

Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode *K-Means*

Titus Kurniawan¹, Mohamad Jajuli²

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Singaperbangsa Karawang

E-mail: ¹titus.kurniawan17207@student.unsika.ac.id, ²mohamad.jajuli@unsika.ac.id

Abstrak – Data merupakan suatu kumpulan fakta yang masih dapat diolah menjadi sebuah informasi yang menarik. Berbagai pengolahan data dilakukan secara komputerisasi, mulai dari penyimpanan data, arsip, membuat laporan, serta menghasilkan informasi, baik yang dibutuhkan secara perorangan maupun perusahaan. Dibutuhkan sebuah metode atau Teknik yang dapat merubah tumpukan data tersebut menjadi sebuah informasi yang berharga atau pengetahuan (*knowledge*) yang bermanfaat untuk mendukung pengambilan keputusan. Manfaat dari penelitian ini adalah untuk melihat bagaimana pola suatu kecelakaan lalu lintas di kecamatan cileungsi itu terjadi. Yang bertujuan untuk mengetahui faktor penyebab, waktu, dan umur korban yang paling sering mengalami kecelakaan. Sehingga pemerintah dapat mengambil suatu keputusan dari hasil penelitian ini. Suatu teknologi yang dapat digunakan untuk mewujudkannya adalah *data mining*. *Data mining* merupakan suatu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan Teknik atau metode tertentu. *Data mining* mampu menganalisa suatu data yang besar atau banyak. Salah satu Teknik atau metode yang dapat digunakan adalah *K-Means Clustering*. Algoritma *K-Means Clustering* merupakan suatu tata cara penganalisaan informasi ataupun tata cara data Mining yang melaksanakan proses pemodelan tanpa *supervise* (*Unsupervised*). Pada penelitian ini akan menghasilkan 3 *cluster*, untuk mengkategorikan lokasi tidak rawan, rawan, dan sangat rawan. *Cluster 1* menghasilkan nilai evaluasi 0.35, pada *cluster 2* menghasilkan nilai evaluasi 0.22, dan pada *cluster 3* menghasilkan nilai evaluasi 0.38. Pada penelitian ini, menggunakan evaluasi *silhouette coefficient*.

Kata Kunci — Data Mining, K-Means Clustering, Silhouette Coefficient

Abstract – Data is a collection of facts that can still be processed into interesting information. Various data processing is carried out in a computerized manner, starting from data storage, archives, making reports, and generating information, both for individuals and companies. It takes a method or technique that can turn the pile of data into valuable information or knowledge that is useful to support decision making. The benefit of this research is to see how the pattern of a traffic accident in Cileungsi sub-district occurs. Which aims to determine the causes, time, and age of the victims who most often experience accidents. So that the government can make a decision from the results of this study. A technology that can be used to make it happen is data mining. Data mining is a process of looking for interesting patterns or information in selected data using certain techniques or methods. Data mining is able to analyze a large or a lot of data. One technique or method that can be used is *K-Means Clustering*. The *K-Means Clustering Algorithm* is a procedure for analyzing information or data mining procedures that carry out the modeling process without supervision (*Unsupervised*). This research will produce 3 clusters, to categorize locations that are not vulnerable, vulnerable, and very vulnerable. Cluster 1 resulted in an evaluation value of 0.35, in cluster 2 an evaluation value of 0.22, and in cluster 3 an evaluation value of 0.38. In this study, using a silhouette coefficient evaluation.

Keywords — Data Mining, K-Means Clustering, Silhouette Coefficient

1. PENDAHULUAN

Kecelakaan merupakan kejadian yang tidak dapat di duga dan dapat terjadi dimana saja. Salah satunya adalah kecelakaan lalu lintas di jalan raya. Kejadian tersebut menjadi penyumbang kematian yang terbesar di dunia. Pengendara motor dan mobil di seluruh dunia yang mengalami kecelakaan dan meninggal dunia mencapai 1,35 juta orang setiap tahunnya, menurut World Health Organization (WHO). Pada usia remaja dan produktif, yaitu rentang usia 5-29 tahun, menjadi korban yang paling banyak meninggal dunia akibat kecelakaan lalu lintas. Tercatat laki-laki cenderung lebih banyak terlibat kecelakaan lalu lintas dibandingkan perempuan (Marsaid, M. Hidayat, Ahsan, 2013).

Kecelakaan lalu lintas ini dapat mengakibatkan sebanyak 80 orang per hari atau 3 orang per jam yang meninggal dunia di jalan raya akibat kecelakaan lalu lintas pada tahun 2020, menurut Kepolisian Republik Indonesia (POLRI). Dan jika dibandingkan dari tahun sebelumnya pada minggu ke-33 dan minggu ke-32, mengalami kenaikan sebanyak 260 peristiwa atau sekitar 26,7 %. Berdasarkan informasi yang di dapatkan dari Korlantas Polri, jumlah angka kecelakaan lalu lintas pada minggu ke-32 tahun 2020 mengalami jumlah kecelakaan sebanyak 1,234 kejadian. Tentunya dari jumlah kejadian yang terjadi menjadi informasi, bahwa jalan raya di Indonesia masih rawan terjadinya kecelakaan. Pengambilan judul penelitian ini dilakukan untuk mengetahui suatu pola kecelakaan lalu lintas yang sering kali terjadi dan sering memaka korban jiwa terkhusus di jalan raya kecamatan cileungsi ini.

Kecamatan cileungsi merupakan salah satu kecamatan yang berada di daerah kota bogor. Dari data sensus penduduk, banyak penduduk di kecamatan cileungsi terus bertambah. Seiring jumlah penduduk di Kecamatan Cileungsi setiap tahunnya semakin bertambah, tentunya akan mempengaruhi tingkat kebutuhan transportasi dan akan menimbulkan suatu resiko tumbuhnya permasalahan pada bidang transportasi itu meningkat.



Gambar 1. Grafik Kecelakaan Lalu Lintas Kecamatan Cileungsi Tahun 2019
(Sumber : Unit Laka Lalu Lintas Polres Bogor)

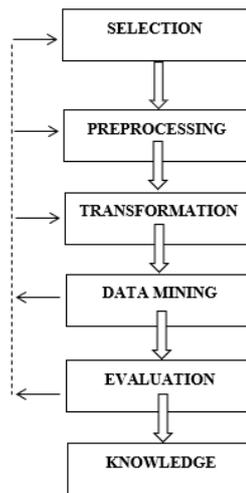


Gambar 2. Grafik Kecelakaan Lalu Lintas Kecamatan Cileungsi Tahun 2020
(Sumber : Unit Laka Lalu Lintas Polres Bogor)

Sebanyak 35 kejadian kecelakaan lalu lintas di Kecamatan Cileungsi pada tahun 2019 dan sebanyak 20 kejadian kecelakaan lalu lintas di Kecamatan Cileungsi pada tahun 2020. Kecamatan cileungsi merupakan kecamatan yang memiliki penduduk cukup banyak. Namun, dengan seiring bertambahnya jumlah penduduk, tidak diimbangi dengan fasilitas atau kondisi jalan yang baik. Kondisi jalanan di kecamatan cileungsi dapat dikatakan cukup buruk, banyak sekali jalan yang berlubang, marka jalan yang sudah pudar warnanya, pembatas jalan yang tidak ada, zebra cross untuk orang menyebrang belum terrealisasi secara menyeluruh. Selain dari faktor jalanan, penduduk di kecamatan cileungsi itu sendiri menurut kepolisian setempat terkhusus bagian unit laka, masih banyak pengendara yang tidak mematuhi peraturan lalu lintas yang ada, sehingga itu dapat menimbulkan terjadinya kecelakaan. Menurut kepolisian setempat penyebab terjadinya kecelakaan di kecamatan cileungsi adalah tingkat kesadaran pengendara yang kurang akan keselamatan dirinya sendiri dan faktor jalanan yang ada. Namun, sejauh ini penentuan penyebab kecelakaan hanya berdasarkan lokasi rawan lalu lintas, sehingga penyebab kecelakaan lalu lintas jalan raya belum diketahui secara pasti. Berdasarkan informasi tersebut, maka diperlukan suatu analisis lebih lanjut mengenai permasalahan kecelakaan lalu lintas untuk mengetahui apa saja faktor penyebab terjadinya kecelakaan. Analisa terhadap data kecelakaan bisa dilakukan memakai data mining, yakni k-means Clustering yang bisa mengklasifikasikan data menjadi beberapa cluster sesuai dengan ciri data.

2. METODE PENELITIAN

Pada penelitian ini menggunakan Metodologi KDD (*Knowledge Discovery in Database*). Terdapat 6 tahapan yang ada pada metodologi tersebut, yaitu : *Selection Data, Preprocessing Data, Transformation Data, Data Mining, Evaluation Data, dan Knowledge*.



Gambar 3. Flowchart Metodologi KDD

2.1. Pengertian Data Mining

Data mining adalah suatu proses yang mencari pola atau informasi menarik dalam data terpilih dengan menggunakan Teknik atau metode tertentu yang nantinya akan menghasilkan suatu pengetahuan (*Knowledge*). Teknik dalam Data Mining yaitu bagaimana menelusuri data yang ada untuk membangun sebuah model. Model tersebut digunakan untuk mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan.

2.2. Tahapan Data Mining

Metodologi yang digunakan pada penelitian ini adalah KDD (*Knowledge Discovery in Database*). Tahapan pada metodologi KDD ada 6, yaitu *Selection Data, Preparation Data, Transformation Data, Data Mining, Evaluation Data, dan Knowledge*. Untuk penjelasan masing-masing tahapan dapat dilihat dibawah ini,

2.2.1. Selection Data

Pada tahap pertama ini dilakukannya terlebih dahulu pemilihan data atau pemilihan attribute apa saja yang dibutuhkan, karena tidak semua attribute yang ada dalam database tersebut dibutuhkan semuanya yang sesuai dengan tujuan dari penelitian ini.

2.2.2. Preprocessing Data

Teknik yang digunakan pada data preprocessing adalah data cleaning. Data cleaning merupakan suatu Langkah yang harus dilakukan untuk menghilangkan *data noise* ataupun *missing values*. Tujuan dari dilakukannya tahap data cleaning adalah menghindari dari suatu data yang masih mempunyai nilainya yang tidak sesuai atau nilai yang error dalam setiap attribute yang ada.

2.2.3. Transformation Data

Pada tahap *data transformation* dilakukannya proses normalisasi. Tujuan dari dilakukannya proses ini adalah untuk mengubah bentuk nilai setiap attribute sesuai dengan algoritma yang nantinya akan digunakan. Karena, tidak semua tipe data dapat dilakukan menggunakan metode yang diinginkan. Maka dari itu tipe data ini harus disesuaikan dengan kebutuhan yang ada dengan cara mentransformasikan data sesuai dengan tujuan kita.

2.2.4. Data Mining

Pada tahap data mining, metode atau algoritma yang digunakan adalah *K-Means Clustering*. Metode *K-Means* adalah salah satu metode dalam fungsi *Clustering* atau pengelompokan. Menurut (Larose, 2005) *clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar kemiripan objek yang diteliti. Sebuah cluster adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain

2.2.5. Evaluation Data

Pada tahap evaluation ini, dilakukan analisis terhadap hasil dari proses pembelajaran data. Tahap ini adalah proses interpretasi hasil pemodelan data mining yang digunakan. Evaluasi ini dilakukan dengan menggunakan metode *Silhouette Coefficient*, yang dimana metode ini bertujuan untuk menguji suatu kualitas dari klaster yang dihasilkan. Evaluasi ini juga dilakukan untuk mengetahui pemodelan yang diterapkan apakah sudah sesuai dan cocok diterapkan pada kasus penelitian ini serta sesuai dengan tujuan yang ingin dicapai dari penelitian ini. Kemudian, setelah melihat hasil dari evaluasi tersebut tentukan langkah-langkah berikutnya apakah bisa dilanjutkan ke tahap berikutnya atau diulang dari awal karena tidak sesuai dengan tujuan

2.2.6. Knowledge

Setelah suatu hasil Analisa sudah didapatkan, maka selanjutnya hasil dari penelitian kita dapat menjadi pengetahuan atau *knowledge* bagi banyak orang, terkhusus untuk pemerintahan setempat yang berada di Kecamatan Cileungsi itu sendiri. Setelah diketeahui, pemerintah setempat dapat langsung bertindak dengan tujuan untuk mengurangi angka kecelakaan yang sering terjadi di jalan raya kecamatan Cileungsi itu sendiri.

3. HASIL DAN PEMBAHASAN

Dataset kecelakaan lalu lintas yang ada pada kecamatan Cileungsi ada sebanyak 55 kejadian. Dari dataset yang diperoleh memiliki 13 *attribute*, namun tidak semua *attribute* yang digunakan, hanya ada 8 *attribute* saja yang digunakan pada penelitian ini. *Attribute* yang digunakan adalah jenis kelamin, bulan, usia, penyebab, jenis kendaraan, hari terjadinya kecelakaan, tahun terjadinya, dan waktu. Setelah menentukan *attribute* apa saja yang digunakan, maka selanjutnya masuk ke dalam

tahap preprocessing data. dalam dataset yang sudah di gunakan terdapat data yang mengalami *missing values* atau datanya hilang. Data yang hilang tersebut dapat diatasi dengan cara menghitung nilai rata-rata pada masing-masing *attribute* yang ada. *Attribute* yang mengalami *missing values* adalah terdapat pada *attribute* hari dan waktu.

RABU	2020	15.00 WIB
?	2020	?
SELASA	2020	14.00 WIB
RABU	2020	15.00 WIB
SENIN	2020	20.00 WIB

Gambar 4. Data *Missing Values*

Setelah dilakukan proses untuk menghilangkan *missing values*, kemudian didapatkan hasil, dan diambil rata-rata dari setiap *attribute* yang ada. Maka didapatkannya hasil sebagai berikut

RABU	15.00 WIB	L
RABU	20.00 WIB	L
SELASA	14.00 WIB	L
RABU	15.00 WIB	L
SENIN	20.00 WIB	L

Gambar 5. Data yang sudah di perbaiki

Setelah *missing values* berhasil diatasi, maka selanjutnya adalah mentransformasikan dataset ke dalam bentuk yang sesuai supaya dapat dilakukan proses data mining. Karena metode yang digunakan adalah *K-Means Clustering*, maka dataset yang ada harus bertipe data numerical. Sedangkan dataset yang didapatkan dari polres adalah data bertipe kategorikal. Maka dataset yang ada harus dilakukan transformasi dengan cara memberikan insialisasi data pada setiap *atributenya*.

Tabel 1. Inisialisasi Jenis Kelamin

No	Jenis Kelamin	Frekuensi	Inisialisasi
1	Laki-laki	43	1
2	Perempuan	12	2

Berikut adalah inisialisasi bulan yang terbagi menjadi 12 bulan dalam 1 tahun.

Tabel 2. Inisialisasi Bulan

No	Bulan	Frekuensi	Inisialisasi
1	April	6	1
2	Mei	6	2
3	Januari	5	3
4	Februari	5	4
5	Juli	5	5
6	Oktober	5	6
7	Desember	5	7
8	Maret	4	8
9	Juni	4	9
10	September	4	10
11	Agustus	3	11
12	November	3	12

Agar inisialisasi tidak terlalu panjang, maka pada inisialisasi umur menggunakan rumus jangkauan nilai untuk menginisialisasi umur.

Tabel 3. Inisialisasi Umur

No	Umur	Frekuensi	Inisialisasi
1	15-26	12	3
2	27-38	14	2
3	39-50	28	1
4	51-60	1	4

Berikut adalah inisialisasi penyebab terjadinya kecelakaan. Pada inisialisasi penyebab di bagi menjadi 3 bagian.

Tabel 4. Inisialisasi Penyebab Kecelakaan

No	Penyebab	Frekuensi	Inisialisasi
1	Tabrak Depan	39	1
2	Tabrak Belakang	14	2
3	Tabrak Samping	2	3

Berikut adalah inisialisasi jenis kendaraan yang mengalami kecelakaan lalu lintas. Pada inisialisasi ini, dibagi menjadi 4 bagian.

Tabel 5. Inisialisasi Jenis Kendaraan

No	Jenis Kendaraan	Frekuensi	Inisialisasi
1	R2	47	1
2	R4	5	2
3	R2 & R4	2	3
4	Lebih	1	4

Berikut adalah inisialisasi hari terjadinya kecelakaan lalu lintas. Pembagian hari terbagi menjadi 7 hari yang berada pada 1 minggu.

Tabel 6. Inisialisasi Hari

No	Hari Terjadinya	Frekuensi	Inisialisasi
1	Senin	7	4
2	Selasa	10	2
3	Rabu	13	1
4	Kamis	7	5
5	Jumat	5	6
6	Sabtu	5	7
7	Minggu	8	3

Berikut adalah inisialisasi tahun terjadinya kecelakaan lalu lintas.

Tabel 7. Inisialisasi Tahun

No	Tahun Terjadinya	Frekuensi	Inisialisasi
1	2019	35	1
2	2020	20	2

Berikut adalah inisialisasi waktu terjadinya kecelakaan. Pada inisialisasi waktu dibagi menjadi 5 bagian waktu, yaitu dini hari, pagi hari, siang hari, sore hari, dan malam hari.

Tabel 8. Inisialisasi Waktu

No	Waktu	Frekuensi	Inisialisasi
1	Dini Hari	1	5
2	Pagi Hari	3	4
3	Siang Hari	8	3
4	Sore Hari	11	2
5	Malam Hari	32	1

Setelah dilakukannya proses inialisasi pada setiap *atributenya*, maka selanjutnya adalah masuk ke dalam proses *data mining*, dimana metode yang digunakan adalah *K-Means Clustering*. Langkah pertama yang dilakukan pada saat proses data mining ini adalah menentukan *centroid* awal terlebih dahulu. *Centroid* awal dapat dipilih secara acak atau *simple random sampling*.

Tabel 9. *Centroid* Awal

Centorid	Data Ke-i	Jenis Kelamin	Bulan	Usia	Penyebab	Jenis Kendaraan	Hari	Tahun	Waktu
C1	10	1	1	3	1	1	2	1	1
C2	25	1	11	3	1	1	1	1	1
C3	48	1	11	1	2	1	7	2	1

Setelah menentukan *centroid* awal, maka tahap selanjutnya yaitu menghitung jarak setiap data ke *centroid* terdekat. Untuk menentukan *cluster* yang diikuti oleh data tersebut menggunakan rumus *Euclidean Distance*. Berikut contoh perhitungan jarak setiap *centroid*

- Data ke-1 ke pusat *cluster* ke-1 $d(x_1 C_1) = \sqrt{(1-1)^2 + (3-1)^2 + (1-3)^2 + (1-1)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2 + (1-1)^2} = 2.83$
.....(1)

Dari perhitungan data ke-1 dengan *cluster* ke-1 maka dihasilkan jarak sebesar 2.83

- Data ke-1 ke pusat *cluster* ke-2 $d(x_1 C_2) = \sqrt{(1-1)^2 + (3-11)^2 + (1-3)^2 + (1-1)^2 + (1-1)^2 + (2-1)^2 + (1-1)^2 + (1-1)^2} = 8.31$
.....(2)

Dari perhitungan data ke-1 dengan *cluster* ke-2 maka dihasilkan jarak sebesar 8.31

- Data ke-1 ke pusat *cluster* ke-3 $d(x_1 C_3) = \sqrt{(1-1)^2 + (3-11)^2 + (1-1)^2 + (1-2)^2 + (1-1)^2 + (2-7)^2 + (1-2)^2 + (1-1)^2} = 9.54$
.....(3)

Dari perhitungan data ke-1 dengan *cluster* ke-3 maka dihasilkan jarak sebesar 9.54

Hasil perhitungan diatas yaitu jarak data pertama dengan *cluster* ke-1 sebesar 2.83, jarak data pertama dengan *cluster* ke-2 sebesar 8.31, dan jarak data pertama dengan *cluster* ke-3 sebesar 9.54. Kemudian cari data yang memiliki jarak terdekat dengan *centroid*. Berdasarkan perhitungan, *cluster* 1 adalah yang paling dekat dengan *Centroid*. Berikut merupakan perhitungan lengkap dari Iterasi 1.

Tabel 10. Hasil Iterasi 1

Data Ke- i	C1	C2	C3	Jarak Terdekat	Kelompok Data
1	2.83	8.31	9.54	2.83	Cluster 1
2	4.00	9.11	8.54	4.00	Cluster 1
3	5.39	10.00	8.37	5.39	Cluster 1
4	3.46	7.14	9.43	3.46	Cluster 1
5	3.87	7.62	8.49	3.87	Cluster 1
6	3.61	7.48	8.25	3.61	Cluster 1
7	4.69	8.19	8.54	4.69	Cluster 1
8	7.87	5.39	4.80	4.80	Cluster 3
9	8.31	6.16	3.46	3.46	Cluster 3
10	0.00	10.05	11.45	0.00	Cluster 1
11	4.12	11.22	10.39	4.12	Cluster 1
12	1.73	10.30	10.95	1.73	Cluster 1
13	5.10	10.82	9.33	5.10	Cluster 1
14	2.83	9.64	9.64	2.83	Cluster 1
15	3.74	10.05	9.33	3.74	Cluster 1
16	4.12	10.10	10.10	4.12	Cluster 1
17	1.73	9.17	10.58	1.73	Cluster 1
18	8.54	4.24	4.90	4.24	Cluster 2
19	8.12	2.24	6.86	2.24	Cluster 2
20	9.11	5.66	2.83	2.83	Cluster 3
21	4.69	6.40	8.89	4.69	Cluster 1
22	5.39	7.48	6.48	5.39	Cluster 1
23	4.36	6.48	7.48	4.36	Cluster 1
24	10.58	4.36	2.65	2.65	Cluster 3
25	10.05	0.00	6.48	0.00	Cluster 2
26	9.43	2.83	6.63	2.83	Cluster 2
27	10.25	5.29	6.00	5.29	Cluster 2
28	6.71	7.35	5.29	5.29	Cluster 1
29	7.28	7.62	7.21	7.21	Cluster 1
30	5.39	5.29	8.00	5.29	Cluster 2
31	11.18	2.45	5.29	2.45	Cluster 2
32	11.79	5.29	2.83	2.83	Cluster 3
33	6.48	4.58	7.42	4.58	Cluster 2
34	6.40	4.69	6.63	4.69	Cluster 2
35	6.24	4.24	7.62	4.24	Cluster 2
36	3.61	8.60	9.70	3.61	Cluster 1
37	3.16	8.54	9.00	3.16	Cluster 1
38	3.61	8.83	8.60	3.61	Cluster 1
39	7.75	4.36	7.14	4.36	Cluster 2
40	7.42	4.24	5.10	4.24	Cluster 2
41	2.45	10.34	11.18	2.45	Cluster 1
42	3.74	11.00	10.25	3.74	Cluster 1
43	3.32	10.58	11.49	3.31	Cluster 1
44	2.24	9.17	10.95	2.24	Cluster 1
45	8.83	5.00	3.00	3.00	Cluster 3
46	4.69	6.56	7.94	4.69	Cluster 1
47	6.86	8.83	6.16	6.16	Cluster 1
48	11.45	6.48	0.00	0.00	Cluster 3
49	9.33	3.16	4.24	3.16	Cluster 2
50	10.58	6.56	1.73	1.73	Cluster 3
51	5.74	5.66	8.60	5.66	Cluster 2
52	5.57	5.48	7.87	5.48	Cluster 2
53	11.40	3.21	5.57	3.21	Cluster 2
54	6.78	6.00	7.55	6.00	Cluster 2
55	6.48	5.20	5.20	5.20	Cluster 3

Pada hasil iterasi ke-1 ditemukan hasil bahwa jumlah data pada *cluster* 1 adalah sebanyak 29 data, pada *cluster* 2 sebanyak 17 data, dan pada *cluster* 3 sebanyak 9 data. Setelah ditemukan hasil dari iterasi ke-1, maka selanjutnya adalah menentukan *Centroid* baru yang terbentuk dari perhitungan iterasi ke-1. Untuk rumus dalam menentukan *centroid* barunya adalah sebagai berikut.

1. Cluster 1 pada Variabel Jenis Kelamin :
 $1+1+1+1+2+1+1+1+1+2+1+2+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+2+2+2+1$
 $= 35/29$ (jumlah cluster 1) = **1.21**
2. Cluster 2 pada Variabel Jenis Kelamin :
 $1+1+1+2+1+1+1+2+1+2+2+1+1+1+1+1+1 = 21/17$ (jumlah cluster 2) = **1.24**
3. Cluster 3 pada Variabel Jenis Kelamin :
 $1+1+2+1+1+1+1+2+1 = 11/9$ (jumlah cluster 3) = **1.22**

Untuk perhitungan pada *variabel* yang lain adalah sama dan semuanya dihitung sampai dengan *variabel* yang terakhir dan untuk hasil perhitungan lebih jelasnya dapat melihat pada Tabel 11 berikut.

Tabel 11. *Centroid* baru yang terbentuk dari Iterasi ke-1

Variabel	Cluster 1	Cluster 2	Cluster 3
Jenis Kelamin	1.21	1.24	1.22
Bulan	3.07	8.53	9.44
Usia	1.79	1.76	1.78
Penyebab	1.31	1.29	1.44
Jenis Kendaraan	1.21	1.35	1.00
Hari	3.62	1.71	5.67
Tahun	1.31	1.41	1.44
Waktu	1.69	1.94	1.11

Setelah mendapatkan *centroid* baru, maka selanjutnya melakukan kembali perhitungan sampai dari setiap datanya tidak ada yang berpindah dari *clusternya*, Pada penelitian ini perhitungan berhenti pada Iterasi ke-3, dimana semua data sudah tidak ada yang berpindah atau dapat dikatakan setiap data pada masing-masing *clusternya* sudah konsisten. Kemudian setelah perhitungan pada Iterasi ke-3 sudah dilakukan, maka selanjutnya adalah pengelompokan data di setiap *cluster* masing-masing.

Setelah data sudah dimasukan ke dalam masing *clusternya*, maka selanjutnya adalah menghitung perhitungan evaluasinya dengan menggunakan *silhouette coefficient*. Sebelum menentukan $S(i)$, maka sebelumnya harus menentukan terlebih dahulu $d(i,C)$ pada masing-masing *clusternya*. Tujuan dari dilakukannya perhitungan $d(i,C)$ pada masing-masing *clusternya* adalah untuk mendapatkan nilai terendah dari masing-masing *clusternya*. Kemudian setelah mendapatkan nilai terendah dari masing-masing *clusternya* maka itu dimasukkan ke dalam nilai $b(i)$.

Setelah menghitung semua nilai $d(i,C)$ dan memasukkannya ke dalam $b(i)$ sebagai nilai terendahnya, kemudian dari semua nilai $b(i)$ dihitung nilai $S(i)$ dengan menggunakan rumus sebagai berikut

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots (4)$$

Setelah didapatkan nilai dari *silhouette coefficient* ini, maka total keseluruhannya dicari nilai rata-ratanya, apakah setiap *cluster* penyebarannya dikatakan baik atau tidaknya. Untuk rata-rata pada *cluster* 1 didapatkan hasil $S(i) = 0.35$, untuk rata-rata pada *cluster* 2 didapatkan hasil $S(i) = 0.22$, sedangkan nilai rata-rata yang didapatkan pada *cluster* 3 adalah $S(i) = 0.38$. Setelah dicari rata-ratanya, maka didapatkan bahwa nilai rata-rata dari *silhouette coefficientnya* keseluruhan adalah bernilai $S(i) = 0.32$. Setelah didapatkan hasil bahwa rata-rata nilai $S(i) = 0.32$, itu artinya masuk ke dalam kategori *weak structure*. Dapat dilihat bahwa dari masing-masing *cluster* menghasilkan nilai

s(i) yang berbeda-beda, namun dari ketiga *cluster* tersebut nilai s(i) yang terbaik adalah pada *cluster* 3 walaupun masih tetap masuk ke dalam *weak structure*.

4. SIMPULAN

Berdasarkan hasil analisa dan pengolahan data yang ada, maka dapat diambil kesimpulan sebagai berikut :

1. Menurut hasil Analisa dan pengolahan data dari dataset kecelakaan lalu lintas di kecamatan cileungsi diketahui bahwa penyebab dari kecalakaan lalu lintas di kecamatan cileungsi ini adalah tabrak depan. Pada setiap *cluster* di dominasi dengan penyebab kecelakaan, yaitu tabrak depan
2. Menurut hasil Analisa dan pengolahan data dari dataset kecelakaan lalu lintas di kecamatan cileungsi juga di ketehau bahwa hari paling sering terjadinya suatu kecelakaan lalu lintas di kecamatan cileungsi pada *cluster* 1 adalah di hari senin, *cluster* 2 yaitu pada hari rabu, dan *cluster* 3 pada hari kamis dan jumat. Dan waktu dari setiap *cluster* di dominasi pada jam pulang kerja, yaitu rentang pada pukul 17 sampai dengan pukul 23.59.
3. Menurut hasil Analisa data dan pengolahan data dari dataset kecelakaan lalu lintas di kecamatan cileungsi diketahui bahwa umur yang paling rentan mengalami kecelakaan pada setiap *clusternya* di dominasi oleh rentan usia 39-50 tahun dan di susul dengan rentang usia 27-38 tahun. Ini diakibatkan karena pada usia tersebut kebanyakan adalah orang yang sedang melakukan pekerjaan dan tentunya menghabiskan waktu sebagai besar di jalan..

5. SARAN

Berdasarkan temuan-temuan dalam penelitian ini, dapat disimpulkan beberapa saran, yaitu :

1. Menggunakan lebih banyak lagi angka kejadian kecelakaan di kecamatan cileungsi, khususnya setiap kecelakaan yang terjadi di wilayah Kecamatan Cileungsi. Sehingga dapat menganalisis pengaruh jumlah kejadian kecelakaan terhadap akurasi klasifikasi penyebab terjadinya kecelakaan.
2. Mengumpulkan lebih lengkap data kecelakaan untuk memetakan kelompok kecelakaan lalu lintas dan dapat merancang suatu metode klasifikasi yang lebih efisien dan akurat

DAFTAR PUSTAKA

- [1] World Health Organization (WHO) .2013. Global Status Report On Road Safety 2013: Supporting A Decade Of Action. Switzerland : Printed in Luxembourg.
- [2] Achmad, Maududie & Wahyu Catur Wibowo. 2014. Perbaikan Inisialisasi KMeans Menggunakan Graf Hutan Yang Minimum. Prosiding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen. Depok: Universitas Gunadarma
- [3] Anshori, Iedam Fardian, and Yeni Nuraini. 2020. "Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means." *Jurnal Responsif* 2(1):118–27.
- [4] Anonim, 2009. Undang-undang Republik Indonesia No.22 Tahun 2009 tentang Lalu lintas dan Angkutan Jalan. Jakarta: Pemerintah Republik Indonesia
- [5] Anonim, 1993. Peraturan Pemerintah Republik Indonesia Nomor 44 Tahun 1993 Tentang Jalan. Jakarta: Pemerintah Republik Indonesia.
- [6] Aprianti, Winda, and Jaka Permadi. 2018. "K-Means Clustering Untuk Data Kecelakaan Lalu Lintas Jalan Raya Di Kecamatan Pelaihari." *Jurnal Teknologi Informasi Dan Ilmu Komputer* 5(5):613. doi: 10.25126/jtiik.2018551113.

- [7] Ariani, Dinda Retno. 2020. "Identifikasi Penyebab Kecelakaan Lalu Lintas Dengan Menggunakan Metode K-Means." 8:320–25.
- [8] Aribowo, Agus Sasmito. 2015. "Analisa Asosiatif Data Mining Untuk Mengetahui Pola Kecelakaan Lalu Lintas." *Telematika* 8(2):2–7. doi: 10.31315/telematika.v8i2.458.
- [9] Azizirrahman, Muhammad, Ellyn Normelani, Deasy Arisanty, Kecelakaan Lalu Lintas, and Daerah Rawan Kecelakaan. 2015. "Faktor Penyebab Terjadinya Kecelakaan Lalu Lintas Pada Daerah Rawan Kecelakaan Di Kecamatan Banjarmasin Tengah Kota Banjarmasin." 2(3).
- [10] Oktavia, Rezki, Sutardi, and Jumadil Nangi. 2017b. "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN DAERAH RAWAN KECELAKAAN MENGGUNAKAN ALGORITMA C4.5 Rezki." *SemanTIK* 3(2):95–104.
- [11] Putu, Ni, Ratindia Apriyanti, I. Ketut Gede, Darma Putra, and I. Made Suwija Putra. 2020. "Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression." *Jurnal Ilmiah Merpati* 8(2):72–80.
- [12] Marsaid, M. Hidayat, and Ahsan, "Faktor Yang Berhubungan Dengan Kejadian Kecelakaan Lalu Lintas pada Pengendara Sepeda Motor di Wilayah Polres Kabupaten Malang," *J. Ilmu Keperawatan*, vol. 1, no. 2, pp. 98–112, 2013.
- [13] Rahmat C.T.I., Brilian, Agum Agidtama Gafar, Nurul Fajriani, Umar Ramdani, Fitria Rihin Uyun, Yuwanda Purnamasari P., and Natalis Ransi. 2017. "Implementasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan Kecelakaan." *Seminar Nasional Riset Kuantitatif Terapan 2017 (April)*:58–60.