

Pengelompokan Jumlah Kecelakaan Lalu Lintas di Padang Menggunakan Analisis *Cluster* dengan Metode Non-Hirarki (*K-Means*)

Tiara Angela Hanami¹, Helma²

^{1,2}Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang (UNP)

Article Info

Article history:

Received April 05, 2023

Revised May 09, 2023

Accepted June 30, 2023

Keywords:

Traffic Accidents

Cluster Analysis

Elbow Method

K-Means

Kata Kunci:

Kecelakaan Lalu Lintas

Analisis Cluster

Metode Elbow

K-Means

ABSTRACT

Traffic Accidents in August and July 2016. Traffic accidents in West Sumatra Province, especially in Padang City, are relatively high. Based on data from the Central Statistics Agency of Padang City, the number of traffic accident cases has increased from 2016-2021. This study aims to make a pattern in recognizing the characteristics of traffic accidents in the Padang City to minimize the increase in the frequency of accidents in the following year. This type of research is applied research. The result showed that optimizing clusters using the elbow method produced 4 clusters with categories: the first cluster included relatively high traffic accidents in 2021; the second cluster includes groups that have relatively moderate numbers in 2018 and cause material losses in 2021; the third cluster is relatively low, namely material losses in 2016 and 2018; the fourth cluster of very low traffic accidents in July and August 2016.

ABSTRAK

Kecelakaan lalu lintas di Provinsi Sumatera Barat tepatnya di Kota Padang relatif meningkat. Berdasarkan dari data Badan Pusat Statistik Kota Padang, menghitung angka kasus jumlah kecelakaan lalu lintas meningkat dari tahun 2016-2021. Penelitian ini bertujuan untuk membuat pola dalam mengenali karakteristik kecelakaan lalu lintas di Kota Padang untuk meminimalisir peningkatan nilai frekuensi kecelakaan pada tahun berikutnya. Jenis penelitian ini merupakan penelitian terapan. Dari hasil analisis diperoleh bahwa pengoptimalan cluster dengan metode *elbow* menghasilkan 4 cluster dengan kategori: *cluster* pertama termasuk kecelakaan lalu lintas yang relatif tinggi pada tahun 2021; *cluster* kedua termasuk kelompok yang memiliki angka relatif sedang pada 2018 serta menyebabkan kerugian material tahun 2021; *cluster* ketiga relatif rendah yaitu kerugian material tahun 2016 dan 2018; *cluster* keempat kecelakaan lalu lintas yang sangat rendah pada bulan Juli dan Agustus tahun 2016.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis pertama

(Tiara Angela Hanami)

Departemen Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Negeri Padang, Jl.Prof.Dr. Hamka, Air Tawar barat, Padang Utara, Padang, 25171 Padang, Sumatera Barat

Email: hanamiangela02@gmail.com



1. PENDAHULUAN

Berdasarkan UU RI Pasal 1 No. 22 tahun 2009 menyebutkan bahwa kecelakaan lalu lintas adalah kecelakaan yang terjadi di jalan raya secara tidak terduga sehingga menimbulkan korban jiwa dan kerugian akan material [1]. Menurut Organisasi Kesehatan Dunia (WHO) korban jiwa yang ditimbulkan sebanyak 1,3 juta jiwa pada kasus kecelakaan lalu lintas setiap tahunnya, tidak menutup kemungkinan akan bertambah pada tahun berikutnya. Kecelakaan dikategorikan menjadi tiga jenis yaitu kecelakaan yang berakibat meninggal dunia, luka berat, dan luka ringan [2].

Sumatera Barat, khususnya Kota Padang, memiliki tingkat kecelakaan lalu lintas relatif naik. Berdasarkan sumber Badan Pusat Statistik (BPS) Kota Padang, pada 2016-2021 tercatat angka kasus kecelakaan lalu lintas meningkat di daerah Padang. Tahun 2016 tercatat sejumlah 932 kasus kecelakaan lalu lintas dan mengalami kerugian material mencapai Rp.4.624.100. Tahun 2018 tercatat kasus kecelakaan sebanyak 1092 serta mengalami kerugian mencapai Rp.1.770.830. Sedangkan pada tahun 2021 mencapai 1100 kasus kecelakaan lalu lintas dan mengalami kerugian material sebesar Rp. 4.498.050 [3].

Sebagai upaya dalam mengurangi tingkat kecelakaan lalu lintas diperlukan pengetahuan untuk menentukan pola yang sewaktu-waktu terjadi dengan melakukan pengelompokan [4]. Pola dikelompokkan berdasarkan kecelakaan dengan kategori luka berat, kematian, luka ringan, kerusakan harta benda dan jumlah kecelakaan lalu lintas. Untuk mengelompokkan kecelakaan lalu lintas di Kota Padang Tahun 2016-2021 memerlukan analisis statistika yakni analisis *cluster* dengan metode non-hierarki. Kelengkapan dan kebenaran data menjadi faktor penting dalam keberhasilan pengelompokan pola kecelakaan lalu lintas [5].

1.1 Analisis Cluster

Analisis cluster dikenal juga sebagai teknik *data mining* dimana teknik ini bertugas dalam mengelompokkan suatu peristiwa dalam bentuk kelompok - kelompok kecil dimana setiap kelompok memiliki kemiripan objek satu sama lain. Jarak *Euclidean* disebut juga sebagai ukuran jarak yang digunakan pada analisis cluster [6].

Jarak *euclidean* antara *cluster* ke-*i* dan ke-*j* dari *p* variabel didefinisikan:

$$\begin{aligned}
 x &= (x_1, x_2, \dots, x_p) \text{ dan } y = (y_1, y_2, \dots, y_p) \text{ maka} \\
 d(x, y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \\
 d(x, y) &= \sqrt{(x - y)'(x - y)} \\
 d(x, y) &= \sqrt{\sum_{i=0}^p (x_i - y_i)^2} \tag{1}
 \end{aligned}$$

Dimana:

$d(x, y)$ = jarak antara objek x ke objek y

x_i = data x pada observasi ke- i

y_i = titik pusat ke y observasi ke- i

p = banyaknya variabel yang diamati.

1.2 Metode K-Means

Pengertian Metode *k-means* sendiri adalah metode *cluster* data non-hierarki yang bertujuan mempartisi data dalam bentuk *cluster*. Data yang telah diperoleh memiliki karakteristik seimbang yang selanjutnya akan dikelompokkan pada satu *cluster* yang sama [7]. Adapun algoritma yang digunakan dalam proses pengelompokkan data adalah sebagai berikut [8]:

- a. Menentukan banyaknya *cluster* (nilai k).
- b. Menentukan *centroid* (titik pusat) secara acak dibagi menjadi beberapa kelompok.
- c. Menghitung pusat kelompok dari data yang ada. Lokasi kelompok diambil dari rata – rata semua nilai dari setiap fiturnya. Nilai n menyatakan banyaknya jumlah data yang didapat dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i adalah sebagai berikut.

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (2)$$

Jarak adalah akar dari jumlah kuadrat dari pengurangan keanggotaan data ke- k dengan *cluster* ke- i dan nilai *centroid cluster* ke- i .

Dimana:

$d(x, y)$ = jarak data ke x ke pusat cluster y

x_i = data pada observasi ke- i

y_i = titik pusat ke y observasi ke- i

n = banyak observasi.

- d. Data kemudian dialokasikan ke *centroid*/rata-rata terdekat. Diantaranya *Eucliden* pengukuran jarak sebagai berikut.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3)$$

Selanjutnya data yang dihasilkan akan dialokasi kembali dengan menerapkan metode *k-means* yang berdasarkan pada perbandingan jarak antara data dengan titik pusat pada setiap kelompok yang diperoleh.

$$a_{i1} = \begin{cases} 1, & d = \min\{D(x_i, C_1)\} \\ 0 & \text{lainnya} \end{cases} \quad (4)$$

Fungsi tujuan yang digunakan dalam metode *k-means* ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok.

- e. Objek yang diperoleh kemudian dihitung kembali dengan menggunakan *centroid* baru. Proses *clustering* dikatakan selesai jika tidak terjadi perubahan pada anggota *cluster* yang ada.

2. METODE PENELITIAN

Penelitian yang dilaksanakan merupakan jenis penelitian terapan dengan memanfaatkan data sekunder berupa data kecelakaan lalu lintas tahun 2016-2021 yang dikumpulkan melalui sumber BPS Kota Padang. Variabel yang digunakan adalah kecelakaan lalu lintas yang menyebabkan luka ringan (X_1, X_2, X_3), luka berat (X_4, X_5, X_6), meninggal (X_7, X_8, X_9), jumlah kecelakaan lalu lintas (X_{10}, X_{11}, X_{12}), dan kerugian material (X_{13}, X_{14}, X_{15}).

Metode yang digunakan yaitu:



- a. Melakukan proses identifikasi terhadap topik permasalahan yaitu kecelakaan lalu lintas di Kota Padang.
- b. Mengumpulkan data dan mengkaji teori-teori yang relevan tentang masalah kecelakaan lalu lintas di Kota Padang.
- c. Menentukan variabel.
- d. Melakukan standarisasi (jika diperlukan).
- e. Memeriksa korelasi yang nantinya akan dicari akar karakteristik (nilai eigen), proporsi keberagaman (%), keragaman kumulatif dan didapatkan skor AKU (Analisis Komponen Utama).
- f. Menentukan jumlah *cluster* dengan dilalukan pengoptimalan dengan metode *elbow* bertujuan untuk menentukan banyak *k* optimal.
- g. Menentukan analisis *cluster k-means* dengan *k* optimal yang sudah didapatkan.
- h. Menginterpretasikan hasil analisis yang diperoleh.
- i. Membuat kesimpulan.

3. HASIL DAN PAMBAHASAN

3.1 Statistik Deskriptif

Setelah selesainya pengumpulan data maka tahapan selanjutnya yaitu mendeskripsikan data pada masing-masing variabel dan mencari karakteristik data, dalam menentukan karakteristik data akan dicari nilai rata-rata, nilai maksimum dan minimum dengan menggunakan R-studio didapatkan data sebagai berikut:

Tabel 1. Statistik Deskriptif Kecelakaan Lalu Lintas tahun 2016 sampai 2021

NO	Variabel	Keterangan	Rata-rata	Nilai Maksimum	Nilai Minimum
1	X_1	Kecelakaan lalu lintas yang menyebabkan luka ringan pada tahun 2016	46,917	112	31
2	X_2	Kecelakaan lalu lintas yang menyebabkan luka ringan pada tahun 2018	76,75	97	56
3	X_3	Kecelakaan lalu lintas yang menyebabkan luka ringan pada tahun 2021	82,083	108	50
4	X_4	Kecelakaan lalu lintas yang menyebabkan luka berat pada tahun 2016	25,833	47	12
5	X_5	Kecelakaan lalu lintas yang menyebabkan luka berat pada tahun 2018	8,333	21	4
6	X_6	Kecelakaan lalu lintas yang menyebabkan luka berat pada tahun 2021	4,667	10	1
7	X_7	Kecelakaan lalu lintas yang menyebabkan korban meninggal tahun 2016	4,917	9	2
8	X_8	Kecelakaan lalu lintas yang menyebabkan korban meninggal tahun 2018	5,917	10	2
9	X_9	Kecelakaan lalu lintas yang menyebabkan korban meninggal tahun 2021	4,917	11	2
10	X_{10}	Jumlah kecelakaan lalu lintas pada tahun 2016	48,25	65	34
11	X_{11}	Jumlah kecelakaan lalu lintas pada tahun 2018	55,25	68	39
12	X_{12}	Jumlah kecelakaan lalu lintas pada tahun 2021	57,25	69	42
13	X_{13}	Kerugian material pada tahun 2016	192670833	1017100000	18900000
14	X_{14}	Kerugian material pada tahun 2018	73784583	127850000	5200000
15	X_{15}	Kerugian material pada tahun 2021	187420833	765500000	49100000

Berdasarkan Tabel 1, hasil nilai rata-rata pada kasus kecelakaan lalu lintas pada kategori luka ringan di tahun 2021 sebesar 82,083 dengan nilai maksimum 108 dan minimum 50. Serta tahun 2016 memiliki rentang data yang besar dengan nilai maksimum sebesar 112 dan minimum 31. Kerugian

material tahun 2016-2021 menunjukkan terjadinya perbedaan yang signifikan. Dikarenakan rentang yang sangat besar antara data akan mempengaruhi hasil pengolahan data sehingga harus dilakukan Standarisasi data terlebih dahulu sebelum mengolah data yang lebih lanjut.

3.2 Standarisasi

Standarisasi dilakukan apabila data mempunyai skala yang berbeda atau pada skala umum dengan rentang nilai yang sangat berbeda. Dapat diketahui bahwa data awal bukan menggunakan satuan data dalam satuan yang sama. Untuk itu diperlukan pembakuan atau standarisasi variabel asal X kedalam bentuk baku X^* dengan menggunakan R-Studio yang dapat dibuktikan hasilnya pada Tabel 2. Dilihat data hasil standarisasi memiliki rentang yang tidak begitu jauh antara data dan variabel sehingga setiap data dan variabel memiliki pengaruh pada saat dilakukannya analisis cluster. Sebelum melakukan analisis *cluster* maka di uji korelasi terlebih dahulu dengan menggunakan R-Studio.

Tabel 2. Standarisasi Data

Bulan	X1	X2	X3	X4	X5	X6	X8	X9	X10	X11	X12	X13	X14	X15
Januari	-0.08886664	-0.6592354	1.247.693.477	-133.354.746	0.33353554	0.5117663	0.03484434	0.0307253	-11.040.354	-0.3658115	0.8815049	-0.4785336	-0.47339953	296.016.502
Februari	301.760.200	0.3615162	0.245362316	0.30526990	0.53365686	0.8955911	-163.768.398	15.055.396	15.095.995	0.3095328	0.5402772	-0.3073309	146.383.487	-0.70829830
Maret	-0.32069266	10.420.173	-1.258.134.426	-114.074.542	0.93389950	-10.235.326	-0.80141982	-0.3379783	15.095.995	0.7597623	-15.070.890	-0.3593779	-0.07945459	-0.45508011
April	0.46751580	0.7017667	0.495945106	-0.27313623	253.487.008	0.1279416	0.45297642	22.429.468	11.490.981	13.225.491	0.6540198	-0.3605493	127.160.057	0.66302286
Mei	-0.18159705	-17.650.497	-0.381094660	-0.08033418	-0.26682843	-10.235.326	0.03484434	-0.3379783	-0.3830327	-18.290.573	-0.4834059	-0.2059146	-128.701.090	-0.48862064
Juni	-0.73797949	-0.2339222	0.746527896	-0.27313623	-0.86719240	20.470.653	-0.38328774	0.0307253	-0.3830327	-12.662.704	12.227.326	27.594.234	0.32532045	-0.01034806
Juli	-0.50615347	-0.2339222	-2.009.882.797	0.20886888	-0.46694975	-14.073.574	-121.955.190	-10.753.854	-0.4731580	-0.1406967	-17.345.741	-0.3789582	-0.29605662	-0.26894298
Agustus	-0.08886664	0.7868294	-0.193157568	0.40167092	-0.06670711	-0.2558832	170.737.266	0.3994289	0.2478447	14.351.065	-0.5971485	12.828.628	115.382.322	-0.54136376
September	-0.50615347	-0.6592354	-0.318448963	204.048.828	-0.86719240	0.1279416	0.87110850	-0.7066819	0.6083461	-0.5909262	-0.1421782	-0.4100860	-185.694.499	-0.53317066
Oktober	-0.59888388	17.225.184	-0.193157568	-0.27313623	-0.06670711	-10.235.326	-0.80141982	-10.753.854	-12.842.861	0.5346475	-0.4834059	-0.5816234	-0.15661906	-0.31605334
November	-0.27432745	0.1913909	1.623.567.662	-104.434.439	-0.86719240	0.5117663	0.87110850	-0.7066819	-11.040.354	0.5346475	0.3127921	-0.4348543	-0.45309309	0.01551143
Desember	-0.18159705	-12.546.739	-0.005220475	146.208.215	-0.86719240	0.5117663	0.87110850	0.0307253	-0.2929074	-0.7034836	13.364.752	-0.5250579	0.38799967	-0.31682144

3.3 Korelasi Antar Variabel

Berdasarkan data hasil standarisasi untuk langkah selanjutnya akan dilakukan uji korelasi antara variabel. Maka dibuktikan pada Tabel 3 terdapat banyak variabel atau peubah, ada beberapa variabel yang mempunyai nilai korelasi cukup tinggi, yang mempunyai hubungan korelasi yang signifikan diantaranya adalah hubungan antara variabel X_1 dengan variabel X_6 , variabel X_2, X_5, X_8, X_{13} , variabel $X_3, X_{10}, X_{11}, X_{12}$, dan variabel X_9, X_{14} . Sedangkan pada variabel X_4, X_7 dan X_{15} terdapat hubungan korelasi yang sedang.

Tabel 3. Uji Korelasi

Berdasarkan Tabel 3 dapat disimpulkan terdapat korelasi antar variabel atau peubah, maka adanya *Multikolinearitas* yang ditandai dengan cukup banyak terjadi korelasi antar variabel asal, maka dari itu, Analisis Komponen Utama (AKU) perlu untuk dilakukan.

	X1	X2	X3	X4	X5	X6	X8	X9	X10	X11	X12	X13	X14	X15
X1	1	0.10567984	0.150003980	0.04584780	0.388298107	0.25184195	-0.37202054	0.67590069	0.55395998	0.228320115	0.22640941	-0.18989578	0.56434524	-0.10467994
X2	0.10567984	1	-0.062371277	-0.34254158	0.439498964	-0.17511814	-0.21421285	0.12188763	0.20611890	0.823183104	-0.30455025	0.02153610	0.47312402	-0.13700510
X3	0.15000398	-0.06237128	1	-0.29637386	-0.007218123	0.72644950	0.36215433	0.31514352	-0.30090426	0.003044852	0.79918763	0.16298644	0.13532416	0.50020321
X4	0.04584780	-0.34254158	-0.296373860	1	-0.365376890	0.05494102	0.28521218	-0.01992585	0.21365012	-0.231316037	0.10017931	-0.02179160	-0.13347149	-0.50347939
X5	0.38829811	0.43949896	-0.007218123	-0.36537689	1	-0.15827827	-0.17242563	0.72667486	0.56075571	0.571319212	-0.04759393	-0.23572400	0.50656594	0.26461775
X6	0.25184195	-0.17511814	0.726449499	0.05494102	-0.158278271	1	0.15562612	0.40739902	0.02515806	-0.192446626	0.86123770	0.53797319	0.30061977	0.22983830
X7	-0.06752141	0.47834345	-0.460289466	-0.10679414	0.029363796	-0.44491688	-0.11196833	-0.17988205	0.15174642	0.273745073	-0.63461608	0.33351729	0.12584434	-0.43975440
X8	-0.37202054	-0.21421285	0.362154326	0.28521218	-0.172425632	0.15562612	1	0.04087756	-0.15673241	0.129425583	0.26914306	0.13379021	-0.10557935	0.07683153
X9	0.67590069	0.12188763	0.315143520	-0.01992585	0.726674860	0.40739902	0.04087756	1	0.61399119	0.355582166	0.46226332	0.10146592	0.73684856	0.14538948
X10	0.55395998	0.20611890	-0.300904263	0.21365012	0.560755711	0.02515806	-0.15673241	0.61399119	1	0.337758133	-0.13582691	-0.02198969	0.39328856	-0.35645997
X11	0.22832011	0.82318310	0.003044852	-0.23131604	0.571319212	-0.19244663	0.12942558	0.35558217	0.33775813	1	-0.24179389	-0.16590025	0.58051262	-0.02054306
X12	0.22640941	-0.30455025	0.799187629	0.10017931	-0.047593927	0.86123770	0.26914306	0.46226332	-0.13582691	-0.241793887	1	0.23495481	0.26236203	0.37904902
X13	-0.18989578	0.02153610	0.162986440	-0.02179160	-0.235724001	0.53797319	0.13379021	0.10146592	-0.02198969	-0.165900246	0.23495481	1	0.26797675	-0.11207335
X14	0.56434524	0.47312402	0.135324161	-0.13347149	0.506565939	0.30061977	-0.10557935	0.73684856	0.39328856	0.580512623	0.26236203	0.26797675	1	-0.05171112
X15	-0.10467994	-0.13700510	0.500203208	-0.50347939	0.264617755	0.22983830	0.07683153	0.14538948	-0.35645997	-0.020543062	0.37904902	-0.11207335	-0.05171112	1



3.4 Analisis Komponen Utama (AKU)

Berdasarkan dari hasil pemeriksaan korelasi antara variabel kasus kecelakaan lalu lintas terdapat *multikolinearitas*, setelah analisis komponen utama maka selanjutnya dilakukan analisis *cluster*. Dari analisis komponen utama maka didapatkan akar-akar karakteristik, nilai proporsi keberagaman dan keragaman kumulatif disajikan pada Tabel 4.

Tabel 4. Nilai Keragaman dan Proporsi Masing-Masing Variabel

NO	Komponen Utama ke-j	Akar karakteristik (nilai eigen)	Proporsi Keberagaman (%)	Keragaman Kumulatif (%)
1	KU1	3,8435	0,9233	0,9233
2	KU2	0,91703	0,05256	0,97586
3	KU3	0,44112	0,01216	0,98802
4	KU4	0,30749	0,00591	0,99393
5	KU5	0,1959	0,0024	0,9963
6	KU6	0,16237	0,00165	0,99797
7	KU7	0,12463	0,00097	0,99895
8	KU8	0,08329	0,00043	0,99938
9	KU9	0,07999	0,0004	0,99978
10	KU10	0,04653	0,00014	0,99991
11	KU11	0,02829	0,00005	0,99996
12	KU12	0,0239	0,00004	1

Hasil diatas menunjukkan nilai eigen pada komponen utama bernilai lebih dari 1 yaitu komponen utama pertama (KU₁) dengan nilai sebesar 3,8435 dan nilai proporsi keberagaman data sebesar 92,33% (Tabel 4). Pada komponen utama yang digunakan sudah mampu menjelaskan keberagaman data. Dapat dilihat bahwa terdapat satu nilai yang jaraknya sangat jauh dari data lainnya, hal ini mengakibatkan data memiliki nilai yang sangat besar dibandingkan dengan nilai pada data-data lainnya.

3.5 Optimasi K Cluster dengan Metode Elbow

Penentuan jumlah *cluster* dilakukan sebelum analisis *k-means cluster*. Pada *k-means* adanya pertimbangan agar dapat menentukan banyak *k* kelompok yang paling efektif. Pendekatan metode paling banyak digunakan untuk cluster yaitu metode *elbow* untuk mencari banyak *k* optimal. Untuk mendapatkan perbandingan *k* optimal maka akan dihitung nilai *SSE* (*Sum of Square Error*). Dikarenakan bertambah besar jumlah *cluster* maka nilai *SSE* akan semakin menurun. Dapat dilihat dalam menghitung nilai *SSE* dengan menggunakan persamaan sebagai berikut:

$$\begin{aligned}
 SSE &= ((45-46,917)^2 + (69-76,50)^2 + (102-82,083)^2 + (12-25,833)^2 + (10-8,333)^2 + (6-4,667)^2 + (3-4,917)^2 \\
 &\quad + (6-5,917)^2 + (5-4,917)^2 + (36-48,20)^2 + (552-55,20)^2 + (65-57,20)^2 + (49700000-192670833)^2 \\
 &\quad + (56300000-73784583)^2 + (765500000-187420833)^2 + \dots + ((43-46,917)^2 + (62-76,50)^2 + \\
 &\quad (82-82,083)^2 + (41-25,833)^2 + (4-8,333)^2 + (6-4,667)^2 + (2-4,917)^2 + (8-5,917)^2 + (5-4,917)^2 + \\
 &\quad (45-48,20)^2 + (49-55,20)^2 + (69-57,20)^2 + (35800000-192670833)^2 + (88115000-73784583)^2 + \\
 &\quad (125550000-187420833)^2). \\
 &= 3,83564E+17.
 \end{aligned}$$

Pada metode *elbow*, banyaknya *k* yang akan dibentuk dapat dilihat dari persamaan diatas dimana *k* mendekati nilai yang berada di 4. Artinya berdasarkan metode yang diperoleh, jumlah *cluster* terbaik sebanyak 4 *cluster*.

3.6 Analisis Cluster K-Means

Setelah melakukan analisis komponen utama untuk melihat multikolinearitas data, tahap selanjutnya membuat analisis *cluster* dengan menggunakan jumlah *cluster* yang telah didapatkan sebanyak $k = 4$. Jarak yang digunakan dalam penelitian ini dikenal sebagai jarak *euclidean distance* dengan persamaan, dilakukan pencarian dengan menggunakan rumus dibawah ini:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$\begin{aligned} d(1,1) &= \sqrt{(-0,2575459 - 0,8125899)^2 - (-0,2595423 - (-0,4452492))^2 \\ &\quad + \dots + (-0,2472272 - 0,12918564)^2} \\ &= 4,45049234 \end{aligned}$$

$$\begin{aligned} d(1,2) &= \sqrt{(-0,2575459 - (-0,32834))^2 - (-0,2595423 - (-0,4732407))^2 \\ &\quad + \dots + (-0,2472272 - 0,03408157)^2} \\ &= 1,122347937 \end{aligned}$$

$$\begin{aligned} d(1,3) &= \sqrt{(-0,2575459 - 0,9778618)^2 - (-0,2595423 - 1,7596919)^2 \\ &\quad + \dots + (-0,2472272 - 0,24391095)^2} \\ &= 4,577332027 \end{aligned}$$

$$\begin{aligned} d(1,4) &= \sqrt{(-0,2575459 - 0,418134)^2 - (-0,2595423 - 1,2357415)^2 \\ &\quad + \dots + (-0,2472272 - (-0,32287456))^2} \\ &= 3,346147885 \end{aligned}$$

Dari hasil perhitungan jarak maka dilakukan perbandingan dengan menentukan jarak terdekat antara pusat *cluster* dengan data yang dihasilkan. Hasil yang diperoleh akan membuktikan data yang dimiliki berada pada kelompok data yang sama dengan pusat *cluster* terdekat.

3.7 Analisis Cluster

Cluster akan terbentuk dari hasil jara terdekat antara pusat *cluster* dan data yang dimiliki, maka terbentuklah *cluster*. Seperti misalnya data pertama jarak terdekat pusat *cluster* 1 maka data pertama berada pada *cluster* 1. Hasil pengelompokkan dapat dilihat pada Tabel 5 berikut ini:

Tabel 5. Data Hasil Pengelompokkan

Bulan	Jan	Feb	Mar	Apr	Mei	Juni	Juli	Agust	Sep	Okt	Nov	Des
Jumlah Cluster	2	2	2	2	2	2	4	4	2	1	3	2

Setiap *cluster* yang terbentuk memiliki karakteristik yang berbeda. Sehingga didapat pengelompokkannya 4 *cluster* yaitu *cluster* pertama, *cluster* kedua, *cluster* ketiga, dan *cluster* keempat. Dari setiap *cluster* yang sudah dikelompokkan kita dapat mencari rata-rata setiap *cluster* tersebut untuk dapat melihat karakteristiknya.

Berikut ini merupakan karakteristik setiap *cluster* berdasarkan kecelakaan lalu lintas di Kota Padang. Perolehan nilai rata – rata pada setiap *cluster* dapat dilihat pada uraian dibawah ini (Tabel 6).



Tabel 6. Rata-Rata Kecelakaan Lalu Lintas di Kota Padang Setiap Cluster

NO	Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Rata-rata
1	X_1	0,8125899	-0,3283400	0,9778618	0,4181340	0.004107
2	X_2	-0,4452492	-0,4732407	1,7596919	1,2357415	-0.01602
3	X_3	1,48552673	-0,3360229	-1,0667003	1,1348081	0.013006
4	X_4	0,92279367	0,05662408	1,13568461	-1,25573546	-0.14997
5	X_5	1,05935291	-0,20059194	0,48400235	0,03069015	-0.03016
6	X_6	-1,0645165	-0,0288620	0,2411470	0,5271328	0.1075
7	X_7	0,83645134	-0,1776246	-0,012604	0,298557487	-0.006230
8	X_8	-0,9088742	0,2511949	0,1627857	-0,6317352	-0.02514
9	X_9	1,24325828	0,05293269	0,3459223	-1,00632104	0.0254
10	X_{10}	-2,0481376	0,1057600	-0,9441751	1,0731163	-0.16250
11	X_{11}	-1,3481639	-0,2840978	1,8477357	0,8866053	-0.06724
12	X_{12}	1,3200895	0,01651065	-2,6470897	0,59745747	-0.07028
13	X_{13}	0,100763	0,07007255	0,0167331	-0,33903825	0.04272
14	X_{14}	-1,1508394	0,2915587	-0,9233998	-0,1291150	-0.02312
15	X_{15}	0,1291856	0,03408157	0,24391095	-0,32287456	-0.002997
	Rata-rata	0,062949	-0,06334	0,1081	0,167828	

3.8 Pembahasan

Berdasarkan perhitungan nilai *SSE* (*Sum of Square Error*) dari masing-masing nilai *cluster* didapatkan hasil optimal, selanjutnya dicari jumlah *k cluster* optimal yang dapat dibentuk. Diawali dengan menghitung jumlah *cluster* dengan metode *elbow* sehingga didapatkan nilai yang mendekati 4, maka $k = 4$. Dengan kata lain pengklusteran kecelakaan lalu lintas di Kota Padang akan dibentuk sebanyak 4 *cluster*. *Cluster* ke-1 diprediksikan sebagai tingkat kecelakaan lalu lintas tertinggi pada bulan Oktober. *Cluster* ke-2 merupakan tingkat kecelakaan lalu lintas sedang di Kota Padang bulan Januari-Juni dan September. *Cluster* ke-3 di prediksi merupakan tingkat kecelakaan lalu lintas rendah pada bulan November. *Cluster* ke-4 terdapat 2 anggota dari bulan Agustus dan Juli diprediksikan merupakan tingkat kecelakaan lalu lintas sangat rendah diantara ke-10 bulan lainnya.

Karakteristik dari masing-masing *cluster* memiliki ciri-ciri yang berbeda. Pada *cluster* ke-1 diprediksikan memiliki tingkat kecelakaan lalu lintas tertinggi bulan Oktober tahun 2016 yang menyebabkan korban mengalami kerugian materi dan mengalami luka ringan pada kasus kecelakaan yang dialami, serta pada Oktober tahun 2021 menyebabkan luka ringan, luka berat dan meninggal. Pada hasil data akumulasi menunjukkan bahwasannya jumlah rata-rata kasus kecelakaan lalu lintas tertinggi di bandingkan dengan *cluster* 2, 3, dan 4. Hasil dari *cluster* ke-2 merupakan kecelakaan lalu lintas dengan taraf sedang pada bulan Januari-Juni, September, Desember dimana kecelakaan lalu lintas yang menyebabkan korban meninggal dan jumlah kecelakaan tahun 2016, serta tahun 2018 menyebabkan luka ringan, berat, meninggal dan mengalami kerugian material tahun 2018 dan 2021, dikarenakan memiliki rata-rata yang sedang dan hampir merata pada setiap anggota *clusternya*.

Hasil dari *cluster* ke-3 menunjukkan tingkat kecelakaan lalu lintas dengan taraf rendah pada bulan September dimana terdapat jumlah kecelakaan lalu lintas tahun 2018 dan tahun 2021. Hasil karakteristik dari *cluster* ke-4 menunjukkan tingkat kecelakaan lalu lintas dengan taraf sangat rendah pada bulan Agustus dan Juli, dimana tahun 2016 menyebabkan korban mengalami luka berat serta memiliki nilai rata-rata sangat rendah. Hasil pengklusteran kecelakaan lalu lintas tahun 2016-2021 mempresentasikan gambaran bagi pihak yang berkepentingan khususnya demi meminimalisir kebijakan kecelakaan lalu lintas di Kota Padang yang dikeluarkan sekaligus sebagai bahan pertimbangan guna memetakan kecelakaan lalu lintas di Kota Padang.

4. KESIMPULAN

Hasil yang diperoleh dari penelitian ini yaitu pengoptimalan *cluster* dengan metode *elbow* sebanyak $k = 4$, sehingga didapatkan 4 *cluster* atau kelompok. *Cluster* pertama mencakup kecelakaan lalu lintas sangat tinggi pada bulan Oktober tahun 2016 merupakan kasus kecelakaan banyak menyebabkan luka ringan dan mengalami kerugian material, serta Oktober tahun 2021

menyebabkan banyak luka ringan, luka berat dan meninggal. *Cluster* kedua terdiri dari bulan Januari-Juni, September, dan Desember menunjukkan kasus kecelakaan lalu lintas yang relatif menengah atau sedang tahun 2016 yang menyebabkan korban meninggal dan memiliki jumlah kecelakaann lalu lintas, serta menimbulkan luka ringan luka berat meninggal dan kerugian material pada tahun 2018 dan 2021. *Cluster* ketiga terdiri dari bulan September yang merupakan jumlah kasus kecelakaan lalu lintas terendah di Kota Padang tahun 2018 dan 2021. *Cluster* keempat terdiri dari bulan Agustus dan Juli menunjukkan tingkat kecelakaan lalu lintas dengan taraf sangat redah yang menyebabkan luka berat tahun 2016, dan memiliki nilai rata-rata sangat rendah dibandingkan kecelakaan lalu lintas ke 10 bulan lainnya yang terjadi di Kota Padang.

REFERENSI

- [1] Indonesia. Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan.
- [2] Jenewa,2021.WHO Mulai Kampanye Kurangi Jutaan Kematian dan Cedera Akibat Kecelakaan di Jalan, Beritatrans.com
- [2] Bolla, M. E., Messah, Y., & Koreh, M. B. (2013). Analisis Daerah Rawan Kecelakaan Lalu Lintas (Studi Kasus Ruas Jalan Timor Raya Kota Kupang), *Jurnal Teknik Sipil Vol. II No, 2, 147-156*.
- [3] *Jumlah Korban Kecelakaan Lalu Lintas di Wilayah Poresta Padang Tahun 2016-2021*. Padangkota.bps.go.id.
- [4] Irawan Valdi, Rizal Adhi., & Purnamasari Intan. (2020). "Penerapan Algoritma K-means Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota di Provinsi Jawa Tengah Tahun 2020", "*Jurnal Ilmiah*
- [5] Afnanda, A. (2021). Analisis cluster covid-19 di Sumatera Barat dengan metode Non-Hirarki (k-means). Skripsi. Padang: Universitas Negeri Padang.
- [6] Irwan, Wahidah Sanusi, dan Febriyanto Saman. (2022). "Pengelompokkan Daerah Rawan Kriminalitas di Sulawesi Selatan Menggunakan Metode K-means Clustering", "*Jurnal of Mathematics Vol. V No, 1, 12-19*."
- [7] Sari Mustika, R. (2023). Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Tahun 2020 Menggunakan Analisis K-Means Cluster. Skripsi. Padang: Universitas Negeri Padang.
- [8] Ediyanto, M. MN., & N. Satyahadewi (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Custer Analysis. Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster),2(2),133-136.
- [9] Gunawan, Indra. dkk. Penerapan Algoritma K-Means dalam Pengelompokkan Jumlah Penduduk Berdasarkan Kelurahan di Kota Pematangsiantar, "*Jurnal ilmu Komputer dan Teknologi Vol. 2 No, 2, pp. 20-26*."
- [10] Latifah, L. L., Hudjimartsu, S. A., & Yanuarsyah, I. (2022). Analisis Daerah Rawan Kecelakaan Lalu Lintas Menggunakan Cluster Analysis Di Kota Bogor Berbasis Webgis. *Jurnal Ilmiah Teknologi Infomasi Terapan, 8(2)*. <https://doi.org/10.33197/jitter.vol8.iss2.2022.760>
- [11] Fayyad, U. P.-S. (1996). *Advances in Knowledge Discovery and Data Mining*, MIT Press.
- [12] Johnson, R. A., & Wincher, D. W. (1998). *Applied Multivariate Statistical Analysis*. In Second Edition. New Jersey: Prentice-Hall International.
- [13] Wahyudi, Imam. dkk. (2021). Analisa Penentuan Cluster Terbaik pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi di Pamekasan. "*Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM) Vol. 2 No, 2*."
- [14] Anton, H., & Rosres, C. (2004). *Aljabar Linear Elementer*. Jakarta: Erlangga
- [15] Nugroho, S. (2008). *Statistika Multivariat Terapan*. In UNIB Press Bengkulu