JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

K-MEANS CLUSTERING AREAS PRONE TO TRAFFIC ACCIDENTS IN ASAHAN REGENCY

Nurul Rahmadani[1*]; Elly Rahayu[2]; Ayu Lestari[3]

Manajemen Informatika [1]; Sistem Informasi [2,3]
Sekolah Tinggi Manajemen Informatika dan Komputer Royal
www.stmikroyal.ac.id
cloudyrara@gmail.com[1]; ellyrahayu68@gmail.com[2]; ayulesstarii27@gmail.com[3]
(*) Corresponding Author

Abstract— Traffic accidents on the highway still contribute to the high mortality rate in Indonesia, so it is of particular concern to the police in this country. Accidents occur in various places with different time events, this makes it difficult to determine which areas have a high level of traffic accident vulnerability. Information about traffic accident-prone areas is needed by the community and law enforcement. This information can be taken into consideration for supervision and anticipatory action, especially for the police. The initial stage of traffic accident prevention is to know the factors that cause traffic accidents obtained through traffic accident data analysis. The information system in this study analyzed traffic accident-prone areas in Asahan Regency. The analysis can be done with data mining, namely K-Means Clustering which can group data into several groups according to the characteristics of the data. The results of this study are the Asahan District Police Satlantas can find out the accident-prone areas in the most vulnerable categories, quite vulnerable and not vulnerable.

Keywords: Accident Prone Areas, K-Means Clustering, Traffic Accident.

Abstrak— *Kecelakaan lalu lintas di jalan raya masih menjadi penyumbang tingginya angka kematian di Indonesia, sehingga menjadi perhatian khusus bagi kepolisian di negara ini. Kecelakaan terjadi di berbagai tempat dengan waktu kejadian yang berbeda, hal ini menyebabkan sulitnya menentukan daerah mana yang memiliki tingkat kerawanan kecelakaan lalu lintas. Informasi mengenai daerah rawan kecelakaan sangat dibutuhkan oleh masyarakat dan penegak hukum. Informasi tersebut dapat dijadikan bahan pertimbangan untuk pengawasan maupun tindakan antisipasi khususnya bagi kepolisian. Tahapan awal pencegahan kecelakaan lalu lintas adalah dengan mengetahui faktor-faktor penyebab kecelakaan lalu lintas yang diperoleh melalui analisa data kecelakaan. Sistem Informasi pada penelitian ini melakukan analisa terhadap wilayah rawan kecelakaan di wilayah Kabupaten Asahan. Analisa tersebut dapat dilakukan dengan data mining, yaitu K-Means Clustering. K-Means Clustering mengelompokkan data menjadi beberapa cluster sesuai karakteristik data tersebut. Hasil dari penelitian ini ialah Satlantas Polres Kabupaten Asahan dapat mengetahui daerah-daerah rawan kecelakaan dalam katagori paling rawan, cukup rawan dan tidak rawan.*

Kata Kunci*: Daerah Rawan Kecelakaan, K-Means Clustering, Kecelakaan Lalu Lintas.*

INTRODUCTION

The rate of population growth and the amount of traffic flow in the Asahan Regency is increasing rapidly [1], so the need for transportation infrastructure continues to grow. This situation greatly affects the service so that if it is not balanced with an increase in inadequate transportation infrastructure, the resulting impact is the emergence of problems with traffic, such as traffic jams and accidents. The development of transportation will indirectly increase the risk of growing traffic problems. Traffic accidents according to RI Law No. 22 of 2009 Article 1 paragraph 24 are an unexpected and unintentional event on a road involving vehicles with or without other road users resulting in human casualties [2] and/or property losses.

All the developments and growth that occur naturally emerge several transformation problems that exist. One of the problems that are most often highlighted is the issue of traffic safety or can be called safety life. Based on Table 1, accidents involve the number of accident victims, starting with death, heavy injuries, minor injuries.

Table 1. Number of Accident Victims in 2015-2018

No	Year	Total Accident	Death	Heavy Injuries	Minor Injuries
1	2015	390	129	223	414
2	2016	453	148	162	542
3	2017	377	122	85	547
4	2018	395	108	91	500
Total		1615	507	561	2003

In order for the resulting policy to be relevant to the problems encountered to prevent accidents, the policy must be supported by information derived from traffic accident data that has been occurring so far. As in Table 2, which contains data on accident victims in Asahan Regency from 2015 to 2018.

Table 2. Accident Victim Data for 2015-2018

No	Sub-district	Total of Accident Victims			
		2015	2016	2017	2018
1	Aek Kuasan	5	7	6	6
2	Aek Ledong	4	6	5	5
3	Aek Songsongan	5	7	4	6
4	Air Batu	24	30	30	24
5	Air Joman	22	32	29	22
6	Bandar Pasir Mandoge	14	18	12	14
7	Bandar Pulau	32	38	27	32
8	Buntu Pane	6	5	4	6
9	Kisaran Barat Kota	37	47	31	37
10	Kisaran Timur Kota	36	35	36	36
11	Meranti	31	44	33	31
12	Pulau Rakyat	6	6	4	7
13	Pulo Bandring	33	29	29	33
14	Rahuning	3	4	4	4
15	Rawang Panca Arga	18	18	14	18
16	Sei Dadap	13	16	15	13
17	Sei Kepayang	12	13	15	12
18	Sei Kepayang Barat	10	16	10	10
19	Sei Kepayang Timur	11	15	11	11
20	Setia Janji	5	6	5	5
21	Silau Laut	6	7	5	6
22	Simpang Empat	13	13	11	13
23	Tanjung Balai	36	30	29	36
24	Teluk Dalam	3	5	4	3
25	Tinggi Raja	5	6	4	5

Therefore, a grouping of areas is carried out, namely the most vulnerable areas (areas that have the highest number of accidents, high risk, and potential accidents on a road segment), areas that are quite vulnerable (areas that have high enough accident rates) and areas that are not vulnerable. This can be seen from the number of various victims of accidents that are accident-prone areas and makes all parties feel the need to take preventative measures and also to find out the factors that trigger accidents [3].

Researches related to the theme of traffic accidents have been carried out. The first journal from Aljofey & Alwagih [4] is to analyze the time of the frequency of traffic accidents for the location of the highway. The k-means algorithm is applied to find out when and where accidents often occur within 24 hours. The second is the journal from Anshori & Nuraini [5] where the research was conducted in Tasikmalaya with 4 clusters based on time grouping, namely night, daytime, evening' and morning. And the results obtained are the most traffic accidents occur in the morning in Tasikmalaya. The third is a journal from Purwaningsih [6] who analyzed traffic accidents in Jakarta City in 2013 by grouping them into 3 clusters through RapidMiner tools as a medium for calculating K-Means Clustering. The fourth journal from Wicaksono, Kusrini, & Lutfi [7] by analyzing the data on traffic accident vulnerability in Bantul Regional Police using K-Means, found that the most vulnerable time for traffic accidents is at night from 19.30 to 23.59 WIB.

Data mining can be used to identify patterns and predict future behavior [3]. One method of data mining is K-Means Clustering which is a method of grouping data into clusters [8] based on the similarity of each of the existing clusters [9], [10], [11]. The purpose of grouping traffic accidents in the Asahan Regency is to find out the accident-prone areas. An accident-prone area is an area where the accident rate is high with repeated accidents occurring in the same period and relatively active space [12].

The application of K-Means Clustering in traffic accident data in this study will determine the initial centroid, K-Means Clustering processing, and display the resulting clusters. Then an analysis of clusters produced by accident-prone areas will be carried out to help reduce the risk of accidents in the Asahan Regency.

MATERIALS AND METHODS

Data Mining is a method of processing data on a large scale, where the data will be stored in a database, data warehouse, or information storage. Data mining plays an important role in several fields, including economics, industry, science and technology, and weather [13]. Data Mining is the process of finding patterns and relationships hidden in a large amount of data to classify, estimate, forecasting, associate rules, sequential patterns, clustering, regression, description, and visualization [12]. Besides the data processed using data mining techniques will produce new knowledge derived from old data, so the results obtained from the data mining process can be used to determine future decisions [14].

K-Means is a method of grouping that is partial [15], [16]. This method partitioned data into groups (clusters) that have the same characteristics [17]. Clustering is one of the non-clustering methods hierarchy which divides data into groups so that data that has the same characteristics are grouped into the same cluster and data that has different characteristics are grouped into clusters [18], [11]. The purpose of grouping this data is to minimize the objective functions established in the grouping process, which generally try to minimize variations within a group and maximize variation between groups [19].

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

The following are data grouping techniques using the K-Means algorithm [20]:

1) Determine the number of K clusters.
2) Initialization of the center point of the K cluster (centroid) can be done randomly and used as the initial cluster.
3) Allocate each data to the closest centroid with the specified matrix distance. To calculate this, the Euclidean distance theory is formulated as follows:

$$(T_{(x,y)}) = \sqrt{(T_{1x} - T_{1y})^2 + (T_{2x} - T_{2y})^2 + (T_{kx} - T_{ky})^2} \quad \text{.....(1)}$$

Where, $T_{(x,y)}$ is the data of x distance to the center of the clustery; T_{kx} is i-data in the k attribute data; T_{ky} is the center of j in the k attribute.

4) Recalculate the cluster center with the new cluster membership. This is calculated by determining the centroid/cluster center.
5) Set each object as the center of the new cluster, if the cluster center is changed, then return to the third step, otherwise, the grouping is complete.
6) Analyze the results in the grouping process.

One of the characteristics of the K-Means algorithm is that it is very sensitive in determining the initial center point of the cluster because K-Means generates random center points of the initial cluster. When the initial random center point generation approaches the final center cluster solution, K-Means has a high possibility to find the right center point of the cluster. Conversely, if the start of the central point is far from the final solution of the center of the cluster, then this is most likely to cause incorrect clustering results. As a result, K-Means does not guarantee unique clustering results. This is what makes the K-Means method difficult to achieve global optimum, but only a local minimum. Besides, the K-Means algorithm can only be used for data whose attributes are numeric.

RESULTS AND DISCUSSION

The purpose of this study is to classify the area of traffic accidents in the Asahan Regency. This research was made based on data from traffic accidents in Asahan Regency from 2015 to 2018 in table 2.

This stage of the analysis is carried out by the K-Means Clustering method in grouping data against a fact or rule. The following is an example of the data used for the calculation of the K-Means Clustering method. For initial determination the following data are assumed:

Taken 23rd data as the center of the 1st cluster.

C1 = 36 30 29 36

Taken the 15th data as the center of the 2nd cluster.

C2 = 18 18 14 18

Taken the 3rd data as the center of the 3rd cluster.

C3 = 5 7 4 6

After knowing the number of clusters and initial centroids, then measuring between centroids using equation 2 and then the distance matrix will be obtained namely C1, C2, C3.

The example of calculating the 1st data distance ie Aek Kuasan in each cluster is the following equation (1), so tha5t the results are [54,03 22,32 2,00]. The same equation and calculation will be implemented in 24 other data to get the distance of each data in each cluster as in table 3.

Table 3. Cluster Center Distance Calculation Results

Data to	C1	C2	C3	Shortest Distance
1	54,03	22,32	2,00	2,00
2	56,01	24,29	2,00	2,00
3	54,91	23,11	0,00	0,00
4	17.00	21,73	43,47	17,00
5	19,90	21,28	42,37	19,90
6	37,43	6,00	18,17	6,00
7	10,00	31,00	53,81	10,00
8	55,23	23,60	2,24	2,24
9	17,18	43,03	65,68	17,18
10	8,60	37,70	60,57	8,60
11	16,19	37,08	59,25	16,19
12	54,24	22,56	1,73	1,73
13	4,36	28,21	51,21	4,36
14	58,43	26,78	4,12	4,12
15	31,89	0,00	23,11	0,00
16	38,08	7,42	17,75	7,42
17	40,46	9,90	15,56	9,90
18	43,69	12,17	12,57	12,17
19	42,41	10,77	13,19	10,77
20	55,44	23,73	1,73	1,73
21	53,90	22,14	1,41	1,41
22	40,88	9,17	14,07	9,17
23	0,00	31,89	54,91	0,00
24	58,55	26,81	4,12	4,12
25	55,88	24,12	1,41	1,41

The next step is to create a grouping table where the smallest value of 3 groups (clusters) is given a value of 1 (one) while the remainder is given a value of 0 (zero), as in Table 4.

Table 4. Iteration-1 Grouping

Data to	C1	C2	C3
1	0	0	1
2	0	0	1
3	0	0	1
4	1	0	0
5	1	0	0
6	0	1	0
7	1	0	0

Data to	C1	C2	C3
8	0	0	1
9	1	0	0
10	1	0	0
11	1	0	0
12	0	0	1
13	1	0	0
14	0	0	1
15	0	1	0
16	0	1	0
17	0	1	0
18	0	1	0
19	0	1	0
20	0	0	1
21	0	0	1
22	0	1	0
23	1	0	0
24	0	0	1
25	0	0	1

After that, the process of this iteration-2 will be calculated by its centroid which is no longer based on the previous sample but from the following new centroid.

C1= 31,38 35,63 30,50 31,38
C2 = 13,00 15,57 12,57 13,00
C3 = 4,80 5,90 4,50 5,30

Repeat the calculation process as in the previous example, until the last grouping of data is equal to the value of the previous data set from clustering results. The results of the calculation after the iteration has stopped (2nd iteration), can be seen in Table 5.

Table 5. Results Distance Between Data and Centroid

Data to	C1	C2	C3	Shortest Distance
1	52,53	15,15	2,00	2,00
2	54,52	17,14	0,99	0,99
3	53,49	16,12	1,41	1,41
4	11,86	27,46	44,15	11,86
5	13,83	26,49	43,08	13,83
6	35,45	2,87	19,05	2,87
7	4,32	37,86	54,68	4,32
8	54,11	16,83	1,73	1,73
9	13,89	49,79	66,58	13,89
10	8,57	44,55	61,28	8,57
11	8,76	43,28	60,09	8,76
12	53,08	15,81	2,14	2,14
13	7,17	35,36	51,93	7,17
14	57,07	19,71	2,96	2,96
15	30,67	7,61	23,92	7,61
16	36,06	2,47	18,41	2,47
17	38,77	3,81	16,04	3,81
18	41,46	4,98	13,47	4,98
19	40,45	3,29	14,00	3,29
20	54,03	16,64	0,62	0,62
21	52,51	15,13	1,84	1,84
22	39,59	3,01	14,81	3,01
23	8,76	39,19	55,65	8,76
24	57,01	19,63	3,10	3,10
25	54,51	17,12	0,62	0,62

The grouping results obtained from the last iteration are iteration-2, where the results are the same as grouping iteration-1 (see Table 4).

To be easier to use, this research was implemented using the Visual Studio 2010 programming language as a tool for the K-Means Clustering system in traffic accident decision making.

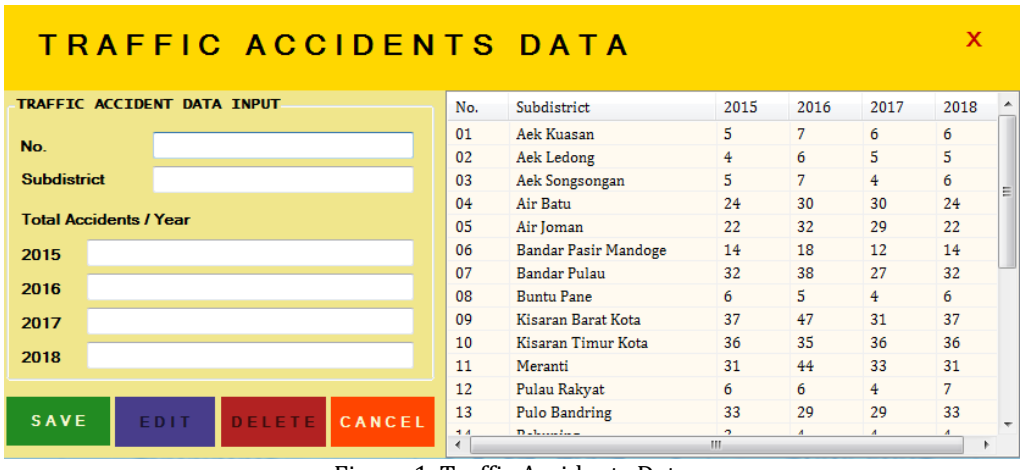

Figure 1. Traffic Accidents Data

In Figure 1, there is a menu for traffic accident data that contains the initial data to be clustered. Users can also save, modify, delete, and cancel traffic accident data in the menu.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

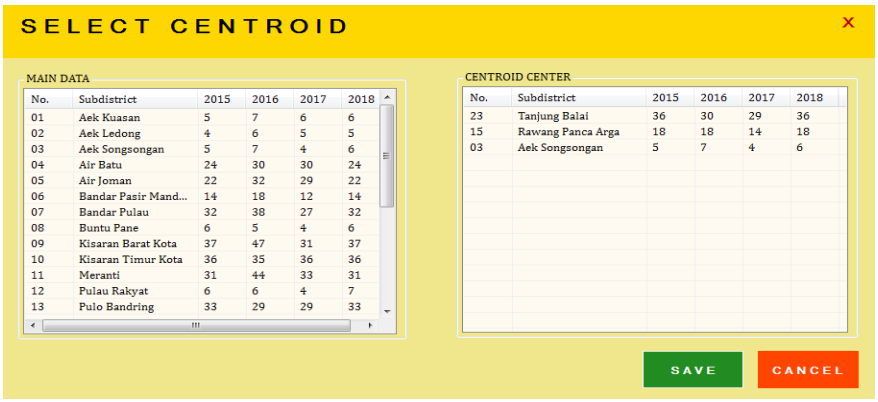

Figure 2. Centroid Data

In Figure 2, the user is intended to select the starting center point of the cluster as many as three central points only and after that, the data will be stored to continue the clustering process. The center point that the user chooses is data 24, 15, and 03.

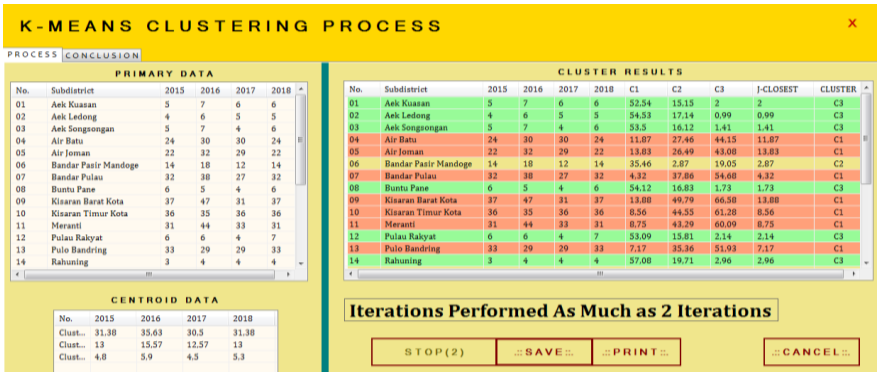

Figure 3. K-Means Clustering Process

In Figure 3, there are 3 tables: initial data containing traffic accident data, initial centroid data tables, and grouped data result in tables. From Figure 3, it can be seen that: "Iterations Performed As Much as 2 Iterations". Besides, users can also see conclusions from the calculation results of traffic accidents by pressing the conclusion button. The following is the display of clustered data (clusters):

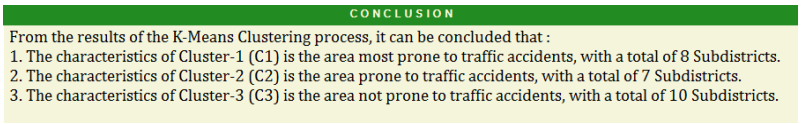

Figure 4. Conclusion of Clustering Results

Based on the results of the calculation and implementation of the group of traffic accident-prone areas in Asahan Regency, the same results can be found:

1. In Cluster-1 (C1), which is the area most prone to traffic accidents, there are 8 Sub-districts: Air Batu, Air Joman, Bandar Pulau, Kisaran Barat, Kisaran Timur, Meranti, Pulo Bandring, and Tanjung Balai.
2. In Cluster-2 (C2), which is an area prone to traffic accidents, there are 7 Sub-districts: Bandar Pasir Mandoge, Rawang Panca Arga, Sei Dadap, Sei Kepayang, Sei Kepayang Barat, Sei Kepayang Timur, and Simpang Empat.
3. In Cluster-3 (C3), which is the area not prone to traffic accidents, there are 10 Sub-districts: Aek Kuasan, Aek Ledong, Aek Songsongan, Buntu Pane, Pulau Rakyat, Rahuning, Setia Janji, Silau Laut, Teluk Dalam, and Tinggi Raja.

CONCLUSION

From the results and discussions that have been carried out, it can be concluded that the K-Means Clustering method can classify traffic accident-prone areas in Asahan Regency with 3 clusters, namely the most accident-prone areas with a total of 8 Subdistricts, the area is quite prone to traffic

accidents with results totaling 7 Subdistricts, and areas not prone to traffic accidents with results totaling 10 Subdistricts.

REFERENCE

[1] F. A. Arisandi, M. Lubis, and M. H. M. Hasibuan, "Penerapan Managemen Lalu Lintas Pada Kabupaten Asahan," vol. 15, no. 2, 2020.

[2] S. Hussain, L. J. Muhammad, F. S. Ishaq, A. Yakubu, and I. A. Mohammed, "Performance evaluation of various data mining algorithms on road traffic accident dataset," in *Smart Innovation, Systems and Technologies*, 2019, vol. 106, pp. 67–78.

[3] A. Almjewail, A. Almjewail, S. Alsenaydi, H. ALSudairy, and I. Al-Turaiki, "Analysis of traffic accident in Riyadh using clustering algorithms," in *Advances in Intelligent Systems and Computing*, 2018, vol. 753, pp. 12–25.

[4] A. M. Aljofey and K. Alwagih, "Analysis of Accident Times for Highway Locations Using K-Means Clustering and Decision Rules Extracted from Decision Trees," 2018.

[5] I. Fardian Anshori and Y. Nuraini, "Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means," *J. Responsif*, vol. 2, no. 1, pp. 118–127, 2020.

[6] E. Purwaningsih, "Analisis Kecelakaan Berlalu Lintas Di Kota Jakarta Dengan Menggunakan Metode K-Means," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 5, no. 1, pp. 1–6, 2019.

[7] E. A. Wicaksono, K. Kusrini, and E. T. Lutfi, "Analisis Data Kerawanan Kecelakaan Lalu Lintas Menggunakan Metode K-Means (Studi Kasus Polres Bantul)," in *Seminar Nasional Teknologi Informasi dan Multimedia*, 2017, pp. 109–113.

[8] S. Nawrin, M. Rahatur, and S. Akhter, "Exploring K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017.

[9] A. A. Mousa, M. A. El-Shorbagy, and M. A. Farag, "K-means-clustering based evolutionary algorithm for multi-objective resource allocation problems," *Appl. Math. Inf. Sci.*, vol. 11, no. 6, pp. 1681–1692, Nov. 2017.

[10] U. R. Raval and C. Jani, "Implementing & Improvisation of K-means Clustering Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 55, no. 5, pp. 191–203, 2016.

[11] V. Soundarya, "Recommendation System for Criminal Behavioral Analysis on Social Network using Genetic Weighted K-Means Clustering," *J. Comput.*, pp. 212–220, 2017.

[12] R. R. Fiska, "SATIN-Sains dan Teknologi Informasi Penerapan Teknik Data Mining Dengan Metode Support Vector Machine (SVM) Untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)," 2017.

[13] R. Wulan Sari, A. Wanto, and A. Perdana Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2018, vol. 2, no. 1, pp. 224–230.

[14] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.

[15] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, Oct. 2017.

[16] D. Xia, F. Ning, and W. He, "Research on Parallel Adaptive Canopy-K-Means Clustering Algorithm for Big Data Mining Based on Cloud Platform," *J. Grid Comput.*, Jun. 2020.

[17] N. Ramadhani, A. F. Rahman, and D. Riskiyati, "Aplikasi Cluster Data Perkara Lalu Lintas Mingguan Di Pengadilan Negeri Pamekasan," *J. LINK*, vol. 26, no. 2, pp. 18–24, 2017.

[18] W. Astuti, D. Djoko, and A. Widodo, "Pemetaan Tindak Kejahatan Jalanan di Kota Semarang Menggunakan Algoritma K-Means Clustering," *J. Tek. Elektro*, vol. 8, no. 1, pp. 5–7, 2016.

[19] H. Priyatman, F. Sajid, and D. Haldivany, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.

[20] Purnawansyah, Haviluddin, A. F. O. Gafar, and I. Tahyudin, "Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities," *Proc. Elev. Int. Conf. Manag. Sci. Eng. Manag.*, pp. 300–310, 2018.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri