

# Penerapan Data Mining untuk Menentukan Penyebab Kematian di Indonesia Menggunakan Metode Clustering K-Means

Lili Rahmawati\*, Alwis Nazir, Fadhilah Syafria, Elvia Budianita, Lola Oktavia, Ihda Syurfi

Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>11950125106@students.uin-suska.ac.id, <sup>2</sup>alwis.nazir@uin-suska.ac.id, <sup>3</sup>fadhilah.syafria@uin-suska.ac.id,

<sup>4</sup>elvia.budianita@uin-suska.ac.id, <sup>5</sup>lola.oktavia@uin-suska.ac.id, <sup>6</sup>ihdasyurfi11@gmail.com

Email Penulis Korespondensi: 11950125106@students.uin-suska.ac.id

Submitted: 03/03/2023; Accepted: 31/03/2023; Published: 31/03/2023

**Abstrak**—Kematian dalam ilmu kedokteran dipelajari pada disiplin ilmu yang disebut ilmu tanatologi. kematian tidak hanya dialami oleh orang yang berusia lanjut, tetapi juga dapat dialami oleh orang yang masih muda, remaja, atau bahkan bayi. Kematian dapat disebabkan oleh berbagai macam faktor yaitu, karena sakit, usia lanjut, kecelakaan, dan sebagainya. Berdasarkan informasi yang diberikan oleh World Health Organization (WHO), ada lima penyebab kematian tertinggi diantaranya yaitu penyakit jantung iskemik, Alzheimer, stroke, gangguan pernapasan, kondisi neonatal. Dalam penelitian ini k-means digunakan untuk pengelompokan penyebab kematian di Indonesia berdasarkan angka kematian yang terjadi untuk menentukan kasus kematian yang paling berdampak atas tingginya kematian di Indonesia. Dengan mengetahui apa saja kasus kematian ini akan memberikan persiapan dini dalamantisipasi penyebab kematian di Indonesia. Tujuan dilakukannya penelitian ini adalah untuk mengelompokkan angka kematian berdasarkan jumlah penyebab kematian manakah yang termasuk dalam cluster rendah, sedang, dan tinggi dengan menerapkan metode K-Means. Pada penelitian ini penulis menggunakan algoritma K-Means clustering untuk mengelompokkan angka kematian pada data penyebab kematian di Indonesia dari tahun 2017-2021. Hasil dari penelitian ini dibentuk 3 cluster yang telah dievaluasi menggunakan Davies Bouldin Index (DBI) di Rapidminer dengan nilai sebesar 0.259. Hasil clustering dari jumlah keseluruhan 21 Kasus didapatkan cluster tinggi, sedang dan rendah. pengelompokkan cluster ini didapatkan menurut banyaknya angka kematian setiap kasus, yaitu cluster pertama (C0) rendah dengan jumlah 17 Kasus, cluster kedua (C1) sedang dengan jumlah 3 Kasus dan cluster ketiga (C2) tinggi dengan jumlah 1 Kasus.

**Kata Kunci:** Clustering; Data Mining; Kematian; K-means; Rapidminer

**Abstract**—Death in medical science is studied in a scientific discipline called tanatology. death is not only experienced by elderly people, but also can be experienced by young people, teenagers, or even babies. Death can be caused by various factors, namely, due to illness, old age, accidents, and so on. Based on information provided by the World Health Organization (WHO), there are five highest causes of death including ischemic heart disease, Alzheimer's, stroke, respiratory disorders, neonatal conditions. In this study, k-means is used to group causes of death in Indonesia based on the number of deaths that occur to determine the cases of death that have the most impact on the high mortality rate in Indonesia. Knowing what these death cases are will provide early preparation in anticipating the causes of death in Indonesia. The purpose of this study was to classify mortality rates based on the number of causes of death which were included in the low, medium, and high clusters by applying the K-Means method. In this study the authors used the K-Means clustering algorithm to classify death rates in data on causes of death in Indonesia from 2017-2021. The results of this study formed 3 clusters which were evaluated using the Davies Bouldin Index (DBI) in Rapidminer with a value of 0.259. Clustering results from a total of 21 cases obtained high, medium and low clusters. This cluster grouping was obtained according to the number of deaths per case, namely the first cluster (C0) was low with 17 cases, the second cluster (C1) was moderate with 3 cases and the third cluster (C2) was high with 1 case.

**Keywords:** Clustering; Data Mining; Mortality; K-Means; Rapidminer

## 1. PENDAHULUAN

Tanatologi adalah bidang ilmiah yang mempelajari kematian dalam kedokteran [1]. Tanatology adalah disiplin kedokteran forensik yang berfokus pada pemahaman kematian, perubahan postmortem, dan faktor-faktor yang mempengaruhi perubahan tersebut [2],[3],[4].

Ada banyak penyebab kematian. sehingga semua kalangan usia, termasuk anak kecil, remaja, bahkan bayi, dapat mengalami kematian. Ada yang meninggal dunia karena sakit, tua, kecelakaan, dan sebagainya [5]. Wajar jika ada yang meninggal, orang-orang yang ditinggalkan merasakan kehilangan. Untuk menjaga keseimbangan antara penduduk dan wilayahnya, suatu bangsa tentu akan mempertimbangkan jumlah penduduknya. Tingkat kelahiran dan kematian digunakan dalam perhitungan negara bagian.

Berdasarkan informasi yang diberikan oleh World Health Organization (WHO), ada lima penyebab tertinggi kematian diantaranya yaitu penyakit jantung iskemik, Alzheimer yang menyebabkan 814.000 meninggal pada tahun 2019, Stroke sekitar 795.000 orang dewasa, gangguan pernapasan, kondisi neonatal [6]. Hampir 55,3 juta orang di seluruh dunia meninggal setiap tahun. Menurut data di laman Ekologi, ada sekitar dua orang meninggal per detik, 151.600 orang meninggal per hari, 6.316 orang meninggal per jam, dan 105 orang meninggal per menit [7]. Ada beberapa penyebab, seperti penyakit, kecelakaan, dan bunuh diri. Perkiraan populasi Indonesia pada tahun 2019 adalah 266.911,9 ribu, namun antara Januari dan Desember tahun itu, bencana merenggut nyawa 589 orang [8]. Menurut hasil perhitungan, bencana tersebut mengakibatkan kematian 0,22 orang. Ini menunjukkan bahwa dari 100.000 orang yang terkena bencana tidak ada yang meninggal.

Metode k-means digunakan pada penelitian ini untuk analisis data mining dalam pengelompokan penyebab kematian di Indonesia. Dalam penelitian ini k-means digunakan untuk pengelompokan penyebab kematian di Indonesia berdasarkan angka kematian yang terjadi untuk menentukan kasus yang paling berdampak atas tingginya kematian di Indonesia. Dengan mengetahui apa saja kasus kematian ini akan memberikan persiapan dini dalamantisipasi penyebab kematian di Indonesia. Adanya kematian yang bisa diantisipasi seperti penyakit AIDS, Malaria, Demam Berdarah dan penyakit berat lainnya. Untuk kematian yang tidak bisa diantisipasi adalah sebuah bencana alam seperti banjir, longsor, angin puting beliung dan bencana alam yang tidak diketahui sebelumnya.

Untuk mengetahui penyebab kematian di Indonesia, penelitian ini akan melakukan pengelompokan penyebab kematian di Indonesia berdasarkan angka kematian dari rata-rata jumlah kematian laki-laki dan perempuan setiap tahunnya. Data yang didapat oleh peneliti bersumber dari website Kaggle, yang berisi data penyebab kematian di Indonesia yang terdiri dari tiga jenis penyebabnya yaitu jenis penyakit, bencana alam, dan bencana sosial. Dengan penelitian ini, akan memungkinkan untuk menentukan penyebab kematian mana yang memiliki pengaruh terbesar terhadap penurunan populasi Indonesia, yang memungkinkan perencanaan dini untuk potensi penyebab kematian di Indonesia. Teknik data mining kemudian akan digunakan untuk memproses hasil pengelompokan angka kematian. Clustering adalah salah satu metode analisis data mining.

Dalam hal melakukan klasterisasi, Algoritma K-means banyak digunakan oleh peneliti sebelumnya. Penelitian Putra May Chandra Abirianto adalah salah satunya, penelitian mengenai Pengelompokan Pasien Penyakit Liver. Seperti yang ditunjukkan dengan pengujian hasil akhir metode dengan tingkat kesesuaian 86% untuk perbandingan perhitungan sistem dengan perhitungan manual, temuan penelitian ini menunjukkan bahwa metode k-means clustering dapat diterapkan pada sistem dan beroperasi dengan baik [9].

Selanjutnya penelitian yang dilakukan oleh Nelia Vitalaya dan Rizki Tri Prasetyo, mengenai Implementasi Algoritma k-Means clustering Untuk Pengelompokan Penyebaran Pneumonia Pada Balita Di Kota Bandung. Hasil dari penelitian ini dibentuk 3 cluster yang telah di evaluasi Dengan menggunakan Davies Bouldin Index (DBI) yang memiliki nilai 0,571 [10].

Penelitian yang dilakukan Rahmat Kurniawan dan kawan-kawan, pada penelitian ini mereka melakukan analisis Persentase Merokok Penduduk Umur 15 Tahun keatas menurut Provinsi, Berdasarkan hasil analisis perhitungan k-means dengan pengujian tools rapidminer diperoleh hasil yang sama [11]. Selanjutnya penelitian yang dilakukan oleh Siti andry Yani Nasution mengenai Penjualan Frozen Food dengan menerapkan algoritma k-means, hasil akhirnya adalah penjualan Frozen Food dapat diterapkan menggunakan algoritma k-means untuk mengelompokkan Penjualan Frozen Food tertinggi dan terendah. Pengujian menggunakan aplikasi Rapidminer dengan perhitungan yang dilakukan secara manual memiliki nilai yang sama[12].

Penelitian yang sudah dilakukan tentang analisis penyebab kematian. Seperti penelitian yang dilakukan oleh Nengah Widya Utami dan Ni Nyoman Saptiari tentang Klasifikasi Penyebab Kematian Menggunakan Algoritma Support Vector Machine, hasil dari analisa yang telah diteliti yang diperoleh dengan menggunakan metode SVM (Support Vector Machine), adalah penyakit dan bencana non-alam menjadi penyebab terbanyak, terhitung hingga 0,9 persen kasus dengan 20 kasus sampel pengganti [13].

Sedangkan pada penelitian ini, dikembangkan suatu sistem clustering dengan menggunakan metode K-Means yang memiliki komponen 21 kasus kematian diambil dari tahun 2017 sampai 2021. Diharapkan penggunaan metode K-Means dapat mengelompokkan hasil yang tepat mengenaiantisipasi bahaya pada penyebab kematian. Penelitian ini untuk mengelompokkan angka kematian berdasarkan jumlah angka kematian manakah yang termasuk dalam cluster rendah, sedang, dan tinggi. Penelitian ini bertujuan untuk membantu mengetahui kasus kematian apakah yang paling berdampak pada menurunnya jumlah penduduk Indonesia. Dan diharapkan hasil dari penelitian ini akan berguna bagi penduduk Indonesia untuk mengantisipasi kasus kematian yang berdampak tinggi.

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

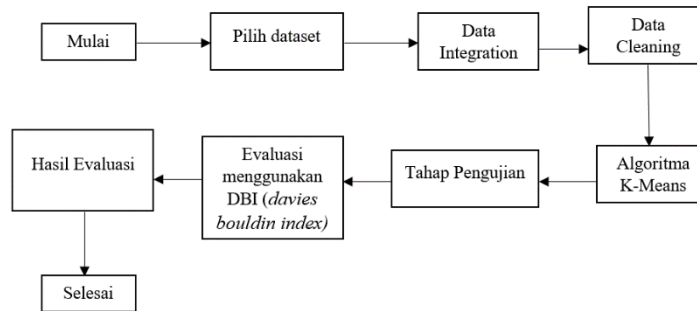
Dalam penelitian ini menggunakan analisa data mining. Data mining merupakan serangkaian proses yang menggunakan statistik, matematika, kecerdasan buatan, dan penelitian basis data untuk menemukan informasi berguna dalam basis data yang tidak dapat ditemukan secara manual [14], [15],[16]. Dengan mengidentifikasi dan mengekstraksi pola yang signifikan atau menarik, informasi yang dihasilkan dapat diperoleh [17]. Clustering adalah salah satu metode analisis yang digunakan dalam data mining.

### 2.2 Clustering

Clustering merupakan pengelompokan berdasarkan atas prinsip kesamaan kelas serta mengurangi kesamaan antar kelas [18],[19],[20]. Dalam pengelompokan, sejumlah algoritma telah dikembangkan dengan kinerja yang baik. Akan ada tiga cluster yang dibentuk dalam penelitian ini: level rendah, sedang, dan tinggi. Algoritma dari clustering yang digunakan penelitian ini yaitu k-means.

### 2.3 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan dalam analisis penyebab kematian menggunakan algoritma k-means, tahapan tersebut sebagai berikut.



**Gambar 1.** Tahapan Penelitian

Tahapan penelitian ini akan menjadi acuan dalam melakukan penelitian ini. Dimana penelitian akan dimulai dengan memilih dataset yaitu data causes of death in indonesia yang kemudian data di integrasi. Dalam tahap integration akan ada penggabungan data dari beberapa basis data menjadi satu basis data baru dikenal sebagai integrasi data. Dalam penelitian ini data penyebab kematian di Indonesia dari tahun 2017 sampai 2021 digabungkan ke dalam satu database. Dataset didapatkan dari sumber kaggle. Setelah data dilakukan data integration, selanjutnya data akan dilakukan proses cleaning. Proses cleaning ini data akan dibersihkan agar dataset bisa digunakan dalam penelitian. Kemudian data akan dilakukan pemodelan menggunakan algoritma k-means dengan nilai  $k = 3$ . Dan tahap selanjutnya akan mendapatkan hasil dari clustering. Kemudian tahap terakhir yaitu melakukan tahap pengujian yang menggunakan software rapidminer dan evaluasi untuk mencari tahu apakah algoritma k-means bisa digunakan dalam memprediksi penyebab kematian berdasarkan nilai Davies-Bouldin Index pada rapidminer.

### 2.4 Pemodelan Algoritma K-Means

Dengan menggunakan algoritma k-means, Langkah pemodelan metode ini, data yang tersedia dihitung nilai rata-ratanya. Teknik non-hierarkis yang disebut k-means berupaya mengelompokkan data ke dalam satu atau beberapa kelompok sehingga yang memiliki kesamaan karakteristik berada dalam kelompok yang sama dan yang memiliki karakteristik berbeda berada dalam kelompok yang berbeda [21], [22]. Data penyebab kematian tahun 2017 hingga 2021 digunakan untuk menghasilkan data uji untuk penelitian ini. Meminimalisasikan fungsi objektif yang diatur dalam proses clustering merupakan tujuan dari clustering k-means, yang pada umumnya untuk meminimalkan variasi antar cluster dan meminimalisasikan variasi dalam suatu cluster [21]. Secara umum metode menggunakan algoritma K-Means sebagai berikut [23] :

- Tentukan  $k$  sebagai jumlah cluster yang di bentuk. Penentuan banyaknya jumlah cluster  $k$  dilakukan dengan beberapa faktor seperti pertimbangan teoritis dan konseptual yang disarankan untuk menentukan berapa banyaknya cluster.
- Pilih secara acak  $k$  centroid pertama. Untuk menentukan centroid awal dilakukan secara acak dari beberapa objek yang tersedia sebanyak  $k$  cluster.

$$v = \sum_k^n = 1 (X_i)/N \tag{1}$$

$$i = 1, 2, 3 \dots n$$

Penjelasan rumus diatas adalah sebagai berikut:

$v$  = nilai centroid pada cluster

$x_i$  = objek ke- $i$ , dimana objek tersebut adalah nilai yang berubah sesuai iterasi

$n$  = banyaknya objek atau jumlah objek yang menjadi anggota cluster

- Hitung jarak setiap objek ke masing- masing centroid dari masing-masing cluster.

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

Penjelasan rumus sebagai berikut :

$d$  = nilai jarak

$x_i$  = objek  $x$  ke- $i$

$y$  = data  $y$  ke- $i$

$n$  = banyaknya objek

- Alokasikan masing-masing objek ke dalam centroid yang paling terdekat.
- Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan.
- Ulangi langkah 5 jika posisi centroid baru tidak sama.

## 2.5 Tahap Evaluasi

Model indeks Davies-Bouldin (DBI) akan digunakan selama fase evaluasi cluster ini. Dengan menggunakan Davies-Bouldin Index, pengukuran ini bertujuan untuk meminimalkan jarak antar titik dalam suatu cluster sekaligus memaksimalkan jarak antar cluster. Cluster yang dibentuk oleh algoritma clustering yang digunakan lebih unggul jika nilai DBI nya lebih kecil atau mendekati nol.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Kumpulan dataset Kaggle berfungsi sebagai sumber informasi yang digunakan dalam penelitian ini. Data tersebut mencakup atribut-atribut berikut antara lain: Causes, Type, Year, Data Redundancy, Jumlah Causes, Total Deaths, Jenis Kelamin, Jumlah Penduduk, Satuan dengan jumlah data yang didapat adalah 613 record. Data yang digunakan dalam penelitian ini adalah data penyebab kematian di Indonesia dari tahun 2017-2021. Algoritma K-means dan perangkat lunak Rapidminer digunakan untuk membuat model aturan dari kumpulan data yang penulis kumpulkan.

**Tabel 1.** Dataset Penelitian

Causes	Angka Kematian Pada Penduduk Indonesia										Rata-rata
	2017		2018		2019		2020		2021		
	L	P	L	P	L	P	L	P	L	P	
AIDS	2	0	0	0	513	101	529	136	397	291	196,9
Angin Siklon Tropis	19	22	0	0	0	1	0	0	127	101	27
Angin Putting Beliung	14	16	3	7	2	2	9	3	1	3	6
Antraks	0	1	0	0	0	0	0	0	0	0	0,1
Aksi Teror dan Sabotase	0	0	10	8	2	1	2	2	1	3	2,9
Avian Influenza (Flu Burung)	0	1	79	89	0	0	318	137	37	44	70,5
Banjir	74	106	16	10	24	19	37	51	63	61	46,1
Banjir Bandang	19	4	7	11	48	61	24	25	62	58	31,9
Banjir disertai Tanah Longsor	21	28	19	24	47	51	32	35	68	38	36,3
Campak	0	1	0	0	7	12	5	3	10	13	5,1
COVID-19	0	0	0	0	0	0	107	1140	6974	5221	14409,4
Demam Berdarah Dengue (DBD)	291	202	301	166	683	236	396	351	483	222	333,1
Diabetes melitus YTT	2919	1578	3918	746	3610	120	421	5422	1684	6825	34707,4
Difteri	24	20	13	16	8	15	2	11	4	21	13,4
Gelombang Pasang	0	0	0	0	0	0	0	0	2	5	0,7
Gempa & Tsunami	0	0	1573	104	227	210	0	0	0	0	305,2
Gempa Bumi	2	3	252	321	29	35	0	0	63	58	76,3
Kebakaran	0	1	7	8	2	1	4	2	25	17	6,7
Kecelakaan Industri	0	0	18	4	0	0	0	0	1	0	2,3
Kecelakaan Lalu Lintas	5	9	123	123	43	60	2	6	91	78	54
Kegagalan Teknologi	0	0	3	2	0	0	1	0	3	4	1,3
Keracunan	4	8	27	18	5	5	1	1	4	2	7,5
KLB Campak	8	6	0	0	7	12	5	3	10	13	6,4
KLB Diare	28	6	14	22	0	0	0	0	0	0	7
Konflik	0	0	2	5	1	3	0	0	0	0	1,1
Kronis Filariasis	0	0	0	0	218	207	327	564	428	189	193,3
Leptospirosis	82	54	74	76	73	49	80	26	37	47	59,8
Letusan Gunung Berapi	37	34	0	0	0	0	0	0	37	14	12,2
Malaria	192	207	242	224	281	249	0	0	0	0	139,5
Neoplasma	1528	6291	1783	444	1693	602	197	3771	1925	1061	13636
Pneumonia	92	8	25	07	15	09	46	68	02	37	1,9
Rabies - Lyssa	0	0	0	0	0	0	0	0	0	0	0
Tanah Longsor	17	91	50	61	83	32	28	12	37	25	43,6
Tetanus Neonatorium	73	90	19	4	18	16	3	17	27	23	29
Tsunami	9	5	3	1	1	1	2	0	4	5	3,1
Tuberkulosis paru	1	1	259	172	0	0	0	0	0	0	43,3
	8433	1075	0	0	5839	615	625	6916	7251	6897	15531,5

Causes	Angka Kematian Pada Penduduk Indonesia										Rata-rata
	2017		2018		2019		2020		2021		
	L	P	L	P	L	P	L	P	L	P	
Tersambar Petir	0	0	1642	221 0	0	0	0	0	0	0	385,2

### 3.2 Penerapan algoritma K-means

Berikut adalah langkah-langkah dari algoritma k-means clustering:

- 1) Tentukan data yang ingin di cluster. Data Penyebab kematian di Indonesia dari tahun 2017-2021 yang digunakan pada proses clustering dengan menggunakan data sebanyak 21 Causes.

**Tabel 2.** Data Sample

Causes	Angka Kematian Pada Penduduk Indonesia										Rata-rata
	2017		2018		2019		2020		2021		
	L	P	L	P	L	P	L	P	L	P	
AIDS	2	0	0	0	513	101	529	136	397	291	196,9
Angin Siklon Tropis	19	22	0	0	0	1	0	0	127	101	27
Angin Putting Beliuang	14	16	3	7	2	2	9	3	1	3	6
Aksi Teror dan Sabotase	0	0	10	8	2	1	2	2	1	3	2,9
Avian Influenza (Flu Burung)	0	1	79	89	0	0	318	137	37	44	70,5
Banjir	74	106	16	10	24	19	37	51	63	61	46,1
Banjir Bandang	19	4	7	11	48	61	24	25	62	58	31,9
Banjir disertai Tanah Longsor	21	28	19	24	47	51	32	35	68	38	36,3
Campak	0	1	0	0	7	12	5	3	10	13	5,1
Demam Berdarah Dengue (DBD)	291	202	301	166	683	236	396	351	483	222	333,1
Difteri	24	20	13	16	8	15	2	11	4	21	13,4
Gempa Bumi	2	3	252	321	29	35	0	0	63	58	76,3
Kebakaran	0	1	7	8	2	1	4	2	25	17	6,7
Kecelakaan Industri	0	0	18	4	0	0	0	0	1	0	2,3
Kecelakaan Lalu Lintas	5	9	123	123	43	60	2	6	91	78	54
Keracunan	4	8	27	18	5	5	1	1	4	2	7,5
Kronis Filariasis	0	0	0	0	218	207	327	564	428	189	193,3
Leptospirosis	82	54	74	76	73	49	80	26	37	47	59,8
Malaria	192	207	242	224	281	249	0	0	0	0	139,5
Rabies - Lyssa	17	91	50	61	83	32	28	12	37	25	43,6
Tanah Longsor	73	90	19	4	18	16	3	17	27	23	29

- 2) Mengembangkan nilai k sebagai pusat cluster awal angka kematian sebanyak 3 cluster. Berikut hasil centroid awal :

**Tabel 3.** Centroid Awal

C0	2,3
C1	196,9
C2	333,1

Dari tabel.2 tersebut, dapat dilihat untuk penentuan centroid awal dilakukan secara acak. masing-masing titik awal Cluster memiliki tingkatan yakni tingkat kasus yang rendah, sedang, tinggi.

- 3) Menghitung jarak/selisih antara setiap cluster dengan setiap data angka kematian untuk menentukan jarak mana yang paling dekat dengan centroid awal setiap data. Berikut adalah perhitungan cara mendapatkan jarak antara setiap data dengan cluster pertama (C0) :

$$D(0.1) = \sqrt{(196,9 - 2,3)^2} = 194,6$$

$$D(0.2) = \sqrt{(27 - 2,3)^2} = 24,7$$

$$D(0.3) = \sqrt{(6 - 2,3)^2} = 3,7$$

$$D(0.4) = \sqrt{(2,9 - 2,3)^2} = 0,6$$

Sampai perhitungan D (0.21)

Mengukur selisih antara setiap titik data dan cluster kedua (C1) :

$$D(1.1) = \sqrt{(196,9 - 196,9)^2} = 0$$

$$D(1.2) = \sqrt{(27 - 196,9)^2} = 169,9$$

$$D(1.3) = \sqrt{(6 - 196,9)^2} = 190,9$$

$$D(1.4) = \sqrt{(2,9 - 196,9)^2} = 194$$

Sampai perhitungan D (1.21)

Mengukur selisih antara setiap titik data dan cluster ketiga (C2) :

$$D(2.1) = \sqrt{(196,9 - 333,1)^2} = 136,2$$

$$D(2.2) = \sqrt{(27 - 333,1)^2} = 306,1$$

$$D(2.3) = \sqrt{(6 - 333,1)^2} = 327,1$$

$$D(2.4) = \sqrt{(2,9 - 333,1)^2} = 330,2$$

Sampai perhitungan D (2.21)

Berikut hasil perhitungan jarak/selisih antara setiap titik data dengan centroid awal iterasi 1 di bawah ini:

**Tabel 4.** Hasil Perhitungan Jarak Data dengan centroid awal

No	Causes	C0	C1	C2	Jarak Terpendek	Cluster
1	AIDS	194,6	0	136,2	0	1
2	Angin Siklon Tropis	24,7	169,9	306,1	24,7	0
3	Angin Putting Beliuang	3,7	190,9	327,1	3,7	0
4	Aksi Teror dan Sabotase	0,6	194	330,2	0,6	0
5	Avian Influenza (Flu Burung)	68,2	126,4	262,6	68,2	0
6	Banjir	44	150,8	287,0	44,0	0
7	Banjir Bandang	29,6	165,0	301,2	29,6	0
8	Banjir disertai Tanah Longsor	34,0	161	297	34	0
9	Campak	2,8	191,8	328,0	2,8	0
10	Demam Berdarah Dengue (DBD)	3331	136,2	0,0	0	2
11	Difteri	11,1	183,5	319,7	11,1	0
12	Gempa Bumi	74,0	120,6	257	74	0
13	Kebakaran	4,4	190,2	326,4	4,4	0
14	Kecelakaan Industri	0,0	194,6	331	0	0
15	Kecelakaan Lalu Lintas	51,7	142,9	279,1	51,7	0
16	Keracunan	5,2	189,4	325,6	5,2	0
17	Kronis Filariasis	191,0	3,6	140	3,6	1
18	Leptospirosis	57,5	137,1	273,3	57,5	0
19	Malaria	137,2	57,4	193,6	57,4	1
20	Rabies - Lyssa	41,3	153,3	289,5	41,3	0
21	Tanah Longsor	26,7	167,9	304,1	26,7	0

- 4) Kelompokkan data sesuai dengan seberapa dekat pusat jarak masing-masing data. Dapat dilihat tabel berikut dan hasilnya, jika nilai terendah (C0) berada di cluster 1 maka masuk ke dalam kelompok tersebut, begitu juga sebaliknya:

**Tabel 5.** Output Cluster Iterasi 1

cluster	Hasil
C0	17
C1	3
C2	1

- 5) Menentukan nilai centroid baru dilakukan setelah hasil iterasi 1 cluster diperoleh.  
 6) Mengulangi langkah 3 - 5 hingga anggota setiap cluster tidak mengalami perubahan.  
 Perhitungan cara menentukan nilai centroid baru untuk iterasi 2 ditunjukkan di bawah ini:

$$C0 = \frac{27 + 6 + 2,9 + 70,5 + 46,1 + 31,9 + 36,3 + 5,1 + 13,4 + 76,3 + 6,7 + 2,3 + 54 + 7,5 + 59,8 + 43,6 + 29}{17} = 30,49$$

$$C1 = \frac{196,9 + 193,3 + 139,5}{3} = 176,57$$

$$C2 = \frac{333,1}{1} = 333,1$$

**Tabel 6.** Centroid Baru Iterasi 2

Cluster	Nilai
C0	30,49
C1	176,57
C2	333,1

Penghitungan ulang selisih setiap data angka kematian untuk setiap pusat klaster dilakukan selanjutnya setelah mendapatkan centroid baru. Perhitungan selisih setiap data cluster pertama (C0):



$$E(0.1) = \sqrt{(196,9 - 30,49)^2} = 166,4$$

$$E(0.2) = \sqrt{(27 - 30,49)^2} = 3,5$$

$$E(0.3) = \sqrt{(6 - 30,49)^2} = 24,5$$

$$E(0.4) = \sqrt{(2,9 - 230,49)^2} = 27,6$$

Sampai perhitungan E (0.21)

Mengukur selisih antara setiap titik data dan cluster kedua (C1) :

$$E(1.1) = \sqrt{(196,9 - 176,57)^2} = 20,3$$

$$E(1.2) = \sqrt{(27 - 176,57)^2} = 149,6$$

$$E(1.3) = \sqrt{(6 - 176,57)^2} = 170,6$$

$$E(1.4) = \sqrt{(2,9 - 176,57)^2} = 173,7$$

Sampai perhitungan E (1.21)

Mengukur selisih antara setiap titik data dan cluster ketiga (C2) :

$$E(2.1) = \sqrt{(196,9 - 333,1)^2} = 136,2$$

$$E(2.2) = \sqrt{(27 - 333,1)^2} = 306,1$$

$$E(2.3) = \sqrt{(6 - 333,1)^2} = 327,1$$

$$E(2.4) = \sqrt{(2,9 - 333,1)^2} = 330,2$$

Sampai perhitungan E (2.21)

Berikut hasil perhitungan jarak/selisih antara setiap data dengan titik pusat iterasi 2 di bawah ini:

**Tabel 7.** Perhitungan Jarak/selisih Data Iterasi 2

No	Causes	C0	C1	C2	Jarak Terpendek	Cluster
1	AIDS	166,4	20,3	136,2	20,3	1
2	Angin Siklon Tropis	3,5	149,6	306,1	3,5	0
3	Angin Putting Beliuang	24,5	170,6	327,1	24,5	0
4	Aksi Teror dan Sabotase	27,6	173,7	330,2	27,6	0
5	Avian Influenza (Flu Burung)	40,0	106,1	262,6	40,0	0
6	Banjir	15,6	130,5	287,0	15,6	0
7	Banjir Bandang	1,4	144,7	301,2	1,4	0
8	Banjir disertai Tanah Longsor	5,8	140,3	297	5,8	0
9	Campak	25,4	171,5	328,0	25,4	0
10	Demam Berdarah Dengue (DBD)	302,6	156,5	0,0	0,0	2
11	Difteri	17,1	163,2	319,7	17,1	0
12	Gempa Bumi	45,8	100,3	257	45,8	0
13	Kebakaran	23,8	169,9	326,4	23,8	0
14	Kecelakaan Industri	28,2	174,3	331	28,2	0
15	Kecelakaan Lalu Lintas	23,5	122,6	279,1	23,5	0
16	Keracunan	23,0	169,1	325,6	23,0	0
17	Kronis Filariasis	162,8	16,7	140	16,7	1
18	Leptospirosis	29,3	116,8	273,3	29,3	0
19	Malaria	109,0	37,1	193,6	37,1	1
20	Rabies - Lyssa	13,1	133,0	289,5	13,1	0
21	Tanah Longsor	1,5	147,6	304,1	1,5	0

Output dari perhitungan data cluster iterasi 2 adalah:

**Tabel 8.** Output Cluster Iterasi 2

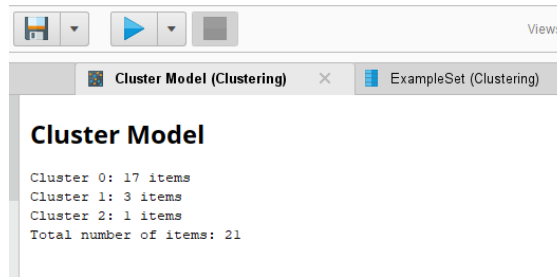
Cluster	Hasil
C0	17
C1	3
C2	1

Setelah diperoleh hasil iterasi 2, bandingkan dengan hasil iterasi 1. Jika tidak ada perbedaan, proses berakhir. jika ada, proses diulang untuk menghasilkan hasil yang sama. Berdasarkan perhitungan manual data angka kematian yang dilakukan diatas, hasil dari iterasi 2 memiliki nilai yang sama, yaitu C0 = 17, C1 = 3, dan C2 = 1. Jika posisi iterasi 2 tidak berubah, maka proses berhenti.

### 3.3 Hasil Pengujian

Menyesuaikan hasil perhitungan manual dengan Rapidminer 10.0 merupakan langkah selanjutnya setelah mendapatkan hasil akhir. Bahwa disimpulkan data yang digunakan adalah benar. Hal ini ditunjukkan dengan

evaluasi perhitungan manual dan menggunakan Rapidminer 10.0, keduanya menghasilkan hasil yang sama. Gambar.2 di bawah ini menunjukkan tampilan cluster model yang ada di Rapidminer:

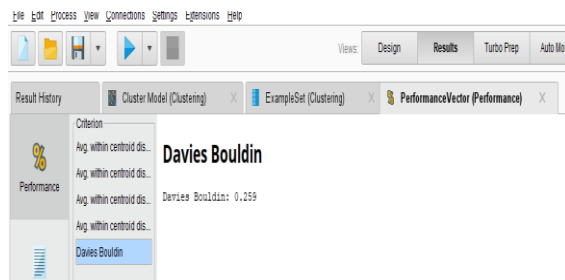


**Gambar 2.** Cluster Model Data Angka Kematian

Gambar diatas menunjukkan bahwa kasus kematian di Indonesia ada 17 kasus kelompok terendah, 3 kasus kelompok sedang, 1 kasus kelompok tertinggi yang menggunakan data sample 21 kasus.

### 3.4 Evaluasi

Pada tahap ini melakukan uji kelayakan apakah algoritma k-means bisa digunakan sebagai algoritma yang tepat pada penelitian ini. Untuk mengetahui hal itu, maka diperlukan software rapidminer. Dalam penelitian ini menggunakan Indeks Davies Bouldin untuk mengevaluasi data. Hasil dari pengujian ini adalah sebagai berikut:



**Gambar 3.** Nilai DBI (Davies Bouldin Index) Pada Rapidminer

Dilihat dari gambar di atas hasil evaluasi cluster pada penelitian ini membentuk tiga cluster dengan nilai 0.259.  $k = 3$  adalah nilai yang paling mendekati 0.

### 3.5 Pembahasan

Dalam analisis kematian di Indonesia, bisa dilihat dari model algoritma k-means yang sudah dibuat. Kualitas hasil cluster yang dihasilkan melalui proses clustering K-Means ditentukan dengan mengevaluasi hasil cluster. Indeks Davies Bouldin (DBI) digunakan dalam penelitian ini untuk mengevaluasi data. Langkah evaluasi aplikasi rapidminer adalah sebagai berikut:

Langkah pertama masukkan cluster distance performance pada panel operators, Setelah itu sambungkan konektor untuk setiap proses dan tekan button running. Performance vector akan menampilkan hasil evaluasi indeks davies-bouldin setelah proses selesai dijalankan, hasil evaluasi cluster pada penelitian ini membentuk tiga cluster dengan nilai 0.259.

Dengan itu, maka algoritma k-means dikatakan baik dilakukan dalam penelitian ini karena pengelompokan data berdasarkan nilai-nilai ini dalam percobaan evaluasi mulai dari  $k = 2$  sampai  $k = 4$ ,  $k = 3$  adalah nilai yang paling mendekati 0. Dengan demikian, penelitian ini sejalan dengan penelitian Nelia Vitalaya dan Rizki Tri Prasetyo [10], bahwa penelitian ini sama-sama melakukan cara atau metode yang sama dengan topik yang berbeda. Pada penelitian mereka algoritma k-means dikatakan baik.

## 4. KESIMPULAN

Berdasarkan penelitian yang dilakukan menggunakan teknik data mining untuk menganalisis kematian di Indonesia dengan menggunakan teknik clustering k-means. Menghasilkan bahwa hasil cluster 0, 1, dan 2 yang menggunakan 21 data uji yang menggunakan perhitungan manual dan aplikasi rapidminer adalah sama yaitu, Kelompok kasus terendah pada cluster 0 terdiri dari 17 kasus, yaitu Angin Siklon Tropis, Angin Putting Beliuang, Aksi Teror dan Sabotase, Avian Influenza(Flu Burung), Banjir, Banjir Bandang, Banjir disertai Tanah Longsor, Campak, Difteri, Gempa Bumi, Kebakaran, Kecelakaan Industri, Kecelakaan Lalu Lintas, Keracunan, Leptospirosis, Rabies-Lyssa, Tanah Longsor. Kelompok kasus sedang pada cluster 1 terdiri dari 3 kasus yaitu Aids, Kronis Filariasis, Malaria. Kelompok kasus tertinggi pada cluster 2 terdiri dari 1 kasus yaitu Demam Berdarah Dengue (DBD). Telah dinilai baik berdasarkan evaluasi Indeks Davies Bouldin yang membentuk 3



cluster, dikarenakan dalam percobaan evaluasi dari pembentukan jumlah  $k=2$  sampai  $k=4$ ,  $k=3$  yang nilainya mendekati 0 yaitu dengan nilai sebesar 0.259. Dengan demikian dapat dikatakan bahwa metode k-means sangat cocok digunakan untuk menganalisis kematian di Indonesia. Untuk itu disarankan kepada penduduk Indonesia untuk lebih waspada terhadap lingkungan sekitar dan mengantisipasi gejala-gejala penyakit yang ada. Disarankan penelitian selanjutnya untuk menambahkan atribut lainnya dalam melakukan analisis kematian di Indonesia. Dan lebih baik lagi untuk menambahkan jumlah data pada dataset yang akan dilakukan penelitian agar hasil analisis kematian di Indonesia semakin akurat.

## REFERENCES

- [1] I. Aflanier, N. Nirmalasari, and M. H. Arizal, Ilmu Kedokteran Forensik dan Medikolegal. 2017.
- [2] Ninla Elmawati Falabiba, "Modul Forensik Tanatologi," Univ. Andalas Padang, no. perubahan pada mayat paska mati, pp. 1-12, 2019.
- [3] Erza, "Fakultas Kedokteran Universitas Andalas," J. Fak. Kedokt. Univ. Andalas 1, vol. 2007, pp. 6-9, 2019, [Online]. Available: [http://scholar.unand.ac.id/61716/2/2. BAB 1 \(Pendahuluan\).pdf](http://scholar.unand.ac.id/61716/2/2. BAB 1 (Pendahuluan).pdf).
- [4] A. G. Parinduri, "Buku ajar kedokteran forensik & medikolegal pedoman bagi mahasiswa kedokteran," pp. 27-32, 2020, [Online]. Available: <http://umsupress.umsu.ac.id/>.
- [5] R. F. Ramadhan and W. S. Ardias, "Konstrual diri (," no. April, pp. 79-90, 2019.
- [6] A. P. Anggraini, "5 Penyebab Tertinggi Kematian Menurut Data WHO," 12 Agustus, 2021. <https://health.kompas.com/read/2021/08/12/120700868/5-penyebab-tertinggi-kematian-menurut-data-who?page=all>.
- [7] N. Zakiah, "7 Penyakit Penyebab Kematian Tertinggi di Indonesia," 15 Maret, 2020. <https://www.idntimes.com/health/medical/nena-zakiah-1/penyakit-penyebab-kematian-tertinggi-di-indonesia>.
- [8] Badan Pusat Statistik, "Jumlah Korban Meninggal, Hilang, dan Terluka Terkena Dampak Bencana Per 100.000 Orang." [https://www.bps.go.id/indikator/indikator/view\\_data/0000/data/1246/sdgs\\_13/1](https://www.bps.go.id/indikator/indikator/view_data/0000/data/1246/sdgs_13/1).
- [9] P. M. C. Abrianto, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Pasien Penyakit Liver," JATI (Jurnal Mhs. Tek. Inform., vol. 2, no. 2, pp. 247-255, 2018.
- [10] N. Vitalaya and R. T. Prasetyo, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Pneumonia Pada Balita Di Kota Bandung," POTENSI (eProsiding Sist. Informasi), vol. 1, no. 1, pp. 108-116, 2020, [Online]. Available: <http://eprosiding.ars.ac.id/index.php/psi%0Ahttp://eprosiding.ars.ac.id/index.php/psi/article/view/291>.
- [11] R. Kurniawan, & S., and R. Dewi, "Penerapan Algoritma K-Means Clustering Dalam Persentase Merokok Pada Penduduk Umur Di Atas 15 Tahun Menurut Provinsi," J. Sist. Komput. dan Inform., vol. 2, no. 2, pp. 178-186, 2021, doi: 10.30865/json.v2i2.2770.
- [12] S. A. Y. Nasution, P. Poningsih, and ..., "Penerapan Algoritma K-Means Pada Penjualan Frozen Food Pada UD Soise Sosis Pematangsiantar," J. Sist. ..., vol. 2, pp. 171-177, 2021, doi: 10.30865/json.v2i2.2768.
- [13] N. W. S. Utami Ni Nyoman, "Penerapan Data Mining Untuk Klasifikasi Penyebab Kematian Menggunakan Algoritma Support Vector Machine," J. Ilm. Ilmu Terap. Univ. Jambi[JITIJ], vol. 4, no. Vol. 4 No. 2 (2020): Volume 4, Nomor 2, Desember 2020, pp. 234-240, 2020, [Online]. Available: <https://online-journal.unja.ac.id/JITIJ/article/view/13268/11199>.
- [14] D. Abdullah and D. Saputra, "Aplikasi Penjadwalan Pengadaan Barang Menggunakan Algoritma Apriori," no. January, 2016.
- [15] R. Yani, A. Nazir, M. Affandes, R. Mai Candra, and A. Akhyar, "Implementasi Data Mining Untuk Menemukan Pola Asosiasi Data Tracer Study Menggunakan Algoritma Apriori," J. Nas. Komputasi dan Teknol. Inf., vol. 5, no. 3, pp. 383-390, 2022, [Online]. Available: <https://ojs.serambimekka.ac.id/jnkti/article/view/4412>.
- [16] A. Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," J. Ilmu Komput., pp. 1-12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>.
- [17] D. Merawati and Rino, "Penerapan data mining penentu minat Dan bakat siswa Smk dengan metode C4 . 5," J. Algor, vol. 1, no. 1, pp. 28-37, 2019.
- [18] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," JUITA J. Inform., vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [19] D. Darmansah, "Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 7, no. 2, pp. 126-134, 2020, doi: 10.35957/jatisi.v7i2.309.
- [20] S. Defiyanti, M. Jajuli, and N. Rohmawati, "Optimalisasi K-MEDOID dalam Pengklasteraan Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION," J. Nas. Teknol. dan Sist. Inf., vol. 3, no. 1, pp. 211-218, 2017, doi: 10.25077/teknosi.v3i1.2017.211-218.
- [21] A. A. Lesmana et al., "Implementasi Algoritma K-Means Untuk Clustering Penyakit Hiv / Aids Di Indonesia Implementation of K-Means Algorithm for Clustering of Hiv / Aids Disease in Indonesia," vol. 6, no. 2, pp. 5564-5580, 2019.
- [22] R. Helilintar, I. N. Farida, and R. H. Irawan, "Penerapan Metode K-Means Clustering Pada Data Penerimaan Mahasiswa Baru," Semin. Nas. Teknol. Inf. dan Komunikasi SENATIK[2018] Literasi Digit. pada Era Revolusi Ind. 4.0, pp. 14-20, 2018.
- [23] R. NOVIANTO, "Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 6, no. 1, pp. 85-95, 2019, doi: 10.35957/jatisi.v6i1.150.