Integra: Journal of Integrated Mathematics and Computer Science 1(3) 97-105

Integra: Journal of Integrated Mathematics and Computer Science

Journal Homepage: https://integrajimcs.com/index.php/integra/index

integra

Journal of Integrated Mathematics and Computer Science

Research Paper

Traffic Violation Modeling Using K-Means Clustering Method: A Case Study in Bandung, Indonesia

Akmal Junaidi[1], Yunita Rosalina Manurung[1], Dewi Asiah Shofiana[1]*, Favorisen Rosyking Lumbanraja[1]

[1] *Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Bandar Lampung, 35145, Indonesia*

***Corresponding author**: dewi.asiah@fmipa.unila.ac.id

Keywords

Cartometric, Clustering, K-Means, Traffic Violations

Abstract

Violations of traffic regulations are both an issue and a problem that persists as a feature of life, especially in metropolitan regions such as Bandung. Traffic violation has both behavioral and environmental patterns, with different types of violations occurring at different times during the day. This negligence stems largely from not properly equipping the vehicle with the necessary documents, especially for drivers who do not pay attention to proper document preparation. With the goal of increasing road safety, law enforcement bodies face the ongoing challenge of managing rising traffic violation rates which results in a growing backlog of violation cases and a corresponding backlog workload for police departments. Comprehensive preventive strategies for the problem are extremely difficult to implement in the absence of streamlined mechanisms for the efficient allocation of limited police resources. Currently, agencies responsible for managing violation records are still using a manual desktop system based on Microsoft Excel spreadsheets. This method impedes the analysis of large datasets to derive actionable insights that could inform targeted, data-driven strategies needed to guide proactive measures. In this regard, this study attempts to implement the K-Means clustering technique in order to identify and classify high-incidence traffic violation areas in Bandung. Using this technique, the research classifies the city into three violation risk clusters: very prone, prone, and moderately prone areas. The map of the classes demonstrates the distribution of these clusters spatially, illustrating clearly and vividly how stakeholders can visualise the pattern of traffic violations. This method improves the understanding of data and at the same time boosts purposeful planning for the safety and public traffic order anticipations.

Received: 25 September 2024, Accepted: 2 November 2024

https://doi.org/10.26554/integra.20241326

1. INTRODUCTION

Around the globe, traffic accidents pose significant public health challenges and are a major concern for healthcare organizations. Each year, approximately 1.35 million individuals lose their lives or suffer disabilities due to road accidents. In 2019, low- and middle-income countries accounted for 93% of all fatalities related to road traffic injuries, contributing to an estimated 1.3 million deaths. This problem continues to escalate, with road traffic injuries projected to become the seventh leading cause of death worldwide by 2030 [1]. The Global Status Report on Road Safety 2023 [2] indicates a slight decrease in annual road traffic fatalities, now estimated at 1.19 million. This suggests that road safety initiatives are making a difference, and that substantial progress is possible when effective strategies are implemented. However, the cost of global mobility remains unacceptably high. Road traffic injuries continue to be the leading cause of death for children and young adults aged 5 to 29. Over half of these deaths involve vulnerable road users, pedestrians, cyclists, and motorcyclist especially in low-and-middle income countries. Swift and decisive action is essential to meet the global target of reducing road traffic deaths and injuries by at least 50% by 2030 [2].

Some studies about the There are several studies related to traffic accidents and the factors that cause them, among others: Yibeltal et al [3] investigate the road traffic accidents and its associated factors among public transportation in Africa; Zainafree et al [4] explored the risk factors of road traffic accidents in Indonesia's rural and urban areas based on the 2018 national survey; Tiruneh et al [5] studies the occurrence and contributing factors of road traffic incidents among children and adolescents in hospitals within the Amhara National Regional State, Ethiopia; Jongrak [6], and Kwangsukstith [7] investigate the factors related with traffic accident in Thailand, and many more.

Road traffic accidents caused by traffic offenses are a major public health concern that result in fatalities and economic expenses. As a result, it is critical to emphasize road safety initiatives that lower the frequency and severity of accidents. An incremental road safety strategy that identifies high-risk areas and common traffic violations in order to prioritize further enforcement us suggested, and in fact, by analyzing data on traffic violations in different districts and comparing them to the overall average using the Kolmogorov-Smirnov (KS) test, risky areas are identified and the most common violations are detected [8]. Wang et al [9] use a combination of random forest and association rules to analyse the factors influencing traffic accident severity.

Traffic accidents have often been viewed as just the tip of the iceberg when it comes to understanding the underlying issues in traffic management and broader transportation systems. In Indonesia, as in many other nations, data on traffic accidents reveal alarmingly high numbers and serious consequences. This pattern is especially noticeable in Jakarta, the country's most densely populated city [10].

Prasetyanto et al [11] investigated the safety-related attitudes and behaviors of student and worker motorcyclists, focusing on the use of safety gear, readiness of documents and vehicles, and perceptions of road conditions, as well as how these factors influence traffic violations. They gathered data from motorcyclists in Bandung City and analyzed it using Structural Equation Modeling (SEM). Notably, they found that having a positive attitude towards safe riding does not necessarily result in better practices regarding vehicle preparation, safety equipment, and documentation. Additionally, high-quality road infrastructure contributes to fewer traffic violations. The study recommends enhancing road infrastructure and actively promoting and regulating safe riding behavior to reduce violations.

Combating traffic violations in Bandung is further complicated by their diverse geographical and categorical distribution, which poses challenges for both the public and the authorities in pinpointing the most susceptible areas for such breaches. Therefore, authorities are proactive when devising timely responses to these infractions, while citizens should also be equally informed about where and when such violations are more commonly committed. With the growing problem, policymakers need structured and thorough plans to solve it, which requires processing and analysing enormous caches of violation data gathered. Based on the existing problems, a system is needed using the right method for grouping traffic violation data in Bandung City.

The presence of data mining is expected to be able to process large amounts of data and generate knowledge to solve problems for the Bandung police. This is in accordance with the definition of data mining, which is finding patterns that dance from large amounts of data that can be stored in databases, data warehouses, and other information storage [12].

Given the considerable volume of data available, it becomes imperative to organise and classify the records. This study addresses the problem by employing a data mining technique, which in this case is clustering, to classify traffic violations in Bandung. Data mining is the process of extracting useful information from large volumes of data stored in databases or data warehouses in order to improve decision making. The used of data mining to analysed the road accident is used by [13, 14].

Clustering is one of the techniques that divides data into clusters based on common attributes. According to Garcia et al [15], clustering is the process of reducing the number of distinct values into a smaller number of groups in such a manner that all the objects belonging to a group are as similar as possible. In this research, clusters will be formed from the traffic violation data for Bandung with consideration to the type of violation, type of vehicle, and its temporal occurrence.

Careful processing of this data is expected to aid the Bandung Metropolitan Police in accurately classifying the traffic violations. This classification is useful in devising policies and measures aimed at managing and minimising contraventions. Furthermore, this research seeks to improve the effectiveness of police in maintaining order and safety for the people of Bandung. In addition to enforcement, this also seeks to improve public cooperation in minimising violations in known high-risk areas, thereby traffic law compliance. This may ultimately help in reducing violation cases and traffic accidents in the city.

2. METHODS

Figure 1 show the steps of this study.

2.1 Data Collection

The data utilized in this study is classified into two categories: spatial data and non-spatial data.

2.1.1 Spatial Data

Geospatial information was obtained from the Bandung City Spatial Planning (PTRW) for the years 2016-2036 and includes:

- The administrative map of the Bandung metropolitan area city and its suburbs which was provided by Regional Development Planning Agency (Bappeda) in 2018.
- The 2018 Road Network Map of Bandung City by Bappeda.
- Traffic violation locations in coordinates obtained through a cartometric extension.

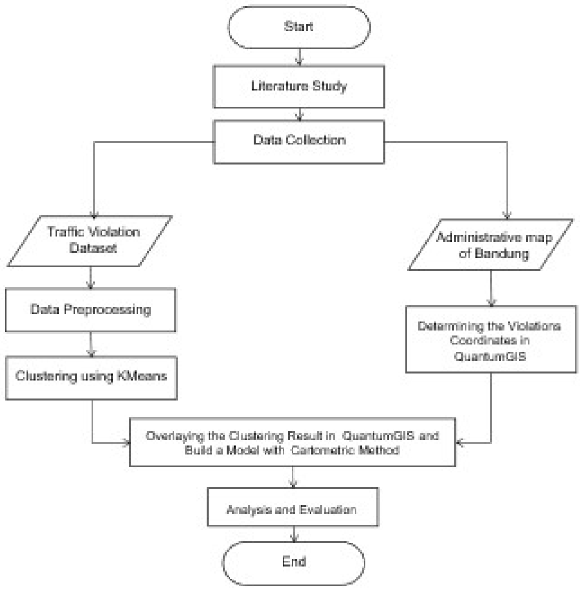

Figure 1. Research Workflow

2.1.2 Non-Spatial Data

Non-spatial data was collected from the Portal of Indonesian Open Data (`https://data.go.id/`). This dataset records traffic violations from the city of Bandung in the year 2018, totaling 2,316 violations. The violations are classified into the following ten categories: failure to wear a helmet, stopping at a zebra crossing, designated stopping area clipping (halting at a stopping point), stopping over the stop line, absence of a driver's license, overloading, mirror-less vehicle, non-bearing vehicle documentation, disregard for pre-ordained traffic directions, and blunt red-light disobedience. The relevant dataset contains the spatial data of the violations.

This study makes use of quantitative data obtained from the Portal of Indonesian Open Data, which falls under the sector of the Secretariat One Indonesian Data of the Minister of National Development Planning (BAPPENAS). The data selected for this analysis pertains to Bandung city and its traffic violations for the year 2018.

2.2 Data Preprocessing

Data preprocessing is one of the most critical steps in the K-means clustering approach. In this particular study, the initial dataset comprised 16 attributes but only 10 of them were selected for clustering. These attributes include: helmet usage, halting at zebra crossings, stopping in designated areas (RHK), breaching the stop line, smoking, overloading a vehicle, other violations, the location of the violation, the time of the violation, two-wheeled and four-wheeled vehicles. The careful selection of these attributes is based on their direct relevance to the traffic violations and their usefulness in providing pertinent information for the clustering process.

The next step after confirming the relevant attributes is to proceed with the data cleaning stage. This step includes deletion of unnecessary data points without useful contributions to the set. Cleaning the data is necessary so that the clustering algorithm applies its computations with minimized redundancy, inefficiency, and improved accuracy. The cleaner the data is, the less noise is introduced, the more precise the data will be. This ensures that the K-means algorithm works on optimal data, thereby enhancing the data preprocessing phase.

The last part of preprocessing is the transformation of data, which involves selecting information relevant to the problem at hand and encoding it numerically for clustering purposes. This is the same as how continuous data may be split into ranges or categories for ease of use by the algorithm. As an example, the attribute "intersection" was more aptly replaced with "region" so as to enhance ease of reference in the analysis and interpretation of data for effective clustering.

In summary, data preprocessing encompasses just about every relevant attribute to be selected, cleaning up duplicate and conflicting entries, and changing such information into K-means friendly formats that the algorithm can digest. These procedures aim to prepare the dataset properly so that the algorithm can make meaningful and reliable clusters.

2.3 K-Means Clustering

Clustering is a basic technique in data mining that helps to group similar objects within a large collection of data [15]. A cluster comprises data points that are related to each other within a certain proximity, and are distinctly separated from other clusters. It stems from the fact that there are no labels provided is called unsupervised learning. Han et al [12] explained this form of learning as a process which does not depend on externally provided labels but rather organizes the data based on inherent similarities between the objects.

The K-means algorithm, which is a subdivision of non-hierarchical clustering, was first proposed by MacQueen in 1976. It represents one of the partitioning methods of clustering. The dataset is split into two or more clusters on the basis of common attributes. Data points that show similarity with certain attributes form a cluster whereas those that possess differing features are separated into another cluster. This method incorporates a distance metric, namely the shortest distance from each data point to the centroid for clustering [16]. This algorithm then proceeds to recalibrate the centroids after each iteration and reassign the data points until a minimal distance is achieved, resulting in optimal partitioning of the dataset. Figure 2 shows the flowchart showing how the K-Means algorithm works.

1. Determine the number of clusters to be formed.
2. Determine the center point (centroid). Calculating the distance of each object to the center (centroid) using the Euclidean distance theory available in Equation (1).

$$Distance = \Sigma_j^k = 1\Sigma_{n\in S_j}|x_n - \mu_j|^2 \quad (1)$$

3. Group data based on the shortest distance between data and centroid.
4. The calculation of the new centroid point is done by calculating the average value of each cluster formed as in

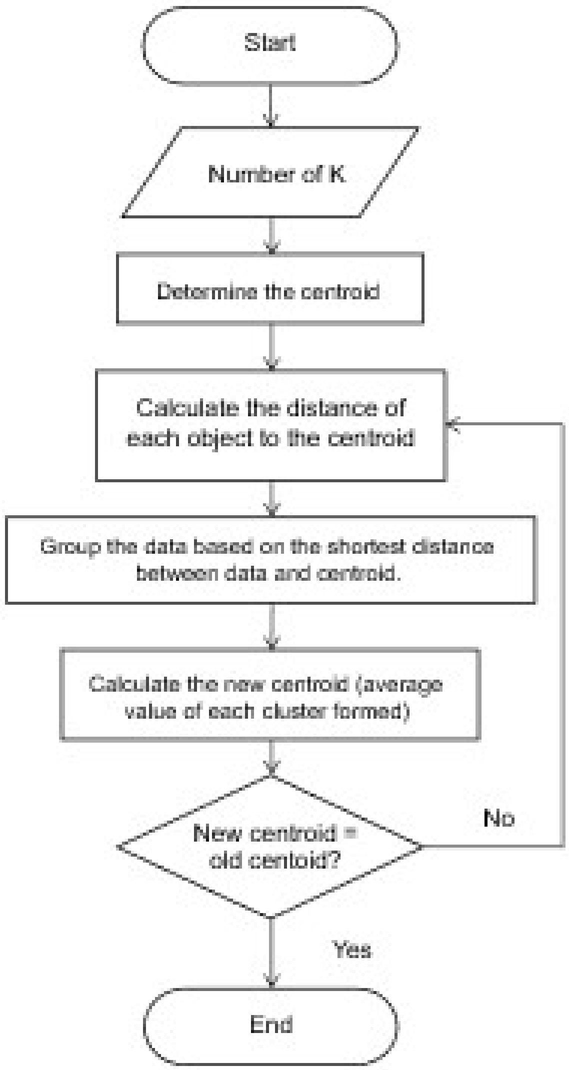

Figure 2. K-Means Algorithm

Equation (2).

$$\mu_k = \frac{1}{N_k} \Sigma_{q=1}^{N_k} x_q \tag{2}$$

This calculation process stops when the centroid point does not change anymore.

Mohiuddin et al [17] present a systematic and synoptic overview of research on the K-means algorithm to address the issue of the algorithm's inability to handle varied data types. Variants of the K-means algorithms, including recent advances, are explained, and their performance is assessed using experimental analysis of a variety of datasets. A data mining strategy that combines K-means clustering and bagging neural networks had been investigated by [18, 19], and many more.

To conclude, the K-means algorithm serves as an effective method to organize data into relevant clusters considering all aspects ranging from its underlying methodology to usability across numerous industries like pattern detection, customer classification, and even outlier identification.

2.4 Cartometric Method

Once the clustering results are obtained, the next step is to perform modeling using the cartometric method. At this stage, the results of clustering are interpreted into three separate clusters: 1) Identification of time-based high-risk areas for traffic violations; 2) Identification of location-based high-risk areas for traffic violations; and 3) Identification of traffic violation areas per category of infraction. Through these clusters, the model sheds light on the key factors and conditions that lead to the highest probability of violations and consequently serves as a basis for more customized strategies and optimal resource distribution.

For the purpose of analyzing and visualizing the spatial information and the three clusters, Geographic Information Systems (GIS) is utilized. GIS is defined as a collection of tools for acquiring, keeping, retrieving, manipulating, analyzing, and displaying data that have a spatial component (geographic or could be mapped). These systems are fundamental in making informed decisions in planning and policy making for management of natural resources, monitoring environmental conditions, dealing with transportation and towns planning, and even in office work and bureaucratic issues [20]. Using GIS does not simply lead to the drawing of the spatial information onto a map, rather, it allows extracting and generating intelligence which assists decision-making.

The cartometric method used in this study centers around delineating boundaries on a map draft and measuring the coordinates of relevant points, lines, distances, and areas within the region of interest. This method uses base maps and other geospatial information to define the boundaries and make spatial measurements [21]. This research effectively combines cartometry with GIS so that diagrammatic representations of traffic violation data can be processed and visually mapped out in a structured manner to highlight socio-geographic problem areas.

2.5 Analysis and Evaluation

After performing cartometric processes, the resultant maps are analysed and evaluated in relation to the traffic violation data sourced from BAPPENAS to assess alignment. The importance of this evaluation step is that it certifies the modelled results correlate with reality in the patterns of traffic violations and in relation to the observed data. Verification of the model output, in comparison with actual violation data of 2018, affirms whether the identified clusters-location, time, and nature of violations-are accurate representation of the phenomena captured by data.

Deficiencies within the modelled clusters and original data are resolved during this stage of evaluation. These gaps and errors may highlight opportunities for enhancement within the model, perhaps through modification of the clustering criteria, or inclusion of other considerations that were not part of the analysis framework. This evaluation is also necessary to establish the extent of certainty and usefulness that can be attributed to the cartometric technique in pinpointing areas prone to traffic violations. It serves as a last validation to ascertain that the model will not only be adequate, but useful for urban development and traffic control strategies.

3. RESULTS AND DISCUSSION

This study aimed to compute and evaluate traffic violation hot spots using the K-Means clustering technique. Following the completion of the clustering calculations, the results were transformed into a map format using cartometric techniques, which included the spatial coordinates of the points where incidents took place. The cluster analysis resulted in the formation of three distinctive clusters with differing levels of proneness: Cluster 1 was classified as "very prone," Cluster 2 as "moderately prone," and Cluster 3 as "prone" in regard to susceptibility to traffic violations. These clusters depict the spatial characteristics of traffic violations in Bandung City.

The outcome of the clustering analysis was derived from three dominant criteria which delineate different angles of vulnerability. The first parameter analyzed the type of violation, for instance, the offenses of speeding, illegal parking, or running a red light. The second parameter analyzed the time frame, categorizing violations into two broad groups: morning and evening. Lastly, the third parameter measured vulnerability from the perspective of the offender's vehicle, categorizing vehicles as two-wheeled, four-wheeled, or larger vehicles.

The modelling procedure utilized an administrative map of Bandung City and the coordinates of the roads in relation to the traffic violations which were extracted from Google Maps. The Bandung City police provided TKP or scene of the accident, or crime scene data, which aided in determining the location points for the traffic violation scenes. This data provided a precise way to retrace the steps an offender would take to commit the violations, considering the sub-districts and districts within the city. After the incident locations were determined, the selected coordinates were plotted in the administrative and road network maps of the Bandung City, illustrating the clusters proportional to their levels of vulnerability.

3.1 K-Means Clustering Process

The K-means algorithm is one of the most popular methods used in the fields of data mining and machine learning because of its straightforwardness and effectiveness. The ability to divide datasets into subsets known as clusters based on likeness is very helpful in revealing the hidden information in the datasets. In this case, traffic violation records of Bandung City are clustered in groups to ascertain behavioral and high-risk area patterns based on certain types of violations.

This approach applies the Euclidean Distance measure as it computes the straight distance between two points in multidimensional space, ideal for continuous data such as metrics for traffic violations. This measure does work effectively especially for the case where clustering of instances that are close to each other in parameters is the goal. It is very useful when the dataset contains numerical information like the geometric location coordinates, the time, or frequency of the traffic violations.

Analyzing 57 study areas with a spectrum of violation types provides an opportunity to advance understanding of traffic patterns and determine locations in Bandung City where certain classes of violations are concentrated. The results would also be helpful to the Department of Urban Planning and the Traffic Department as they work on planning and managing the city's traffic systems by indicating areas considered as hotspots where counteractive measures would be highly useful. Figure 3 shows the visualization of the distance calculation between each traffic data violation locations.

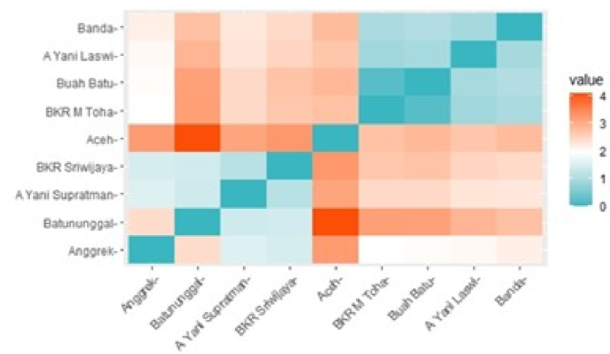

Figure 3. Visualization of Euclidean Distance Result

In the visualization in Figure 3, the color red is used to indicate differences, while the color blue represents similarities or similarities between data points. The visualization illustrates that the more intense the red color on the plot, the greater the distance or gap between the data points. For example, the distance between the intersection of Ahmad Yani Laswi and Supratman is 305.15078, as shown by the red coloration on the plot, which signifies a larger spatial gap between these two locations.

There are different methods to determine the optimal number of clusters, in this case, we have the Elbow Method, which is what we utilize here. The Elbow Method works by checking what percentage of the ratio between the clusters contributes to a bend in the curve forming an "elbow." This indicates the ideal number of clusters since the decrease in the rate of Sum of Squares Error (SSE) significantly reduces at this point.

The comparison results are obtained from the Elbow Method is performed on the data sets to obtain SSE for each potential number of clusters. Each data point is assigned to a certain cluster centroid which then gets assigned to a squared distance. The more the clusters the lesser the SSE, but after a point the amount of reduction SSE exhibits starts to become minimal and that is what is referred to 'elbow' point. This point is best assumed to consider in the case referring optimal number clusters because it offers a good balance between model complexity and goodness of fit. Figure 4 demonstrates the results obtained from the analysis conducted to find the optimal number of clusters. The maximum number of identified clusters where data can be grouped is three.

The next step is the random initialization of each cluster's centroid points. The distance from the centroid to every data

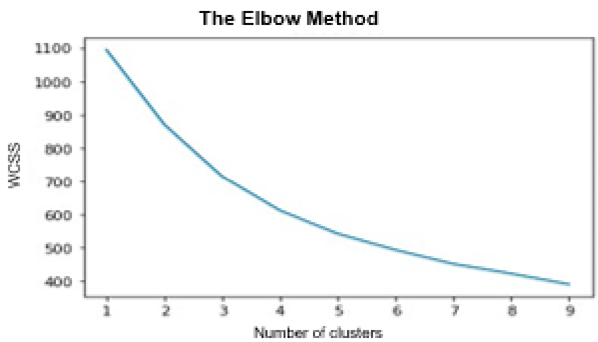

Figure 4. The Plot of Elbow Method

point in the dataset is computed using Euclidean Distance as shown in (1). This distance calculation is critical with regards to how the grouping will be performed based on the proximity of the data points to the centroids. The clustering steps follow the follows the K-Means algorithm available in Figure 2. Figure 5 displays the final result of the clustering process

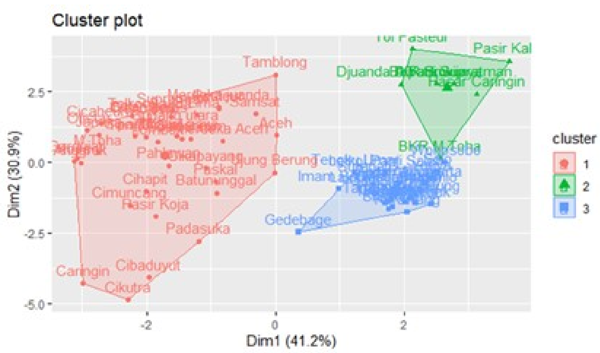

Figure 5. Scatter Plot of the K-Means Result

3.2 Modelling Based on Time Event

Traffic infractions are not bound by hour or address; they pop up whenever and wherever a vehicle meets an inattentive driver. The recorded data from Bandung City in 2018 splits the day into a morning shift (06.00-12.00) and an afternoon shift (12.01-18.00), a division designed to spotlight changing risk profiles rather than to mimic the neat timetables of public transit. In the first stretch, surveyors logged 1,014 breaches, the bulk of which fell into a pedigree the analysts label very vulnerable. Three neighborhoods-Astana Anyar, Sumur Bandung, and Sukajadi, accounted for nearly all of that morning cluster.

When the afternoon period is scanned, a nearly identical tally emerges-1,019 violations-though the trouble centers shift to Coblong, Bandung Wetan, Dipatiukur, and Bandung Kidul. The afternoon hotspots thus differ from those of the morning, a finding that speaks to the patchwork nature of city mobility and underlines why universal deterrents seldom hit the mark. Vulnerability maps, rough sketches of pinpricks across an outline of Bandung, appear as Figure 6 makes the narrative concrete by showing exactly where enforcement resources could make the most difference if they were aimed at the right hour and the right intersection.

3.3 Modelling Based on Violation Category

Data collected throughout 2018 reveal that the most widespread traffic infraction was simply the failure to wear a helmet, a breach almost exclusively recorded among riders of two-wheeled vehicles. This study has grouped this sort of misconduct into what they term the highly vulnerable category, the first cluster identified in the overall study. Districts where the helmet-related offenses were particularly numerous include Kopo Street, Pasir Koja, and Jalan Mohamad Hatta, all of which cut through the Kopo and Soreang administrative areas.

A second cluster, labeled vulnerable, centers on drivers who do not carry the full set of vehicle documents. This shortcoming appears in the records of both motorcycle and car operators alike. Streets such as Cicaheum, Pahlawan, and Ujung Berung, which run through Gede Bage and Arcamanik, each show a steady stream of stops for paper violations.

The moderately vulnerable group is most clearly marked by overload citations. Those fines pile up on four-wheeled traffic and are especially common on Cibaduyut, Gede Bage, and Padasuka Roads, which link neighborhoods in Gede Bage, Arcamanik, and Kopo. Figure 6 offers a color-coded map that lays out these patterns of vulnerability by violation type.

3.4 Modelling Based on Vehicle Type

Statistics from Bandung City traffic enforcement reveal that in 2018 motorcycles accounted for a striking share of the rule-breaking. Riders of two-wheeled vehicles logged 33,456 separate infractions, which works out to 65 percent of the total. By comparison, drivers of cars and larger vehicles committed 9,314 offenses, or the remaining 35 percent. Patrols labeled six spots as especially risky: Tamansari Street, Ujung Berung, Supratman, Telkom, Pasir Koja, and Sulanjana. The imbalance in the data largely stems from scooter and motorbike users who skip helmets, lights, or proper plates and who often ignore posted signs. A vulnerability-map model, which distinguishes violation patterns by vehicle type, is presented as Figure 6.

3.5 Overall Traffic Violation Model

The final modelling of traffic violation vulnerability areas in Bandung City was carried out through an overlay process, combining the three classifications based on the time of occurrence, type of violation, and type of vehicle involved. Figure 7 presents the resulting map of traffic violation vulnerability areas in Bandung City, obtained from the overlay of the previously conducted classifications.

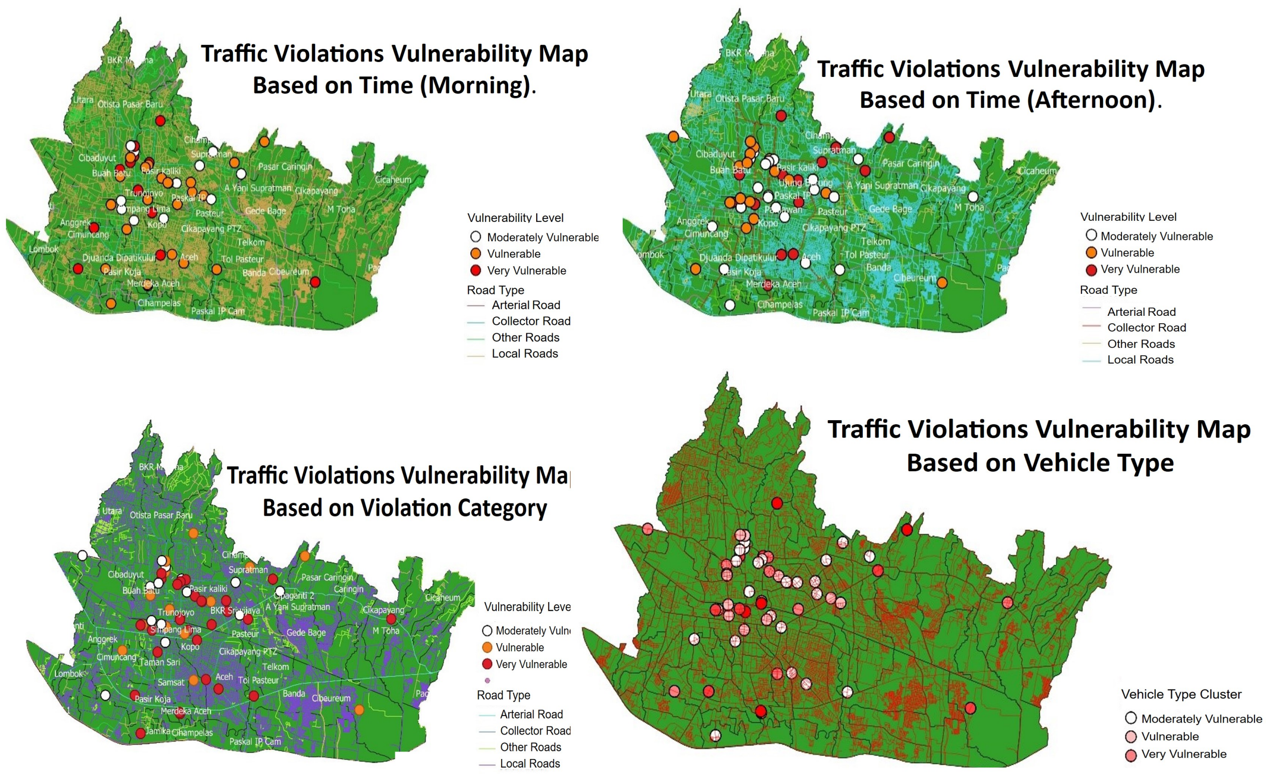

Figure 6. Traffic Violations Vulnerability Map Based on Time Violation Category, and Vehicle Type

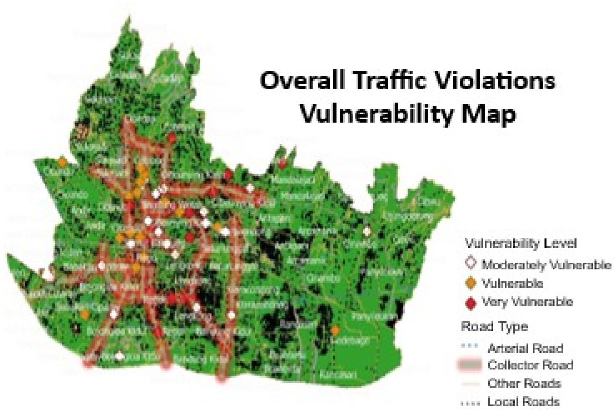

Figure 7. Overall Traffic Violations Vulnerability Map

3.5.1 Moderately Vulnerable Areas are indicated by the color white, with nearly all roads in Bandung City classified as moderately vulnerable to violations. The following districts are categorized as moderately vulnerable:

a. Andir District
b. Kebon Jati District
c. Gedebage District
d. Arcamanik District
e. Kopo District
f. Soreang District
g. Braga District

3.5.2 Vulnerable Areas are shown in yellow, and the areas falling into this category include:

a. Ahmad Yani Laswi Street
b. Ahmad Yani Supratman Street
c. Aceh Street
d. Gedebage Street
e. Ciumuncang Street
f. Cikapayang Street
g. Cihapit Street

3.5.3 Highly Vulnerable Areas are indicated in red, and the areas falling into this category include:

a. Caringin Street
b. Tamansari Street
c. Taman Sari Street
d. Otista Street
e. PasKal Street
f. Pasteur Street
g. Cikutra Street
h. Cipaneglit Street
i. Pahlawan Street

3.6 Evaluation

A thorough evaluation stage compared the predicted map to firsthand records of traffic offenses, confirming how closely the model corresponded to real-world events. Observers tallied 603 violations in regions flagged as moderately vulnerable, 498 in those marked vulnerable, and 930 in areas designated as very vulnerable. The resulting accuracy coefficient was computed according to the formula presented in Equation (3).

$$\text{Accuracy} = \frac{a+b}{n} \times 100\% \tag{3}$$

Where:
a = number of highly vulnerable violation cases in 2018
b = number of vulnerable violation cases in 2018
n = total number of violation cases in 2018

By substituting the values into the equation:

$$\text{Accuracy} = \frac{930+498}{2031} \times 100\% = 71.5\%$$

Based on the calculation, the accuracy level of the vulnerability mapping for traffic violations was found to be 71.5%. This accuracy level was obtained by summing the number of cases in the vulnerable and very vulnerable categories.

4. CONCLUSIONS

The analysis conducted in this study applied a three-cluster system and identified five districts of Bandung City that rank as highly vulnerable to traffic infractions-Gedebage, Kopo, Soreang, Arcamanik, and the paired neighborhoods of Braga and Cihampelas. Four other districts-Andir, Sarijadi, Kebon Jati, and Gunung Batu-placed into the vulnerable category, while Lembang, Dipatiukur, and Ujung Berung settled into the moderately vulnerable category. Cartometric tracing anchored to Google Maps imagery later confirmed, to within 71.5% accuracy, the precise hotspots where violations clustered. This spatial portrait supplies Bandung officials, as well as residents, with an actionable roadmap for directing enforcement where it matters most.

Future research in the field of traffic data analytics might benefit from the adoption of alternative clustering techniques, including Fuzzy -means and Mixture Modelling, both of which promise to enhance the stability and precision of the final output. Moreover, regionally dispersed mapping of traffic violations would allow researchers to observe variations in vulnerability over time and across different locales. Integrating a Geographic Information System into the vulnerability mapping process and making that system accessible through web or mobile interfaces would democratize the information and empower community engagement. Finally, the Bandung City Police could bolster compliance by enforcing stricter penalties and by running educational campaigns in schools, workplaces, and neighborhood gatherings on the dangers of overlooking traffic rules and signage.

5. ACKNOWLEDGEMENT

Financial support for this research was provided by Universitas Lampung, and it has been greatly appreciated in making this study possible.

REFERENCES

[1] Sirwan K Ahmed, Mona G Mohammed, Salar O Abdulqadir, Rabab G Abd El-Kader, Nahed A El-Shall, Deepak Chandran, Mohammad E Ur Rehman, and Kuldeep Dhama. Road traffic accidental injuries and deaths: A neglected global health issue. *Health Science Reports*, 6(5):e1240, 2023.

[2] World Health Organization. *Global status report on road safety 2023: summary*. World Health Organization, 2023.

[3] Yibeltal Assefa Atalay, Bersufekad Wubie Alemie, Belete Gelaw, and Kelemu Abebe Gelaw. Epidemiology of road traffic accidents and its associated factors among public transportation in africa: systematic review and meta-analysis. *Frontiers in Public Health*, 13:1511715, 2025.

[4] Intan Zainafree, Nadia Syukria, Silfia Addina, and Muhamad Zakki Saefurrohim. 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):82–88, 2022.

[5] Bewket Tadesse Tiruneh, Berhanu Boru Bifftu, and Berihun Assefa Dachew. Prevalence and factors associated with road traffic incident among adolescents and children in the hospitals of amhara national regional state, ethiopia. *BMC Emergency Medicine*, 19:1–6, 2019.

[6] Sarun Jongrak. Food delivery rider is not worker: the distortion of the thai legal system. *CMU Journal of Law and Social Sciences*, 14(2):38–58, 2021.

[7] Siriaran Kwangsukstith, Vithawat Surawattanasakul, Chollada Mahakkanukrauh, Jinjuta Panumasvivat, Wachiranun Sirikul, Amornphat Kitro, and Penprapa Siviroj. Major injuries and associated factors in traffic accidents among motorcycle food delivery riders during the covid-19 pandemic in thailand. *Heliyon*, 10(20), 2024.

[8] Omar Elfahim, Marouane El Midaoui, Mohamed Youssfi, Omar Bouattane, et al. Traffic violations analysis: Identifying risky areas and common violations. *Heliyon*, 9(9):e19058, 2023.

[9] Jianyu Wang, Shuo Ma, Pengpeng Jiao, Lanxin Ji, Xu Sun, and Huapu Lu. Analyzing the risk factors of traffic accident severity using a combination of random forest and association rules. *Applied Sciences*, 13(14):8559, 2023.

[10] Sutanto Soehodho. Public transportation development and traffic accident prevention in indonesia. *IATSS Research*, 40(2):76–80, 2017.

[11] Dwi Prasetyanto, Muhammad Rizki, and Andrean Maulana. The role of attitude, behavior, and road conditions in traffic violation among workers and students in bandung city, indonesia. *The Open Transportation Journal*, 17:2667–1212, 2023.

[12] Han Jiawei, Kamber Micheline, and Pei Jian. *Data Mining: Concepts and Techniques.-3rd*. Morgan kaufmann, 2012.

[13] S Nagendra Babu and J Jebamalar Tamilselvi. A data mining framework to analyze road accident data using map reduce methods ccmf and tcamp algorithms. *International Journal*

of Simulation–Systems, Science & Technology, 19(5):p1–7, 2018.

[14] Sachin Kumar and Durga Toshniwal. A data mining framework to analyze road accident data. *Journal of Big Data*, 2:1–18, 2015.

[15] Salvador García, Julián Luengo, Francisco Herrera, et al. *Data preprocessing in data mining*, volume 72. Springer, 2015.

[16] Asroni Asroni and Ronald Adrian. Penerapan metode k-means untuk clustering mahasiswa berdasarkan nilai akademik dengan weka interface studi kasus pada jurusan teknik informatika umm magelang. *Semesta Teknika*, 18(1):76–82, 2015 (in Indonesia).

[17] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8):1295, 2020.

[18] Wenbin Wu and Mugen Peng. A data mining approach combining k-means clustering with bagging neural network for short-term wind power forecasting. *IEEE Internet of Things Journal*, 4(4):979–986, 2017.

[19] Emi Sukiyah and Nur Khoirullah. The bandung city spatial planning policies in geological perspective. *Journal of Geological Sciences and Applied Geology*, 4(1):48–54, 2020.

[20] Nurpilihan Bafdal, Kharista Amaru, and Boy Macklin Pareira. *Buku Ajar Sistem Informasi Geografis*. Teknik Manajemen Industri Pertanian; Universitas Padjajaran, 2011.

[21] MA Khafid, AP Wicaksono, DR Saputra, FW Andita, and DS Wibowo. Geospatial modeling analysis of potential inundation impact of sea level rise: A case study of south coast yogyakarta. In *IOP Conference Series: Materials Science and Engineering*, volume 830, page 032075. IOP Publishing, 2020.