 - 4)^2} \\
 &= 1,000
 \end{aligned}$$

Table 6: Iteration Calculation Results 1

Case ID	Suspect	Cluster 1	Cluster 2	Cluster 3	Closest Distance (Smallest Value)	Smallest Centroid Distance	Square Results
P-001	Sahrial Effendi	1,414	1,414	1,000	C3	1,000	1,000
P-002	Elshadai Manalu	1,414	0,000	1,000	C2	0,000	0,000
P-003	Rahmat Riyadi	1,414	0,000	1,000	C2	0,000	0,000

P-004	Wanda Fahroji As Wanda	1,414	1,414	1,000	C3	1,000	1,000
P-005	Chairul Hamdi As Am	1,414	1,414	1,000	C3	1,000	1,000
P-006	Ayu Pratiwi	1,414	1,414	1,732	C1	1,414	2,000
P-007	Rilia Wati Dewi Wahyuni Syahputri	0,000	1,414	1,000	C1	0,000	0,000
P-008	Doni Afrizal Harahap	1,000	1,000	0,000	C3	0,000	0,000
P-009	Kelengi Tarigan	1,000	1,000	0,000	C3	0,000	0,000
P-010	Deni Mickael Manihuruk	1,414	0,000	1,000	C2	0,000	0,000
P-011	Budiman As Akiong Suheri Kapenia	1,414	1,414	1,000	C3	1,000	1,000
P-012	Sembiring Alias Blek	1,414	1,414	1,000	C3	1,000	1,000
P-013	Willi Gunawan Pane	1,732	1,000	1,414	C2	1,000	1,000
P-014	Muhammad Nasri	1,732	1,000	1,414	C2	1,000	1,000
P-015	Akhirudin Sembiring	1,000	1,000	0,000	C3	0,000	0,000
P-016	Kabul Evidence	1,414	1,414	1,000	C3	1,000	1,000
P-017	Saiful Amri	1,000	1,000	0,000	C3	0,000	0,000
P-018	Rudi Harjah As Gogon	1,000	1,000	0,000	C3	0,000	0,000
P-019	Lukman Hakim Als Kolok Jumadil Tsyani Hidayat Berry Prime	1,414	1,414	1,000	C3	1,000	1,000
P-020	Mr. AZIZI	1,732	1,000	1,414	C2	1,000	1,000
Within Cluster Variation (WCV)							12

Within Cluster Variation (WCV) is obtained by summing all the results of the minimum square of the distance, as shown in table 3.6 with total results = 12

Based on the results of iteration 1 calculation in table 3.6, the grouping of cases / cases is shown in table 3.7 below:

Table 7: Iteration Grouping Results 1

Cluster	Case ID	Types of Things		K1 (Gender)	K2 (Age)	K3 (Criminal Type)	
C1	P-006	Embezzlement	1	Woman	2	28	3 Oharda
	P-007	Narcotics Crime	1	Woman	3	42	4 Drugs
Remarks : in Cluster 1 (C1) there are two data grouped based on the least sex, namely women.							
C2	P-002	Narcotics Crime	2	Man	2	28	4 Drugs
	P-003	Narcotics Crime	2	Man	2	21	4 Drugs
	P-010	Narcotics Crime	2	Man	2	24	4 Drugs
	P-013	Deceit	2	Man	2	26	3 Oharda
	P-014	Deceit	2	Man	2	23	3 Oharda
	P-020	Deceit	2	Man	2	25	3 Oharda
Description: in Cluster 2 (C2) there are six data grouped based on male sex and age grouping categories 16 to 30 years and also grouping based on the type of crime where there are two types of crimes, namely Narcotics and Oharda.							
C3	P-001	Theft	2	Man	3	32	3 Oharda
	P-004	Embezzlement	2	Man	3	34	3 Oharda
	P-005	Theft	2	Man	3	36	3 Oharda
	P-008	Narcotics Crime	2	Man	3	37	4 Drugs
	P-009	Narcotics Crime	2	Man	3	35	4 Drugs
	P-011	Theft	2	Man	3	36	3 Oharda
	P-012	Theft	2	Man	3	37	3 Oharda
	P-015	Narcotics Crime	2	Man	3	43	4 Drugs
	P-016	Crimes against the origin of marriage	2	Man	3	44	3 Oharda
	P-017	Narcotics Crime	2	Man	3	37	4 Drugs
	P-018	Narcotics Crime	2	Man	3	42	4 Drugs
	P-019	Theft	2	Man	3	33	3 Oharda
Description: in Cluster 3 (C3) there are twelve data grouped based on the sex of men who commit the most crimes and also grouping with the most age categories of 31 to 45 years and the most criminal type categories with seven cases and Narcotics with five cases.							

Calculate Between Cluster Variation (BCV)

Centroid m1 (1, 3, 4)

Centroid m2 (2, 2, 4)

Centroid m3 (2, 3, 4)

D (m1, m2)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(1-2)^2 + (3-2)^2 + (4-4)^2}$$

$$= 1,414$$

D(m1, m3)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(1-2)^2 + (3-3)^2 + (4-4)^2}$$

$$= 1$$

D(m2, m3)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(2-2)^2 + (2-3)^2 + (4-4)^2}$$

$$= 1$$

Between Cluster Variation (BCV) = 1,414 + 1 + 1

= 3,414

Calculate Ratio

Next calculate the ratio value by comparing the BCV and WCV values.

$$\text{Ratio} = \text{BCV}/\text{WCV}$$

$$= 3,414/12$$

$$= 0,284$$

4. Menghitung Between Cluster Variation (BCV)

Centroid m1	=	1	2,500	3,500
Centroid m2	=	2	2	3,500
Centroid m3	=	2	3	3,417

D (m1,m2)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(1-2)^2 + (2,5-2)^2 + (3,5-3,5)^2}$$

$$= 1,118$$

D (m1, m3)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(1-2)^2 + (2,5-3)^2 + (3,5-3,417)^2}$$

$$= 1,121$$

D (m2, m3)

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{(2-2)^2 + (2-3)^2 + (3,5-3,417)^2}$$

$$= 1,003$$

Between Cluster Variation (BCV) = 1,118 + 1,121 + 1,003

= 3,242

5. Calculating Ratios

Next calculate the ratio value by comparing the BCV and WCV values.

$$\text{Ratio} = \text{BCV}/\text{WCV}$$

$$= 3,242/5,417$$

$$= 0,598$$

Based on calculations in iterations 1 and 2, the ratio 1 and 2 is as follows:

Iteration rate 1 (Previous)	= 0.284
Iteration rate 2 (Present)	= 0.598

1. If the current ratio is > from the previous ratio then the iteration continues and take steps as in the previous iteration.
2. Iteration stops if the current ratio ≤ is from the previous ratio or
3. If no data is moved cluster.

Therefore, iteration stops until iteration stage 2 because there is no data moving clusters, the cluster results for grouping cases into 3 clusters are:

1. High cluster (C3), where the highest number of cases / cases with a total of 12 cases grouped by male sex commit the most crimes and also grouping with the most age categories of 31 to 45 years and the most criminal type categories are Oharda as many as seven cases and Narcotics as many as five cases.
2. Medium cluster (C2), where the number of cases / cases is medium with a total of 6 cases grouped based on male sex and age grouping categories 16 to 30 years and also grouping based on the type of crime where there are two types of crimes, namely Narcotics and Oharda.
3. Low cluster (C1), where the number of cases / cases is low with the number of 2 cases grouped by the least gender, namely women.

4. Conclusions

From the results of the analysis that has been carried out by researchers, the conclusions can be drawn in this study are as follows:

1. Using the K-Means Clustering method, building a system for grouping general criminal cases at the Binjai State Prosecutor's Office based on case types using PHP programming language and MySQL databases was successfully created according to what was designed by testing 20 (twenty) general crime case data in 2022 which resulted in 3 (three) clusters, *namely* high, medium and low clusters based on the calculation results.
2. Applying a clustering algorithm in grouping general criminal cases based on case types by determining the criteria and values of each assessment criterion, including gender, age and type of crime.
3. With the criteria data set and the initial centroid determination of 3 (three) centroids, the system produces 2 (two) iterations in the calculation and grouping cases based on the highest type of case, namely 12 (twelve) cases grouped by gender, men commit the most crimes with the most age categories of 31 (thirty-one) to 45 (forty-five) years and the most criminal type categories are Oharda as many as 7 (seven) cases and Narcotics as many as 5 (five) cases, medium case types as many as 6 (six) cases grouped by male gender with age categories 16 (sixteen) to 30 (thirty) years and Oharda criminal type categories as many as 3 (three) cases and Narcotics as many as 3 (three) cases, low case types as many as 2 (two) cases grouped by female gender and Oharda criminal type category 1 (one) case and Narcotics as many as 1 (one) case.

References

- [1] Hutagalung, J. (2021). *Kombinasi K-Means Clustering dan Metode MOORA*. Yogyakarta: Deepublish.
- [2] Munandar, T. A. (2022). Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten. *JSiI (Jurnal Sistem Informasi)*, 9.2, 109-14.
- [3] Purnama, J. J. (2019). Analisa Algoritma K-Means Clustering Pemetaan Jumlah Tindak Pidana. *Klik - Kumpulan Jurnal Ilmu Komputer*, 6.2, 128.
- [4] Suryadi, U. T. (2019). Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means. *Jurnal Teknologi Dan Komunikasi STMIK Subang*, 12.1, 15-27.
- [5] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," *Reg. Dev. Ind. Heal. Sci. Technol. Art Life*, pp. 394-403, 2019, [Online]. Available: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>
- [6] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574-583, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7052276>
- [7] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [8] W. M. P. Duhita, "Clustering Menggunakan Metode K-Means Untuk," *J. Inform.*, vol. 15, no. 2, pp. 160-174, 2016.