

## **ANALISIS KECELAKAAN BERLALU LINTAS DI KOTA JAKARTA DENGAN MENGGUNAKAN METODE K-MEANS**

**Esty Purwaningsih**

Program Studi Sistem Informasi  
Universitas Bina Sarana Informatika  
www.bsi.ac.id  
esty.epw@bsi.ac.id

**Abstract**— *In the Jakarta city area, accidents in traffic enter problems that almost always occur at any time. Not a few victims suffered property damage, but victims also suffered minor injuries, serious injuries and even lost their lives. The purpose of this study is to determine the level of traffic accident events especially in the Jakarta city area and determine the performance of the K-Means method in clustering the level of traffic accident events. From the clustering of traffic accidents data obtained accuracy level in cluster 0, cluster 1 and cluster 2 get the same accuracy value that is 33.33% where in cluster 0 there are 2 data, cluster 1 there are 2 data and cluster 2 there are 2 data.*

**Keyword:** *Traffic Accident, K-Means, Rapid Miner*

**Intisari**— Di wilayah kota Jakarta, kecelakaan dalam berlalu lintas masuk kedalam masalah yang hampir sering terjadi setiap saat. Tidak sedikit korban mengalami kerugian harta benda, tetapi korban juga mengalami luka ringan, luka berat bahkan sampai kehilangan nyawa. Tujuan dari penelitian ini adalah untuk mengetahui tingkat kejadian kecelakaan lalu lintas khususnya di wilayah kota Jakarta dan mengetahui kinerja dari metode K-Means dalam mengkluster tingkat kejadian kecelakaan berlalu lintas. Dari pengklusteran data kecelakaan berlalu lintas didapat tingkat nilai akurasi pada *cluster 0*, *cluster 1* dan *cluster 2* mendapatkan nilai akurasi yang sama yaitu sebesar 33.33% dimana pada *cluster 0* terdapat 2 data, *cluster 1* terdapat 2 data dan *cluster 2* terdapat 2 data.

**Kata Kunci:** Kecelakaan Lalu Lintas, K-Means, Rapid Miner.

### **PENDAHULUAN**

Beberapa kali kita mendapatkan informasi atau bahkan melihat secara langsung adanya kejadian kecelakaan lalu lintas bagian wilayah yang ramai kendaraan. Kecelakaan lalu lintas adalah indikator utama dalam tingkat keselamatan di jalan raya. Keselamatan jalan sangat diperhatikan oleh

negara-negara maju dalam mengurangi jumlah kecelakaan lalu lintas dan jumlah korban dari kecelakaan lalu lintas yang terjadi (P. G. G. Saragih, 2013).

Berdasarkan Undang-Undang Republik Indonesia No. 22 tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, dinyatakan bahwa kecelakaan lalu lintas merupakan peristiwa yang terjadi di jalan raya baik yang tidak diduga maupun tidak disengaja yang dapat mengakibatkan korban juga kerugian harta benda.

Menurut Hobs dalam (Azzirrahman, Normelani, & Arisanty, 2015), Faktor penyebab kecelakaan lalu lintas dapat dikelompokkan menjadi 3 kelompok, diantaranya faktor manusia seperti menggunakan kendaraan berkecepatan tinggi, ugal-ugalan, membawa muatan yang berlebihan dan tidak tertib pada aturan. Faktor yang kedua adalah kendaraan, seperti kaca spion, lampu kendaraan dan faktor yang ketiga faktor lingkungan fisik jalan seperti jalan rusak, berlubang, gelap, tidak ada rambu lalu lintas serta tikungan tajam.

Penelitian dilakukan (Iswari & Ayu, 2015) dengan mengambil dataset kecelakaan lalu lintas dan penyajian informasi melalui sejumlah ruas jalan utama di Kabupaten Sleman Daerah Istimewa Yogyakarta (DIY). Pada penelitiannya menggunakan metode klasterisasi yang populer yaitu metode K-Means dengan tujuan dapat membantu memonitoring terhadap keselamatan pengendara di jalan raya. Hasil dari penelitiannya berupa sistem klasterisasi dengan mengelompokkan objek jalan dengan dasar jumlah korban, kendaraan yang terlibat, dan jumlah kecelakaan yang terjadi pada rentang waktu tertentu. Visualisasi dari hasil pengelompokkan dengan metode K-Means berupa peta yang efektif dalam menunjukkan jalan yang memiliki tingkat kerawanan yang sama.

Analisis data kecelakaan lalu lintas dapat membantu mengidentifikasi penyebab terjadinya kecelakaan yang membuat kecelakaan lalu lintas lebih sering terjadi. Pada penelitian ini diambil dari basis data Resor Kendari yang tercatat sebanyak 2620 data kecelakaan, namun data diseleksi

menjadi 500 data. Metode yang dipakai menggunakan algoritma K-Means Clustering yang diolah dengan melalui aplikasi Rapid Miner. Hasil penelitian menunjukkan frekuensi tingkat kecelakaan di tiap lokasi dan waktu rawan yang dianggap berpotensi terjadi kecelakaan (Rahmat C.T. I et al., 2017).

Penelitian mengenai kecelakaan lalu lintas pernah dilakukan (Aprianti & Permadi, 2018), Data diambil dari Kepolisian Resor (Polres) Tanah Laut. Peneliti melakukan analisa kasus kecelakaan lalu lintas dengan pengelompokan (klusterisasi) melalui metode K-Means. Tujuan dari pengklusteran dengan metode K-means yaitu untuk mencari hasil kluster dengan kualitas terbaik. Dan menghasilkan informasi terkait dengan kecelakaan lalu lintas yang sering terjadi di area Polres Tanah Laut.

Penelitian yang dilakukan (R. Saragih & Sitompul, 2019) dengan menganalisis perbandingan dari metode apriori dan metode k-means clustering terhadap faktor-faktor yang berkaitan dengan kejadian kecelakaan berlalu lintas di wilayah Polresta Medan. Tujuan dari penelitian yang dilakukan (R. Saragih & Sitompul, 2019) yaitu untuk menghasilkan pola asosiasi dan klusterisasi yang terdapat pada data kecelakaan yang ada di wilayah tersebut. Hasil dari penelitiannya berupa aplikasi identifikasi keterkaitan antara variabel penyebab kecelakaan lalu lintas dengan variabel pelaku kecelakaan lalu lintas serta memberikan informasi mengenai variabel yang berpengaruh terhadap angka kecelakaan lalu lintas.

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstrasi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data (Vulandari, 2017).

Menurut Aggarwal & Reddy dalam (Ayu, Wati, Puspitasari, & Purwaningsih, 2019), Pengelompokan atau *clustering* berguna pada pembelajaran mesin dan tugas penambangan data. Bisa dilihat sebagai tugas eksplorasi atau langkah sebelum pemrosesan.

Algoritma K-Means membagi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain (Eko, 2016).

Menurut (Vulandari, 2017), Algoritma k-means menetapkan nilai-nilai cluster (K) secara acak, hasil sementara dari nilai tersebut menjadi pusat cluster atau bisa disebut sebagai centroid,

mean, atau "means". Selanjutnya menghitung jarak setiap data dengan masing-masing centroid menggunakan rumus Euclidian sampai ditemukan jarak terdekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut sampai nilai centroid tidak berubah.

Peneliti menganalisis kejadian kecelakaan lalu lintas khususnya di wilayah Jakarta dengan menggunakan metode K-Means. Dimana pada metode K-Means merupakan algoritma clustering yang cukup sederhana untuk mempartisi dataset kedalam beberapa cluster k dan algoritmanya cukup mudah untuk diimplementasikan, dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan.

Tujuan dari penelitian pada studi kasus ini adalah untuk mengetahui tingkat kejadian kecelakaan lalu lintas khususnya di wilayah Jakarta dan mengetahui kinerja dari metode K-Means dalam mengkluster tingkat kejadian kecelakaan lalu lintas di sekitar wilayah Jakarta.

## BAHAN DAN METODE

Peneliti menggunakan data sekunder dari sumber Badan Pusat Statistik Provinsi DKI Jakarta yang diambil melalui data.jakarta.go.id. Data yang diambil adalah dataset jumlah kejadian kecelakaan lalu lintas, korban dan kerugiannya tahun 2013. Pada dataset tersebut terdiri dari subsatuan kerja diantaranya subdit bin gakkum, satlantas wilayah Jakarta pusat, satlantas wilayah Jakarta utara, satlantas wilayah Jakarta barat, satlantas wilayah Jakarta selatan, satlantas wilayah Jakarta timur, satlantas wilayah Kepulauan Seribu. Dataset memiliki 6 (enam) variabel diantaranya jumlah kejadian, korban mati, korban luka berat, korban luka ringan, kerugian benda dan kerugian uang.

Namun pada penelitian ini, peneliti hanya mengambil sampel dengan wilayah subdit bin gakkum, satlantas wilayah Jakarta pusat, satlantas wilayah Jakarta utara, satlantas wilayah Jakarta barat, satlantas wilayah Jakarta selatan, satlantas wilayah Jakarta timur. Sedangkan variabel yang digunakan diantaranya jumlah kejadian, korban mati, korban luka berat, korban luka ringan, dan kerugian benda.

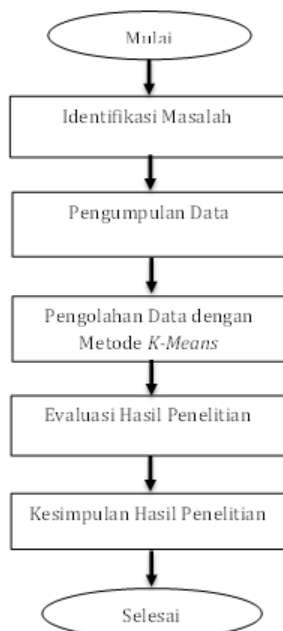
Tabel 1. Sampel Data Kecelakaan Lalu Lintas Wilayah Jakarta

tahun	subsatuan kerja	jumlah kejadian	korban mati	korban luka berat	korban luka ringan	kerugian benda
2013	Subdit Bin Gakkum	590	41	73	455	942
2013	Satlantas Wil Jakarta Pusat	458	22	138	301	692
2013	Satlantas	747	155	32	709	1020

	Wil Jakarta Utara				
2013	Satlantas Wil Jakarta Barat	566	74	427	117 775
2013	Satlantas Wil Jakarta Selatan	511	33	105	696 810
2013	Satlantas Wil Jakarta Timur	838	101	201	679 1201

Sumber: (BPS Provinsi DKI Jakarta, 2013)

Peneliti menggunakan metode clustering dengan algoritma K-Means untuk analisis kejadian kecelakaan lalu lintas di wilayah Jakarta. Agar penelitian menjadi lebih terarah dengan baik. Berikut adalah metodologi penelitian yang dilakukan peneliti:



Sumber: (Purwaningsih, 2019)  
Gambar 1. Tahapan Penelitian

Langkah-langkah yang dilakukan dalam tahapan penelitian dapat dijelaskan sebagai berikut:

1. Identifikasi Masalah  
Identifikasi masalah pada penelitian ini adalah menganalisa kejadian kecelakaan lalu lintas di wilayah Jakarta dengan menggunakan metode K-Means.
2. Pengumpulan Data  
Data yang dikumpulkan bersumber dari Badan Pusat Statistik Provinsi DKI Jakarta pada tahun 2013.
3. Pengolahan Data  
Pengolahan data pada penelitian ini menggunakan metode clustering dengan algoritma K-means pada data kejadian kecelakaan lalu lintas di wilayah Jakarta. Sehingga hasil dari penelitian ini dapat

berguna untuk dijadikan evaluasi dan agar dapat meminimalisir kejadian kecelakaan lalu lintas.

4. Evaluasi Hasil Penelitian  
Pada tahap ini penulis melakukan uji hasil penelitian menggunakan *software RapidMiner* dalam menghubungkan *database* yang akan diuji. Sehingga informasi yang dihasilkan oleh proses *data mining* dapat ditampilkan kedalam bentuk yang mudah dipahami bagi penelitian.
5. Kesimpulan Hasil Penelitian  
Pada tahap ini penelitian yang telah selesai akan diberikan kesimpulan dari permasalahan yang ada agar dapat digunakan oleh pihak yang berwenang.

Menurut Sugiyono dalam (Endra, 2017) *Sampling Jenuh* Suatu teknik penentuan sampel jika semua anggota populasi digunakan sebagai sampel. Hal ini sering kali dilakukan jika jumlah populasi relatif kecil atau sedikit, yaitu kurang dari 30 orang, atau penelitian yang ingin membuat generalisasi, dengan kesalahan yang relatif kecil. Istilah lain sampel jenuh adalah sensus, dimana semua anggota populasi dijadikan sampel. Sampel jenuh dikenal pula dengan total sampel.

Maka dari itu penelitian ini menggunakan sampel yang berasal dari populasi tersebut. Dan pada penelitian ini sampel yang digunakan adalah data kejadian kecelakaan lalu lintas di wilayah Jakarta pada tahun 2013 sebanyak 5 data yang diambil dari Jakarta Open Data melalui [data.jakarta.go.id](http://data.jakarta.go.id).

Dalam pengolahan data, penulis menggunakan metode clustering dengan algoritma K-Means untuk melakukan perhitungan pada kejadian kecelakaan lalu lintas di wilayah Jakarta.

a. Menghitung Jarak Terdekat

Untuk menghitung jarak terdekat, pada penelitian ini digunakan  $L_2$  (Euclidian) distance space, jarak antara dua titik hitung dengan rumus sebagai berikut :

$$D_{L2}(X_2, X_1) = ||X_2 - X_1||^2 = \sqrt{\sum_{j=i}^p (X_{2j} - X_{1j})^2} \dots\dots\dots(1)$$

- b. Pengukuran Nilai Akurasi (Accuracy Rate)  
Nilai akurasi merupakan perbandingan antara jumlah cluster yang sesuai dengan label pada data asli dibandingkan dengan jumlah data keseluruhan.

**Accuracy Rate =**

$$\frac{\text{Jumlah data yang sesuai label aslinya}}{\text{Jumlah data keseluruhan}} \times 100\% \dots\dots\dots(2)$$

**HASIL DAN PEMBAHASAN**

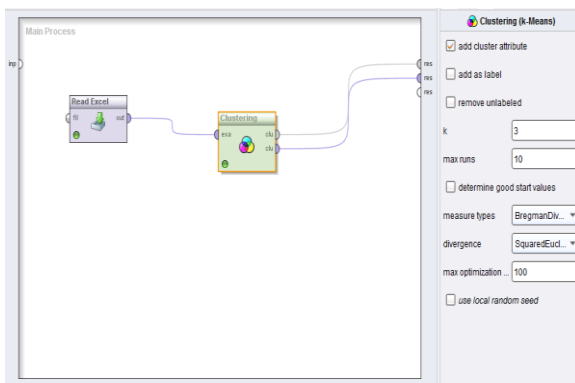
Dalam penelitian ini penulis menggunakan *RapidMiner* sebagai *softwarre* untuk implementasinya. Adapun langkah-langkah dan hasil implementasi menggunakan *RapidMiner* adalah sebagai berikut :

a. Memilih Operator

Operator yang digunakan pada penelitian ini adalah :

- 1) *Read Excel* : berguna sebagai pembaca data, data disini di olah dengan *Microsoft Excel* yang berarti memiliki fungsi format .xlsx.
- 2) *Clustering* : berguna sebagai implementasi algoritma *k-means*.

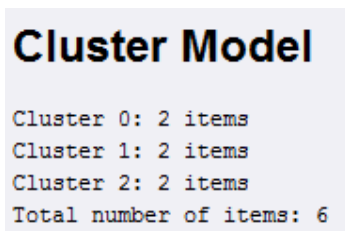
Pada langkah ini, peneliti memasukkan data yang akan diolah berupa excel dengan menggunakan *Rapid Miner* dan memilih operator clustering untuk mengelompokkan data dengan mengisi nilai parameter dan jumlah cluster (k) yang akan digunakan.



Sumber: (Purwaningsih, 2019)  
Gambar 2. Pemilihan Operator

b. Hasil Olahan

Setelah dilakukan pemilihan operator kemudian dijalankan dan menampilkan hasil example set.



Sumber: (Purwaningsih, 2019)  
Gambar 3. Hasil Cluster Model

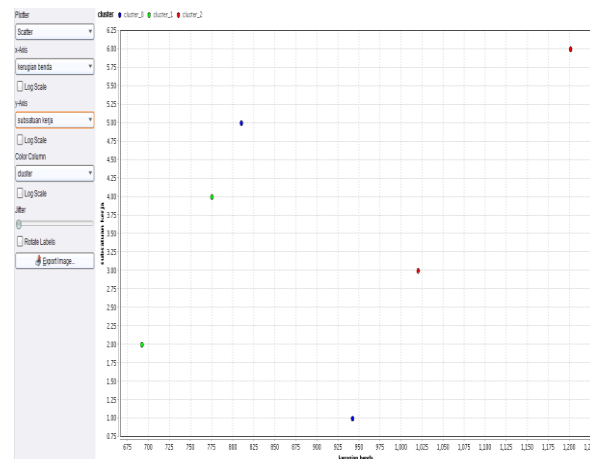
Pada hasil data *ExampleSet (Clustering)* data yang disajikan yaitu berupa penjabaran setiap cluster pada semua data .

id	cluster	subsatuan ...	jumlah keja...	korban mati	korban luka...	korban luka...	kerugian be...
1	cluster_0	1	590	41	73	455	942
2	cluster_1	2	458	22	138	301	692
3	cluster_2	3	747	155	32	709	1020
4	cluster_1	4	566	74	427	117	775
5	cluster_0	5	511	33	105	696	810
6	cluster_2	6	838	101	201	679	1201

Sumber: (Purwaningsih, 2019)  
Gambar 4. Pusat Cluster Pada Data 1-6

Hasil statistic *ExampleSet (Clustering)* menyajikan data yang telah dikluster, dimana pada cluster 0 ada pada wilayah Subdit Bin Gakkum dan wilayah Satlantas Jakarta Selatan, cluster 1 ada pada wilayah Satlantas Jakarta Pusat dan wilayah Satlantas Jakarta Barat serta cluster 2 ada pada wilayah Satlantas Jakarta Utara dan wilayah Satlantas Jakarta Timur.

*Scatter* (Grafik Penyebaran), grafik ini menggambarkan penyebaran yang dilakukan dengan sampel kerugian benda (x) dan subsatuan kerja (y).



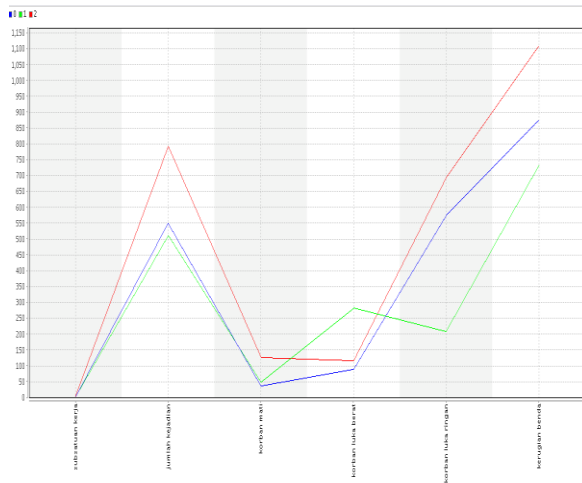
Sumber: (Purwaningsih, 2019)  
Gambar 5. Scatter Chart

*Centroid Table*, tabel yang menampilkan nilai *centroid* dalam masing-masing *cluster*.

Attribute	cluster_0	cluster_1	cluster_2
subsatuan kerja	3	3	4.500
jumlah kejadian	550.500	512	792.500
korban mati	37	48	128
korban luka berat	89	282.500	116.500
korban luka ringan	575.500	209	694
kerugian benda	876	733.500	1110.500

Sumber: (Purwaningsih, 2019)  
Gambar 6. Table Centroid

Plot atau grafik yang menggambarkan titik *centroid* dengan tiap *cluster*. Dimana garis berwarna biru menunjukkan grafik untuk *cluster* 0 dan garis berwarna hijau pada grafik menunjukan *cluster* 1, dan garis berwarna merah pada grafik menunjukkan *cluster* 2.



Sumber: (Purwaningsih, 2019)

Gambar 7. Grafik Hasil Plot

Adapun pengukuran nilai akurasi (*Accuration Rate*) dengan rincian sebagai berikut:

1) Data yang ada pada *cluster* 0, maka nilai akurasinya adalah sebagai berikut :

$$Accuration Rate = \frac{2}{6} \times 100\% = 33.33\%$$

Jadi nilai akurasi pada cluster 0 adalah 20%

2) Data yang ada pada *cluster* 1, maka nilai akurasinya adalah sebagai berikut :

$$Accuration Rate = \frac{2}{6} \times 100\% = 20\%$$

Jadi nilai akurasi pada cluster 1 adalah 33.33%

3) Data yang ada pada *cluster* 2, maka nilai akurasinya adalah sebagai berikut :

$$Accuration Rate = \frac{2}{6} \times 100\% = 33.33\%$$

Jadi nilai akurasi pada cluster 2 adalah 33.33%

### KESIMPULAN

Data yang diambil dari data.jakarta.go.id merupakan data kejadian kecelakaan lalu lintas pada tahun 2013 dan data diolah dengan menggunakan metode *K-Means* melalui tools *RapidMiner*. Dari pengklusteran data kejadian kecelakaan berlalu lintas didapat tingkat nilai akurasi pada masing-masing cluster yaitu *cluster* 0 sebesar 33.33%, *cluster* 1 mendapat nilai akurasi sebesar 33.33% dan *cluster* 2 dengan nilai akurasi sebesar 33.33%. Sehingga dapat disimpulkan bahwa hasil perhitungan di setiap *cluster* memiliki hasil akurasi yang sama dimana *cluster* 0 terdapat

2 data, *cluster* 1 terdapat 2 data dan *cluster* 2 terdapat 2 data.

### REFERENSI

Aprianti, W., & Permadi, J. (2018). K-Means Clustering Untuk Data Kecelakaan Lalu Lintas K-Means Clustering for Highway Traffic Accident Data in Pelaihari Sub District, 5(5), 613–620.

<https://doi.org/10.25126/jtiik2018551113>

Ayu, D., Wati, M., Puspitasari, D., & Purwaningsih, E. (2019). Metode Clustering Pada Model Algoritma K-Means Untuk Pemilihan Alat Kontrasepsi, 3(2), 129–138.

Azzirrahman, M., Normelani, E., & Arisanty, D. (2015). Faktor Penyebab Terjadinya Kecelakaan Lalu Lintas pada Daerah Rawan Kecelakaan di Kecamatan Banjarmasin Tengah Kota Banjarmasin. *Jurnal Pendidikan Geografi*, 2(3), 20–37.

BPS Provinsi DKI Jakarta. (2013). Data Jumlah Kejadian Kecelakaan Lalu Lintas, Korban dan Kerugian di Provinsi DKI Jakarta. Retrieved from <http://data.jakarta.go.id/dataset/data-jumlah-kejadian-kecelakaan-lalu-lintas-korban-dan-kerugiannya>

Eko, W. A. (2016). Implementasi data mining dalam pengelompokan data peserta didik di sekolah untuk memprediksi calon penerima beasiswa dengan menggunakan algoritma k-means (studi kasus sman 16 bekasi), 21(3).

Endra, F. (2017). *Pengantar Metodologi Penelitian (Statiska Praktis)*. zifatama jawara.

Iswari, L., & Ayu, E. G. (2015). Pemanfaatan Algoritma K-Means Untuk Pemetaan Hasil Klasterisasi Data Kecelakaan Lalu Lintas. *Pemanfaatan Algoritma K-Means Untuk Pemetaan Hasil Klasterisasi Data Kecelakaan Lalu Lintas*, 21(1), 1–13. <https://doi.org/10.20885/teknoin.vol21.iss1.art7>

Purwaningsih, E. (2019). *Laporan Akhir Penelitian PDY: Analisis Kecelakaan Berlalu Lintas Di Kota Jakarta Dengan Menggunakan Metode K-Means*. Jakarta.

Rahmat C.T. I, B., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., P, Y. P., & Ransi, N. (2017). IMPLEMETASI K-MEANS CLUSTERING PADA RAPIDMINER UNTUK ANALISIS DAERAH RAWAN KECELAKAAN. In *Prosiding Seminar*

*Nasional Riset Kuantitatif Terapan 2017* (pp. 58–62). Kendari: Lembaga Pengembangan Sistem Informasi.

Saragih, P. G. G. (2013). ANALISA KECELAKAAN LALU LINTAS DI KOTA PEMATANG SIANTAR. *JURNAL TEKNIK SIPIL USU*, 2(3). Retrieved from <https://jurnal.usu.ac.id/index.php/jts/article/view/5676>

Saragih, R., & Sitompul, J. N. (2019). Perbandingan

Data Mining Mengidentifikasi Pola Keterkaitan Variabel Kecelakaan Lalu Lintas Di Polresta Kota Medan. *Journal Information System Development (ISD)*, 4(1), 39–45. Retrieved from <https://ejournal.medan.uph.edu/index.php/isd/article/view/281>

Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gaava Media.