Journal of Physics: Conference Series

PAPER • OPEN ACCESS

K-Means method for analysis of accident-prone areas in Palangka Raya

To cite this article: H Sitohang *et al* 2020 *J. Phys.: Conf. Ser.* **1566** 012008

K-Means method for analysis of accident-prone areas in Palangka Raya

H Sitohang[1*], Rosmiati[1*] and S Merni[1*]

[1] College of Informatics and Computer Management (STMIK Palangkaraya), Jl. G. Obos Number.114, Central Borneo, Indonesia

*Email: hotmiansitohang@gmail.com, fayadhah@gmail.com, merni@gmail.com

Abstract. Traffic accidents are the most feared thing for everyone. Even though people are careful using vehicles, accidents can occur. This can be caused by humans not obeying traffic rules, drowsiness, damaged roads, and bad weather. In recent years the accident rate in Palangka Raya has increased sharply and even resulted in death. Accidents often occur in different places and times, making it difficult to determine accident-prone areas. By utilizing data mining techniques using the K-Means method, it can explore information on areas where accidents occur frequently. So that it helps the police in providing signs of accident-prone areas. The results of the study stated the K-Means Clustering algorithm in determining the accident-prone point of success obtained by 68%.

1. Introduction

The rapid development of transportation will indirectly increase the risk of growing problems in traffic and create a high level of risk of accidents. According to RI Law No. 22 of 2009, a traffic accident is an unexpected and unintentional event on the Road involving vehicles with or without other road users resulting in human casualties. Traffic accidents generally occur because of various factors including careless drivers, damaged road conditions, vehicle conditions, and weather. Traffic violations are quite high and the ownership of private vehicles is increasingly increasing, this will indirectly trigger a traffic accident. Palangka Raya is the capital of Central Kalimantan Province and consists of Five Districts and one that has a very wide administrative area. Palangka Raya is experiencing very rapid development in various fields. This development is supported by an increase in population, and an increase in community activity, an increase in community income, and an increase in the number of vehicles. With the increase in the number of vehicles in Palangka Raya, the level of traffic accidents that occur is also very large. Based on data revealed by Palangka Raya Regional Police from 2016 to 2018 there were accidents that were so increasing as in table 1.

Table 1. Accident data in Palangka Raya city.

Street Name	Accidents Per Year		
	2016	**2017**	**2018**
Cilik Riwut	59	44	39
Yos Sudarso	4	8	4
Imam Bonjol	3	2	2
Pangeran Diponegoro	8	2	1
Goerge Obos	20	11	13
Jendral Ahmad Yani	2	-	2
Mahir Mahar	17	17	16
Trans Kalimantan	4	1	1
Piere Tendean	7	8	1
RTA Milono	43	32	15

From table 1 shows from year to year traffic accidents are increasing. This should be a concern of the Palangka Raya district police to determine policies to reduce traffic accidents.

The application of the K-means method can be used for the analysis of accident-prone areas, as conducted by Brilian et al [1], conducting research into the implementation of K-Means clustering on rapid miner for analysis of accident-prone areas. From 2620 accident data recorded in Kendari resort database, the data was selected into 500 data. The data is then analyzed using the K-Means clustering algorithm with the help of the RapidMiner Studio application. The results of the analysis show the frequency of accident rates at each location along with the act-time-prone potential for accident cases. Likewise, research conducted by Dwi et. al [2], made a mapping of accident-prone areas using the Fuzzy C-Means method. The results of the grouping of accident data were visualized using a map that illustrates the mapping of the results of clustering in the Pasuruan City Police area.

From the above problems the K-Means method is truly tested to determine accident-prone areas, but each data of each accident-prone area is very different. So that the K-Means algorithm in the analysis of accident data in the City of Palangka Raya is very feasible to use for grouping accident-prone areas.

2. Research Method

2.1. Dataset

The dataset used in this study was from Palangka Raya Regional Police accident data from 2016 to 2018 as many as 386 accident cases with different locations. Data used for analysis as follows:

Table 2. List of Palangka Raya district police accident data.

Street Name	2016		2017		2018	
	Category	**Amount**	**Category**	**Amount**	**Category**	**Amount**
Cilik Riwut	D	53	D	34	D	31
	SI	3	SI	5	SI	0
	MI	3	MI	5	MI	8
Yos Sudarso	D	4	D	8	D	3
	SI	0	SI	0	SI	0
	MI	0	MI	0	MI	1
Imam Bonjol	D	3	D	2	D	2
	SI	0	SI	0	SI	0
	MI	0	MI	0	MI	0
Pangeran Diponegoro	D	8	D	1	D	1
	SI	0	SI	1	SI	0

	MI	0	MI	0	MI	0
Goerge Obos	D	18	D	10	D	11
	SI	2	SI	0	SI	0
	MI	0	MI	1	MI	2
Jendral Ahmad Yani	D	2	D	0	D	2
	SI	0	SI	0	SI	0
	MI	0	MI	0	MI	0
Mahir Mahar	D	11	D	12	D	13
	SI	1	SI	0	SI	0
	MI	5	MI	5	MI	3
Trans Kalimantan	D	2	D	1	D	1
	SI	1	SI	0	SI	0
	MI	1	MI	0	MI	0
Piere Tendean	D	5	D	7	D	1
	SI	0	SI	1	SI	0
	MI	2	MI	0	MI	0
RTA Milono	D	36	D	28	D	13
	SI	4	SI	2	SI	0
	MI	3	MI	2	MI	2

Explanations:
D: Death
SI: Seriously Injured
MI: Minor Injuries

2.2. K-Means

K-Means is a data analysis method that groups data by system partition. The K-Means algorithm is simple to implement, relatively fast, adaptable and has a general use in practice. A number of representative cluster K is also referred to as cluster means one cluster centroid (or centroid only). For data sets in X grouped based on the concept of closeness or similarity. Although the concept intended for data gathered in one cluster is similar data, the quantity used to measure it is dissimilarity. Because K-Means categorizes explicitly and only on one cluster, then from the value of a before the data in all clusters, only one has a value of 1, while the other 0 is like the following equation:

$$aij = \begin{cases} 1 \arg \min_{j} \{d(Xi, Cj)\} \\ 2 \; the \; other \end{cases} \tag{1}$$

d(x*i*,c *j*) expressed dissatisfaction (data distance to-*i* to *cluster cj*. While relocating centroids to get C centroid points is obtained by calculating the average of each feature of all data incorporated in each cluster. A feature average of all data in a cluster is expressed by the following equation.

$$cj = \frac{1}{Nk} \sum_{i=1}^{Nk} Xji \tag{2}$$

Nk is the amount of data joined in a cluster. If we look at the steps that always choose the closest cluster, then K-Means actually tries to minimize the objective function / non-negative cost function, as stated by the following equation:

$$J = \sum_{i=1}^{N} \sum_{I=1}^{K} a_{ic} \, d(Xi, CI)^2 \tag{3}$$

In other words, K-Means tries to minimize the total squared distance between each X point and the closest Cj cluster representation.

K-Means method is a method of grouping nonhierarchical data (partition) that seeks to keep the existing data in the form of two or more groups [3]. This method partitions the data into groups so that the same character data is entered into the same group and different characteristic data are grouped into other groups. In grouping data by the K-Means method generally done with the algorithm-algorithm as follows:

- Determine the number of groups

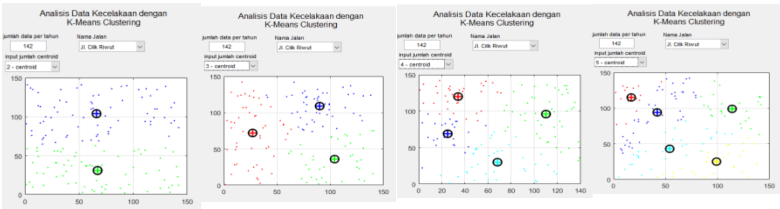
- Allocate data into groups randomly
- Calculate the center of the group (centroid/average) of the data in each group
- Allocate each data to the nearest centroid/average
- Return to step 3, if there is still data that moves between groups, or if there is a change in the centroid value above the specified threshold value, or if the change in the value of the objective function used is still above the specified threshold value [3].

3. Result and discussion

Matlab (matrix laboratory) is a high-level language that is interactive and allows users to do computing intensively [4]. Matlab has developed into an advanced programming environment that contains built-in functions for managing signals, linear algebra, and other mathematical calculations. MatLab also contains a toolbox that includes additional functions for special applications. The results of the analysis of accident data from each road that occurred in a traffic accident are as follows:

3.1. On Jalan Cilik Riwut

With a total data of 142 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 1.

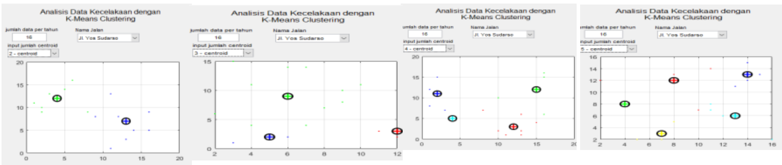


(a) (b) (c) (d)

Figure 1. Results of the crash analysis on Jalan Cilik Riwut.

3.2. On Jalan Yos Sudarso

With a total data of 16 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 2.

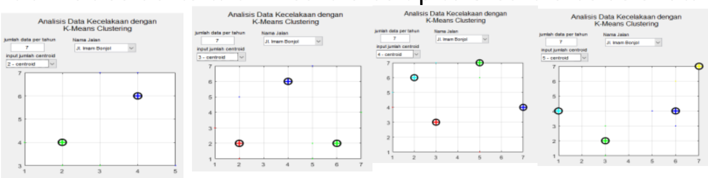


(a) (b) (c) (d)

Figure 2. Results of the crash analysis on Jalan Yos Sudarso.

3.3. On Jalan Iman Bonjol

With a total of 7 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 3.

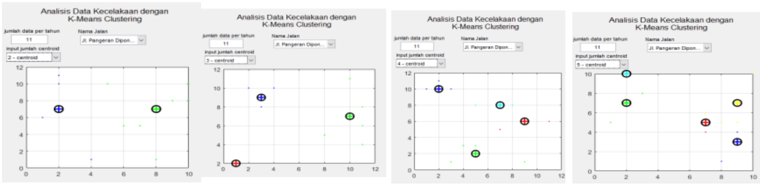


(a) (b) (c) (d)

Figure 3. Results of the crash analysis on Jalan Imam Bonjol.

3.4. On Jalan Pangeran Diponegoro

With a total of 11 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 4.

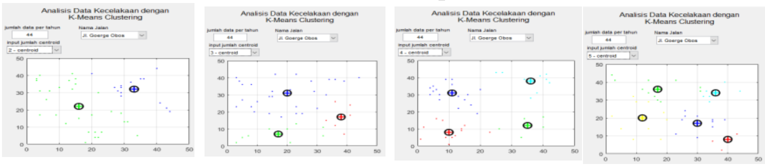


(a) (b) (c) (d)

Figure 4. Results of the crash analysis on Jalan Pangeran Diponegoro.

3.5. On Jalan George Obos

With a total of 44 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 5.

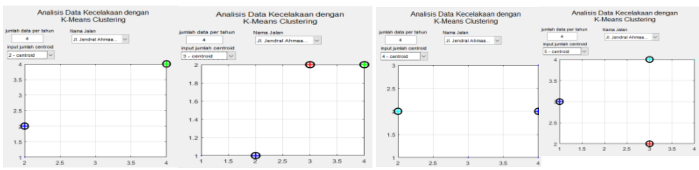


(a) (b) (c) (d)

Figure 5. Results of the crash analysis on Jalan George Obos.

3.6. On Jalan Ahmad Yani

With the amount of data 4 times the incidence of traffic accidents with 2 centroids up to 5 centroids as shown in Figure 6.

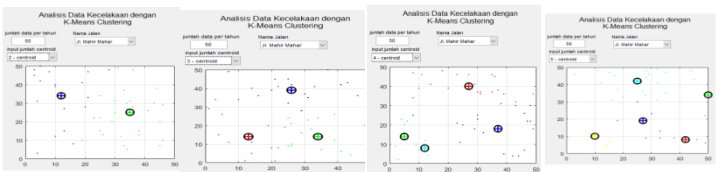


(a) (b) (c) (d)

Figure 6. Results of the crash analysis on Jalan Ahmad Yani.

3.7. On Jalan Mahir Mahar

With a total of 50 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 7.

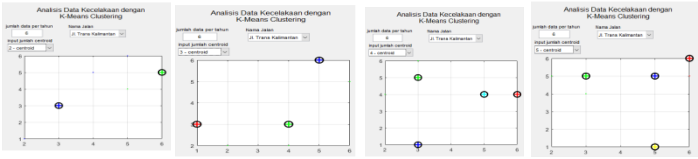


(a) (b) (c) (d)

Figure 7. Results of the crash analysis on Jalan Mahir Mahar.

3.8. On Jalan Trans Kalimantan

With a total of 6 Kalimantan data, traffic accidents with 2 centroids up to 5 centroids are shown in Figure 8.



(a) (b) (c) (d)

Figure 8. Results of Crash Analysis on Jalan Trans Kalimantan.

3.9. On Jalan Piere Tendean

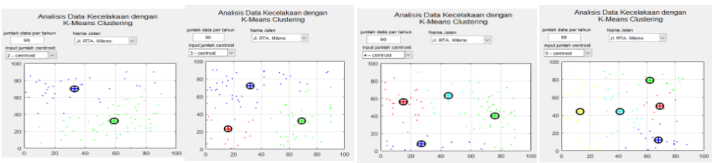
With a total of 16 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 9.

(a) (b) (c) (d)

Figure 9. Results of the crash analysis on Jalan Piere Tendean.

3.10. On Jalan RTA Milono

With the total data of 90 traffic accidents with 2 centroids up to 5 centroids as shown in Figure 10.



(a) (b) (c) (d)

Figure 10. Results of the crash analysis on Jalan RTA Milono.

From the results of K-Means clustering analysis with MatLab can be concluded in table 3.

Table 3. Results of the K-Means clustering analysis.

Street Name	2-centroid	3-centroid	4-centroid	5-centroid
Cilik Riwut	Prone	Very vulnerable	Not Prone	Prone
Yos Sudarso	Prone	Not Prone	Very vulnerable	Prone
Imam Bonjol	Prone	Prone	Not Prone	Prone
Diponegoro	Not Prone	Not Prone	Not Prone	Prone
Goerge Obos	Not Prone	Prone	Prone	Not Prone
Ahmad Yani	Not Prone	Not Prone	Prone	Very Vulnerable
Mahir Mahar	Not Prone	Prone	Very Vulnerable	Very Vulnerable
Transkalimantan	Prone	Prone	Prone	Very Vulnerable
Piere Tendean	Prone	Prone	Verys Vulnerable	Very Vulnerable
RTA Milono	Prone	Prone	Not Prone	Prone

The new centroid is calculated using the average value of data in each cluster, if the new centroid is different from the previous centroid, then the process continues to the next step. But if the new centroid is calculated the same as the previous centroid, then the clustering process is completed.

4. Conclusion

From this study, it can be concluded that the K-Means Clustering algorithm in determining accident-prone points of success obtained by 68%. With this analysis, Palangka Raya Resort Police Traffic Unit can determine which areas are prone to accidents. So that the police can make a decision by making signs or billboards on every accident-prone road.

References

[1] Brilliant Rahmat et al., *"Implementation of K-Means Clustering at Rapidminers for Analysis of Accident Prone Areas,"* in the National Seminar on Applied Quantitative Research 2017, Makassar, 2017, pp. 58-62.
[2] Dwi Puspitasari, Yan Watequlis Syaifudin, and Resqy Dwi Nofyandi, *"Mapping of Accident Prone Areas Using Fuzzy C-Means Method,"* Informatics Polinema, vol. 5, no. 2, pp. 90-95, February 2019.
[3] Eko Prasetyo, Data Mining processes data into information using matlab, Andi. Jogyakarta, 2014.
[4] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann, 2012.
[5] Lizda Iswari and Ervina Gita Ayu, *"Utilization of K-MEANS Algorithm for Results Mapping," Teknoin*, vol. 21, no. 1, p. 13, Maret 2015.
[6] Nameirakpam Dhanacahndra, Khumanthem Manglem, and Yambem Jina Chanu, *"Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm," Elsevier*, no. 54, pp. 764-771, 2015.
[7] Setyaningsih, "Analysis of Voice Data Recording from the Blackbox System on Railways," Informatics, vol. 9, no. 1, pp. 993-998, January 2015.
[8] Winda Aprianti and Jaka Permadi, "K-MEANS Clustering for Traffic Accident Data," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 5, no. 5, pp. 613-619, Oktober 2018.