Geographic Information Systems for Crime Prone Areas Clustering

Heti Mulyani[1], Jajang Nurjaman[2], Muhammad Nugraha[3]
[1,3] Politeknik Enjinering Indorama, Indonesia
[2]Sekolah Tinggi Teknologi Indonesia Tanjungpinang, Indonesia

Article Info	ABSTRACT
	Crime is one of the problems that is quite complicated and very disturbing to the community. Crimes can occur at different times and places, making it difficult to track which areas are prone to such actions. K-means algorithm is used to cluster prone areas and Geographic Information System is used to map crime-prone areas. Web-based application is developed with the PHP programming language. The data used is quantitative data in the form of the number of crimes committed and the coordinates of the cases. The attributes of the crime used consist of five parameters: theft, mistreatment, rape, women and child protection cases and fraud. The results of this study are clustering areas into 3 cluster and mapping prone areas that is safe area, safe enough area and prone area. From the overall crime data for 2019 in Purwakarta district, it was found that 68.75% was safe, 18.75% was quite safe and 12.5% was prone area.

Corresponding Author:

Heti Mulyani,
Politeknik Enjinering Indorama,
Indonesia
Email: heti.mulyani22@gmail.com

1. INTRODUCTION

Crime is one of the problems that is quite complicated and very disturbing to the community. Crimes can occur at different times and places, making it difficult to trace which areas are prone to such actions. Information about crime is very much needed by the community, both local and migrants as an anticipation material, especially in Purwakarta Regency. The population density of Purwakarta Regency has increased from the number of 2017 as many as 971 people per Km2, with an unemployment rate of 9.89% [1]. As the population growth rate and high unemployment rate, causing crime rates in Purwakarta is also still relatively high. At the end of 2019 Purwakarta Regional Police reported 1140 criminal cases. Crime is dominated by theft and fraud. Data mining is a discipline that is expected to help solve the problem of crime in Purwakarta Regency. The technique used is a cluster with k-means. This method is used because k-means is a simple and fast cluster process. With this method the determination of vulnerable areas can be done more easily and accurately. In addition to cluster techniques, this study also utilizes Geographic Information Systems (GIS) where GIS is a tool for entering, retrieving, manipulating, analyzing and displaying geographic or geospatial data attribute data, making the mapping process easier.

Several studies of crime have been carried out including K-means for clustering crime data as made research on geographical information systems for the classification of crime-prone areas using the K-Means method [2]. In this study, the application can display vulnerability in the Cilacap district. This application is able to display reports of crime-prone areas per year. Subsequent research crime prediction using K-means Clustering [3], this research show common crime patterns and crime prone areas for future prediction. The next research crime analysis using K-means Clustering [4], this research was made cluster and trend crime data using rapidminer. The next research is discussed the analysis of the k-means method in grouping crime by region he next research is discussed the analysis of the k-means method in grouping crime by region [5], in this study the grouping of regions was based on 2 clusters, namely high and low, but in this study the mapping process was not carried out and the data used were still static . Subsequent research, conducted a study with the title grouping criminal data on the police to determine patterns of crime-prone areas using K-means clustering data mining algorithm [6]. In this study, researchers grouped the level of vulnerability of the police

in the North Sumatra Regional Police, but data processing is still static, so that if there is new crime data can not be input into the system. The next research is application of k-means algorithm to know the level of crime action in the region of Pematangsiantar [7]. In this study, researchers grouping area with k-means at pematangsiantar becomes two area, but still using static data.

In this study used are quantitative data, that is data that can be calculated in the form of numbers or nominal. The aplication will be depelove for dynamic clustering crime areas. Crime data is quantitative data so that it can be calculated using k-means. There are 6 types of criminal parameters used, namely narcotics, theft, persecution, rape, gambling and fraud. This study aims to provide information to the public and the police regarding the level of vulnerability in Purwakarta regency. For migrants, information on this crime can be used as reference material for choosing a place to live, especially in Purwakarta district, while for the police, this information can be used to find out which areas require extra supervision. In this study the next will discuss the research methodology in section 2, results and discussion in section 3, conclusions in section 4.

2. METHOD

In this study, there are several methodological steps used, namely: Crime Data Analysis, K-Means Clustering Analysis, and Design Model Analysis.

2.1 Crime Data Analysis

In this study, the data used are Purwakarta regency crime data obtained from Purwakarta Police Precinct. The parameters used consisted of 6 attributes, namely: theft, fraud, mistreatment, rape, and women and child protection cases. The data purwakarta district crime in 2019 can be seen from table 3.1

Table 2.1 Crime data quantity at Purwakarta Regency (2019)

No	Districts	Theft	Fraud	Mistreatment	Rape	PPA
1	BabakanCikao	75	31	10	0	0
2	Bojong	2	3	3	0	0
3	Bungursari	28	15	6	0	0
4	Campaka	8	5	3	0	0
5	Cibatu	3	2	3	0	0
6	Darangdan	31	19	5	0	1
7	Jatiluhur	43	29	15	0	0
8	Kiarapedes	1	0	0	0	0
9	Pasawahan	19	13	4	0	0
10	Plered	15	3	4	2	12
11	Pondoksalam	11	1	4	0	0
12	Purwakarta	78	32	4	0	21
13	Sukasari	7	1	3	0	0
14	Sukatani	2	2	4	5	15
15	Tegalwaru/Maniis	5	0	3	0	0
16	Wanayasa	1	0	0	0	1

Source: Resort Police of Purwakarta

2.2 K-Means clustering Analysis

Clustering is a technique of portioning object into group [8]. One of the clustering methods is K-Means algorithm. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem [9]. K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters [10]. The data grouping method used is K-Means clustering. Following are the steps of the K-Means Clustering algorithm [6]:

1. Randomly select 'K' cluster centers
2. Calculate the distance between each data point and cluster centers with Euclidean formula
$$D(ik) = \sqrt{\Sigma(Cij - Ckj)^2} \qquad (1)$$
3. Grouping the data into clusters with the shortest distance using equations
$$Min \sum_{k=1}^{k} dik = \sqrt{\sum_{j}^{m}(Cij - Ckj)^2} \qquad (2)$$
4. Recalculate the new cluster center using equation
$$Ckj = (\sum_{i=1}^{p} xij)/p \qquad (3)$$
5. Recalculate the distance between each data point and new obtained cluster centers.

2.2. Design Model Analysis

The proposed design model uses Unified Manipulation Language (UML). Unified Modeling Language (UML) is a language based on graphics or images to visualize, specify, construct, and document

Sistem Informasi Geografis for Crime Prones Areas in Regency Purwakar
(Heti Mulyani[1], Jajang nurjaman[2], Muhammad Nugraha[3])

228

from an OO (Object-Oriented) based software development system [11]. Design was depelove to answer the requirement analysis. Design consists of use case diagram, sequence diagram and class diagram. Each diagram will be defined in the subsection as follows.

2.2.1. Use case Diagram

Use case diagrams illustrate the interaction between one or more actors with the system created. Use case diagram of GIS grouping of areas prone to crime in Purwakarta can be seen in Figure 1.

Figure 1. Use Case Diagram

Figure 1 show interaction admin, police and public. Admin and police access data crime and clustering. Public only access data GIS crime prone areas.

2.2.2. Sequence Diagram

Sequence diagrams illustrate messages sent between objects and interactions between objects. Figure 2 shows the data input process by the admin, then the system will do clustering with K-Means and provide output in the form of maps of crime grouping results to the police and the community.
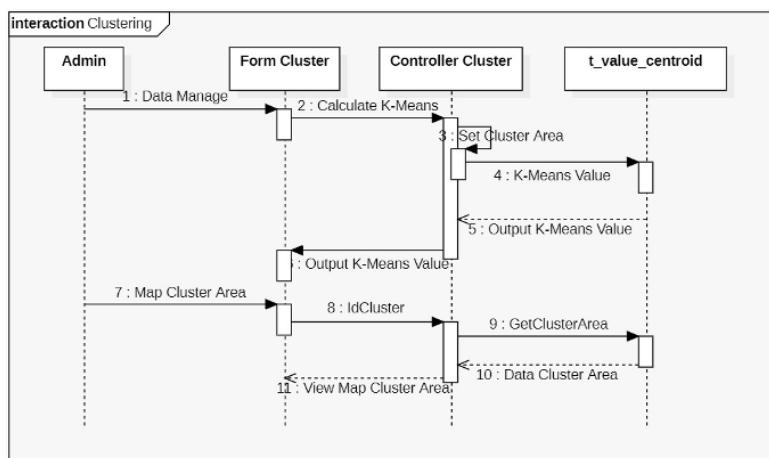
Figure 2. Sequence Diagram

Figure 2 shows a sequence clustering area diagram. Admin manages the data displayed on the form cluster. K-Means calculations are carried out in order to obtain the cluster area. The results of the k-means calculated are then displayed in map cluster area.

2.2.3. Class Diagram

Class diagram is an illustration the overall system by showing its classes and describes the relationship between classes [12]. Figure 3 shows the class diagram Geographic Information System for Crime Prone Areas Clustering in Purwakarta Regency
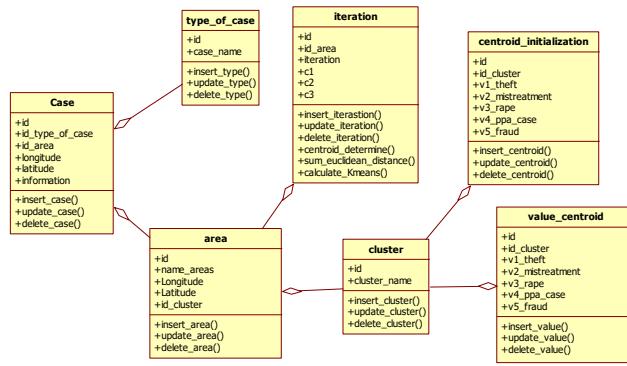
Figure 3. Class diagram

3. RESULTS AND DISCUSSION

This application was implemented using PHP programming language. It's can insert areas, data crime case, calculate k-mean and divide areas according of areas. The application divides groups of regions based on 5 determined attributes, namely theft, mistreatment, rape, fraud and women and child protection cases. Figure 4 shows the implementation for insert district areas in purwakarta regency. Figure 5 shows implementation for insert of data crime cases. Figure 6 shows calculated clustering areas using k-means clustering. In this study using k=3 for clustering areas is safe (C1), middle (C2) and prone (C3). Figure 7 shows map cluster areas.

3.1. Insert District Areas
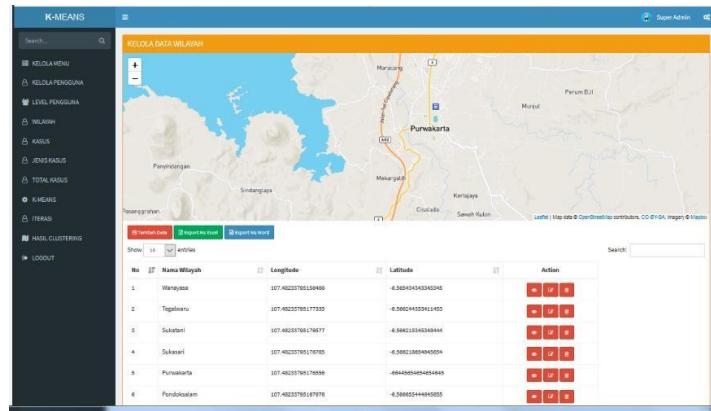
Figure 4 Insert district areas

Figure 4 shows implementation for insert district areas. It can insert areas data, longitude, latitude and shows map of district areas in Purwakarta Regency.

3.2. Insert Crime Data
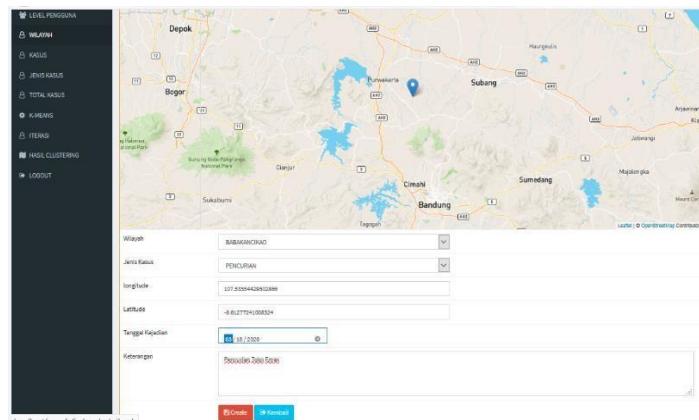
Figure 5 Insert crime data case

Sistem Informasi Geografis for Crime Prones Areas in Regency Purwakar
(Heti Mulyani[1], Jajang nurjaman[2], Muhammad Nugraha[3])

230

Figure 5 shows implementation for insert crime data case. The case crime others is theft, mistreatment, rape, fraud and WCP case.

3.3. Clustering with K-Means

Figure 6 calculated cluster areas with k-means

Based on Figure 6 shows the regional cluster, which is divided into 3 clusters, namely cluster 1 (safe), cluster 2 (safe enough), cluster 3 (vulnerable). From the overall crime data for 2019 in Purwakarta district, it was found that 68.75% was safe, 18.75% was quite safe and 12.5% was prone.

3.4. Map Cluster Areas
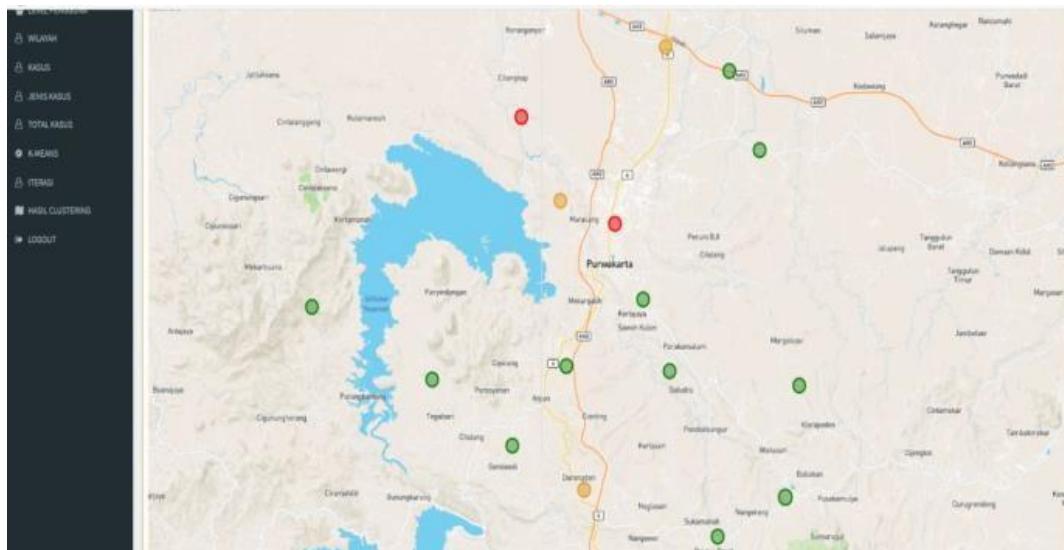
Figure 7 Map cluster areas

Figure 7 shows cluster areas. Areas districts are represented by dots. For safe areas are marked with a green dot, are quite safe areas marked in yellow, prone areas to be marked in the red. Based on map, safe cluster area consists of Wanayasa, Tegalwaru, Sukatani, Sukasari, Pondoksalam, Plered, Pasawahan, Kiarapedes, Cibatu, Campaka. Safe enough cluster area consists of Jatiluhur, Darangdan, Bojong. Prone cluster area consists of Purwakarta and Babakancikao.

4. CONCLUSION

Above discussion indicated that the application was successfully developed by observating parameter attribute crime data. The application is able to divide the areas into 3 clusters which are safe, middle and prone. From the overall crime data for 2019 in Purwakarta district, it was found that 68.75% area was safe consists of Wanayasa, Tegalwaru, Sukatani, Sukasari, Pondoksalam, Plered, Pasawahan, Kiarapedes, Cibatu, Campaka, 18.75% was quite safe consists of Jatiluhur, Darangdan, Bojong and 12.5% area was prone consists of Purwakarta and Babakncikao. The implementation Geographic Information System for cluster areas can be accessed by the police that can be used to manage crime data, to find out the level of vulnerability, and can be accessed by the public to find out the map of crime vulnerability. For further research conduct analysis on a small area, such as the village level, in order to obtain more detailed results, Comparing the results of mapping crime-prone areas with other clustering methods as K-Medoids etc.

ACKNOWLEDGEMENTS

REFERENCES

[1]   BPS, "Kabupaten Purwakarta dalam Angka 2019," *Badan Pus. Stat. Kabupaten Purwakarta*, 2019.

[2]   F. D. Yuliansyah, "Sistem Informasi Geografis Untuk Klasifikasi Daerah Rawan Kriminalitas Menggunakan Metode K-Means," *Univ. Islam Indones.*, p. Yogyakarta, 2016.

[3]   V. Jain, Y. Sharma, A. Bhatia, and V. Arora, "Crime Prediction using K-means Algorithm," *GRD Journals - Glob. Res. Dev. J. Eng.*, vol. 2, no. 5, pp. 206–209, 2017, [Online]. Available: www.grdjournals.com.

[4]   Chetan G. Wadhai, Tiksha P. Kakade, Dnyaneshwari S. Tumsare, Khushabu A. Bokde, and Prof. Deepa Bhattacharya, "Crime Detection Technique Using Data Mining and K-Means," *Int. J. Eng. Res.*, vol. V7, no. 02, pp. 223–226, 2018, doi: 10.17577/ijertv7is020110.

[5]   S. M. Dewi, A. P. Windarto, I. S. Damanik, and H. Satria, "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah," *Semin. Nas. Sains Teknol. Inf.*, pp. 620–625, 2019.

[6]   L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 151, 2020, doi: 10.30865/json.v1i2.1955.

[7]   D. Gultom *et al.*, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020.

[8]   H. Rohul, M. Tambun, and I. Pendahuluan, "Rancang Bangun Aplikasi Data Mining Untuk," vol. 1, no. 1, pp. 31–38, 2015.

[9]   M. S. Davis, "Crime Rate," *Concise Dict. Crime Justice*, pp. 252–258, 2012, doi: 10.4135/9781452229300.n522.

[10]   D. Kartini, "Rancang Bangun Aplikasi K-Means untuk Klasifikasi Kelulusan Siswa Sekolah Kepolisian Negara Daerah Kalimantan Selatan," *ProTekInfo*, vol. 3, no. 1, pp. 14–21, 2016.

[11]   Suendri, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 1–9, 2018, [Online]. Available: http://jurnal.uinsu.ac.id/index.php/algoritma/article/download/3148/1871.

[12]   H. Mulyani, T. Djatna, and I. S. Sitanggang, "Agent based modeling on dynamic spreading dengue fever epidemic," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 3, pp. 1380–1388, 2017, doi: 10.12928/TELKOMNIKA.v15i3.4511.