INDOTECH
Indonesian Journal of Education And Computer Science
Vol. 1, No. 2, Agustus 2023
Hal 66-72
E-ISSN : 2987-2650
P-ISSN : 2987-7644
Site : https://jurnal.intekom.id/index.php/indotech

Application of the K-Means Algorithm in Traffic Violations In Langkat District (Case Study: Langkat Police)

Elisa Puspita Sari[1], Yani Maulita[2], Milli Alfhi Syari[3]
[1,2,3] Program Studi Teknik Informatika, STMIK Kaputama, Binjai, Indonesia

Article Info

Article history:

Received August 24, 2023
Revised August 24, 2023
Accepted August 31, 2023

Kata Kunci:

Aktivitas Lalu Lintas,
Preferensi Kendaraan,
Pelanggaran Lalu Lintas,
Klaster Data Mining

Keywords:

Traffic Activities
Vehicle Preference
Traffic Violations
Data Mining Clusters

ABSTRAK

Aktivitas masyarakat berhubungan dengan lalu lintas dan masyarakat lebih memilih menggunakan kendaraan. Rendahnya edukasi serta minim pemahaman tentang peraturan lalu lintas menyebabkan banyak pelanggaran. Meningkatnya jumlah pelanggar lalu lintas menyebabkan meningkatnya data pelanggaran lalu lintas. Banyaknya data pelanggaran lalu lintas menyebabkan terjadinya penumpukan data pada instansi. Maka diperlukan suatu pengolahan data dengan data *mining* menggunakan Algoritma K-*Means.* Hasil penelitian diketahui kelompok data pelanggaran lalu lintas yang memiliki kelompok paling tinggi dan paling sering muncul saat diproses yaitu usia 17-25 tahun, dengan kendaraan Honda Vario 150 dan bukti pelanggaran SIM dan STNK. Hasil pengujian 3 cluster dari 502 data pelanggaran diketahui yaitu *cluster* 1 kelompok data pelanggaran lalu lintas usia 26-45 tahun jenis kendaraan Honda CBR 250 dan bukti pelanggaran SIM dan STNK. *Cluster* 2 kelompok data pelanggaran lalu lintas usia 26-45 tahun dengan jenis kendaraan Suzuki Nex dengan bukti pelanggaran SIM dan boncengan lebih dari 1. *Cluster* 3 yaitu kelompok data pelanggaran lalu lintas usia 17-25 tahun, dengan jenis kendaraan Honda Vario 150 dan bukti pelanggaran SIM.

ABSTRACT

Societal activities are intertwined with traffic, and people prefer using vehicles. The lack of education and limited understanding of traffic regulations have led to numerous violations. The increasing number of traffic violations has resulted in a rise in traffic violation data. The abundance of traffic violation data has led to data accumulation within institutions. Therefore, data processing through data mining utilizing the K-Means Algorithm is deemed necessary. Research findings have unveiled a cluster of traffic violation data that stands out as the highest and most frequent during processing: the age group of 17 to 25 years, involving Honda Vario 150 vehicles, and evidence of violations related to driver's licenses (SIM) and vehicle registration certificates (STNK). Test results on three clusters from a dataset of 502 traffic violation records reveal the following: Cluster 1 comprises traffic violation data pertaining to individuals aged 26 to 45 years, using Honda CBR 250 vehicles, and violations tied to driver's licenses (SIM) and vehicle registration certificates (STNK). Cluster 2 includes traffic violation data concerning individuals aged 26 to 45 years, utilizing Suzuki Nex vehicles, and violations involving driver's licenses (SIM) as well as carrying more than one passenger. Cluster 3 involves traffic violation data associated with individuals aged 17 to 25 years, employing Honda Vario 150 vehicles, and violations linked to driver's licenses (SIM).

Corresponding Author:

Elisa Puspita Sari
Program Studi Teknik Informatika, STMIK Kaputama
Binjai, Indonesia
Email: elisapuspitasari47@gmail.com

1. INTRODUCTION

In the modern era which is full of dynamics and increased mobility, the traffic sector has a crucial role in maintaining the smoothness, safety and governance of transportation as a whole. Langkat district, as an important part of Indonesia that is experiencing rapid growth in the number of vehicles and road infrastructure, faces serious challenges regarding traffic regulation and law enforcement. Increasing traffic violations are an issue that needs to be addressed with innovative approaches, one of which is through the application of the K-Means algorithm [1].

The K-Means algorithm, as a method in data analysis and clustering, appears as a promising solution to overcome the complexity of traffic violation patterns [2]. This method allows data to be grouped based on certain characteristics, which in turn can assist the authorities, including the Langkat Police, in taking more targeted law enforcement actions [3].

The purpose of this research is to apply the K-Means algorithm to data on traffic violations in Langkat District, by exploring a case study at the Langkat Police Station. Thus, this study aims to identify patterns of violations that may be difficult to recognize manually, and support better law enforcement decisions [4]. The application of the K-Means algorithm in the analysis of traffic violations has great potential to provide valuable information for the Langkat Police in designing a more effective law enforcement strategy. By grouping traffic violations into similar groups, police can allocate their resources more intelligently and focus on the areas most prone to violations [5]. It should be emphasized that implementation of the K-Means algorithm on a larger scale may face certain technical and computational constraints. Processing large and complex data volumes requires adequate hardware and software to ensure accurate and efficient analysis [6].

In addition, this research is also faced with challenges in terms of the availability of accurate and complete data on traffic violations. The quality of data taken from existing sources greatly affects the validity and results of the analysis performed [7]. However, the potential offered by the K-Means algorithm in the analysis of traffic violations in Langkat Regency is very interesting to be explored further. By combining data on traffic violations with relevant variables such as time, location and traffic conditions, law enforcement can build a more complete understanding of the patterns of violations that occur [8].

Furthermore, the application of the K-Means algorithm has the potential to have a wider impact on the development of transportation policies in Langkat Regency. Information generated from data analysis can assist authorities in identifying priority areas that require infrastructure improvements or stricter law enforcement actions [9]. In conclusion, the application of the K-Means algorithm in the analysis of traffic violations in Langkat District is a promising step towards increasing the effectiveness of law enforcement and traffic management as a whole. Through a more careful and precise analysis, it is hoped that this research can make a significant contribution in creating a safer and more orderly traffic environment in this region [10].

2. METHOD

The research stages that will be used are interviews, literature study, field research in the form of observations, and implementation of coding. The design process that will be built in grouping data on traffic violations using the clustering method will have a flowchart description as follows:

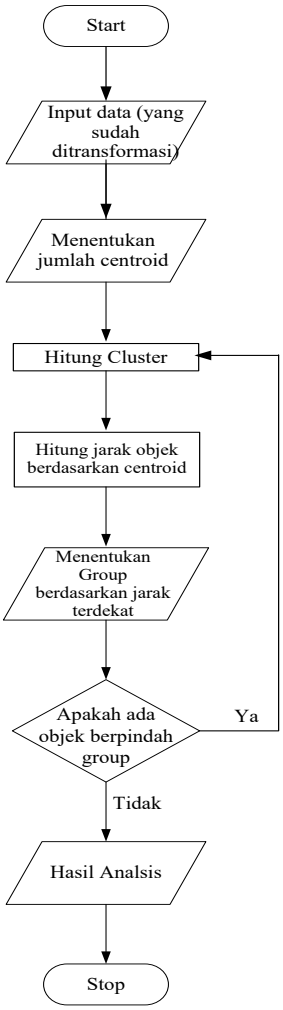

Figure 1. Flowchart for the system to be designed.

In understanding the flow of data mining activities in the clustering process in the system that will later be built, you can see the Unified Modeling Language (UML) system process, such as the following clustering process activity diagram.:

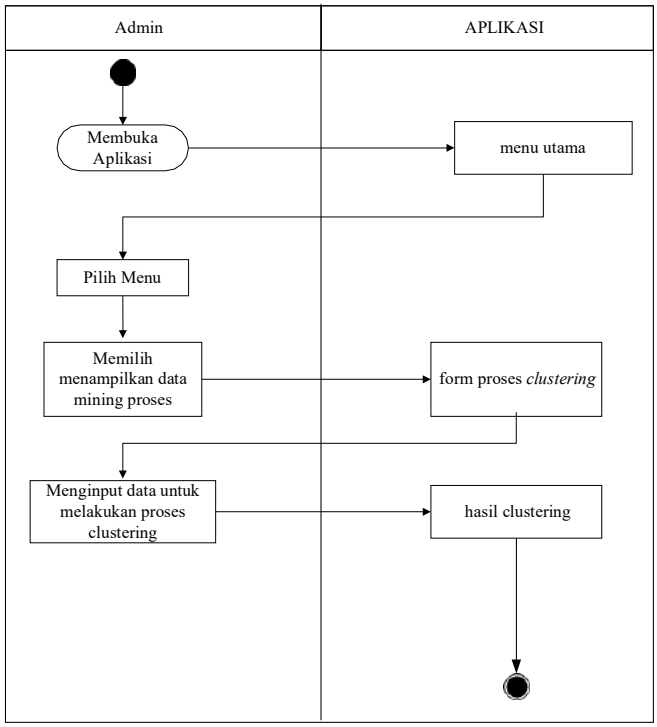

Figure 2. Clustering Process Activity Diagram

3. RESULTS AND DISCUSSION

In this discussion, we will explain the results of testing the k-means algorithm development software using Matlab (matrix laboratory) to obtain grouping results from predetermined variables.

Figure 3. Test Results 1

Of the 502 data on traffic violations obtained by 3 clusters, where cluster 1 has 155 data, cluster 2 has 130 data and cluster 3 has 217 data. Here is the description of the cluster center on the graph:

1. 1. (2.54) (26.14) (10.61)

It can be seen that in cluster 1. The traffic violation data group is aged 26-45 years, with the type of motorized vehicle, namely the Honda CBR 250 and evidence of SIM and STNK violations.

2. 2. (2.68) (50.61) (10.36)

It can be seen that in cluster 2, the traffic violation data group is aged 26-45 years, with the type of motorized vehicle, namely Suzuki Nex and more than 1 evidence of SIM and pillion violations.

3. 3. (2.45) (9.33) (8.47)

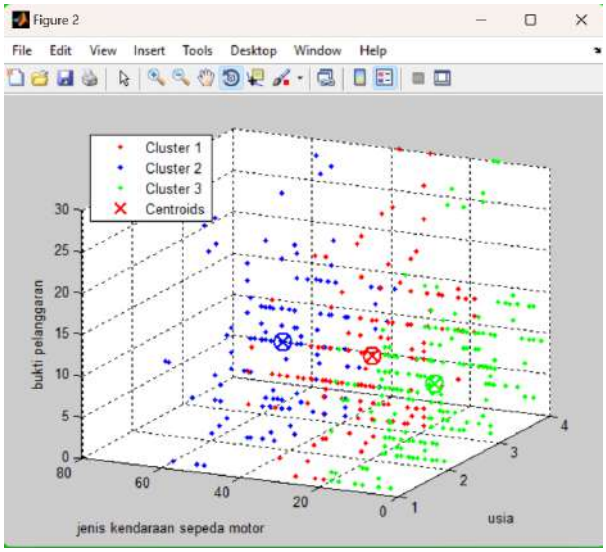

Figure 4. Graph of Test Results 1

It can be seen that in cluster 3. The traffic violation data group is aged 17-25 years, with the type of motorized vehicle, namely the Honda Vario 150 and evidence of a driver's license violation

Figure 5. Test Results 2

Of the 502 data on traffic violations obtained by 3 clusters, where cluster 1 has 216 data, cluster 2 has 138 data and cluster 3 has 148 data. Here is the description of the cluster center on the graph:

1. (2.45) (9.29) (8.44)

It can be seen that in cluster 1. The traffic violation data group is aged 17-25 years, with the type of motorized vehicle, namely the Honda Vario 150 and evidence of a helmet violation and more than 1 pillion

2. (2.66) (49.88) (10.48)
It can be seen that in cluster 2, the traffic violation data group is aged 26-45 years, with the type of motorized vehicle, namely the Kawasaki Ninja 300 and evidence of a driver's license violation and more than 1 pillion.

3. (2.55) (25.44) (10.55)
It can be seen that in cluster 3. The traffic violation data group is aged 17-25 years, with the type of motorized vehicle, namely the Honda Vario 150 and evidence of SIM and STNK violations

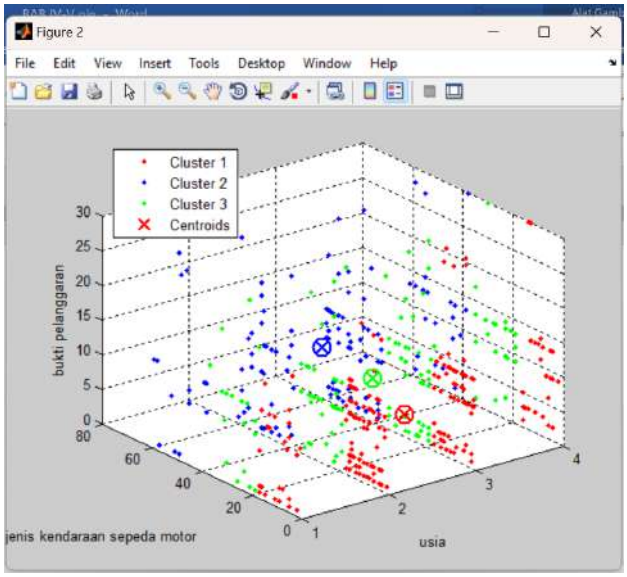

Figure 6. Graph of 3 Cluster Results

3. CONCLUSION

In cluster 3 from the criteria for age, type of motorized vehicle, and evidence of violations, the group that has the highest set/value and the most in the job seeker data, namely in cluster 3, which totals 217 traffic violation data with ages 17-25 years, with the type motorized vehicles, namely the Honda Vario 150 and proof of SIM and STNK violations.

From the tests carried out using the clustering method with the K-Means algorithm, it can be seen that the traffic violation data group has the highest group and occurs most often during processing, namely at the age of 17-25 years, with the type of motorized vehicle, namely the Honda Vario 150 and evidence SIM and STNK violations. From the tests carried out on 502 traffic violation data obtained 3 clusters, where cluster 1 totaled 155 data, cluster 2 totaled 130 data and cluster 3 totaled 217 data. It can be seen that in cluster 1. The traffic violation data group is aged 26-45 years, with the type of motorized vehicle, namely the Honda CBR 250 and evidence of SIM and STNK violations. Cluster 2. Traffic violation data group with age 26-45 years, with the type of motorized vehicle, namely Suzuki Nex and more than 1 proof of SIM and pillion violations. Cluster 3. Traffic violation data group with age 17-25 years, with type of vehicle motorbike, namely the Honda Vario 150 and evidence of a driver's license violation.

REFERENCES

[1] Sukandarrumidi, H., & Kusworo, A. Analysis of Traffic Offense Violation Using K-Means Clustering. *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, no. 2, pp 109-115. 2019
[2] Sulistyo, B., Budiharjo, E., & Arymurthy, A. M. Road Accident Clustering Analysis Using K-Means Clustering Method. *Jurnal Ilmiah Teknologi Informasi Terapan*, vol. 3, no. 1, pp. 27-33. 2017
[3] Adriyanto, L., Suryanto, S., & Marwanto, A. A. K-Means Clustering and Hierarchical Clustering for Data Traffic Accidents in the City of Bandar Lampung. Jurnal Ilmiah Sistem Informasi, vol. 6, no. 2, pp. 191-196. 2016
[4] Widianto, D. B., & Nursani, R. Road Traffic Accident Cluster Based on the K-Means Algorithm. *IOP Conference Series: Materials Science and Engineering*, vol. 18, no. 1, 012102. 2017
[5] Ardiansyah, F., Indriani, D., & Firdausy, C. M. Application of K-Means Clustering for Identifying Types of Road Violations. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 786-791. 2020
[6] Laksana, R. A., & Lestari, A. K-Means Clustering Method in Determining Traffic Accident Distribution. Jurnal Teknik ITS, vol. 6, no. 1, pp. C76-C80. 2017
[7] Herawati, D., Mukhlason, A., & Nofriansyah, F. K-Means Clustering Method on Road Accident Data in Indonesia. *IOP Conference Series: Materials Science and Engineering*, vol. 702, no. 1, 012007. 2020.
[8] Wiranegara, H., & Akbar, A. R. Application of K-Means Clustering Algorithm for Traffic Accident Data Analysis. Jurnal Penelitian Ilmu Komputer dan Teknologi Informasi, vol. 6, no. 2 , pp.124-131. 2019
[9] Yunus, M. M., Pratama, B. A., & Hermanto, D. Y. Clustering Analysis of Traffic Accident Data using K-Means Algorithm in Jember City. Jurnal Teknologi Informasi dan Ilmu Komputer, vol. no. 1, pp. 25-32. 2021
[10] Wahyudi, A., Firmansyah, M. Z., & Fatichah, C. Analysis of Traffic Accidents with K-Means Clustering Method. Jurnal Ilmiah Teknik Sipil Dan Teknik Kimia, vol. 5, no. 2, pp. 43-49. 2019