RESEARCH ARTICLE OPEN ACCESS

Determination of Accident Prone Level with Fuzzy C-Means Algorithm on Bakauheni-Terbanggi Besar Toll Road Section

Muhamad Imi*, Reva Ragam Santika**
*Universitas Budi Luhur, Indonesia
Email: muhamadimi8@gmail.com
**Universitas Budi Luhur, Indonesia
Email: reva.ragam@budiluhur.ac.id

--************************----------------------------------

Abstract:

Toll roads are referred to as freeways that have many differences from public roads. Although toll roads are designed to improve traffic safety with features such as separated lanes, well-designed geometry, and set speed limits, there is still the potential for accidents or other problems to occur. Currently, accidents have been the cause of many fatalities in Indonesia, especially on the Bakauheni-Terbanggi Besar toll road. Therefore, this research aims to build an accident vulnerability level clustering system using the Fuzzy C-Means clustering algorithm on the Bakauheni-Terbanggi Besar toll road section based on the CodeIgniter Framework. Clustering the level of accident vulnerability based on the Bakauheni-Terbanggi Besar toll road section into 3 (three) groups, namely 1) Prone, 2) Moderately Prone, and 3) Safe. From 300 accident data, 142 data samples were used with data grouping indicators using the Fuzzy C-Means algorithm and based on three criteria, namely the number of accidents, the number of minor injuries, and the number of serious injuries. The results of the objective function calculation, until the stop condition is reached at the 28th iteration, the level of accident vulnerability is determined by 3 (three) clusters, namely (i) road sections classified as prone clusters as many as 15 points, (ii) road sections classified as moderately prone clusters as many as 35 points, (iii) road sections classified as safe clusters as many as 92 locations based on 3 criteria with the calculation of the fuzzy C-means clustering algorithm.

Keywords — **Put your keywords here, keywords are separated by comma.**

--************************----------------------------------

I. INTRODUCTION

The Bakauheni-Terbanggi Besar Toll Road is a toll road in Lampung Province, Indonesia. This toll road is part of the Trans Sumatra Toll Road that connects Bakauheni at the southern tip of the island of Sumatra with Terbanggi Besar in central Lampung Province. This toll road has a length of approximately 140 kilometers and is one of the most important toll roads on the island of Sumatra. Travel time between Bakauheni and Terbanggi Besar can be shorter and more efficient through this toll road.

The purpose of the Bakauheni-Terbanggi Besar toll road is to improve connectivity and mobility between the southern and northern parts of the island of Sumatra and support economic and tourism development in the region. In addition, this toll road facilitates access to Bakauheni port, the main port connected to Java Island by ferry. In its construction, this toll road is equipped with facilities such as toll gates, rest areas, and bridges. This toll road also has several toll gates along its route, so that toll road users can pay tolls according to the distance traveled.

The Bakauheni-Terbanggi Besar Toll Road managed by PT Hutama Karya (Persero) is a transportation artery between Sumatra Island and Java Island, where people move to and from Sumatra Island through the Bakauheni ferry port. As a result, there is an increase in flow density which causes congestion and even traffic accidents. Then with the increasing number of vehicles passing through the Bakauheni-Terbanggi Besar toll road, it will affect the risk of traffic accidents. Based on information obtained from Hutama Karya, researchers used 142 data samples from 300 accidents in 2019-2021. This is certainly an important thing that must be done to improve safety and reduce mortality in traffic accidents. Traffic accidents are a problem, especially on toll roads. Because toll roads are designed as toll roads with a high level of comfort, smoothness, and safety, the number of accidents on Indonesian toll roads remains high (Darmawan & Arifin, 2020).

Over time, there are problems that occur on toll roads, these problems are traffic accidents. Traffic accidents are one of the problems that really need more attention, especially on toll roads. Law No. 22 of 2009 concerning Road Traffic and Transportation defines a traffic accident as an unexpected and unwanted event on the road involving vehicles and other road users with or without human casualties and property/material losses as a result. In toll road planning, all aspects related to road traffic safety to prevent accidents and losses have been taken into account as described in Government Regulation No. 15 of 2015 concerning Toll Roads, namely compared to public highways, toll roads have a superior level of security and comfort services and can accommodate long-distance traffic flows with high mobility.

The length of the bakauheni-terbanggi besar toll road allows the need for clustering or grouping by region. In an effort to handle it, many factors are needed so that the grouping is appropriate. So an effort or method is needed to anticipate and evaluate the occurrence of these accidents by clustering a location so that priority handling, supervision and further evaluation can be carried out by related parties. Clustering techniques in data mining can be used to cluster accidents on the Bakauheni-terbanggi Besar toll road.

Based on the explanation previously described, the authors are interested in conducting research on mapping accident-prone areas on the Bakauheni-Terbanggi Besar Toll Road section by utilizing Geographic Information Systems and using the Cluster Analysis method. The purpose of this study is to identify accident-prone locations in the form of points prone to traffic accidents on the Bakauheni-Terbanggi Besar Toll Road and after applying clustering with the Fuzzy C-means algorithm method, the results of this study are expected to help related parties to follow up on the accident rate that has the highest number, so that in the following year the accident rate will be minimized.

II. RESEARCH METHODOLOGY

The type of research used in this research uses quantitative descriptive methods. Quantitative descriptive research is a method that studies the current state of a group of people, subjects, a set of conditions, a system of thought, or a class of events. The purpose of this descriptive research is to describe, describe or paint systematically and the relationship between the phenomena studied [9]. This research will build a web clustering information system for the accident-prone level of the bakauheni-terbanggi Besar toll road using the Fuzzy C-Means method.

The research stages carried out in using the Fuzzy C-Means algorithm to group the Bakauheni-Terbanggi Besar section accidents based on the accident number index have the following stages:

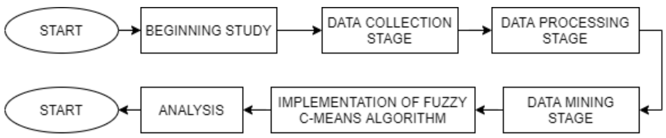

Fig. 1 Research Flow Chart

1) *Beginning Study*

The initial study in this research is to find and study the problem of accidents that occur on the bakauheni-terbanggi Besar toll road, then determine the scope of the problem, background, and make observations by reading several articles and news related to the problem of accidents in the bakauheni-terbanggi Besar toll road. To achieve the goal, the author searches and studies data clustering methods by collecting factors that cause accidents, then researchers use data clustering methods as one of the solutions to accident problems.

2) Data Collection Stage

The data used in this study are secondary data from related agencies, in this case, the data is obtained from PT Hutama Karya (Persero).

3) Data Processing Stage

The collected data is processed for clustering using the K-Means data mining algorithm. The data will be grouped into 3 groups namely Prone, Moderately Prone and Safe.

4) Data Mining Stage

Data mining is the process of searching or mining new information by looking for interesting and hidden patterns from a large data set stored in a database, data warehouse, or other data storage (Tan et al., 2006).

5) Implementation Of Fuzzy C-Means Algorithm

In the Fuzzy K-Means algorithm, the first step is to determine the cluster center, which will mark the average position of each cluster. In the initial condition, these cluster centers are still not accurate. Each data has a membership degree for each cluster. By iterating to improve the cluster center and the membership degree of each data, it can be seen that the cluster center will go to the right location, this iteration is based on the minimization of the objective function which describes the distance from the data point which describes the distance from the given data point to the cluster center weighted by the membership degree of the data point. The output of Fuzzy K-Means is not a fuzzy inference system, but an array of cluster centers and some degree of membership for each data point. Here is the Fuzzy K-Means algorithm.

a) Enter the data to be clustered, in the form of an *n x m* matrix (*Xij* = *i-th* sample data (*i* = 1,2,...n) and *j-th* attribute (*j* = 1,2,...m).

b) Determine:

1. Number of clusters (c): 2
2. Rank (w) : 2
3. Maximum iterations (maxiter): 1000
4. Smallest error (Ɛ) : 0.00001
5. Initial objective function $P0 = 0$: 0
6. Initial iteration (t = 1) : 1

c) Generate random numbers µik *k* (*i* = 1,2,...,n; *k* = 1,2,...,c) as the elements of the initial partition matrix *U*.

Calculate the sum of each column with the formula: $Q_i = \sum_k^c = 1\mu_{ik}$

d) Calculate the kth cluster center value (Vkj with *k* = 1,2,...,c; *j* = 1,2,...,m)

$$V_{kj} = \frac{\sum_{i-1}^{n}\left((\mu_{ik})^{w} * X_{ij}\right)}{\sum_{i=1}^{n}(\mu_{ik})w}$$

Description

Vkj = *k-th* cluster center at the jth attribute.

Xij = *i-th* samel data (*i* = 1,2,...,n), *j-th* attribute (*j* = 1,2,...,m).

e) Calculate the objective function at the tth iteration with the formula:

$$Vkj \sum_{i}^{n} = 1 \sum_{k}^{c} = 1([\sum_{j}^{m} = 1(Xij - Vkj)2](\mu_{ik})W$$

f) Calculate the change in partition matrix U with the formula:

$$\mu_{ik(t)} = \frac{\left[\sum_{j=1}^{m} (X_{ij} - V_{kj})^{2}\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m} (Xij - V_{kj})^{2}\right]^{\frac{-1}{w-1}}}$$

g) Stop condition check:

1. If (|Pt - Pt-1| < Ɛ) or (t > Maxiter) then stop.
2. Otherwise: t = t + 1, repeat step 4.
3.

6) Analysis

In this analysis stage, the data samples used amounted to 142 data sets of 4 variables, namely Sustainable Transport Award (STA), Latitude, Longitude, Number of Accidents and Details containing complete accident information. The data obtained from the results of data collection consists of accumulated accidents with the number of accidents from 2019 - 2021.

TABLE I
ACCIDENT DATA TABLE

No	STA	Latitude	Longitude	Total accidents	Action
1	GT Tegineneng Barat	-5.158.694.611	1.051.880.811	1	Details
2	KM 00+200	-5.868.509.743	1.057.503.515	2	Details
3	KM 00+600	-5.866.313.222	1.057.470.201	3	Details
4	KM 00+800	-5.865.309.782	1.057.455.069	1	Details
5	KM 01+000	-5.864.393.454	1.057.439.778	3	Details
6	KM 01+200	-5.863.536.244	1.057.423.778	5	Details
…	…	…	…	…	…
…	…	…	…	…	…
137	KM 126+800	50.039	1.052.288	2	Details
138	KM 127+000	50.002	1.052.291	2	Details
139	KM 127+800	49.947	105.229	1	Details
140	KM 128+000	-4.991.182.075	1.052.291.619	3	Details
141	KM 129+200	49.766	1.052.284	2	Details
142	KM 129+600	4.978	1.052.283	2	Details

III. RESULTS AND DISCUSSION

A. *Flowchart of Fuzzy C-Means*

The Fuzzy C-Means algorithm flowchart used is based on data on the number of accidents on the Bakauheni-Terbanggi Besar toll road, in general the Fuzzy C-Means flowchart is as follows in Figure 3:

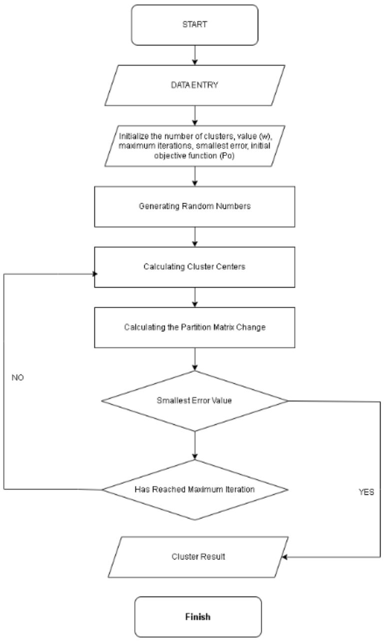

Fig. 3 Flowchart of Fuzzy C-Means

B. *System Implementation*

The implementation of the Accident Analysis Clustering system uses the PHP programming language with the CodeIgnier framework and uses the SQL database, the implementation of the accident-prone area traffic mapping system can be seen as follows:

1) *Login View*

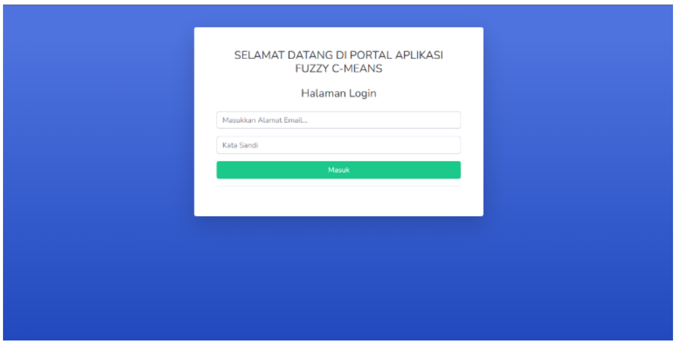

Fig. 4 Login View

The login view is used by the admin and user/officer to access the system, where the admin is in charge of managing accident data, while the user/officer provides accident data info to the admin regarding the data needed to cluster accident-prone points.

2) *Dashboard view*

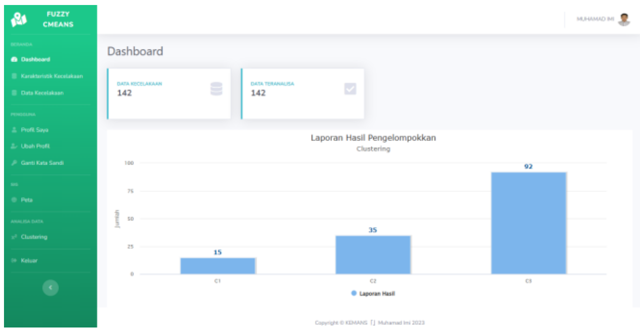

Fig. 5 Dashboard View

The dashboard page displays the amount of accident data and analyzed data, then in this feature you can also see the results of the following number of clusters / groupings: the level of accident prone is determined by 3 (three) clusters, namely (i) road sections classified as prone clusters there are 15 points, (ii) road sections classified as moderately prone clusters there are 35 points, (iii) road sections classified as safe clusters there are 92 location points based on 3 criteria with the calculation of the Fuzzy C-Means clustering algorithm.

3) Accident Characteristics Display

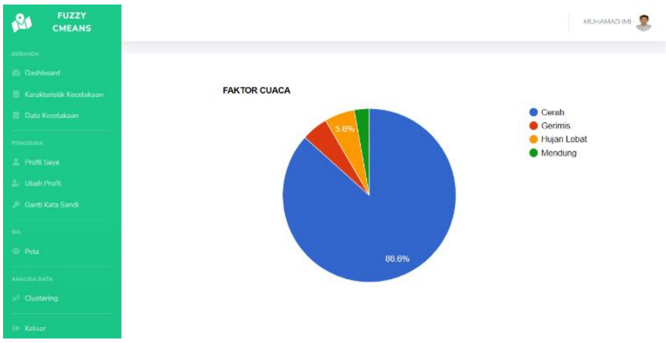

Fig. 6 Accident Characteristics Display

The accident characteristics page contains information on the results of data analysis and processing, namely weather factors, factors causing accidents, vehicle types, gender and the last contains the number of accident location points from the most to the least using bar charts, with this feature companies and related agencies can more easily find out accident patterns and evaluate what methods will be used so that the number of accidents can be minimized.

4) Accident Data Menu Display

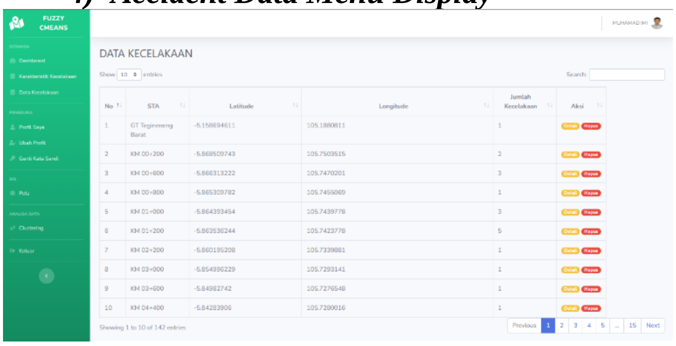

Fig. 7 Accident Data Menu Display

The accident data menu page is a page where the admin gets road data information along with latitude, longitude, number of accidents, and details containing complete accident information in tabular form.

5) User Menu Display

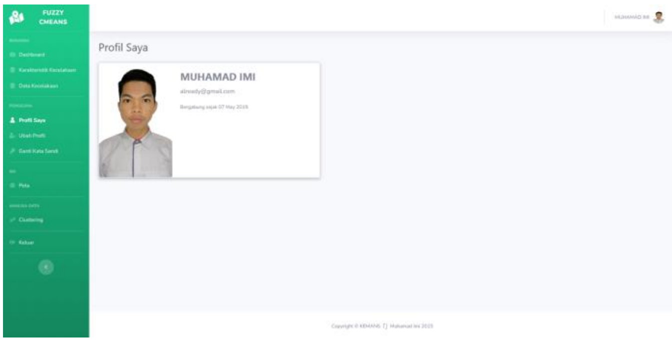

Fig. 8 User Menu Display

On the user menu page there are 3 submenus, namely profile to display admin / user information, Change Profile to change data / user, Change Password to change user / admin password.

6) Map Page View

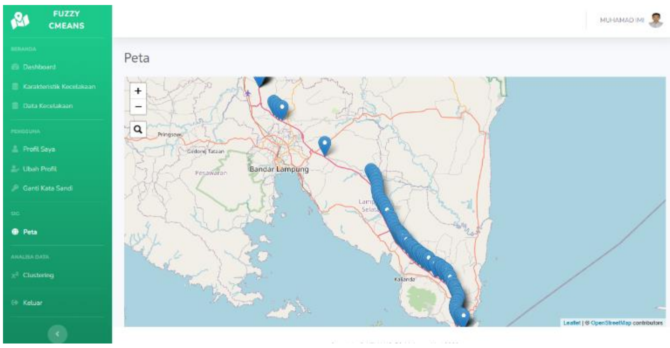

Fig. 9 Map Page View

The map page displays a visualization of accident-prone points, if the coordinate icon is clicked, information on the location of the accident and the number of accidents that occurred at that location point will appear.

7) Clustering Menu Display

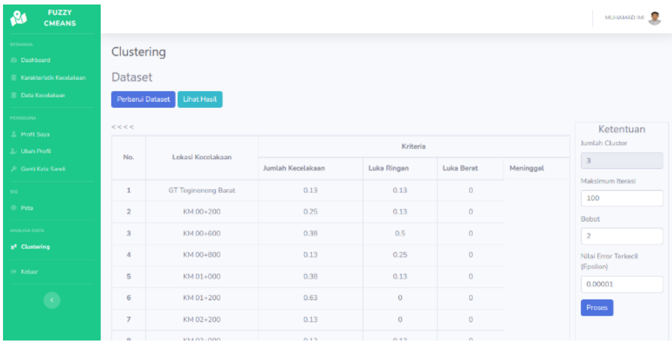

Fig. 10 Clustering Menu Display

The clustering menu page contains accident location data and criteria then before the data mining process is carried out, data preprocessing is carried out by handling missing values. Furthermore, data transformation is carried out by normalizing data with Min-Max Normalization.

8) Display Display of Clustering Result Page

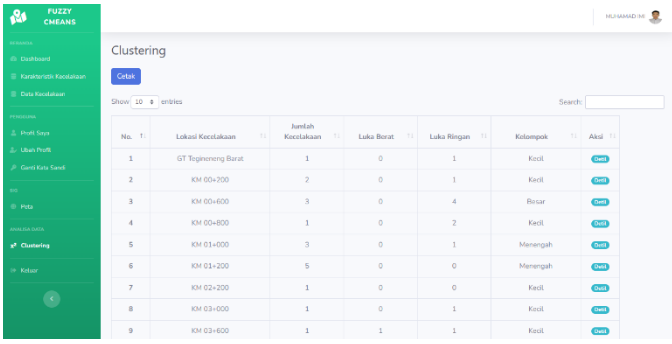

Fig. 11 Display Display of Clustering Result Page

In the algorithm calculation process, the criteria greatly affect the final result of the calculation. The Fuzzy C-Means algorithm has 3 criteria, namely the number of accidents, the number of serious injuries, the number of minor injuries. From these criteria, the algorithm will be calculated. The following is the calculation performed. Enter the data first as in the table below:

TABLE III
ACCIDENT DATA TABLE

No	STA	Number Of Accident	Minor Injuries	Severe injuries
1	GT Teginenen g Barat	1	0	1
2	KM 00+200	2	0	1
3	KM 00+600	3	0	4
4	KM 00+800	1	0	2
5	KM 01+000	3	0	1
…	…			
…	…			
138	KM 127+000	2	0	0
139	KM 127+800	1	0	0
140	KM 128+000	3	2	4
141	KM 129+200	2	0	2
142	KM 129+600	2	0	0

In this calculation process, only a sample of 10 data is used for efficiency reasons. However, these 10 data are still accumulated from all available data. All calculation formulas consider the existence of other data, so that the results are still relevant to the output of the system..

a. Determining the Initial Values

After entering the criteria data into the table, the next step is to determine the initial values used. The value determination in this study by setting the value of the number of clusters (c) = 3, the number of criteria (j) = 3, the rank value (w) = 3, the maximum iteration (MaxIter) = 28, the smallest error (ε) = 0.001, the initial objective function (P0) = 0, the initial iteration (initial t) = 0.

b. Generating Random Numbers

The next step is to generate random numbers. This number is used as the initial membership degree value of each data. In this system, random numbers are generated using a special function. This is because the random numbers generated have special requirements, namely the total number of clusters equal to 1. There are three random numbers that will be generated in each data represented by the variables Ui1 (first cluster), Ui2 (second cluster), and Ui3 (third cluster). By using sample data with equations in the Fuzzy C-Means clustering algorithm.

c. Calculating Cluster Centers

After getting a random number, the next step is to calculate the cluster center of each criterion in each cluster. The cluster center is calculated by multiplying the multiplication of random numbers by each criterion in each cluster. The results of each criterion will be summed up, then divided by the number of random number departures for each cluster.

d. Calculating the Partition Change Matrix

The next step is to calculate the change in the partition matrix (new membership degree). The calculation of the new partition matrix value is done by multiplying the result of the power of 3 criteria difference values with the cluster center. As is done when calculating the objective function. The difference is that in the change in the partition matrix, the power value is again multiplied by -1/(w-1) for each cluster.Hasil Clustering

TABLE IIIII

CLUSTERING RESULT TABLE

No	STA	Cluster 1	Cluster 2	Cluster 3	Cluster	Description
1	GT Tegineneng Barat	0,0117085	0,0564643	0,9318272	C3	Safe
2	KM 00+200	0,0264604	0,2707245	0,7028151	C3	Safe
3	KM 00+600	0,5226488	0,3409564	0,1363948	C1	highly vulnerable
4	KM 00+800	0,0755604	0,2754225	0,6490171	C3	Safe
5	KM 01+000	0,0284654	0,809707	0,1618276	C2	vulnerable enough
6	KM 01+200	0,1242168	0,6140565	0,2617268	C2	vulnerable enough
…	…	…	…	…	…	…
…	…	…	…	…	…	…
138	KM 127+000	0,0220012	0,1578823	0,8201165	C3	Safe
139	KM 127+800	0,0092469	0,0430885	0,9476646	C3	Safe
140	KM 128+000	0,7371345	0,1875839	0,0752815	C1	highly vulnerable
141	KM 129+200	0,0710613	0,5379507	0,3909881	C2	vulnerable enough
142	KM 129+600	0,0220012	0,1578823	0,8201165	C3	Safe

Description (28th iteration):

a. Red = Group 1 (cluster 1) location points classified as very
vulnerable which amounted to 15 data.

b. Yellow = Group 2 (cluster 2) location points classified as moderately prone totaling 35 data.

c. Green = Group 3 (cluster 3) location points classified as safe, totaling 92 data.

IV. CONCLUSION

This research has produced a Fuzzy C-Means implementation system for clustering accident-prone levels in the Bakauheni- Terbanggi Besar Toll Road with the codeigniter framework. The system has successfully displayed clustering of accident-prone levels based on location points with the Fuzzy C-Means clustering algorithm. Of the 142 indicator data for 2020 and 2021, it can be optimally clustered in 3 clusters. Clustering the level of accident prone which is classified as very prone (1st Cluster) there are 15 road points. Clustering the level of accident prone which is classified as moderately prone (2nd Cluster) there are 35 road points. Clustering the accident-prone level classified as safe (3rd Cluster) 92 road points.

Based on the results of the analysis and processing of existing data, it can be concluded that the factors causing the highest accident in the Bakauheni - Terbanggi Besar Toll Road Section are drowsy drivers by 37%, lack of anticipation/caution by 33%, and tire rupture by 21%. Then researchers analyzed other factors such as weather factors. Researchers hope that the future can develop this application with more data such as adding data on the day and time of the accident, so that the grouping or clustering of accident-prone points can be even better, for example by clustering time and day data, toll road managers and related agencies can minimize and even reduce the number of accidents in the Bakauhebi - Terbanggi Besar Toll Road Section.

Based on the data and accident analysis, the researcher provides handling suggestions to the authorities for the blackspot area, it is necessary to install warning signs of dangerous areas where it can be seen that the kilometer often occurs traffic accidents and is placed before entering the kilometer which is the location of the blackspot. There needs to be a shock line/rumble strip every certain distance to avoid drivers who are drowsy or lack of anticipation. Better street lighting and delineators, especially at all blackspot points. Repairing holes and clarifying faded road marking lines so that traffic can be more directed.

REFERENCES

[1] Rendy Novianto, R. N., & Leonard Goeirmanto, L. G. (2019). "Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi." Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi),6(1).

[2] Darmansah, D. D. (2021). "Analisis Penyebaran Penularan Virus Covid-19 di Provinsi Jawa Barat Menggunakan Algoritma K-Means Clustering." JATISI (Jurnal Teknik Informatika Dan Sistem Informasi),8(3),1188-1199.

[3] Lucyana, N., Sari, P., Kurniawan, D., & Buchari, M. A. (2022). "ANALISIS PENYEBAB KECELAKAAN PESAWAT DI INDONESIA MENGGUNAKAN METODE K-MEANS." Jusikom: Jurnal Sistem Komputer Musirawas, 7(2), 89-105.

[4] Susanto, E., Putri, F. R., & Kurniawati, N. A. (2022). "Analisis Cluster Pasien Covid-19 Berdasarkan Jumlah Rumah Terpapar di Kelurahan Kota Surakarta." JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 9(2), 817-826.

[5] Werdiningsih, D., Rahmawati, D., Cholissodin, I., & Santoso, N. (2019). "Optimasi K-Means untuk Pengelompokan Data Kinerja Akademik Dosen menggunakan Particle Swarm Optimization." J. Pengemb. Teknol. Inf. dan Ilmu Komput, 3(4), 4102-4110.

[6] Aziz, F. N. R. F. J., Setiawan, B. D., & Arwani, I. (2018). "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa." Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.

[7] Askia, S., Hidayat, N., & Dewi, R. K. "Kelompok Penyakit Tanaman Apel Menggunakan Metode K-Means Berbasis Web." Pengemb. Teknol. Inf. dan Ilmu Komput, 2(8), 2435-2441.

[8] Triningsih, A., & Supriyono, H. (2019). "Aplikasi Data Mining Berbasis Web Menggunakan Metode K-Means Clustering Untuk Pengelompokan Penjualan Terlaris Produk Kacamata." Jurnal INSYPRO (Information System and Processing).

[9] Irawan, V., Rizal, A., & Purnamasari, I. (2022). "Penerapan Algoritma K-Mean Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2020." Jurnal Ilmiah Wahana Pendidikan, 8(10), 293-300.

[10] Manurung, R. H., Lubis, M. R., & Saifullah, S. (2021). "Penerapan Algoritma K-Means Dalam Pengelompokan Penerimaan Imunisasi Dasar Lengkap Menurut Provinsi." Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen), 2(1), 24-30.

[11] Lumbantobing, C. D. H., Saifullah, S., & Rizki, F. (2020). "Penerapan Algoritma K-Means pada Ketersediaan Lapangan Olahraga Setiap Kelurahan Di Indonesia." Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen), 1(4), 112-119.

[12] Oktavia, R., Hardinata, J. T., & Irawan, I. (2020). "Penerapan Metode Algoritma K-means Dalam Pengelompokan Angka Harapan Hidup Saat Lahir Menurut Provinsi." Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen), 1(4), 154-161.