# Improving Road Safety in Indonesia: A Clustering Analysis of Traffic Accidents Using K-Medoids

Handrizal*, Hayatunnufus, Maryo Christopher Davinci Nababan
Department of Computer Science-Faculty of Computer Science and Information Technology,
Universitas Sumatera Utara, Medan, Indonesia

***Abstract*—Traffic accidents pose a significant public health and safety challenge in Indonesia, ranking fifth globally in terms of traffic fatality rates. This study aims to identify patterns in traffic accident data to inform effective mitigation strategies. Utilizing the K-Medoids algorithm, we clustered traffic accident data from the Indonesian Central Bureau of Statistics for the period 1992–2022. Prior to clustering, rigorous data preprocessing was conducted to ensure accuracy. The K-Medoids algorithm successfully partitioned the data into distinct clusters, revealing variations in accident patterns across different regions of Indonesia, including disparities in accident frequency and severity. This research provides valuable insights for policymakers and transportation authorities to develop targeted interventions and improve road safety in Indonesia. Additionally, this study successfully applied the K-Medoids algorithm to cluster traffic accident data in Indonesia using data from 2018 to 2022.**

***Keywords*—Traffic accidents; K-Medoids; clustering; data mining**

## I. Introduction

Traffic accidents can involve single vehicles or collisions between multiple vehicles, such as cars, motorcycles, bicycles, and others. External factors, such as collisions with inanimate objects like trees, poles, walls, or traffic lights, also contribute to accidents. According to study [1], each year, 20 to 50 million people sustain serious injuries, and approximately 1.3 million fatalities occur due to traffic accidents worldwide.

Contributing factors to traffic accidents include adverse weather conditions and road damage caused by construction [2]. Moreover, the significant increase in vehicle ownership has led to severe traffic congestion, further elevating the risk of accidents [3]. Indonesia ranks fifth globally in traffic fatalities [4].

The consequences of traffic accidents are severe, including fatalities, serious injuries, minor injuries, and material losses [5]. This study utilizes Indonesian traffic accident data to identify potential patterns and insights that can contribute to accident prevention strategies [6].

Data mining techniques, such as clustering, are crucial for extracting valuable information from large datasets [7]. Clustering, a method for grouping similar data points, has proven effective in solving complex problems in computer science and statistics [8]. The K-Medoids algorithm is a prominent partitioning method in clustering, known for its ability to efficiently group large datasets [9]. It identifies representative data points (medoids) within each cluster, effectively summarizing the data and enabling the identification of underlying patterns [10].

Xiangrun [11] developed a one-stop evaluation framework, EWM-GRA-Kmeans, to evaluate the road safety development of the ASEAN community over the past decade (2009–2020). While this approach effectively identifies road safety trends, it has limitations in handling non-linear relationships, data sparsity, and the need for extensive parameter tuning to achieve optimal clustering results.

## II. Materials and Methods

### A. Data Mining

Data mining is the process of extracting meaningful patterns and insights from large datasets. It involves identifying significant relationships and trends within the data to uncover hidden knowledge [12]. This process often requires analyzing vast amounts of information to discover previously unknown patterns and gain valuable insights [13]. Key characteristics of data mining include:

- Discover previously unknown patterns.
- Utilize large datasets for analysis.
- Generate reliable and actionable insights.

Data mining is a crucial component of Knowledge Discovery in Databases (KDD), a multi-step process that includes data cleaning, integration, selection, transformation, and, ultimately, data mining itself. The ultimate goal of KDD is to extract useful knowledge and insights from raw data [14].

Clustering is a fundamental technique in data mining that groups similar data points together. Its goal is to uncover underlying structures and patterns within the data. Common clustering algorithms include K-Means, K-Medoids, Hierarchical Clustering, and Fuzzy C-Means [15].

The K-Medoids algorithm, also known as Partitioning Around Medoids (PAM), is a popular clustering method. Unlike K-Means, which uses the mean of data points as cluster centers, K-Medoids selects actual data points as cluster representatives. This approach is more robust to outliers and noise in the data [16].

Clustering has a wide range of applications across various fields, including psychology, population studies, healthcare, economics, and social sciences [17]. In general, the k-medoids algorithm operates as follows [18]:

*Corresponding Author, email-Handrizal@usu.ac.id

*1)* Determine the number of k values (clusters).

*2)* Randomly select k centroid values (center points) from the n available data points.

*3)* Calculate the distance of each data point to the assigned centroid using the Euclidean Distance formula:

$$d_{ab} = \sqrt{(x_{1a} - x_{1b})^2 + \cdots + (x_{ia} - x_{ib})^2}$$

*4)* Assign each data point to the cluster with the closest centroid.

*5)* Compute the total cost based on the smallest value within the cluster.

*6)* Recalculate the centroid values.

*7)* Repeat steps 3 to 5.

*8)* Compute the total deviation (S) by subtracting the initial total cost from the new total cost. If S<0, swap the object with the new cluster data to establish a new centroid value.

*9)* Repeat steps 3 to 5 until the centroid values remain unchanged.

A traffic accident is an unintentional event that can occur anywhere. According to the Indonesian National Police, in 2020, an average of three people per hour and 80 people per day died due to traffic accidents in Indonesia. The victims were primarily between the ages of 5 and 29, with men being more frequently affected than women.

Traffic accidents can be caused by various factors. Fatigue and stress from work, conflicts between work and family, overtime hours, lack of motivation for safe driving, and irregular working hours are some of the potential causes. Other contributing factors include adverse weather conditions, such as fog, and road damage due to construction.

### *B. Data Collection Stage*

In this study, researchers collected traffic accident data in Indonesia from multiple relevant sources to ensure accuracy and completeness. The data was obtained from the National Statistics Agency website and included information on the year of the accident, the number of victims with minor injuries, and the number of victims with severe injuries. The data then underwent a pre-processing stage to facilitate clustering analysis using the K-Medoids method, aiming to provide accurate insights into accident patterns across various regions in Indonesia.

### *C. Data Pre-processing Stage*

Data pre-processing for traffic accident datasets in Indonesia is crucial before conducting any analysis. This stage aims to improve data quality, reduce noise, and ensure consistency, ultimately leading to more accurate analytical results. In this study, researchers used Microsoft Excel for data pre-processing.

### *D. Clustering Stage*

The K-Medoids method is a clustering technique that partitions data into multiple groups or clusters based on similarities among data points. Unlike the K-Means method, which determines cluster centers using the average of the data, K-Medoids selects specific data points as cluster centers, known as medoids. One key advantage of the K-Medoids method is its robustness against outliers, as the chosen medoid better represents the cluster compared to the mean, which can be influenced by extreme values.

### *E. Analysis Stage*

The analysis stage in clustering traffic accident data in Indonesia consists of a series of systematic processes to categorize data based on specific patterns or characteristics. It begins with the collection of accident data from the Central Bureau of Statistics website, followed by data pre-processing to remove irrelevant or incomplete information. Subsequently, the data is processed using the K-Medoids clustering algorithm, which classifies accident years based on their level of vulnerability. The results of this analysis help identify high-risk years for accidents, serving as a foundation for developing more effective road safety strategies in the future.

### *F. System Architecture*

The system architecture in this study is designed to support the analysis of traffic accident data in Indonesia using the K-Medoids clustering method. The collected data is processed and analyzed using software such as Microsoft Excel for initial data processing, Google Colab for modeling and visualization, and Visual Studio Code for developing a dashboard interface that presents the analysis results to users. The system architecture of this study is shown in Fig. 1.

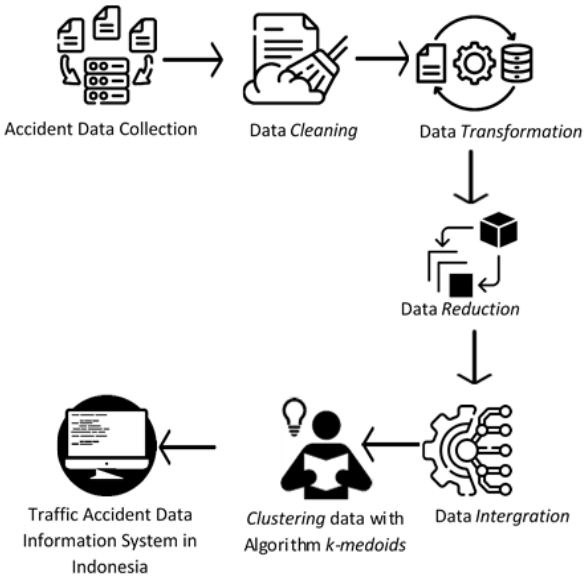


Fig. 1. System architecture.

The explanation of Fig. 1 is as follows:

*1)* Collecting traffic accident data from all regions in Indonesia through the websites of the National and Provincial Central Bureau of Statistics.

*2)* The collected data then undergoes a pre-processing stage, which includes data cleaning, transformation, reduction, and integration.

*3)* The processed data is then input into a data processing system using the K-Medoids algorithm, which is implemented in Google Colab using the Python programming language.

*4)* After data processing, the next step is to design a system that presents information on clustering results and visualization

patterns of traffic accident data in Indonesia, using Visual Studio Code with HTML and CSS.

## III. Results and Discussion

Prior to clustering, the data underwent a preprocessing stage based on accident years. This crucial step ensured data quality and prepared the data for subsequent analysis. Data preprocessing resulted in a cleaner and more structured dataset, as presented in Table I, which summarizes the number of accidents, fatalities, serious injuries, and minor injuries from 1992 to 2022. This preprocessed data served as the foundation for the clustering analysis, enabling the identification of complex accident patterns. Accurate clustering results are essential for effective accident mitigation efforts and informed strategic decision-making to improve road safety in Indonesia.

TABLE I. Traffic Accident Data in Indonesia (1992-2022)

| Year | Number of Accidents | Death Victim (Person) | Serious Injury (Person) | Minor Injury (Person) |
|---|---|---|---|---|
| 1992 | 19920 | 9819 | 13363 | 14846 |
| 1993 | 17323 | 10038 | 11453 | 13037 |
| 1994 | 17469 | 11004 | 11055 | 12215 |
| 1995 | 16510 | 10990 | 9952 | 11873 |
| 1996 | 15291 | 10869 | 8968 | 10374 |
| 1997 | 17101 | 12308 | 9913 | 12699 |
| 1998 | 14858 | 11694 | 8878 | 10609 |
| 1999 | 12675 | 9917 | 7329 | 9385 |
| 2000 | 12649 | 9536 | 7100 | 9518 |
| 2001 | 12791 | 9522 | 6656 | 9181 |
| 2002 | 12267 | 8762 | 6012 | 8929 |
| 2003 | 13399 | 9856 | 6142 | 8694 |
| 2004 | 17732 | 11204 | 8983 | 12084 |
| 2005 | 91623 | 16115 | 35891 | 51317 |
| 2006 | 87020 | 15762 | 33282 | 52310 |
| 2007 | 49553 | 16955 | 20181 | 46827 |
| 2008 | 59164 | 20188 | 23440 | 55731 |
| 2009 | 62960 | 19979 | 23469 | 62936 |
| 2010 | 66488 | 19873 | 26196 | 63809 |
| 2011 | 108696 | 31195 | 35285 | 108945 |
| 2012 | 117949 | 29544 | 39704 | 128312 |
| 2013 | 100106 | 26416 | 28438 | 110448 |
| 2014 | 95906 | 28297 | 26840 | 109741 |
| 2015 | 96233 | 24275 | 22454 | 107743 |
| 2016 | 106644 | 31262 | 20075 | 120532 |
| 2017 | 104327 | 30694 | 14559 | 121575 |
| 2018 | 109215 | 29472 | 13315 | 130571 |
| 2019 | 116411 | 25671 | 12475 | 137342 |
| 2020 | 100028 | 23529 | 10751 | 113518 |
| 2021 | 103645 | 25266 | 10553 | 117913 |
| 2022 | 139258 | 28131 | 13364 | 160449 |

### A. Determining the Number of Clusters

This stage represents the initial phase of K-Medoids clustering. In this study, the number of K values (clusters) is set to three. Here, C1 represents years with a very high accident risk, C2 represents years with a high accident risk, and C3 represents years with a low accident risk.

### B. Medoids Initialization

At this stage, the initial medoids are randomly selected to represent each cluster in the dataset, based on the predetermined number of clusters. These medoids serve as the initial centers for the formation of clusters. The medoids in this dataset are shown in Table II."

### C. Assignment of Cluster Members

At this stage, the distance of each data point in the dataset to each medoid is calculated using the Euclidean Distance formula, and the data is assigned to the cluster with the nearest medoid. This process groups data into clusters that align with the criteria of each medoid. The assignment of cluster members in the dataset is determined by calculating the nearest distance using the Euclidean Distance formula. The shortest distance is from the first data point to the third cluster, meaning the first data point in the dataset belongs to Cluster 3. The complete distance calculations for each data point are shown in Table III.

TABLE II. Initial Medoids

| Name | Year | Number of Accidents | Death Victim (Person) | Serious Injury (Person) | Minor Injury (Person) |
|---|---|---|---|---|---|
| C1 | 2017 | 104327 | 30694 | 14559 | 121575 |
| C2 | 2010 | 66488 | 19873 | 26196 | 63809 |
| C3 | 2002 | 12267 | 8762 | 6012 | 8929 |

TABLE III. DISTANCE CALCULATION RESULTS OF ACCIDENT DATA IN INDONESIA TO INITIAL MEDOIDS

| Year | C1 | C2 | C3 | Closest Distance | Cluster |
|---|---|---|---|---|---|
| 1992 | 137669,231 | 69510,595 | 12195,645 | 12195,645 | C3 |
| 1993 | 140664,534 | 72863,410 | 8583,208 | 8583,208 | C3 |
| 1994 | 141081,168 | 73298,787 | 8265,411 | 8265,411 | C3 |
| 1995 | 141971,217 | 74417,228 | 6867,151 | 6867,151 | C3 |
| 1996 | 143935,269 | 76513,450 | 4940,646 | 4940,646 | C3 |
| 1997 | 140790,585 | 73305,342 | 8085,318 | 8085,318 | C3 |
| 1998 | 143914,658 | 76568,310 | 5132,861 | 5132,861 | C3 |
| 1999 | 146528,638 | 79453,737 | 1855,509 | 1855,509 | C3 |
| 2000 | 146509,128 | 79483,718 | 1508,531 | 1508,531 | C3 |
| 2001 | 146703,652 | 79727,919 | 1153,437 | 1153,437 | C3 |
| 2002 | 147371,058 | 80514,467 | 0,000 | 0,000 | C3 |
| 2003 | 146680,229 | 79740,883 | 1596,992 | 1596,992 | C3 |
| 2004 | 141060,005 | 73648,232 | 7389,889 | 7389,889 | C3 |
| 2005 | 75928,780 | 29932,155 | 95083,850 | 29932,155 | C2 |
| 2006 | 75303,970 | 24917,900 | 90898,708 | 24917,900 | C2 |
| 2007 | 93849,995 | 24897,339 | 55627,242 | 24897,339 | C2 |
| 2008 | 81020,844 | 11251,214 | 69455,342 | 11251,214 | C2 |
| 2009 | 73102,396 | 4544,962 | 76922,716 | 4544,962 | C2 |
| 2010 | 70860,528 | 0,000 | 80514,467 | 0,000 | C2 |
| 2011 | 24666,235 | 63478,905 | 143742,479 | 24666,235 | C1 |
| 2012 | 29403,054 | 84171,645 | 164280,334 | 29403,054 | C1 |
| 2013 | 18776,445 | 57906,853 | 137245,716 | 18776,445 | C1 |
| 2014 | 19170,951 | 55195,524 | 134067,011 | 19170,951 | C1 |
| 2015 | 18983,457 | 53369,856 | 131626,321 | 18983,457 | C1 |
| 2016 | 6099,608 | 70690,728 | 148547,183 | 6099,608 | C1 |
| 2017 | 0,000 | 70860,528 | 147371,058 | 0,000 | C1 |
| 2018 | 10385,633 | 80875,349 | 157092,103 | 10385,633 | C1 |
| 2019 | 20595,993 | 90118,204 | 166323,648 | 20595,993 | C1 |
| 2020 | 12216,167 | 62030,885 | 137409,514 | 12216,167 | C1 |
| 2021 | 7706,269 | 67688,059 | 143249,621 | 7706,269 | C1 |
| 2022 | 52338,892 | 121932,839 | 198781,877 | 52338,892 | C1 |

The total cost of the closest distance from the dataset to the initial medoids is 383,460.873.

### *D. Update of Medoids*

Once all the data have been assigned to their respective clusters, the next step is to evaluate whether better medoids can be identified by replacing the previously selected ones. The goal of this phase is to minimize the total distance between the data points and the medoids within each cluster. The new medoids for this dataset are presented in Table IV.

### *E. Iteration*

The final stage is the iteration stage, during which steps 2 and 3 are repeated until there are no significant changes in the selection of medoids or clustering. If the difference between the total distance of the old medoids to the data and the total distance of the new medoids to the data exceeds 0, the clustering process is halted. The determination of new cluster members in the dataset is performed by calculating the Euclidean distance to identify the closest one. For instance, if the shortest distance is between the 1st data point and the 3rd cluster, it means that the 1st data point in the dataset belongs to cluster 3. The complete distance calculations for each data point are presented in Table V.

The total cost of the closest distance from the dataset to the new medoids is 389,372.706. The calculated cost difference is 5,911.833.

TABLE IV. NEW MEDOIDS

| Name | Year | Number of Accidents | Death Victim (Person) | Serious Injury (Person) | Minor Injury (Person) |
|---|---|---|---|---|---|
| C1 | 2021 | 103645 | 25266 | 10553 | 117913 |
| C2 | 2009 | 62960 | 19979 | 23469 | 62936 |
| C3 | 2000 | 12649 | 9536 | 7100 | 9518 |

TABLE V. DISTANCE CALCULATION RESULTS OF ACCIDENT DATA IN INDONESIA TO NEW MEDOID

| Year | C1 | C2 | C3 | Closest Distance | Cluster |
|---|---|---|---|---|---|
| 1992 | 133713,081 | 66109,353 | 10979,995 | 10979,995 | C3 |
| 1993 | 136686,375 | 69396,352 | 7309,600 | 7309,600 | C3 |
| 1994 | 137120,483 | 69833,437 | 6950,055 | 6950,055 | C3 |
| 1995 | 137989,692 | 70912,227 | 5540,881 | 5540,881 | C3 |
| 1996 | 139931,596 | 72995,675 | 3602,667 | 3602,667 | C3 |
| 1997 | 136850,901 | 69781,182 | 6748,038 | 6748,038 | C3 |
| 1998 | 139943,862 | 73030,513 | 3726,689 | 3726,689 | C3 |
| 1999 | 142477,555 | 75881,712 | 464,722 | 464,722 | C3 |
| 2000 | 142439,826 | 75905,730 | 0,000 | 0,000 | C3 |
| 2001 | 142618,923 | 76148,240 | 575,382 | 575,382 | C3 |
| 2002 | 143249,621 | 76922,716 | 1508,531 | 1508,531 | C3 |
| 2003 | 142583,504 | 76165,526 | 1503,875 | 1503,875 | C3 |
| 2004 | 137043,882 | 70130,897 | 6224,882 | 6224,882 | C3 |
| 2005 | 72837,564 | 33553,022 | 94107,672 | 33553,022 | C2 |
| 2006 | 72021,370 | 28387,912 | 89924,752 | 28387,912 | C2 |
| 2007 | 90227,130 | 21429,024 | 54589,567 | 21429,024 | C2 |
| 2008 | 77698,271 | 8146,543 | 68408,679 | 8146,543 | C2 |
| 2009 | 69803,404 | 0,000 | 75905,730 | 0,000 | C2 |
| 2010 | 67688,059 | 4544,962 | 79483,718 | 4544,962 | C2 |
| 2011 | 27436,517 | 66888,163 | 142738,436 | 27436,517 | C1 |
| 2012 | 34363,145 | 87480,718 | 163289,453 | 34363,145 | C1 |
| 2013 | 19734,398 | 60855,085 | 136293,200 | 19734,398 | C1 |
| 2014 | 20028,156 | 57937,032 | 133109,055 | 20028,156 | C1 |
| 2015 | 17348,848 | 55984,334 | 130718,855 | 17348,848 | C1 |
| 2016 | 11936,233 | 73242,176 | 147646,666 | 11936,233 | C1 |
| 2017 | 7706,269 | 73102,396 | 146509,128 | 7706,269 | C1 |
| 2018 | 14716,282 | 83109,801 | 156252,653 | 14716,282 | C1 |
| 2019 | 23330,557 | 92447,440 | 165513,614 | 23330,557 | C1 |
| 2020 | 5954,417 | 64085,298 | 136602,429 | 5954,417 | C1 |
| 2021 | 0,000 | 69803,404 | 142439,826 | 0,000 | C1 |
| 2022 | 55621,102 | 124493,920 | 197977,315 | 55621,102 | C1 |

Since the total deviation value (S) is greater than 0, the clustering process is stopped. Thus, the members of each cluster are obtained, as shown in Table VI.

TABLE VI. ACCIDENT DATA GROUPING RESULTS IN INDONESIA (1992-2022

| Year | Cluster | Category |
|---|---|---|
| 1992 | C3 | Non-Prone |
| 1993 | C3 | Non-Prone |
| 1994 | C3 | Non-Prone |
| 1995 | C3 | Non-Prone |
| 1996 | C3 | Non-Prone |
| 1997 | C3 | Non-Prone |
| 1998 | C3 | Non-Prone |
| 1999 | C3 | Non-Prone |
| 2000 | C3 | Non-Prone |
| 2001 | C3 | Non-Prone |
| 2002 | C3 | Non-Prone |
| 2003 | C3 | Non-Prone |
| 2004 | C3 | Non-Prone |
| 2005 | C2 | Prone |
| 2006 | C2 | Prone |
| 2007 | C2 | Prone |
| 2008 | C2 | Prone |
| 2009 | C2 | Prone |
| 2010 | C2 | Prone |
| 2011 | C1 | Very Prone |
| 2012 | C1 | Very Prone |
| 2013 | C1 | Very Prone |
| 2014 | C1 | Very Prone |
| 2015 | C1 | Very Prone |
| 2016 | C1 | Very Prone |
| 2017 | C1 | Very Prone |
| 2018 | C1 | Very Prone |
| 2019 | C1 | Very Prone |
| 2020 | C1 | Very Prone |
| 2021 | C1 | Very Prone |
| 2022 | C1 | Very Prone |

## IV. CONCLUSION

This study successfully applied the K-Medoids algorithm to cluster traffic accident data in Indonesia using data from 1992 to 2022. The algorithm facilitates the identification of distinct traffic accident patterns each year, enhancing the understanding of accident characteristics in Indonesia. The clustering results reveal variations in both the number of accidents and the severity of victims across different clusters. This research provides valuable insights to support accident mitigation efforts and the development of traffic safety policies in Indonesia.

For future research, incorporating data from all Indonesian provinces is crucial for obtaining comprehensive and nationally representative results. Analyzing data from each province will provide more detailed insights into traffic accident patterns, including regional variations. Additionally, integrating external factors such as weather conditions, traffic density, and environmental influences will further enhance the analysis. Furthermore, developing a mobile application that provides real-time information about accident-prone areas on digital maps can empower drivers to make informed decisions and improve road safety.

## REFERENCES

[1] M. Amoadu, E.W. Ansah, and J.O. Sarfo, "Psychosocial work factors, road traffic accidents and risky driving behaviours in low-and middle-income countries: A scoping review", *IATSS Research*, 2023.

[2] Dabiri, and B. Kulcsár, "Incident indicators for freeway traffic flow models", *Communications in Transportation Research*, Vol. 2, No. 100060, 2022.

[3] S. Basu, and P. Saha, "Evaluation of risk factors for road accidents under mixed traffic: Case study on Indian highways", *IATSS Research*, Vol. 46, No. 4, 2022, pp. 559-573.

[4] Zainafree, Intan, et al. "Risk factors of road traffic accidents in Rural and Urban areas of Indonesia based on the national survey of year 2018." Nigerian postgraduate medical journal 29.2 (2022): 82-88.

[5] Iranmanesh, M., Seyedabrishami, S., & Moridpour, S. (2022). Identifying high crash risk segments in rural roads using ensemble decision tree-based models. Scientific reports, 12(1), 20024.

[6] Kusumastutie, N. S., Patria, B., Kusrohmaniah, S., & Hastjarjo, T. D. (2024). A review of accident data for traffic safety studies in Indonesia. In IOP Conference Series: Earth and Environmental Science (Vol. 1294, No. 1, p. 012012). IOP Publishing.

[7] A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency", *In Journal of Physics: Conference Series*, Vol. 1751, No. 1, 2021, p. 012038.

[8] E. Esenturk, D. Turley, A. Wallace, S. Khastgir, and P. Jennings, "A data mining approach for traffic accidents, pattern extraction and test scenario generation for autonomous vehicles", *International Journal of Transportation Science and Technology*, Vol.12, No. 4, 2023, pp. 955-972.

[9] M. A. Ahmed, H. Baharin, and P.N. Nohuddin, "Analysis of K-means, DBSCAN and OPTICS Cluster algorithms on Al-Quran verses", *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 8, 2020, pp. 248-254.

[10] M. Nazari, A. Hussain, and P. Musilek, "Applications of Clustering Methods for Different Aspects of Electric Vehicles", *Electronics*, Vol. 12, No. 4, 2023, p. 790.

[11] Xiangrun Chen at all (2024).Road Safety Development Evaluation for ASEAN Community Using EWM-GRA-Kmeans DOI: 10.4108/eai.12-1-2024.2347145

[12] Aziz, M. A., Hidayat, Y. A., Febrianti, D. R., Aida, A. N., Amalia, L., Tahyudin, I., & Darmayanti, I. (2022, August). Comparison of K-Medoids Algorithm with K-Means on Number of Student Dropped Out. In 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS) (pp. 53-58). IEEE

[13] Henderi, H., Fitriana, L., Iskandar, I., Astuti, R., Arifandy, M. I., Hayadi, B. H, & Kurniawan, A. (2024, September). Optimization of Davies-Bouldin Index with k-medoids algorithm. In AIP Conference Proceedings (Vol. 3065, No. 1). AIP Publishing.

[14] Rahman, S. N., Jamhur, A. I., Elva, Y., & Rianti, E. (2021, November). Comparison of the Effectiveness of C. 45 Algorithm with Naive Bayes Algorithm in Determining Scholarship Recipients. In 2021 International Conference on Computer Science and Engineering (IC2SE) (Vol. 1, pp. 1-5). IEEE.

[15] Raj, S., Ramesh, D., & Sethi, K. K. (2021). A Spark-based Apriori algorithm with reduced shuffle overhead. The Journal of Supercomputing, 77(1), 133-151.

[16] Edastama, P., Bist, A. S., & Prambudi, A. (2021). Implementation of data mining on glasses sales using the apriori algorithm. International Journal of Cyber and IT Service Management, 1(2), 159-172.

[17] Viet, T. N., Le Minh, H., Hieu, L. C., & Anh, T. H. (2021). The Naïve Bayes algorithm for learning data analytics. Indian Journal of Computer Science and Engineering, 12(4), 1038-1043.

[18] Kaur, N. K., Kaur, U., & Singh, D. (2014). K-Medoid clustering algorithm-a review. Int. J. Comput. Appl. Technol, 1(1), 42-45.