

## PENGELOMPOKKAN DATA BENCANA ALAM BERDASARKAN WILAYAH MENGGUNAKAN ALGORITMA K-MEANS

Rif'atul Amaliah<sup>1</sup>, Edi Tohidi<sup>2</sup>, Edi Wahyudin<sup>3</sup>, Ade Rizki Rinaldi<sup>4</sup>, Iin<sup>5</sup>

<sup>1</sup> Teknik Informatika, STMIK IKMI Cirebon

<sup>2,3</sup> Komputerisasi Akuntansi, STMIK IKMI Cirebon

<sup>4,5</sup> Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia

rifatulamaliah5@gmail.com

### ABSTRAK

Bencana alam adalah rangkaian kejadian yang ditimbulkan oleh alam. Namun tidak menutup kemungkinan bencana alam yang terjadi disebabkan oleh ulah manusia. Bencana alam memiliki potensi untuk menyebabkan dampak yang signifikan pada infrastruktur dan lingkungan, bahkan dapat mengancam keselamatan manusia. Penelitian ini bertujuan untuk mengidentifikasi wilayah-wilayah dimana jenis bencana alam paling dominan atau sering terjadi untuk memberikan wawasan yang lebih mendalam tentang resiko bencana alam di berbagai wilayah sehingga dapat membantu meningkatkan upaya mitigasi bencana. Metode yang diterapkan melibatkan pengelompokan data menggunakan algoritma K-Means, yang memungkinkan peneliti untuk mengelompokkan wilayah dengan karakteristik serupa ke dalam *cluster-cluster* yang berbeda. Hasil penelitian ini berhasil mengelompokkan data bencana alam menggunakan Algoritma K-Means membentuk 3 *cluster* berdasarkan *MeasureTypes Bregman Divergences* dengan parameter *SequeredEuclideanDistance* sampai menghasilkan nilai DBI terbaik. Nilai *Davies-Bouldin-Index* (DBI) yang dihasilkan pada penelitian ini yaitu 0,012. *Cluster 0* memiliki 884 data dengan tingkat kejadian bencana alam rendah. Kejadian bencana alam yang dominan terjadi yaitu kekeringan. Sedangkan wilayah yang termasuk kedalam *cluster 1* dengan memiliki 2 data adalah wilayah dengan tingkat kejadian sedang. Kejadian bencana yang paling dominan adalah Letusan Gunung Api. Dan *cluster 2* memiliki 1 data yaitu wilayah dengan tingkat kejadian tinggi. Kejadian bencana alam yang paling dominan adalah Puting Beliung.

**Kata kunci:** *Clustering, Algoritma K-Means, Bencana Alam*

### 1. PENDAHULUAN

Indonesia merupakan negara yang rawan akan bencana baik bencana alam maupun non alam, bencana alam merupakan ancaman yang dapat mengganggu kehidupan dan penghidupan masyarakat. Dampak yang terjadi akibat bencana alam diantaranya banyak menelan korban jiwa, rusaknya lingkungan hidup, fasilitas dan infrastruktur, serta kehilangan barang berharga [1]. Teknologi informatika memungkinkan peningkatan kualitas pengambilan keputusan berdasarkan analisis data dalam kerangka pengelolaan lingkungan. Data yang terhimpun dan diolah memiliki potensi untuk mendukung pembuatan kebijakan yang lebih optimal dan efisien oleh pihak-pihak yang kepentingan dalam mengelola sumber daya alam serta merespons peristiwa bencana alam. Pentingnya pemanfaatan sistem informasi layanan berbasis digital juga diungkapkan oleh Nurdyawan dkk, bahwasannya Pelayanan online bertujuan untuk meningkatkan keterjangkauan warga negara terhadap layanan publik pemerintah, memperluas akses masyarakat terhadap sumber informasi yang dimiliki oleh pemerintah, menangani pengaduan masyarakat, serta menjamin keseragaman kualitas pelayanan yang dapat dinikmati oleh seluruh warga negara [2]. Pemanfaatan teknologi website sebagai medium online juga digunakan untuk memperluas penyebaran informasi [3]. Hal ini tentu akan memudahkan

masyarakat dalam mengakses informasi mengenai bencana alam yang terjadi.

Bencana alam memiliki potensi untuk menyebabkan dampak yang signifikan pada infrastruktur dan lingkungan, bahkan dapat mengancam keselamatan manusia. Keberagaman karakteristik bencana di wilayah Indonesia memberikan tantangan dalam merancang strategi mitigasi yang tepat dan disesuaikan dengan karakteristik unik masing-masing daerah tersebut.

Dalam implementasinya, Algoritma K-Means memerlukan tiga parameter yang harus ditentukan oleh pengguna, yakni jumlah kluster ( $k$ ), inisialisasi kluster, dan ukuran jarak sistem. Algoritma ini bertujuan untuk mengelompokkan sejumlah data atau objek kedalam kelompok atau *cluster*, sehingga setiap *cluster* berisi data yang memiliki kemiripan [4]. Uji kinerja K-Means clustering menggunakan *Davies-Bouldin Index* juga telah dilakukan untuk mengelompokan data prestasi siswa. Apabila nilai DBI tidak mendekati nol, maka *cluster cluster* tersebut di anggap belum optimal [5]. *Knowledge Discovery in Database* (KDD) digunakan sebagai proses pengolahan data mining yang mencakup pengumpulan data dan pemanfaatan data historis untuk menemukan pola, keteraturan, dan hubungan dalam kumpulan data yang berskala besar. Hasil dari data *mining* dapat digunakan untuk meningkatkan kualitas pengambilan keputusan di masa mendatang [6]. Data mining perlu di standarisasi

menggunakan tahapan-tahapan agar data layak diolah [7].

Tujuan dari penelitian ini yaitu memberikan hasil informasi dalam bentuk pola pengelompokan daerah rawan bencana mulai dari tahun 2021-2022 yang akan dijadikan sebagai acuan di dalam pengambilan keputusan pihak-pihak terkait untuk mengidentifikasi wilayah-wilayah dimana jenis bencana alam paling dominan atau sering terjadi. Hasil penelitian ini memiliki implikasi yang signifikan bagi sejumlah pihak terkait penanggulangan bencana di Indonesia. Informasi yang dihasilkan dari pengelompokan daerah dengan menggunakan Algoritma K-Means *clustering* dapat disampaikan kepada pemerintah daerah, lembaga kemanusiaan, dan peneliti dibidang mitigasi bencana. Para pembuat kebijakan dapat menggunakan temuan ini sebagai dasar untuk merancang strategi mitigasi yang lebih spesifik dan sesuai dengan karakteristik setiap wilayah.

Pada penelitian ini, *Knowledge Discovery in Database* (KDD) digunakan untuk menganalisis data bencana alam pada tahun 2021-2022 dengan total 9.991 data. Data Mining merupakan konsep yang digunakan untuk mengeksplorasi pengetahuan di dalam kumpulan data besar. Tujuannya adalah mengidentifikasi informasi yang bernilai melalui penerapan teknik statistik, matematika, kecerdasan buatan, dan machine learning pada dataset yang memiliki skala besar [8]. Metode *Data Mining* yang diterapkan pada penelitian ini adalah Algoritma K-Means *clustering* dengan *tools* RapidMiner untuk mengidentifikasi pola atau kelompok yang signifikan dalam dataset bencana alam. Untuk mengevaluasi hasil dari proses pengelompokan yang dilakukan oleh Algoritma K-Means, digunakan *Davies Bouldin Index* (DBI).

Hasil penelitian ini memiliki implikasi yang signifikan bagi sejumlah pihak terkait penanggulangan bencana di Indonesia. Informasi yang dihasilkan dari pengelompokan daerah dengan menggunakan Algoritma K-Means *clustering* dapat disampaikan kepada pemerintah daerah, lembaga kemanusiaan, dan peneliti dibidang mitigasi bencana.

## 2. TINJAUAN PUSTAKA

Penelitian menurut Erliyana dkk yang berjudul "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana" Data Mining merupakan suatu tahapan ekstraksi atau penambangan data yang sebelumnya tidak diketahui, namun dapat dipahami dan memiliki nilai signifikan dari suatu database yang memiliki volume besar. Proses ini digunakan untuk mendapatkan wawasan atau pengetahuan baru, sering disebut sebagai "Data atau *Knowledge Discovery*" yang bertujuan untuk mengidentifikasi pola yang tersembunyi dalam data. Penggunaan mobil di Indonesia mengalami peningkatan setiap tahunnya, mendorong perusahaan otomotif untuk bersaing dalam meningkatkan penjualan. Dalam penelitian ini, dilakukan analisis

data mining dengan menerapkan Teknik *Clustering* menggunakan Algoritma K-Means untuk mengelompokkan data penjualan ke dalam sebuah cluster. Hasilnya, metode *Clustering* memiliki potensi untuk mendukung perusahaan otomotif dalam mengevaluasi kinerja penjualan mobil mereka dengan perbandingan terhadap pesaing. Hasil pengujian menggunakan perangkat lunak RapidMiner v9.6 menunjukkan bahwa *Cluster 0* memiliki 90 anggota (Tingkat Penjualan Tertinggi), *Cluster 1* memiliki 88 anggota (Penjualan Cukup Baik), dan *Cluster 2* memiliki 89 anggota (Tingkat Penjualan Rendah) [9].

Dalam penelitian yang dilakukan Herviany dkk dengan judul "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat" Data Mining merujuk pada kegiatan ekstraksi informasi dari dataset yang besar dengan tujuan mengungkapkan pengetahuan tersembunyi dan memfasilitasi penerapannya secara *real-time* dalam aplikasi. Data mining melibatkan berbagai algoritma untuk menganalisis data. Dalam penelitiannya, Herviany dkk membahas mengenai perbandingan antara algoritma K-Means dan metode K-Medoids untuk mengidentifikasi daerah yang rentan terhadap bencana tanah longsor di Provinsi Jawa Barat pada tahun 2019. Tujuan penelitian ini adalah untuk memberikan panduan yang lebih efektif dalam penanggulangan bencana tanah longsor di wilayah tersebut. Hasil penelitian menunjukkan bahwa metode K-Means lebih optimal daripada K-Medoids, dengan jumlah kluster yang paling optimal adalah 6. Cluster 2 merupakan yang paling dominan, dan kluster 5 memiliki jumlah daerah dan kejadian tanah longsor terbanyak dengan jumlah 4 daerah dan jumlah kejadian sebanyak 106 kejadian [10].

Penelitian menurut Hermanto dan Muhyidin yang berjudul "Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana" mengungkapkan bahwa Algoritma K-Means merupakan suatu model yang digunakan untuk membentuk klaster untuk menilai sejauh mana jarak yang ada dengan kumpulan data. Peningkatan jumlah bencana di Kabupaten Purwakarta mengundang perlunya penelitian lebih lanjut guna mengurangi dampak terhadap korban. Dalam konteks data bencana alam, penggunaan metode *clustering* pada data mining menjadi krusial untuk mengelompokkan data bencana berdasarkan karakteristik yang serupa. Pendekatan ini dapat menjadi dasar untuk mengklasifikasikan kemungkinan kejadian bencana di masa depan. Hasil analisis data menggunakan metode K-Means Clustering menunjukkan bahwa kekeringan dan krisis air bersih termasuk dalam cluster 0. Selain itu, cluster 1 menunjukkan daerah tersebut sering mengalami bencana tanah longsor, sedangkan cluster 2 menggambarkan kecenderungan terjadinya bencana angin puting beliung [11].

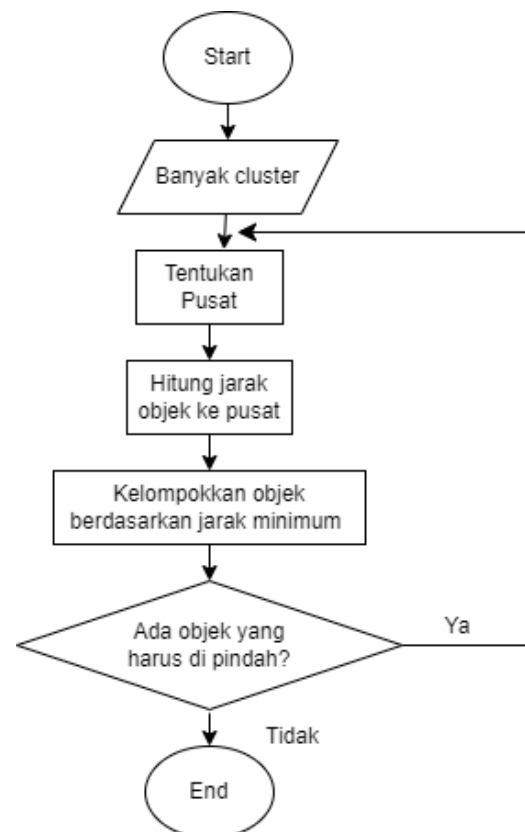
Penelitian yang dilakukan Adha dkk dengan judul "Perbandingan Algoritma DBSCAN dan K-

Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia” *Clustering* merupakan suatu metode dalam data mining yang memiliki tujuan untuk mengelompokkan data (objek) kedalam sejumlah klaster atau kelompok, dimana objek yang memiliki kesamaan dikelompokkan bersama dalam satu klaster, sedangkan objek yang berbeda seharusnya menjadi bagian dari klaster yang berbeda. Lebih dari 180.000 kasus Covid-19 telah tercatat di seluruh dunia, dengan lebih dari 7.000 kematian. Upaya beragam telah dilakukan oleh negara-negara yang terkena dampak Covid-19 untuk mengatasi penyebaran virus ini. Penelitian ini bertujuan untuk mengelompokkan negara-negara berdasarkan pola kasus Covid-19 di seluruh dunia. Diharapkan bahwa hasil pengelompokan ini dapat menjadi pedoman, sehingga negara-negara dengan tingkat pemulihan yang rendah dapat mempelajari proses pemulihan yang berhasil dari negara-negara dengan tingkat pemulihan tinggi dalam kelompok yang sama. Dalam penelitian ini, klasterisasi dilakukan menggunakan algoritma DBSCAN dan K-Means. Setelah melalui serangkaian eksperimen, ditemukan bahwa K-Means lebih efektif dibandingkan DBSCAN dalam mengelompokkan kasus Covid-19. Algoritma K-Means menunjukkan nilai Silhouette Index (SI) terbaik sebesar 0,6902, yang tercapai pada eksperimen dengan nilai  $k = 8$  [12].

Penelitian yang dilakukan Suarna & Wijaya dengan judul “Analisa Penerapan Metode Clustering K-Means Untuk Pengelompokkan Data Transaksi Konsumen (Studi Kasus: Cv. Mitra Indexindo Pratama)” mengungkapkan bahwa *Knowledge Discovery in Database* (KDD) adalah proses yang menghasilkan informasi berharga dari basis data. Seluruh tahapan KDD umumnya terdiri dari serangkaian langkah-langkah, dimulai dengan menetapkan cakupan aplikasi, diikuti dengan menentukan tujuan data dari data mentah yang tersimpan dalam basis data, dan kemudian melakukan pembersihan dan pemrosesan data. Seiring berjalannya waktu, volume data transaksi penjualan terus meningkat, mengakibatkan akumulasi data yang memerlukan pemrosesan melalui aplikasi *RapidMiner* untuk merancang strategi penjualan yang lebih efektif. Penelitian ini bertujuan untuk mengelompokkan data transaksi konsumen dengan menggunakan metode *clustering* menggunakan algoritma K-Means. Dalam penelitian ini, pembentukan empat kelompok dilakukan berdasarkan *Measure Types Bregman Divergences* dengan parameter *Squared Euclidean Distance*, dengan tujuan mencapai nilai DBI yang optimal, yakni nilai *Davies Bouldin Index* (DBI) mendekati nol. Hasil penelitian menunjukkan bahwa proses *clustering* K-Means menghasilkan *centroid* akhir pada *Cluster\_2* (175.000.000), *Cluster\_0* (25.000.000), dan *Cluster\_1* (7.500.000), dengan nilai terendah (0) pada *Cluster\_3*. Pengujian kinerja menggunakan *Davies-Bouldin Index* (DBI) untuk mengukur efisiensi cluster menunjukkan bahwa nilai DBI untuk konfigurasi (K4) adalah 0,261 [13].

Penelitian yang dilakukan Sari dkk dengan judul “Penerapan Algoritma K-Means Untuk *Clustering* Data Kemiskinan Provinsi Banten Menggunakan *RapidMiner*” mengungkapkan bahwa *RapidMiner* merupakan perangkat lunak *open source* yang dapat diakses oleh siapa saja [14]. Penggunaan *RapidMiner* juga dilakukan Luchia dkk dalam penelitiannya yang berjudul “Perbandingan K-Means dan K-Medoids Pada Pengelompokkan Data Miskin di Indonesia” untuk memberikan informasi kepada pemerintah mengenai pola wilayah yang memiliki tingkat kemiskinan tertinggi. Penelitian ini melibatkan perbandingan antara dua algoritma, yaitu K-Means dan K-Medoids, dengan tujuan menilai kinerja algoritma yang paling optimal. Hasil penelitian menunjukkan bahwa K-Means memberikan hasil yang lebih baik dibandingkan dengan K-Medoids dalam mengklasifikasikan data kemiskinan di Indonesia, sebagaimana terbukti dari nilai terbaik DBI K-Means sebesar 0.041 pada eksperimen dengan nilai  $K=8$  karena memiliki nilai yang paling mendekati 0 [15].

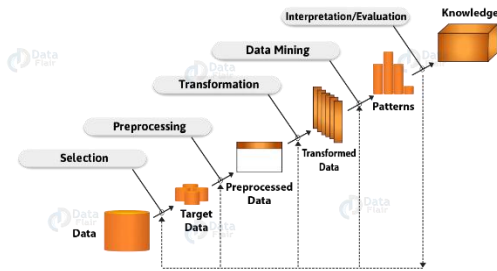
Penelitian yang telah dilakukan oleh () mengungkapkan bahwa dalam melakukan identifikasi kluster berdasarkan data yang telah terkumpul, diperlukan suatu *flowchart* yang dapat mempermudah penentuan langkah-langkah perhitungan sebagai pedoman untuk meraih hasil dari pelaksanaan kluster terhadap data yang sedang diolah. Berikut ini adalah *flowchart* yang menggambarkan proses penentuan kluster menggunakan metode K-Means [16].



Gambar 1. Flowchart k-means clustering

### 3. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini yaitu KDD (*Knowledge Discovery in Database*). Proses data mining melibatkan serangkaian langkah awal yang dimulai dengan pemilihan data, pembersihan data, transformasi data, proses data mining yang mencakup pencarian pola atau informasi dari dataset yang telah terpilih, dan tahapan akhir yaitu interpretasi atau evaluasi [17].



Gambar 2. Tahapan *knowledge discovery in database*

Dari gambar 2 berikut langkah-langkah proses penerapan metode KDD yang dilakukan.

#### a. Selection

Pada tahap ini, peneliti mengimpor dataset kejadian bencana alam yang diperoleh dari website Geoportal Data Bencana Indonesia <https://gis.bnpb.go.id> dengan rentang waktu 2021-2022. Data yang didapatkan sejumlah 9.991 data dengan 16 atribut.

#### b. Preprocessing

Tahap kedua adalah data *preprocessing*. Pada tahap ini, peneliti memilih atribut yang akan digunakan, mengubah atribut sesuai kebutuhan analisis data, mengganti nama atribut, mengubah struktur dataset, mengganti nilai yang hilang pada dataset, dan mengubah skala nilai atribut.

#### c. Transformation

Tahap ketiga adalah data *transformation*. Pada tahap ini, peneliti menetapkan atribut Kabupaten sebagai target dan mengubahnya menjadi label karena fokus penelitian untuk mengelompokkan bencana alam berdasarkan wilayah. Peneliti juga melakukan transformasi pada atribut tanggal kejadian sebagai id karena terdapat beberapa kejadian yang terjadi pada tanggal yang sama.

#### d. Data Mining

Tahap keempat adalah data mining. Pada tahap ini, peneliti menggunakan operator K-Means pada rapidminer untuk mengelompokkan data bencana alam berdasarkan pola yang ada.

#### e. Interpretation/Evaluasi

Tahap kelima adalah *interpretation/evaluasi*. Pada tahap ini, peneliti menggunakan operator *Cluster Distance Performance* pada RapidMiner untuk mengevaluasi hasil kinerja Algoritma K-Means menggunakan *Davies-Bouldin Index* (DBI).

### 4. HASIL DAN PEMBAHASAN

#### 4.1. Selection

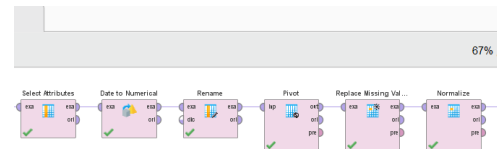
Pada penelitian ini, dataset yang digunakan mengenai data bencana alam yang diperoleh dari website resmi Geoportal Data Bencana Indonesia. Berikut adalah dataset kejadian bencana alam 2021-2022 yang diperoleh dari pengumpulan data kemudian, data akan diimpor menggunakan operator *Read Excel* melalui tools RapidMiner dan memuat ke tab proses proses untuk melihat hasilnya.

No.	Kode Id.	ID Kabu.	Tanggal.	Kejadian	Lokasi	Kabupaten.	Provinsi	Kronologi	Penyebab	Menunggu	Hilang
1	1.000	3201.105	3201.000	Dec 31.2	PUTRANG	Kec. Suk.	BOGOR	JAWA.B.	Dokumen	● Dipicu	
2	2.000	3201.105	3201.000	Dec 31.2	PUTRANG	Kec. Suk.	BOGOR	JAWA.B.	Dokumen	● Dipicu	
3	3.000	3214.105	3214.000	Dec 31.2	PUTRANG	Kec. Suk.	BOGOR	JAWA.B.	Dokumen	● Dipicu	
4	4.000	3208.105	3208.000	Dec 31.2	SANUR	Kec. Tana.	PEKALAN	JAWA.T.	Dokumen	● Dipicu	
5	5.000	3203.105	3203.000	Dec 31.2	SANUR	Kec. Koro.	TANGER	SANTEN	Dokumen	● Dipicu	
6	6.000	3209.105	3209.000	Dec 31.2	PUTRANG	Kec. Ajo.	CIRESON	JAWA.B.	Dokumen	● Dipicu	
7	7.000	3221.105	3221.000	Dec 31.2	SANUR	Kec. Di.	DEMAK	JAWA.T.	Dokumen	● Dipicu	
8	8.000	3204.105	3204.000	Dec 31.2	SANUR	Kec. Koro.	KENDAL	JAWA.T.	Dokumen	● Dipicu	
9	9.000	3219.105	3219.000	Dec 31.2	SANUR	Kec. Koro.	KUDUS	JAWA.T.	Dokumen	● Dipicu	
10	10.000	3203.105	3203.000	Dec 31.2	PUTRANG	Kec. Tana.	KORAN	SULAWA.	Dokumen	● Dipicu	
11	11.000	3217.105	3217.000	Dec 31.2	PUTRANG	Kec. Cio.	SANDAN	JAWA.B.	Dokumen	● Dipicu	
12	12.000	3218.105	3218.000	Dec 31.2	SANUR	Kec. Gub.	PATI	JAWA.T.	Dokumen	● Dipicu	
13	13.000	3214.105	3214.000	Dec 31.2	SANUR	Kec. Hagi.	KOTA.S.	JAWA.T.	Dokumen	● Dipicu	6.000
14	14.000	3215.105	3215.000	Dec 31.2	SANUR	Kec. Ndi.	GROBO.	JAWA.T.	Dokumen	● Dipicu	1.000
15	15.000	3207.105	3207.000	Dec 30.2	PUTRANG	Kec. Ran.	CARAS	JAWA.B.	Dokumen	● Dipicu	
16	16.000	1306.105	1306.000	Dec 30.2	PUTRANG	Kec. Suk.	PADANG	SUMATE.	Dokumen	● Dipicu	
17	17.000	3209.105	3209.000	Dec 30.2	PUTRANG	Kec. Man.	DOMPUS	MUSATE.	Dokumen	● Dipicu	
18	18.000	7306.105	7306.000	Dec 30.2	PUTRANG	Kec. Baga.	GOWA	SULAWA.	Dokumen	● Dipicu	
19	19.000	3212.105	3212.000	Dec 30.2	PUTRANG	Kec. Cio.	KOTA.S.	JAWA.B.	Dokumen	● Dipicu	

Gambar 3. Hasil read excel

Gambar 3 menunjukkan hasil dari operator *Read Excel*.

#### 4.2. Preprocessing



Gambar 4. Tahapan proses preprocessing

Gambar 4 menunjukkan tahapan proses *preprocessing*. Penjelasan mengenai tahapan *preprocessing* adalah sebagai berikut:

##### a. Select Attributes

Setelah melewati proses data *selection*, dataset yang telah diimpor kemudian melalui proses pemilihan atribut yang akan diolah lebih lanjut menggunakan operator *select attributes*. Atribut yang dipilih yaitu Kabupaten, Tanggal Kejadian, dan Tanggal Kejadian.

##### b. Date to Numerical

Setelah menetapkan atribut yang akan digunakan, tahap *preprocessing* selanjutnya yaitu mengubah atribut tanggal ke format numerik agar data dapat diolah menggunakan operator *Date to Numerical* pada RapidMiner.

##### c. Rename

Setelah melakukan proses *Date to Numerical*, langkah *preprocessing* selanjutnya yaitu mengubah nama atribut yang semula Tanggal Kejadian menjadi Tahun Kejadian.

3576

Hasil dari exampleset 3 menunjukkan Nilai uji yang kedua dari jumlah  $k=3$ , dihasilkan *performance vector* Avg. within centroid distance sebesar 0.666, Avg. within centroid distance\_cluster\_0 sebesar 0.668, Avg. within centroid distance\_cluster\_1 sebesar 0.149 dan Avg. within centroid distance\_cluster\_2 0.000. Secara hasil evaluasi *performance* menggunakan Davies-Bouldin Index (DBI) nilai  $k=3$  yaitu 0.012.

#### 4.6. Pembahasan

Pengujian kemampuan algoritma K-Means Clustering dengan Davies-Bouldin Index (DBI) dimulai dari  $k=2$  sampai  $k=10$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.799
Avg. within centroid distance_cluster_0: 0.729
Avg. within centroid distance_cluster_1: 21.457
Davies Bouldin: 0.082
```

Gambar 10. Hasil cluster distance performance exampleset ( $k=2$ )

Gambar 10 menunjukkan pengujian dengan exampleset  $k=2$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.570
Avg. within centroid distance_cluster_0: 0.527
Avg. within centroid distance_cluster_1: 0.149
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 7.069
Davies Bouldin: 0.064
```

Gambar 11 memperlihatkan pengujian dengan exampleset  $k=4$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.483
Avg. within centroid distance_cluster_0: 0.381
Avg. within centroid distance_cluster_1: 1.261
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 0.149
Avg. within centroid distance_cluster_4: 8.392
Davies Bouldin: 0.092
```

Gambar 12 menampilkan pengujian dengan exampleset  $k=5$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.431
Avg. within centroid distance_cluster_0: 0.398
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 5.643
Avg. within centroid distance_cluster_3: 0.149
Avg. within centroid distance_cluster_4: 1.085
Avg. within centroid distance_cluster_5: 1.261
Davies Bouldin: 0.065
```

Gambar 13 menjabarkan hasil pengujian dari exampleset  $k=6$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.344
Avg. within centroid distance_cluster_0: 0.273
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 1.261
Avg. within centroid distance_cluster_3: 0.149
Avg. within centroid distance_cluster_4: 5.643
Avg. within centroid distance_cluster_5: 2.680
Avg. within centroid distance_cluster_6: 0.738
Davies Bouldin: 0.077
```

Gambar 14 memberikan hasil pengujian dari exampleset  $k=7$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.318
Avg. within centroid distance_cluster_0: 0.258
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.149
Avg. within centroid distance_cluster_3: 1.261
Avg. within centroid distance_cluster_4: 1.085
Avg. within centroid distance_cluster_5: 0.738
Avg. within centroid distance_cluster_6: 3.768
Avg. within centroid distance_cluster_7: 3.000
Davies Bouldin: 0.085
```

Gambar 15 memperlihatkan hasil pengujian dengan exampleset  $k=8$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.264
Avg. within centroid distance_cluster_0: 0.167
Avg. within centroid distance_cluster_1: 0.149
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 1.085
Avg. within centroid distance_cluster_4: 0.000
Avg. within centroid distance_cluster_5: 0.692
Avg. within centroid distance_cluster_6: 3.000
Avg. within centroid distance_cluster_7: 1.261
Avg. within centroid distance_cluster_8: 0.650
Davies Bouldin: 0.081
```

Gambar 16 menunjukkan hasil pengujian exampleset  $k=9$ .

##### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.255
Avg. within centroid distance_cluster_0: 0.201
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 1.261
Avg. within centroid distance_cluster_3: 0.149
Avg. within centroid distance_cluster_4: 0.000
Avg. within centroid distance_cluster_5: 0.650
Avg. within centroid distance_cluster_6: 2.989
Avg. within centroid distance_cluster_7: 0.962
Avg. within centroid distance_cluster_8: 0.664
Avg. within centroid distance_cluster_9: 1.085
Davies Bouldin: 0.074
```

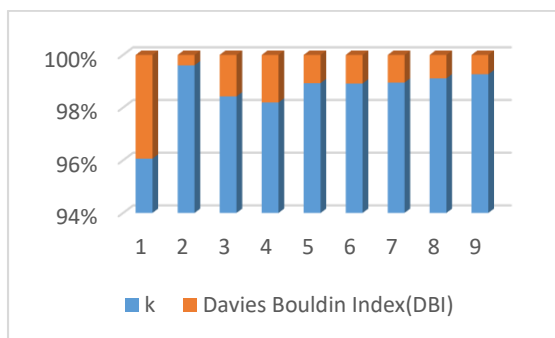
Gambar 17 menampilkan hasil pengujian dengan exampleset  $k=10$ . Dari uraian sebelumnya, maka dapat disimpulkan hasilnya pada tabel 1.



Tabel 1. Hasil DBI k=2, 3, 4, 5, 6, 7, 8, 9, 10

Measure Type	k	Davies-Bouldin Index (DBI)
Bregman Divergences	2	0.082
	3	0.012
	4	0.064
	5	0.092
	6	0.065
	7	0.077
	8	0.085
	9	0.081
	10	0.074

Dari tabel 1 didapatkan nilai uji performance berdasarkan Davies Bouldin Index (DBI) untuk setiap nilai k=2 sampai k=10 maka grafik nilainya clusternya bisa dilihat pada gambar 9.



Gambar 11. Grafik nilai k=2, 3, 4, 5, 6, 7, 8, 9, 10

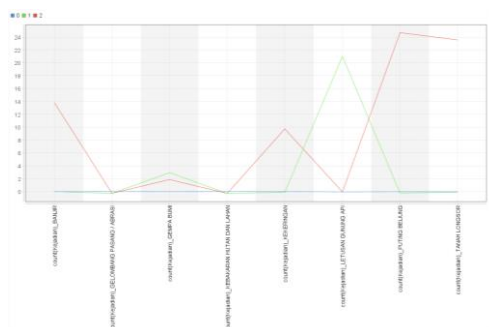
Dari gambar 9 Temuan ini menunjukkan bahwa k=3 memberikan hasil terbaik ditandai dengan nilai *Index Davies Bouldin* (DBI) sebesar 0,012 yang mendekati nol.

#### Cluster Model

Cluster 0: 884 items  
Cluster 1: 2 items  
Cluster 2: 1 items  
Total number of items: 887

Gambar 12. Hasil cluster model

Gambar 10 menunjukkan hasil cluster k=3 yaitu cluster 0 memiliki 884 items, cluster 1 memiliki 2 items, dan cluster 2 memiliki 1 items. Total items keseluruhan adalah 887 items.



Gambar 13 Hasil grafik plot K-Means clustering

Gambar 11 menunjukkan bahwa *cluster 0* merupakan wilayah dengan tingkat kejadian bencana rendah, *cluster 1* adalah wilayah dengan tingkat kejadian bencana sedang, sedangkan *cluster 2* memiliki tingkat kejadian bencana tinggi.

## 5. KESIMPULAN DAN SARAN

Dalam penelitian yang sudah dilakukan, penggunaan algoritma k-means dengan parameter k=3 menghasilkan 3 *cluster*. *Cluster 0* memiliki 884 data dengan tingkat kejadian bencana alam rendah. Kejadian bencana alam yang dominan terjadi yaitu kekeringan. Sedangkan wilayah yang termasuk kedalam *cluster 1* dengan memiliki 2 data adalah wilayah dengan tingkat kejadian sedang. Kejadian bencana yang paling dominan adalah Letusan Gunung Api. Dan *cluster 2* memiliki 1 data yaitu wilayah dengan tingkat kejadian tinggi. Kejadian bencana alam yang paling dominan adalah Puting Beliung. Kolaborasi antarintansi, perkembangan model prediktif, melibatkan masyarakat, dan penelitian lanjutan juga diusulkan untuk lebih mendalami pemahaman mengenai faktor-faktor penyebab bencana alam di Indonesia serta potensi perubahan dalam pola kejadian dengan melibatkan analisis lebih rinci terkait variabel-variabel seperti perubahan iklim, topografi, dan aktivitas manusia. Melalui penerapan saran ini, diharapkan pemahaman dan respons terhadap bencana alam di Indonesia dapat ditingkatkan, melindungi masyarakat dan mengurangi dampak yang mungkin terjadi.

## DAFTAR PUSTAKA

- [1] A. Taryana, M. R. El Mahmudi, dan H. Bekti, "Analisis Kesiapsiagaan Bencana Banjir Di Jakarta," *JANE - J. Adm. Negara*, vol. 13, no. 2, hal. 302, 2022, doi: 10.24198/jane.v13i2.37997.
- [2] O. Nurdiawan, A. Faqih, dan ..., "Pemanfaatan Sistem Informasi Layanan Berbasis Digital Desa Sinarancang," *AMMA J. ...*, vol. 2, no. 8, hal. 922–928, 2023, [Daring]. Tersedia pada: <https://journal.mediapublikasi.id/index.php/amma/article/view/3561%0Ahttps://journal.mediapublikasi.id/index.php/amma/article/download/3561/1875>
- [3] O. Nurdiawan, A. Faqih, A. I. Purnamasari, dan ..., "Pelatihan Dan Pendampingan Tata Kelola Website SMK Cendikia Kota Cirebon," *Prax. J. ...*, vol. 1, no. 1, 2021, [Daring]. Tersedia pada: <http://www.pijarpemikiran.com/index.php/praxis/article/view/561%0Ahttp://www.pijarpemikiran.com/index.php/praxis/article/download/561/521>
- [4] A. M. Suwardi, "Analisa Spasial Clustering Zonasi Rawan Bencana Tanah Longsor Wilayah Bogor Selatan Berbasis Web-Gis," *J. Inov. Inov. Teknol. Inf. dan Inform.*, vol. 3, no. 1, hal. 61, 2020, doi: 10.32832/inova-tif.v3i1.4063.
- [5] I. T. Umagapi dan B. Umaternate, "Uji Kinerja K-Means Clustering Menggunakan Davies-

- Bouldin Index Pada Pengelompokan Data Prestasi Siswa,” hal. 303–308, 2023.
- [6] B. Harahap, “Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung),” *Reg. Dev. Ind. Heal. Sci. Technol. Art Life*, hal. 394–403, 2019, [Daring]. Tersedia pada: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>
- [7] Syafi’i, O. Nurdiawan, dan G. Dwilestari, “Penerapan Machine Learning Untuk Menentukan Kelayakan Kredit Menggunakan Metode Support Vektor Machine,” *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, hal. 1–6, 2022.
- [8] M. Walid dan F. Halimiyah, “Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode Naive Bayes,” *J. ICT Inf. Commun. Technol.*, vol. 22, hal. 190–197, 2022.
- [9] O. Nurdiawan, A. Irma Purnamasari, dan I. Ali, “Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana,” *J. Data Sci. dan Inform.*, vol. 1, no. 2, hal. 32–35, 2021.
- [10] M. Herviany, S. Putri Delima, T. Nurhidayah, dan K. Kasini, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, hal. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.
- [11] T. I. Hermanto dan Y. Muhyidin, “Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana,” *J-SAKTI (Jurnal Sains Komput. dan ...)*, 2021, [Daring]. Tersedia pada: <http://www.tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/332>
- [12] R. Adha, N. Nurhaliza, U. Sholeha, dan ..., “Perbandingan algoritma DBSCAN dan k-means clustering untuk pengelompokan kasus Covid-19 di dunia,” *SITEKIN J. Sains ...*, 2021, [Daring]. Tersedia pada: <http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12469>
- [13] N. Suarna dan Y. A. Wijaya, “ANALISA PENERAPAN METODE CLUSTERING K-MEANS UNTUK PENGELOMPOKAN DATA TRANSAKSI KONSUMEN ( Studi Kasus : Cv . Mitra Indexindo Pratama ),” vol. 7, no. 2, hal. 1322–1328, 2023.
- [14] Y. R. Sari, A. Sudewa, D. A. Lestari, dan T. I. Jaya, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, hal. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- [15] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, dan S. F. Octavia, “Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, hal. 35–41, 2022, doi: 10.57152/malcom.v2i2.422.
- [16] N. A. Manihuruk, M. Zarlis, E. Irawan, dan ..., “Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means,” *KOMIK ...*, 2020, [Daring]. Tersedia pada: <http://www.ejurnal.stmik-budidarma.ac.id/index.php/komik/article/view/2575>
- [17] A. Winarta dan W. J. Kurniawan, “Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python,” *J. Tek. Inform. Kaputama*, vol. 5, no. 1, hal. 113–119, 2021.