(2025), 3 (4): 176-189

Scientica

Jurnal Ilmiah Sain dan Teknologi

3021-8209

APPLICATION OF CLUSTERING METHOD TO IDENTIFY CRIME-PRONE AREAS IN MENTAWAI ISLANDS DISTRICT USING K-MEANS ALGORITHM AND GEOPANDAS LIBRARY

Dika Alkadri, Yeka Hendriyani, Khairi Budayawan, Syafrijon

Department of Electronic Enginering, Universitas Negeri Padang

Dikaalkadri2002@gmail.com

Abstract

Crime is a serious social problem that has a significant impact on the quality of life of the community, including in Mentawai Islands Regency. Based on data collected by the Mentawai Resort Police between 2021 and 2023, there are six types of crime that occur most frequently, namely theft, sexual abuse, maltreatment, narcotics, gambling, and domestic violence. This research uses K-Means clustering method to identify crime-prone areas by considering the number and density of population. The data used in this study includes information about the location of the incident, the number of criminal incidents, population and population density, which is then analyzed using the Geopandas library and Folium library for visualization. After clustering the data of crime-prone areas using the K-Means clustering method, three clusters with different characteristics were formed. The first cluster includes six sub-district areas that are classified as not prone to crime. The second cluster consists of three sub-district areas categorized as crime-prone areas, while the third cluster includes one sub-district area that is very prone to crime. This research is expected to provide useful information for the authorities in designing more targeted security policies.

Article History

Submitted: 31 January 2025
Accepted: 11 Febuary 2025
Published: 12 February 2025

Key Words

Crime, Clustering, K-Means, Geopandas, Folium, Population, Population Density

Abstrak

Kejahatan merupakan masalah sosial yang serius dan memiliki dampak signifikan terhadap kualitas hidup masyarakat, termasuk di Kabupaten Kepulauan Mentawai. Berdasarkan data yang dikumpulkan oleh Kepolisian Resor Mentawai antara tahun 2021 hingga 2023, terdapat enam jenis kejahatan yang paling sering terjadi, yaitu pencurian, pelecehan seksual, penganiayaan, narkotika, perjudian, dan kekerasan dalam rumah tangga. Penelitian ini menggunakan metode K-Means clustering untuk mengidentifikasi daerah rawan kejahatan dengan mempertimbangkan jumlah dan kepadatan penduduk. Data yang digunakan dalam penelitian ini mencakup informasi tentang lokasi kejadian, jumlah insiden kejahatan, jumlah penduduk, dan kepadatan penduduk, yang kemudian dianalisis menggunakan pustaka Geopandas dan Folium untuk visualisasi. Setelah dilakukan pengelompokan data daerah rawan kejahatan menggunakan metode K-Means clustering, terbentuk tiga klaster dengan karakteristik yang berbeda. Klaster pertama mencakup enam wilayah kecamatan yang diklasifikasikan sebagai tidak rawan kejahatan. Klaster kedua terdiri dari tiga wilayah kecamatan yang dikategorikan sebagai daerah rawan kejahatan, sedangkan klaster ketiga mencakup satu wilayah kecamatan yang sangat rawan kejahatan. Penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi pihak berwenang dalam merancang kebijakan keamanan yang lebih terarah.

Sejarah Artikel

Submitted: 31 January 2025
Accepted: 11 Febuary 2025
Published: 12 February 2025

Kata Kunci

Kejahatan, Clustering, K-Means, Geopanda, Folium, Populasi, Kepadatan Penduduk

Pendahuluan

Criminality is one of the most serious social problems and has a significant impact on the quality of life of the community, especially in areas with high levels of vulnerability. The Mentawai Islands Regency, as one of the regions located on the coast of West Sumatra, is not free

from this problem. Although generally known for its natural beauty, the region also faces great challenges in maintaining security and order. With a variety of unique regional characteristics, ranging from coastal areas to remote hills, Mentawai Islands Regency is vulnerable to various forms of criminality that can disrupt the social and economic stability of the community. Therefore, it is important to understand the dynamics of criminality in this region so that appropriate preventative measures can be implemented.

Data from the Mentawai Resort Police from 2021 to 2023 shows that six types of crime - theft, sexual abuse, maltreatment, narcotics, gambling, and domestic violence - occur more frequently in some sub-districts. This research analyzes data from ten sub-districts in Mentawai Islands Regency, focusing on characteristics such as population and population density that influence crime patterns. The analysis aims to provide a clear picture of the distribution of crime to help authorities formulate more effective prevention strategies

Identifying crime-prone areas is an important step in crime prevention and security efforts. By conducting this research, a solution was born, namely knowing the locations that have a high level of vulnerability, the authorities can allocate resources more effectively, such as placing security personnel or installing surveillance cameras at strategic points. In addition, 's information about crime-prone areas also helps people to be more vigilant in carrying out their daily activities in these areas.

The K-Means Clustering method is effective for grouping crime data based on similar characteristics. The goal is to find groups of similar data, so the greater the difference between groups, the better the quality of the cluster. The use of GeoPandas and Folium libraries in Python allows interactive spatial analysis to map crime-prone areas in Mentawai Islands Regency, facilitating the identification of high-risk areas based on clustering results.

By using K-Means Clustering, crime data can be grouped based on the similarity of criminal cases in each region, making it easier to identify areas that have a high risk. GeoPandas and Folium will facilitate the visualization of clustering results on a map, so that authorities can clearly see areas that require special attention. With this approach, crime prevention and response strategies can be more targeted and effective, contributing to improved security across the sub-district.

The main objective of this research is to apply K-Means Clustering to map areas with high crime rates in Mentawai Islands Regency, utilizing GeoPandas and Folium for spatial analysis. By clustering data based on the number of crimes, population, and population density, it is expected to identify vulnerable areas in each sub-district. Visualization of clustering results in the form of an interactive map is expected to help prevent crime and increase public safety.

METHODS

A. Location and Time of Research

This research was conducted at the Mentawai Islands District Police, which is the main location for collecting data and information related to crime rates. The main focus of this research is the crime data that occurred during the period 2021 to 2023, with the aim of identifying crime-prone areas in the region.

B. Population and Sample

The population in this study is all criminal incidents recorded and reported to the Mentawai Islands Regency Resort Police during the period 2021 to 2023, including various types of criminal acts in the sub-district administrative boundaries, as well as data on the number and population density obtained from the Central Bureau of Statistics. The sample of this study used a saturated sampling technique, where all reported criminal incidents were sampled. This data was then spatially analyzed using shapefile data of Mentawai Islands Regency to map and cluster the location of criminal incidents in each sub-district.

C. Independent Variable

In this study, the variables analyzed to understand the dynamics of criminality in Mentawai Islands Regency consist of several factors that can influence the occurrence of crime, namely: (a) type of crime, (b) location of crime, (c) population, and (e) population density.

D. Dependent Variable

Criminality Risk Cluster: Cluster categories resulting from K-Means analysis that divides the area in Mentawai Islands Regency, while the cluster categories are divided into three based on the level of crime risk, namely: (a) Not Prone, (b) Prone, and (c) Very Prone.

E. Data Collection Technique

In this research, the data collection techniques used include documentation studies to collect secondary data from the official archives of the Mentawai Islands Regency Police and the Mentawai Islands Regency Statistics Agency related to crime, population, and population density during the 2021-2023 period, as well as geographic data collection through Geographic Information System (GIS)-based digital maps supported by GeoPandas and Folium libraries to map the location of crime incidents.

F. Data Analysis Technique

Data analysis techniques in this study include data preprocessing, clustering process with the K-Means algorithm, and visualization of clustering results using Python and with the help of the GeoPandas Library and Folium. it is hoped that this research can produce valid findings related to crime rates.

G. Clustering technique using K-Means algorithm

K-Means clustering works by finding the cluster center or centroid that represents the group. This process involves iterating between calculating the distance between each data object and the existing cluster center and updating the location of the cluster center based on the average of the data objects contained in the cluster. This iteration is done repeatedly until a stable or converged cluster center is obtained. Data clustering using the K-Means algorithm is generally done with the following basic algorithm:

1. Determine the number of clusters
2. Randomly allocate data into clusters

3. Calculate the distance of each data to the centroid: Calculate the distance between each data and each centroid. The data will be allocated to the cluster with the centroid that has the closest distance. For this distance calculation, we can use the Euclidean distance theory which is formulated as follows:

$$D_{(i,j)} = \sqrt{(x_{1i} - x_{2j})^2 + (x_{2i} - x_{2j})^2 + \ldots + (x_{ni} - x_{ni})^2}$$

Description:

$D_{(i,j)}$ = Euclidean distance between two data points i and j.

$x_{(ki)}$ = ith data on kth data attribute

$x_{(kj)}$ = The jth center point on the kth attribute

4. Updating the centroid position: After all the data has been clustered, recalculate the new centroid position for each cluster by calculating the average (mean) of all the data in the cluster. Alternatively, the median of the cluster can also be used.
5. Return to Step 3, if there is still data that moves clusters or if the change in centroid value is above the specified threshold value or if the change in the value of the objective function used is above the specified threshold value.

H. Data Validation and Reliability

Data validity was maintained by ensuring the accuracy of crime data extraction from the Mentawai Islands District Police and spatial data from official maps, which were then checked to avoid errors, while reliability was tested by repeating clustering using K-Means on a subset of the data to ensure consistency of results, with the help of GeoPandas, Folium, and reliable analysis tools.

RESULTS

A. Data Management

Data preprocessing in the analysis of crime in Mentawai Islands Regency begins with the selection of important attributes, namely data on the sub-district where the crime occurred, the type of crime, and the number of crimes. Next, data from the Mentawai Islands Regency Central Bureau of Statistics is merged, which includes population and population density. After data merging, the next step is data normalization to ensure that all attributes are on the same scale, so that the data can be analyzed with the K-Means Clustering method.

Table 1
Attribute Initialization

No.	Attributes	Initialize

1	Pencurian	PCR
2	Pencabulan	PCB
3	Penganiayaan	PNG
4	Narkotika	NKT
5	Perjudian	PJD
6	Kekerasan Dalam Rumah Tangga	KDRT
7	Jumlah Penduduk	JP
8	Kepadatan Penduduk	KP

Table 2

Mentawai Islands Regency Criminality Data

No.	District	PCR	PCB	PNG	NKT	PJD	KDRT	JP	KP
1	Pagai Selatan	2	4	1	0	0	0	9456	11.1
2	Pagai Utara	1	1	2	0	0	0	6181	17.61
3	Siberut Barat	1	3	1	0	0	0	7999	6.87
4	Siberut Barat Daya	2	2	1	0	0	0	7240	7.14
5	Siberut Selatan	3	1	5	3	0	1	10212	31.13
6	Siberut Tengah	2	1	0	0	0	0	7278	12.34
7	Siberut Utara	4	3	4	0	0	0	8414	10.75
8	Sikakap	13	4	7	4	0	1	10318	33
9	Sipora Selatan	11	2	5	5	4	1	10060	28.88
10	Sipora Utara	20	5	8	13	3	2	12577	46.17

After obtaining the criminal data, data normalization is carried out, where the method used for normalizing this data is MinMaxScaler. This normalization process changes the data values into a range between 0.1 to 0.9, so that each variable is on a uniform scale. This allows for a fairer comparison between variables.

Table 3

Data Normalization

No.	District	PCR	PCB	PNG	NKT	PJD	KDRT	JP	KP
1	Pagai Selatan	0.142	0.7	0.2	0.1	0.1	0.1	0.51	0.186
2	Pagai Utara	0.1	0.1	0.3	0.1	0.1	0.1	0.1	0.319
3	Siberut Barat	0.1	0.5	0.2	0.1	0.1	0.1	0.327	0.1
4	Siberut Barat Daya	0.142	0.3	0.2	0.1	0.1	0.1	0.232	0.105
5	Siberut Selatan	0.184	0.1	0.6	0.285	0.1	0.5	0.604	0.594
6	Siberut Tengah	0.142	0.1	0.1	0.1	0.1	0.1	0.237	0.211
7	Siberut Utara	0.226	0.5	0.5	0.1	0.1	0.1	0.379	0.179
8	Sikakap	0.605	0.7	0.8	0.346	0.1	0.5	0.617	0.632
9	Sipora Selatan	0.521	0.3	0.6	0.408	0.9	0.5	0.585	0.548
10	Sipora Utara	0.9	0.9	0.9	0.9	0.7	0.9	0.9	0.9

B. Clustering Process

The elbow method is used to determine the optimum number of K clusters. Basically, the elbow method is the SSE (Sum Square Error) method. The determination of the optimum K clusters can be seen in the graph of the SSE calculation results which are elbow-shaped or commonly called elbow points.

Figure 1 shows the graph of SSE (Sum of Square Error) against the number of K clusters. Based on the graph, the elbow point is located at K = 3. This point shows a significant decrease in SSE before stabilizing at a larger K value. Thus, it can be concluded that the optimum number of K clusters for this crime data is K = 3.

Scientica
Jurnal Ilmiah Sain dan Teknologi
(2025), 3 (4): 176-189 3021-8209

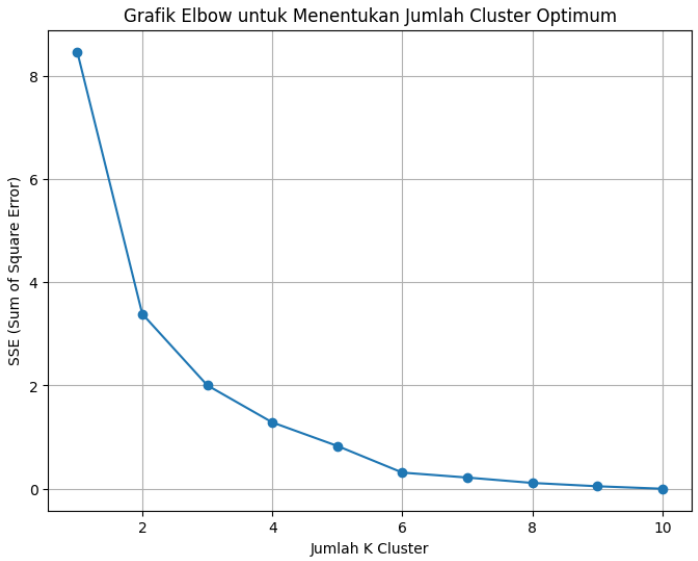

Figure 1 SEE Graph on Criminality Data

Based on the results of the analysis using the elbow method, the initial data centroid is obtained which shows K = 3 as the optimum number of clusters, then this crime data is grouped into three cluster centers, with randomly generated initial data centroids.

Table 4

Initial Data Centroid

Attributes	PCR	PCB	PNG	NKT	PJD	KDRT	JP	KP
C1	0.1	0.7	0.2	0.1	0.1	0.1	0.5	0.1
C2	0.1	0.6	0.2	0.1	0.1	0.5	0.6	0.5
C3	0.9	0.9	0.9	0.9	0.7	0.9	0.9	0.9

After determining the centroid of the initial cluster data, calculations are carried out to determine the results of data grouping on the clusters formed by doing the following calculations.

Distance of the second data to the first cluster center

$$d_{(1,1)}=\sqrt{\begin{matrix}((0.142-0.1)^2+(0.7-0.7)^2+(0.2-0.2)^2+(0.1-0.1)^2\\+(0.1-0.1)^2+(0.1-0.1)^2+(0.51-0.5)^2+(0.186-0.1)^2)\end{matrix}}=0.096$$

Distance of the second data to the first cluster center

$$d_{(1,2)} = \sqrt{\begin{matrix}((0.142-0.1)^2+(0.7-0.6)^2+(0.2-0.2)^2+(0.1-0.1)^2 \\ +(0.1-0.1)^2+(0.1-0.1)^2+(0.51-0.6)^2+(0.186-0.5)^2)\end{matrix}} = 0.528$$

Distance of the third data to the first cluster center

$$d_{(1,3)} = \sqrt{\begin{matrix}((0.142-0.9)^2+(0.7-0.9)^2+(0.2-0.9)^2+(0.1-0.9)^2 \\ +(0.1-0.7)^2+(0.1-0.9)^2+(0.51-0.9)^2+(0.186-0.9)^2)\end{matrix}} = 1.846$$

The following is the distance of the data to the cluster obtained after doing the calculation

Table 5

First Iteration Data Calculation Results

No.	District	C1	C2	C3	Cluster Member
1	Pagai Selatan	0.096	0.528	1.846	C1
2	Pagai Utara	0.760	0.838	2.063	C1
3	Siberut Barat	0.264	0.636	1.974	C1
4	Siberut Barat Daya	0.483	0.737	2.035	C1
5	Siberut Selatan	0.988	0.678	1.524	C2
6	Siberut Tengah	0.673	0.798	2.100	C1
7	Siberut Utara	0.408	0.654	1.790	C1
8	Sikakap	1.064	0.839	1.056	C2
9	Sipora Selatan	1.265	1.079	1.122	C2
10	Sipora Utara	1.900	1.622	0	C3

After obtaining the first iteration data clustering results, a new centroid is found for each cluster, which is then used to perform the next clustering. This process continues to be repeated until there is no cluster movement anymore, indicating that the clustering has stabilized.

Table 6

New Data Centroid

	PCR	PCB	PNG	NKT	PJD	KDRT	JP	KP
C1	0.142	0.367	0.25	0.1	0.1	0.1	0.298	0.183

C2	0.437	0.367	0.667	0.346	0.367	0.5	0.602	0.591
C3	0.9	0.9	0.9	0.9	0.7	0.9	0.9	0.9

After determining the centroid of the new data, calculations are carried out to determine the results of grouping new data against the clusters formed by doing the following calculations

Distance of the first data to the first cluster center

$$d_{(1,1)}=\sqrt{\begin{matrix}((0.142-0.142)^2+(0.7-0.367)^2+(0.2-0.25)^2+(0.1-0.1)^2\\+(0.1-0.1)^2+(0.1-0.1)^2+(0.51-0.298)^2+(0.186-0.183)^2)\end{matrix}}=0.398$$

Distance of the second data to the first cluster center

$$d_{(1,2)}=\sqrt{\begin{matrix}((0.142-0.437)^2+(0.7-0.367)^2+(0.2-0.667)^2+(0.1-0.346)^2\\+(0.1-0.367)^2+(0.1-0.5)^2+(0.51-0.602)^2+(0.186-0.591)^2)\end{matrix}}=0.938$$

Distance of the third data to the first cluster center

$$d_{(1,3)}=\sqrt{\begin{matrix}((0.142-0.9)^2+(0.7-0.9)^2+(0.2-0.9)^2+(0.1-0.9)^2\\+(0.1-0.7)^2+(0.1-0.9)^2+(0.51-0.9)^2+(0.186-0.9)^2)\end{matrix}}=1.846$$

The following is the distance of the data to the cluster obtained after doing the calculation

Table 7

Second Iteration Data Calculation Results

No.	District	C1	C2	C3	Cluster Member
1	Pagai Selatan	0.398	0.938	1.846	C1
2	Pagai Utara	0.364	0.968	2.063	C1
3	Siberut Barat	0.173	0.979	1.974	C1
4	Siberut Barat Daya	0.132	0.987	2.035	C1
5	Siberut Selatan	0.808	0.463	1.524	C2
6	Siberut Tengah	0.313	1.024	2.100	C1
7	Siberut Utara	0.307	0.776	1.790	C1
8	Sikakap	1.073	0.480	1.056	C2

9	Sipora Selatan	1.175	0.554	1.122	C2
10	Sipora Utara	1.949	1.147	0	C3

After searching for a new centroid in the second iteration, the clustering results show that there is no cluster movement anymore, which indicates that the clustering has stabilized and the clustering process can be stopped. Thus, the results in this second iteration can be considered as the final result of data clustering.

C. Testing with Rapidminer

Testing is done using RapidMiner Studio with the K-Means and Euclidean Distance algorithms to group data based on the similarity of its characteristics. This process aims to identify patterns in the data by forming several clusters that have similar characteristics.

Attribute	cluster_0	cluster_1	cluster_2
PCR	0.142	0.437	0.900
PCBL	0.367	0.367	0.900
PGN	0.250	0.667	0.900
NKT	0.100	0.346	0.900
PJD	0.100	0.367	0.700
KDRT	0.100	0.500	0.900
JP	0.298	0.602	0.900
KP	0.183	0.591	0.900

Figure 2 Centroid

Figure 2 shows the centroid data obtained from processing using RapidMiner with the K-Means and Euclidean Distance algorithms. This centroid table contains the average value of each attribute in each cluster formed, which represents the main characteristics of each data group.

Row No.	Kecamatan	cluster	PCR	PCBL	PGN	NKT	PJD	KDRT	JP	KP
1	Pagai Selatan	cluster_0	0.142	0.700	0.200	0.100	0.100	0.100	0.510	0.186
2	Pagai Utara	cluster_0	0.100	0.100	0.300	0.100	0.100	0.100	0.100	0.319
3	Siberut Barat	cluster_0	0.100	0.500	0.200	0.100	0.100	0.100	0.327	0.100
4	Siberut Barat ...	cluster_0	0.142	0.300	0.200	0.100	0.100	0.100	0.232	0.105
5	Siberut Selatan	cluster_1	0.184	0.100	0.600	0.285	0.100	0.500	0.604	0.594
6	Siberut Tengah	cluster_0	0.142	0.100	0.100	0.100	0.100	0.100	0.237	0.211
7	Siberut Utara	cluster_0	0.226	0.500	0.500	0.100	0.100	0.100	0.379	0.179
8	Sikakap	cluster_1	0.605	0.700	0.800	0.346	0.100	0.500	0.617	0.632
9	Sipora Selatan	cluster_1	0.521	0.300	0.600	0.408	0.900	0.500	0.585	0.548
10	Sipora Utara	cluster_2	0.900	0.900	0.900	0.900	0.700	0.900	0.900	0.900

Figure 3 Calculation Results

Figure 3 shows the results of data clustering using RapidMiner with the K-Means and Euclidean Distance algorithms, where each sub-district is grouped into three clusters based on the similarity of its attribute values, according to the centroid that has been calculated previously.

D. Visualization

This visualization displays the clustering results in the form of an interactive geographic map, which maps the sub-district area based on the cluster level that has been processed using the K-Means Clustering algorithm, thus providing a clearer picture of the distribution of crime in the Mentawai Islands Regency and facilitating analysis and location-based decision making.

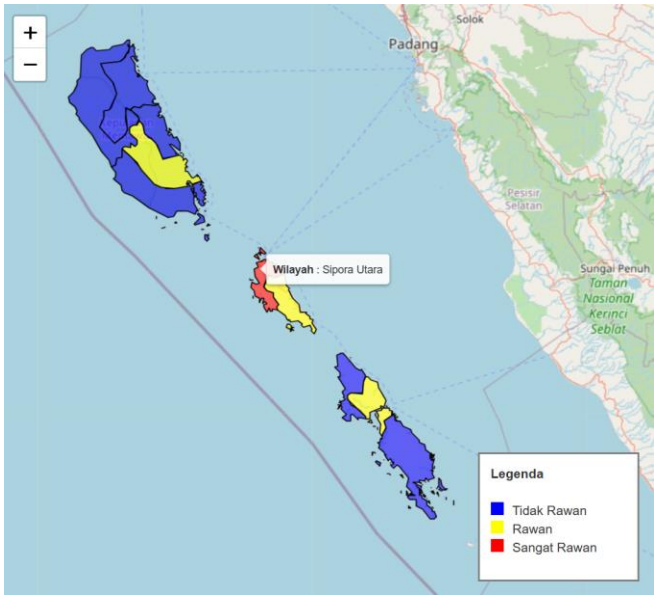

Figure 2 Visualization of GeoPandas and Folium

The map visualization will display the identity of the region when the mouse pointer is directed to a specific region on the map, as shown in Figure 2.

Scientica
(2025), 3 (4): 176-189 3021-8209
Jurnal Ilmiah Sain dan Teknologi

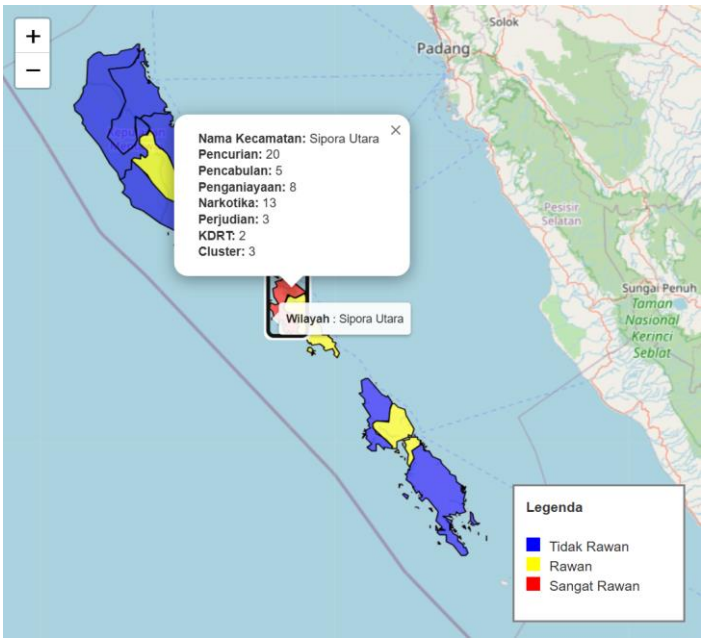

Figure 3 Visualization Description of GeoPandas and Folium

When a region object is clicked, information related to the region object will be displayed, as shown in Figure 3.

CONCLUSIONS

This research aims to identify, classify, and visualize the level of crime vulnerability in the Mentawai Islands Regency using the K-Means Clustering method and the Euclidean distance approach, based on crime data from 2021 to 2023. Based on the analysis result, the grouping of Mentawai Islands Regency area is done into three categories, namely: (a) Not Crime Prone: A total of 6 areas were identified in the Crime Prone category, these areas have a low crime rate and relatively low population density, (b) Crime Prone: 3 areas are classified in the Crime Prone category, these areas have a medium level of risk with a higher population density, and (c) Highly Crime Prone: 1 area is classified into the highly crime-prone category, this area has a very high risk with a significant number of crime cases and a very high population density.

To facilitate the understanding of the distribution of crime vulnerability in Mentawai Islands Regency, GeoPandas and Folium were used to visualize the clustering results in the form of interactive geographic maps. The use of these two tools proved effective in integrating spatial data and clustering analysis results, resulting in a more informative visualization that can support location-based decision making for crime control and prevention efforts in the region.

REFERENCE

Ali, M. I., & Rais, Muh. (2018). *Spatial pattern of crime with geographic information system (GIS) in Makassar, Indonesia*. Center for Open Science. http://dx.doi.org/10.31227/osf.io/fscv8

Aswan, Y., Defit, S., & Nurcahyo, G. W. (2021). Algoritma *K-Means Clustering* dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai). *Jurnal Sistim Informasi Dan Teknologi*, 245–250. https://doi.org/10.37034/jsisfotek.v3i4.73

Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis *K-Means Clustering* pada Data Sepeda Motor. *INFORMAL: Informatics Journal*, *5*(1), 10. https://doi.org/10.19184/isj.v5i1.17071

Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, *166*, 114060. https://doi.org/10.1016/j.eswa.2020.114060

Putra, A. D., Stevi Martha, G., Fikram, M., Yuhan, R. J., & Stis, P. S. (2020). *Faktor-Faktor yang Memengaruhi Tingkat Kriminalitas di Indonesia Tahun 2018*

Dony Novaliendry, Tegar Wibowo, Noper Ardi, Tiolina Evi, & Dwi Admojo. (2023). Optimizing patient medical records grouping through data mining and *K- Means clustering* algorithm: A case study at RSUD mohammad natsir solok. *International Journal of Online and Biomedical Engineering (iJOE)*, *19*(12), 144–155. https://doi.org/10.3991/ijoe.v19i12.42147

Fakhri, D. A., Defit, S., & Sumijan. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode *K-Means Clustering*. *Jurnal Informasi Dan Teknologi*, 160–166. https://doi.org/10.37034/jidt.v3i3.137

Irwan, I., Sanusi, W., & Saman, F. (2022). Pengelompokan Daerah Rawan Kriminalitas di Sulawesi Selatan Menggunakan Metode *K-Means Clustering*. *Journal of Mathematics Computations and Statistics*, *5*(1), 12. https://doi.org/10.35580/jmathcos.v5i1.32719

Puspitasari, D. I., Zaenuddin, Z., & Yuridka, F. (2019). Sistem Informasi Geografi (SIG) Pencarian Lokasi Tambal Ban dengan Pemanfaatan Teknologi GPS. *Elinvo (Electronics, Informatics, and Vocational Education)*, *4*(1), 30–38. https://doi.org/10.21831/elinvo.v4i1.23146

Rosiana, P. S., Mohsa, A. A., Fadila, M. A., & Jaman, J. H. (2023). VISUALISASI DATA TINDAK KEJAHATAN BERDASARKAN JENIS KRIMINALITAS DI KABUPATEN KARAWANG DENGAN MENGGUNAKAN ALGORITMA *CLUSTERING K-MEANS*. *Jurnal Informatika Dan Teknik Elektro Terapan*, *11*(3s1). https://doi.org/10.23960/jitet.v11i3s1.3347

Siahaan, V., & Sianipar, R. H. (2019). *TEORI DAN PRAKTEK PEMROGRAMAN PYTHON*. SPARTA PUBSLISHING.

Simanjuntak, A. Y., Simatupang, I. S. E. S., & Anita, A. (2022). IMPLEMENTASI DATA MINING MENGGUNAKAN METODE NAÃVE BAYES CLASSIFIER UNTUK DATA KENAIKAN PANGKAT DINAS KETENAGAKERJAAN KOTA MEDAN. *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, *5*(1), 85. https://doi.org/10.54314/jssr.v5i1.804

Stedila, S., Astuti, R., & M Basysyar, F. (2024). PENERAPAN DATA MINING *CLUSTERING* MENGGUNAKAN METODE *K-MEANS* PADA DATA TINDAK

KRIMINALITAS DI POLRES KABUPATEN KUNINGAN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1629–1636. https://doi.org/10.36040/jati.v8i2.8790

Surbakti, N. K. (2021). Data mining pengelompokan pasien rawat inap peserta BPJS menggunakan metode *clustering* (Studi Kasus : RSU.Bangkatan). *Journal of Information and Technology*, *1*(2), 47–53. https://doi.org/10.32938/jitu.v1i2.1470

Suryoputro, A. L., & Prasetyo, S. Y. J. (2023). Prediksi dan visualisasi penyakit COVID-19 menggunakan kombinasi Prophet dan *GeoPandas*. *AITI*, *20*(2), 135–149. https://doi.org/10.24246/aiti.v20i2.135-149

Wicaksono, D. A., & Susetyo, Y. A. (2023). *CLUSTERING* ZONASI DAERAH RAWAN BENCANA ALAM DI PROVINSI SUMATERA BARAT MENGGUNAKAN ALGORITMA *K-MEANS* DAN *LIBRARY GEOPANDAS*. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *4*(2), 426–438. https://doi.org/10.35870/jimik.v4i2.225

Yan, X., Liu, S., Liu, W., & Dai, Q. (2022). An improved coverage-oriented retrieval algorithm for large-area remote sensing data. *International Journal of Digital Earth*, *15*(1), 606–625. https://doi.org/10.1080/17538947.2022.2030816

Zahra, F., Ridla, M. A., & Azise, N. (2024). IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN BARANG (STUDI KASUS : TOKO SINAR HARAHAP). JUSTIFY : Jurnal Sistem Informasi Ibrahimy, 3(1), 55–65. https://doi.org/10.35316/justify.v3i1.5335

Maori, N. A. (2023). METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING. Jurnal SIMETRIS, 14.

Rahman Salam, R., Fajri Jamil, M., & Ibrahim, Y. (2023). MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak. Menggunakan Support Vector Machine. 3, 27–35

Rahmansyah, N., & Lusinia, S. A. (2022). Analisa Sistem Pendukung Keputusan Kemiskinan Menurut Kabupaten Kota Provinsi Sumatera Barat. Jurnal Karya Ilmiah Multidisiplin (JURKIM), 2(1), 76–82.