Application Of Clustering Method In Identifying Traffic Accident Patterns Using K-Means Algorithm In Padang City

*Adeng Pratama [1], Yeka Hendriyani[1], Khairi Budayawan [1], Syafrijon[1]

[#] *Department Electronical Engineering, Universitas Negeri Padang, Jl.Prof. Dr. Hamka Air Tawar, Padang, Indonesia*

**Corresponding Author : adengpratama2000@gmail.com*

***Abstract* - Traffic safety in Padang City has become a major concern due to the increasing number of accidents over the past few years. To identify accident patterns and accident-prone areas, this study applies a clustering method using the K-Means algorithm. Accident data from 2021 to 2023 were analyzed based on variables such as the number of accidents, the number of vehicles involved, and the number of victims. The results of the analysis show that K-Means Clustering with Chebyshev distance is able to group accident locations into three clusters based on the level of vulnerability. The first cluster includes roads that are not prone to accidents, the second cluster includes roads that are prone to accidents, and the third cluster identifies roads with the highest accident rates.**

***Keywords*— K-Means Clustering, Chebyshev Distance, traffic safety, accident pattern identification, Padang City**

I. INTRODUCTION

Traffic safety in Padang City has become an important issue with a significant increase in accidents in the last five years. Padang is recorded as the city with the highest number of accidents in West Sumatra, causing material losses and casualties. To overcome this problem, accident data analysis using information technology, specifically the K-Means Clustering method, can help identify patterns and risk factors.[1]

K-Means Clustering, a data clustering algorithm, can group accident incidents based on variables such as location, time, vehicle type, and weather conditions. With this, authorities can know the areas and times of high risk and the contributing factors, so that they can formulate more effective prevention policies and strategies.[2]

DOI: 10.24036/int.j.emerg.technol.eng.educ..v1i1.12

Corresponding author: Adeng Pratama
Universitas Negeri Padang
Email: antoniilwandy@gmail.com

Received: 19-08-2024
Revised: 20-08-2024
Accepted: 04-10-2024
Published:06-10-2024

For all articles published in IJETEED. https://ijeteed.ppj.unp.ac.id/, © copyright is retained by the authors. This is an open-access article under the CC BY-SA license

Research in other cities, such as Tasikmalaya and Malang, shows that this method helps understand accident trends and design infrastructure improvements and better placement of traffic officers. The application of this method in Padang City is expected to provide in- depth insights to improve traffic safety and design more effective policies.[3]

This study will use accident data related to location, time, vehicle type, and weather conditions, analyzed with K-Means Clustering to provide practical recommendations for the Padang City government. It is hoped that the results of this study can be a model for other cities in Indonesia to improve traffic safety more effectively.[4]

This research will use accident data related to location, time, vehicle type, and weather conditions, which will be analyzed with K-Means Clustering to provide practical recommendations to the Padang City government. It is hoped that the results of this study can also be a reference for other cities in Indonesia in improving traffic safety more effectively.[5]

Data mining is one of the vital aspects of Machine Learning that focuses on identifying patterns and relationships in large transaction datasets to extract new knowledge [6].

Data mining is an integral part of Knowledge Discovery in Databases (KDD). KDD is the overall process of transforming raw data into meaningful information through a series of stages [7].

This entire process is often repeated to improve and increase the results obtained. The knowledge gained from KDD supports decision-making [8]

RapidMiner can be used to perform analysis processes on data mining, text mining, and prediction analysis. RapidMiner uses a variety of descriptive and predictive methods and techniques in the creation of models that can be used in decision-making [9].

Cluster Analysis is one of the techniques used to understand traffic accident patterns by grouping events based on certain characteristics. This technique is important in identifying accident trends and can be utilized to improve road safety [5].

For example, using the K-Means Clustering method, we can group accidents based on the time of occurrence, the type of vehicle involved, weather conditions, and location. This allows for the identification of specific accident hotspots, such as certain intersections where accidents occur frequently during rush hours or certain dangerous roads [5].By identifying accurate accident segmentation and a deep understanding of accident patterns, authorities can strategically design interventions to reduce accidents in vulnerable areas [10].

The algorithm starts with the initial cluster formation and iterates to fix the cluster until there are no more significant changes, indicating algorithm convergence [11]. The algorithm starts with the initial cluster formation and iterates to fix the cluster until there are no more significant changes, indicating algorithm convergence. [12]. The entire process is often repeated to improve and improve the results obtained. The knowledge gained from KDD supports decision-making [8]. With this method, the dynamic behavior of the process model from start to finish can be analyzed, providing the necessary information to improve and optimize the workflow [13].

II. METHODS

This study uses a quantitative method with an exploratory approach to analyze traffic accident data in Padang City. This study aims to understand patterns and relationships in accident data that are not widely known. The K-Means Clustering technique is applied to group accident incidents based on variables such as the number of accidents, the number of vehicles involved, the number of victims. In this way, the study aims to identify accident hotspots.[14]

The location of the accident, the time of the incident, the type of vehicle, the number of victims, and weather conditions. The study population is the entire accident data reported during the period, while the sample is taken by purposive sampling, i.e. selecting the data that is most relevant to the research objectives. The focus of the research is on independent and dependent variables that will be measured and analyzed to identify accidentpatterns[9].

Research instruments include the extraction of accident data from Police records, data processing with python and RapidMiner software. The validity of the data is ensured by the accuracy of extraction from official sources and the examination of the data for errors or shortcomings. Data reliability is tested by repeating the analysis to ensure consistency of results. The use of data analysis software such as RapidMiner helps maintain data integrity and detect errors, with data being divided into subsets to ensure the stability of clustering results.[15]

III. RESULTS AND DISCUSSION

A. Data Management

Preprocessing is carried out by deleting incomplete or irrelevant traffic accident data, such as incidents less than 3 years old and data without the identity of the victim, age, or occupation. The remaining data is complete and ready to use, then processed using Ms. Excel for simplification and generalization.

TABLE I
DATA COLLECTED

No	Street Name	Total Accident	Total Vehicles Engage	Total Victims
1	JL BY PASS	326	670	432
2	JL PADANG SOLOK	93	186	102
3	JL ADINEGORO	146	292	234
4	JL INDARUNG	133	266	189
5	JL GAJAH MADA	59	118	60
6	JL RAYA AMPANG	21	42	27
7	JL RAYA BELIMBING	8	16	8
8	JL M HATTA	81	162	72
9	JL DR SUTOMO	20	40	19
10	JL KHATIB SULAIMAN	16	32	21
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
129	JL ULU GADUT	1	2	1
130	JL VETERAN	15	28	17
131	JL ZAINUL ARIFIN	3	6	4
132	JL SILUNGKANG	1	2	2
133	JL SUTOMO	1	2	2

B. Clustering Process

Chebychev Distance K-Means, The process of clustering areas prone to traffic accidents using the k-means algorithm and Chebychev distance begins by determining the number of clusters to be created, in this case k=3. The initial cluster center is randomly determined, for example C1=(2, 2, 2), C2=(3, 3, 3), and C3=(4, 4, 24). Furthermore, the distance of each data to the center of the cluster is calculated using the Chebychev distance. For example, the distance of the first data is calculated against the center of the first cluster. Chebyshev Distance

$$d_{ij} = \max_{k} |xij - xjk| \qquad (1)$$

Where:

$d_{c1} = \max(|326-2|, |670-2|, |432-2|) = 668$
$d_{c2} = \max(|326-3|, |670-3|, |432-3|) = 667$
$d_{c3} = \max(|326-4|, |670-4|, |432-4|) = 666$

TABLE 2
CHEBYSHEV DISTANCE CALCULATION RESULTS

No	Street Name	C1	C2	C3	Member Cluster
1	JL BY PASS	668	667	666	C3
2	JL PADANG SOLOK	184	183	182	C3
3	JL ADINEGORO	290	289	288	C3
4	JL INDARUNG	264	263	262	C3
5	JL GAJAH MADA	116	115	114	C3
6	JL RAYA AMPANG	40	39	38	C3
7	JL RAYA BELIMBING	14	13	12	C3
8	JL M HATTA	160	159	158	C3
9	JL DR SUTOMO	38	37	36	C3
10	JL KHATIB SULAIMAN	30	29	28	C3
-	-				
-	-				
-	-				
-	-				
129	JL ULU GADUT	1	2	3	C1
130	JL VETERAN	25	24	23	C3
131	JL ZAINUL ARIFIN	4	3	2	C3
132	JL SILUNGKANG	1	2	3	C1
133	JL SUTOMO	1	2	3	C1

Calculate the new cluster center by finding the average of all the data in a given cluster.

TABLE 3
NEW CENTROID FOR LITERACY

	Total Accident	Total Vehicles Engage	Total Victims
C1	1.23	2.00	3.00
C2	2.04	1.04	2.00
C3	50.6	24.24	26.7

Furthermore, the process of calculating the distance and updating the position of the center of the cluster is carried out again with a new center point of the cluster, and repeated until the position of the cluster is stable and unchanged.

131	JL ZAINUL ARIFIN	2,368852	152,4	664	C1
132	JL SILUNGKANG	6,368852	156,4	668	C1
133	JL SUTOMO	6,368852	156,4	668	C1

TABLE 4
FINAL CENTROID INFORMATION

	Total Accident	Total Vehicles Engage	Total Victims
C1	4.21311475	8.36885246	4.51639344
C2	79.2	158.4	102.2
C3	326	670	432

TABLE 5
CALCULATION RESULTS:

No	Street Name	C1	C2	C3	Member Cluster
1	JL BY PASS	661,631	511,6	0	C3
2	JL PADANG SOLOK	177,631	27,6	484	C2
3	JL ADINEGORO	283,6311	133,6	378	C2
4	JL INDARUNG	257,6311	107,6	404	C2
5	JL GAJAH MADA	109,6311	42,2	552	C2
6	JL RAYA AMPANG	33,63115	116,4	628	C1
7	JL RAYA BELIMBING	7,631148	142,4	654	C1
8	JL DR M HATTA	153,6311	30,2	508	C2
9	JL DR SUTOMO	31,63115	118,4	630	C1
10	JL KHATIB SULAIMAN	23,63115	126,4	638	C1
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
123	JL SUNGAI PISANG	3,631148	146,4	658	C1
124	JL TAMAN SISWA	6,368852	156,4	668	C1
125	JL TANJUNG BEROK	3,368852	153,4	665	C1
126	JL TEKNOLOGI	3,368852	153,4	665	C1
127	JL TUT WURI	6,368852	156,4	668	C1
128	JL UJUNG TANAH	6,368852	156,4	668	C1

129	JL ULU GADUT DEKAT	6,368852	156,4	668	C1
130	JL VETERAN	18,63115	131,4	643	C1

The results of clustering identified three road groups based on the level of vulnerability to traffic accidents. The first cluster (C1), with a central point at (4,213, 8,369, 4,516), represents a group of roads that are not prone to accidents, consisting of 122 roads. The second cluster (C2), with a central point at (79.2, 158.4, 102.2), covers accident-prone roads, with a total of 10 roads. The third cluster (C3), which has a center point at (326, 670, 432), shows a group of roads that are very prone to accidents, with only 1 road.

C. Process of Using Rapidminer

With the help of Rapidminer Studio 9.10 tools. The data in TABLE 1 will be processed using rapidminer with Chebychev K-Means, then the centroid table and the clustering result table are obtained as follows:

TABLE 6
CENTROID RESULTS

	Cluster_0	Cluster_1	Cluster_2
Total Accident	4.213	79.200	326
Vehicles Quantity	8.377	158.400	670
Total Victims	4.516	102.200	432

TABLE 7
CLUSTERING RESULTS

id	cluster	Total Accident	92	cluster_0
1	cluster_2	326,0	670,0	432,0
2	cluster_1	93,0	186,0	102,0
3	cluster_1	146,0	292,0	234,0
4	cluster_1	133,0	266,0	189,0
5	cluster_1	59,0	118,0	60,0
6	cluster_0	21,0	42,0	27,0
7	cluster_0	8,0	16,0	8,0
8	cluster_1	81,0	162,0	72,0
9	cluster_0	20,0	40,0	19,0
10	cluster_0	16,0	32,0	21,0
11	cluster_0	6,0	12,0	8,0
12	cluster_0	21,0	42,0	18,0
13	cluster_0	4,0	8,0	3,0
14	cluster_0	4,0	8,0	4,0
15	cluster_0	2,0	4,0	2,0
16	cluster_0	3,0	6,0	2,0
17	cluster_0	3,0	6,0	3,0
18	cluster_0	2,0	4,0	2,0
19	cluster_0	8,0	16,0	8,0
20	cluster_0	12,0	24,0	11,0
21	cluster_0	16,0	32,0	15,0
22	cluster_0	2,0	4,0	2,0
23	cluster_0	4,0	8,0	5,0
24	cluster_0	6,0	12,0	7,0
25	cluster_0	2,0	4,0	2,0
26	cluster_0	2,0	4,0	1,0
27	cluster_0	2,0	4,0	4,0
28	cluster_0	2,0	4,0	3,0
29	cluster_0	2,0	4,0	1,0
30	cluster_0	2,0	4,0	4,0
31	cluster_0	4,0	8,0	4,0
32	cluster_0	1,0	2,0	1,0
33	cluster_0	4,0	8,0	2,0
34	cluster_0	3,0	6,0	2,0
35	cluster_0	2,0	4,0	2,0
36	cluster_0	1,0	2,0	1,0
37	cluster_0	1,0	2,0	1,0
38	cluster_0	6,0	12,0	6,0
39	cluster_0	2,0	4,0	3,0
40	cluster_0	3,0	6,0	3,0
41	cluster_0	1,0	2,0	1,0
42	cluster_0	1,0	2,0	1,0
43	cluster_0	1,0	2,0	2,0
44	cluster_0	1,0	2,0	2,0
45	cluster_0	1,0	2,0	1,0
46	cluster_0	6,0	12,0	7,0
47	cluster_0	11,0	22,0	11,0
48	cluster_0	1,0	2,0	1,0
49	cluster_0	2,0	4,0	2,0
50	cluster_0	1,0	2,0	1,0
51	cluster_0	1,0	2,0	1,0
52	cluster_0	1,0	2,0	1,0
53	cluster_0	6,0	12,0	6,0
54	cluster_1	57,0	114,0	77,0
55	cluster_0	1,0	2,0	2,0
56	cluster_0	2,0	4,0	2,0
57	cluster_0	2,0	4,0	2,0
58	cluster_1	55,0	110,0	86,0
59	cluster_0	2,0	4,0	2,0
60	cluster_0	3,0	6,0	4,0
61	cluster_0	1,0	2,0	1,0
62	cluster_0	1,0	2,0	1,0
63	cluster_0	2,0	3,0	2,0
64	cluster_0	1,0	2,0	1,0
65	cluster_0	1,0	2,0	1,0
66	cluster_0	1,0	2,0	2,0
67	cluster_0	3,0	6,0	4,0
68	cluster_0	1,0	2,0	1,0
69	cluster_0	12,0	24,0	13,0
70	cluster_0	1,0	2,0	1,0

71	cluster_0	2,0	4,0	2,0
72	cluster_0	2,0	4,0	2,0
73	cluster_0	1,0	2,0	1,0
74	cluster_0	7,0	14,0	7,0
75	cluster_0	2,0	4,0	2,0
76	cluster_0	6,0	12,0	2,0
77	cluster_0	3,0	6,0	2,0
78	cluster_1	66,0	132,0	78,0
79	cluster_0	14,0	28,0	8,0
80	cluster_0	5,0	10,0	6,0
81	cluster_0	2,0	4,0	2,0
82	cluster_1	47,0	94,0	56,0
83	cluster_0	1,0	2,0	1,0
84	cluster_0	6,0	12,0	4,0
85	cluster_0	1,0	2,0	2,0
86	cluster_0	9,0	18,0	6,0
87	cluster_0	2,0	4,0	4,0
88	cluster_0	1,0	2,0	2,0
89	cluster_0	1,0	2,0	2,0
90	cluster_0	1,0	2,0	2,0
91	cluster_0	13,0	26,0	16,0
92	cluster_0	8,0	16,0	6,0
93	cluster_0	1,0	2,0	1,0
94	cluster_0	1,0	2,0	2,0
95	cluster_0	2,0	4,0	4,0
96	cluster_0	2,0	4,0	4,0
97	cluster_0	1,0	2,0	2,0
98	cluster_0	1,0	2,0	2,0
99	cluster_0	1,0	2,0	2,0
100	cluster_0	4,0	8,0	4,0
101	cluster_0	2,0	4,0	2,0
102	cluster_0	3,0	6,0	4,0
103	cluster_0	4,0	8,0	4,0
104	cluster_0	6,0	12,0	6,0
105	cluster_0	5,0	10,0	6,0
106	cluster_0	9,0	18,0	8,0
107	cluster_0	14,0	28,0	14,0
108	cluster_0	1,0	2,0	2,0
109	cluster_0	3,0	6,0	4,0
110	cluster_0	1,0	2,0	2,0
111	cluster_0	4,0	8,0	6,0
112	cluster_0	3,0	6,0	4,0
113	cluster_0	11,0	21,0	9,0
114	cluster_0	5,0	10,0	6,0
115	cluster_0	1,0	2,0	2,0
116	cluster_0	6,0	12,0	6,0
117	cluster_0	17,0	34,0	20,0
118	cluster_0	2,0	4,0	2,0
119	cluster_0	1,0	2,0	2,0
120	cluster_0	1,0	2,0	2,0
121	cluster_1	55,0	110,0	68,0
122	cluster_0	14,0	28,0	16,0
123	cluster_0	6,0	12,0	6,0
124	cluster_0	1,0	2,0	2,0
125	cluster_0	3,0	5,0	2,0
126	cluster_0	3,0	5,0	3,0
127	cluster_0	1,0	2,0	1,0
128	cluster_0	1,0	2,0	3,0
129	cluster_0	1,0	2,0	1,0
130	cluster_0	15,0	28,0	17,0
131	cluster_0	3,0	6,0	4,0
132	cluster_0	1,0	2,0	2,0
133	cluster_0	1,0	2,0	2,0

TABLE 8
NAME OF ACCIDENT-PRONE AREA

Not Vulnerable	JL RAYA AMPANG, JL RAYA BELIMBING, JL DR SUTOMO, JL KHATIB SULAIMAN, JL SUNGAI PISANG, JL JEND SUDIRMAN, JL KOTO KACIAK, JL KAMPUNG KELAWI, JL MARAPALAM, JL SAWAHAN, JL BY PASS KM 9 DPN, JL BY PASS KM 9, JL BY PASS KM 07, JL BY PASS KM 08, JL BY PASS KM 10, JL BY PASS KM 12, JL BY PASS KM 16, JL BY PASS KM 19, JL BY PASS KM 25, JL KENANGA AIR, JL PONDOK DEPAN, JL PARAK ANAU RAYA, JL SD NEGERI 14, JL BARU, JL DAMAR, JL BUNDO KANDUNG, JL PEMUDA, JL TANAH SIRAH, JL SIMP PERUMDAK, JL SEI PISANG, JL BANDAR DAMAR, JL PROKLAMASI, JL BHAKTI, JL RAYA PADANG SARAI, JL RAYA KORONG, JL KOTO PULAI, JL AIR CAMAR, JL TAMSIS, JL RAYA SEI LAREH, JL RAYA KAYU KALEK, JL SEBRANG PADANG, JL PIAI ATAS PADNG, JL HILIGOO, JL KAMPUNG BARU, JL BUKIT GADO GADO, JL BATANG ARAU, JL AIR DINGIN, JL AIR PAKU, JL ALAI TIMUR, JL AMPANG KARANG, JL AREA KOTO LUA, JL BADAR BEKALI, JL BADAR PURUS, JL BALAI BARU, JL BANDAR BEKALI, JL BATUNG TABA, JL BELAKANG OLO, JL BHAYANGKARA, JL BY PASS AIR PACAH, JL BY PASS DEPAN, JL BY PASS KM 06, JL CUPAK TANGAH, JL DR WAHIDIN, JL FLAMBOYAN, JL GANTING, JL GURUN LAWEH, JL HANGTUAH, JL KAMPUNG KELAWI TIMUR, JL KOTO LUAR DEKAT, JL KOTO PANJANG, JL KURANJI, JL KHATIB SULAIMAN, JL KOTO TINGGI, JL LAMBUNG BUKIT, JL LINGKAR UNIVESITAS, JL MARANSI, JL M YUNUS DEPAN, JL MUARO, JL NURTANIO, JL PAGANG, JL PAGANG RAYA, JL PERMINDO, JL PERTANIAN, JL RADEN SALEH, JL RASUNA SAID, JL RAYA SITEBA, JL RIMBO KALUANG, JL RIMBO TAROK, JL SAMUDRA, JL SIMP 3 RAMBUTAN, JL SISINGAMANGARAJA, JL S PARMAN, JL TAMAN SISWA, JL TANJUNG BEROK, JL TEKNOLOGI, JL TUT WURI, JL UJUNG TANAH, JL ULU GADUT DEKAT, JL VETERAN, JL ZAINUL ARIFIN, JL SILUNGKANG, dan JL SUTOMO.
Vulnerable	- JL PADANG SOLOK - JL ADINEGORO - JL INDARUNG - JL GAJAH MADA - JL RAYA PADANG BUNGUS - JL PADANG PAINAN - JL ST SYAHRIL - JL DR M HATTA - JL GAJAH MADA - JL JHONI ANWAR

Very Vulnerable	JL BY PASS

The results of clusterization analysis show that the clusters formed each represent different levels of vulnerability to traffic accidents. The first cluster (C1), with the initial center point (2), has the lowest score and consists of 122 roads that are considered not prone to accidents. The second cluster (C2), with an initial center point (3), has a higher value than C1 but lower than C3, and includes 10 roads that are categorized as accident-prone to accidents. The third cluster (C3), with the initial center point (4), has the highest score and includes only 1 road that is very prone to accidents.

IV. Conclusion

This study aims to identify and classify the risk of traffic accidents in Padang City using the K-means clustering method and the Chebyshev Distance formula. Based on accident data from 2021 to 2023, the results of the analysis show that of the total roads analyzed, 122 roads were identified as not prone to accidents. These roads have relatively safe characteristics with a low frequency of accidents.

In addition, as many as 10 roads are categorized as accident-prone roads, indicating a higher level of risk compared to roads that are not prone to accidents, but are still in the medium risk category. Meanwhile, 1 road is classified as very accident-prone with a very high frequency of accidents compared to other roads, indicating a very high level of risk.

REFERENCES

[1] "Badan Pusat Statistik Provinsi Sumatera Barat." Accessed: Jun. 07, 2024. [Online]. Available: https://sumbar.bps.go.id/indicator/34/536/1/data-kecelakaan-lalu-lintas-yang-terjadi-di-provinsi-sumatera-barat.html

[2] B. A. Fauzan, M. Jamaris, and J. Junadhi, "Implementation of K-Means Clustering Algorithm for Grouping Traffic Violation Levels in Siak," *J. Teknol. Dan Open Source*, vol. 5, no. 1, pp. 81–88, 2022, doi: 10.36378/jtos.v5i1.2427.

[3] A. A. Vernanda, A. Faisol, and N. Vendyansyah, "Penerapan Metode K-Means Clustering Untuk Pemetaan Daerah Rawan Kecelakaan Lalu Lintas Di Kota Malang Berbasis Website," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 836–844, 2021, doi: 10.36040/jati.v5i2.3791.

[4] S. Novianti, T. Nurkholifa, M. Suryana, and E. Susanto, "Penggunaan Geographical Information System (GIS) untuk Visualisasi Analisis Perilaku Spasial Wisatawan," *J. Indones. Tour. Hosp. Recreat.*, vol. 4, no. 2, pp. 213–223, 2021, doi: 10.17509/jithor.v4i2.37168.

[5] I. F. Anshori and Y. Nuraini, "Pengelompokan Data Kecelakaan Lalu Lintas di Kota Tasikmalaya Menggunakan Algoritma K-Means," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 1, pp. 118–127, 2020, doi: 10.51977/jti.v2i1.198.

[6] N. Wahidi and R. Ismailova, "Association rule mining algorithm implementation for e-commerce in the retail sector," *J. Appl. Res. Technol. Eng.*, vol. 5, no. 2, pp. 1–6, 2024, doi: 10.4995/jarte.2024.20753.

[7] S. K. Yuliarnis, Y. Hendriyani, D. Kurniadi, and M. Giatman, "Application of Data Mining for Analysis of Consumer Purchase Data on Sales Transaction Data At Halal Mart Hni Hpai Dharmasraya," *J. Pendidik. Teknol. Kejuru.*, vol. 3, no. 1, pp. 68–75, 2020, doi: 10.24036/jptk.v3i1.6923.

[8] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: 10.1016/j.eswa.2020.114060.

[9] M. Sholeh, E. K. Nurnawati, and U. Lestari, "Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner," vol. 8, no. 1, pp. 10–21, 2023.

[10] M. A. Anggara, A. Sonita, M. H. Rifqo, and A. Wijaya, "Classification Of Accident Proposed Road Data Traffic In Bengkulu City Using K-Means Algorithm," *J. Komputer, Inf. dan Teknol.*, vol. 3, no. 2, pp. 423–436, 2023, doi: 10.53697/jkomitek.v3i2.1188.

[11] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1785–1792, 2022, doi: 10.1016/j.jksuci.2019.12.011.

[12] S. Vijayalakshmi and K. K. Nivethithaa, "Survey on Data Mining Techniques, Process and Algorithms," *J. Phys. Conf. Ser.*, vol. 1947, no. 1, 2021, doi: 10.1088/1742-6596/1947/1/012052.

[13] Y. Pan and L. Zhang, "A BIM-data mining integrated digital twin framework for advanced project management," *Autom. Constr.*, vol. 124, no. July 2020, p. 103564, 2021, doi: 10.1016/j.autcon.2021.103564.

[14] R. Rahim, U. H. Salsabila, A. A. Dani, E. M. S. S. Ciptaningsih, and M. Mohzana, "Data Mapping using Combining Clustering Methods and C.45 Classification," *SSRG Int. J. Electron. Commun. Eng.*, vol. 10, no. 5, pp. 95–103, 2023, doi: 10.14445/23488549/IJECE-V10I5P109.

[15] M. Afdal, R. P. Disastra, P. Studi, S. Informasi, and D. Mining, “ANALISIS POLA KECELAKAAN LALU LINTAS MENGGUNAKAN ALGORITMA K -MEANS DAN FP - GROWTH STUDI KASUS : POLRES SOLOK,” vol. 8, no. 1, pp. 31–40, 2022.