CLUSTERING BENCANA ALAM MENGGUNAKAN *K-MEANS* PADA WILAYAH JAWA BARAT

Dede Rohman[1], Rhima Annisa[2], Deny Indriyana Efendi[3], Dodi Solahudin[4]
Teknik Informatika, STMIK IKMI Cirebon
Jln Perjuangan No 10 B Karyamulya Kesambi Kota Cirebon Jawa Barat Indonesia
rhimaannisa09@gmail.com

ABSTRAK

Wilayah Jawa Barat memiliki risiko bencana alam yang signifikan, dengan hampir semua jenis bencana, termasuk gempa bumi, tanah longsor, banjir, dan lainnya, telah terjadi di berbagai bagian wilayah tersebut. Oleh karena itu, penting untuk mengkaji lebih mendalam informasi mengenai tingkat frekuensi bencana alam di berbagai daerah, guna meningkatkan kewaspadaan dan kesiapsiagaan di masa yang akan datang. Tujuan dari penelitian ini adalah untuk mengelompokkan insiden-insiden bencana alam di Jawa Barat ke dalam kelompok-kelompok berdasarkan jenis bencananya. Rangkaian tahap metodologi melibatkan akuisisi data insiden bencana, pengolahan data, penentuan jumlah klaster yang optimal melalui metode *elbow*, implementasi algoritma *K-Means*, serta analisis hasil klastering. Dengan menggunakan tools *Rapidminer* diperoleh 3 *cluster* dengan nilai *Davies Bouldin Index* yaitu 0.004. *Cluster* 0 merupakan daerah dengan kejadian tinggi, *Cluster* 2 dengan daerah kejadian sedang dan *Cluster* 1 dengan kejadian rendah.

Kata kunci : *Clustering, K-Means, Bencana Alam, Jawa Barat, Data Mining*

1. PENDAHULUAN

Indonesia, sebagai kawasan yang sering menghadapi ancaman bencana alam, khususnya Jawa Barat, memiliki risiko yang cukup besar. Wilayah ini ditandai oleh keragaman geografis dan fluktuasi cuaca yang berubah-ubah, yang memicu berbagai jenis bencana seperti gempa bumi, banjir, tanah longsor, dan lainnya. Perubahan iklim global dan peningkatan dampak bencana alam di Jawa Barat memerlukan penelitian mendalam untuk memahami pola kejadian. Faktor-faktor seperti kenaikan suhu laut dan hujan ekstrem meningkatkan risiko banjir dan longsor. Oleh karena itu, tindakan mitigasi dan peningkatan kesiapsiagaan sangat penting untuk melindungi penduduk dan sumber daya alam di wilayah tersebut [1].

Jawa Barat sering mengalami bencana alam, dengan data kebencanaan mencapai ribuan kejadian setiap tahunnya. Gubernur Provinsi Jawa Barat dan forkopimda telah melakukan upaya mitigasi dini untuk meningkatkan kesiapsiagaan. Pengelompokan wilayah berdasarkan data statistik kebencanaan di Jawa Barat menjadi fokus strategis, namun belum dilakukan secara ilmiah. Penelitian lebih lanjut dengan menggunakan algoritma *K-Means* dapat membantu mengidentifikasi pola kebencanaan, memungkinkan pengambilan keputusan yang terinformasi, dan merancang strategi mitigasi yang efektif.

Dalam penelitian sebelumnya berjudul "Penerapan Algoritma *K-Means Clustering* untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan dengan Upaya Antisipasi/Mitigasi Bencana Alam" [2] dijelaskan bahwa bencana alam merupakan kejadian atau kerusakan yang diakibatkan oleh alam, umumnya terjadi secara tiba-tiba atau mendadak, dan berpotensi menimbulkan kerugian materi bahkan korban jiwa. Penelitian lain yang berjudul "Penerapan Algoritma *Clustering K-Means* untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam" [3] membahas penerapan algoritma *K-Means* untuk mengidentifikasi empat tingkatan prioritas penerima bantuan rumah akibat bencana, yakni sangat prioritas (C0), prioritas (C1), kurang prioritas (C2), dan tidak prioritas (C3). Studi serupa yang dilakukan oleh [4] dengan judul "Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban, dan Kerusakan Fasilitas dengan Algoritma *K-Means*" merinci bahwa hasil penelitian ini mencakup pembagian data bencana alam ke dalam kelompok-kelompok, yang dapat menjadi dasar untuk pemantauan, analisis, dan prediksi peristiwa bencana di masa depan, serta identifikasi daerah-daerah yang rentan terhadap bencana.

Penelitian ini menerapkan metode *K-Means* untuk mengidentifikasi pola kejadian bencana alam di Jawa Barat, dengan fokus pada perancangan respons dan mitigasi yang akurat. Temuan penelitian berpotensi memberikan kontribusi penting bagi pemerintah dan lembaga terkait dalam mengurangi risiko bencana di wilayah tersebut serta memberikan wawasan berharga bagi masyarakat dan akademisi. Implementasi K-Means dalam *clustering* data bencana alam memungkinkan identifikasi pola yang seragam, menjadi dasar analisis mendalam untuk merancang strategi mitigasi yang lebih terfokus dan efisien. Studi ini juga memberikan pemahaman mendalam mengenai sifat dan pola kejadian bencana alam di Jawa Barat, berkontribusi dalam pengembangan strategi mitigasi dan pengelolaan risiko bencana.

2. *LITERATURE REVIEW*

2.1. Metode *Literature Review*

Penelitian ini menerapkan metode literature review untuk mengeksplorasi *Clustering* Kejadian Bencana Alam di Jawa Barat berdasarkan jenis bencana menggunakan metode *K-Means*. Tahapan *literature review* melibatkan empat langkah, yaitu merumuskan permasalahan, pencarian literatur, evaluasi data, dan analisis serta interpretasi informasi. Proses formulasi dimulai dengan memilih topik penelitian, fokus pada *Data Mining* dan *Analisis Big Data* terkait bencana alam di Jawa Barat. Pencarian literatur dilakukan dengan alat seperti *Publish or Perish*, mengambil informasi dari *Google Scholar*, Sinta, dan *Scopus* dengan kata kunci *"Data Mining"* dan "Algoritma *K-Means*". Evaluasi data melibatkan pemilahan artikel jurnal berdasarkan relevansinya dan kekinian informasi. Sepuluh artikel terpilih kemudian ditinjau dengan teknik pencarian kesamaan, perbandingan, kritikan, sintesis, dan ringkasan [5].

2.2. Hasil *Literature Review*

Hasil *literature review* terkait *data mining* dan analisis *big data* menghadirkan berbagai penemuan menarik dari sepuluh *paper* yang relevan dengan topik tersebut.

Paper 1 berjudul, "Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat" [6] membahas perbandingan antara *K-Means* dan *K-Medoids* dalam menganalisis daerah rawan tanah longsor di Jawa Barat, mengevaluasi keakuratan, kinerja komputasi, dan kemampuan keduanya membentuk kelompok.

Paper 2 berjudul, "Analisis Data untuk Prediksi Daerah Rawan Bencana di Jawa Barat Menggunakan Algoritma *K-Means Clustering*" [1] berfokus pada analisis data untuk memprediksi daerah rawan bencana di Jawa Barat dengan *K-Means Clustering*, bertujuan menemukan pola kompleks yang sulit terdeteksi secara manual.

Paper 3 yang berjudul, "Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban, dan Kerusakan Fasilitas dengan Algoritma *K-Means*" [7] memanfaatkan *K-Means* untuk mengelompokkan data bencana alam berdasarkan variabel seperti lokasi geografis, waktu kejadian, jumlah korban, dan kerusakan fasilitas.

Paper 4 berjudul, "Penerapan Algoritma *Clustering K-Means* untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam" [3] berfokus pada penerapan *K-Means* untuk menetapkan prioritas penerima bantuan rumah setelah bencana alam, mempertimbangkan kualitas data dan implementasi algoritma.

Paper 5 berjudul, "Penerapan Algoritma *K-Means Clustering* Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam" [2] memiliki hasil penelitian menunjukkan bahwa sekitar 14.71% provinsi memiliki tingkat upaya mitigasi tinggi, sementara 85.29% tingkat rendah, berdasarkan pengelompokan menggunakan *K-Means Clustering.*

Paper 6 yang berjudul, "Klasterisasi Wilayah Tanah Longsor Berdasarkan Dampak Wilayah dan Geografis Menggunakan Metode *K-Means*" [8] mengembangkan sistem pengelompokan wilayah terdampak tanah longsor dengan *K-Means*, menunjukkan efektivitas dalam analisis korelasi antara dampak wilayah dan faktor geografis.

Paper 7 dengan judul, "Penerapan Metode Algoritma *K-means* untuk *Clustering* Daerah Rawan Tanah Longsor di Provinsi Jawa Tengah" [9] berhasil mengelompokkan daerah rawan tanah longsor berdasarkan frekuensi kejadian dengan menggunakan algoritma *K-Means*, memberikan informasi berharga bagi BPBD dalam mengelola risiko tanah longsor.

Paper 8 berjudul, "*Clustering* Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma *K-Means*" [10] berfokus pada pengelompokan daerah rentan banjir di Kabupaten Mandailing Natal menggunakan *K-Means*, menghasilkan model yang mengklasifikasikan wilayah ke dalam tiga tingkat kerentanan berbeda.

Paper 9 berjudul, "*Clustering K-Means* untuk Analisis Pola Persebaran Bencana Alam di Indonesia" [11] menggunakan *K-Means* untuk mengelompokkan data bencana alam di Indonesia, memberikan efisiensi dalam organisasi data dan pemahaman tentang kelompok-kelompok dengan karakteristik dan tingkat risiko serupa.

Paper 10 dengan judul, "Pengelompokan Daerah Bencana Alam Menggunakan Algoritma *K-Means Clustering*" [12] memprediksi potensi bencana alam di Jawa Barat dengan *K-Means Clustering*, mengidentifikasi tiga tingkatan potensi bencana alam dan menyoroti nilai informasi dalam manajemen bencana.

Berdasarkan *literature review*, penelitian ini menyoroti pemanfaatan algoritma *K-Means clustering* dalam penanganan berbagai aspek bencana alam di Jawa Barat. Fokusnya meliputi analisis dan pengelompokan kejadian berdasarkan jenis, penentuan prioritas bantuan perumahan pasca-bencana, pemahaman pola kejadian dengan mempertimbangkan faktor-faktor tertentu, prediksi lokasi rentan, dan perbandingan kinerja antara *K-Means* dan *K-Medoids* untuk pengelompokan daerah rawan longsor. Tujuannya adalah untuk meningkatkan pemahaman, pengelolaan risiko, dan respons terhadap bencana di wilayah tersebut.

2.3. Landasan Teori

2.3.1. Bencana Alam

Bencana alam adalah peristiwa yang diakibatkan oleh kekuatan atau fenomena alam, menyebabkan kerusakan lingkungan, kehilangan nyawa, dan kerugian materiil. Contoh bencana alam di Jawa Barat meliputi gempa bumi, banjir, tanah longsor, dan letusan gunung api, disebabkan oleh

aktivitas seismik, kondisi geografis, perubahan iklim, atau ulah manusia. Indikator seperti magnitudo gempa dan curah hujan ekstrim digunakan untuk memahami dan mengukur potensi bencana, mendukung mitigasi, terutama di wilayah rentan seperti Jawa Barat [13].

2.3.2. Data Mining

Data mining adalah proses identifikasi pola, tren, dan hubungan baru dalam dataset besar menggunakan teknologi pengenalan pola, statistika, dan matematika. Tujuannya adalah mengenali pola yang signifikan dalam basis data luas untuk mendukung pengambilan keputusan di masa depan [14].

2.3.3. Clustering

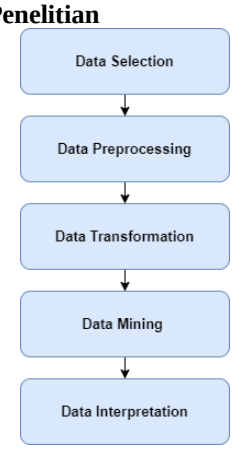
Clustering adalah proses mengelompokkan data ke dalam dua atau lebih kelompok berdasarkan kemiripan antar data dalam satu kelompok. Metode ini merupakan bagian dari pembelajaran tanpa supervisi dan tidak membutuhkan label pada setiap kelompok. Clustering digunakan untuk mengelompokkan data berdasarkan kesamaan atribut di antara kelompok data [8].

2.3.4. K-Means

K-Means adalah algoritma klasterisasi yang membagi data ke dalam kelompok berdasarkan centroid terdekat. Tujuannya adalah mengelompokkan data sehingga kesamaan di dalam satu kelompok maksimal dan perbedaan antar kelompok minimal. Prosesnya melibatkan langkah-langkah menentukan jumlah klaster, inisialisasi centroid, pengelompokan data, dan pembaruan nilai centroid sampai konvergensi [2].

3. METODE PENELITIAN

3.1. Metode Penelitian



Gambar 1. Proses KDD[13]
Sumber:https://www.ncbi.nlm.nih.gov/books/NBK558907/

Gambar 1 menggambarkan proses *Knowledge Discovery in Database* (KDD) sebagai metode analisis data, yang merupakan suatu pendekatan yang melibatkan tahapan-tahapan seperti pengumpulan data, pemilihan, transformasi, penambahan informasi relevan, pemodelan, evaluasi, dan interpretasi hasil [15]. Pendekatan ini menjadi dasar untuk mengungkap pola-pola dan wawasan yang dapat digunakan dalam mengambil keputusan serta memahami lebih dalam konteks data yang dianalisis.

Tabel 1. Deskripsi tahapan KDD

Tahapan	Aktivitas	Deskripsi Aktivitas
Data Selection	Melakukan Proses Pemilihan Data	Data bencana alam Jawa Barat bersumber dari lembaga riset, mencakup informasi Kabupaten, Jenis Bencana, Jumlah Kejadian, dan Tahun kejadian, memberikan informasi komprehensif dari sumber yang kredibel.
Pre-Processing	Membersihkan, Menormalkan, dan Mempersiapkan data untuk analisis	Membersihkan data dengan menangani nilai-nilai yang hilang, normalisasi, dan konversi format untuk mendukung analisis K-Means tanpa bias.
Transformation	Mengubah data ke format atau representasi yang sesuai.	Menyiapkan data untuk klasterisasi K-Means dengan menetapkan atribut relevan dan memilih metrik jarak, seperti Euclidean Distance, untuk perbandingan dan pengelompokan data berdasarkan kesamaan atau kedekatannya.
Data Mining	Menerapkan teknik dan algoritma analisis untuk menemukan pola atau informasi yang tersembunyi dalam data.	Menerapkan algoritma *K-Means* untuk mengelompokkan data bencana alam di Jawa Barat. Menentukan jumlah klaster optimal dengan metode *Elbow* untuk mencerminkan karakteristik data secara efektif.
Interpretation/Evaluation	Menganalisis hasil data mining dan mengevaluasi keberhasilan model atau temuan.	Mengevaluasi hasil klasterisasi bencana alam dengan analisis statistik dan visualisasi untuk mengidentifikasi pola atau kemiripan antara kelompok. Melibatkan analisis mendalam terhadap karakteristik masing-masing klaster dengan tujuan memberikan interpretasi yang signifikan terhadap hasil klasterisasi.
Knowledge Presentation	Menyajikan informasi atau wawasan yang ditemukan dari analisis data.	Menyusun laporan atau visualisasi terstruktur dan informatif untuk menyampaikan hasil klasterisasi secara jelas. Presentasi temuan dan interpretasi melalui berbagai format seperti grafik, peta, tabel, atau narasi, bertujuan memudahkan pemahaman dan mendukung proses pengambilan keputusan.

Tabel 1 menyajikan rinciannya aktivitas proses *Knowledge Discovery in Database* (KDD) yang telah diimplementasikan dalam rangka penelitian ini.

3.2. Sumber Data

Data diperoleh dari situs web https://opendata.jabarprov.go.id/. Data ini mencakup informasi tentang kejadian bencana di wilayah Jawa Barat dalam rentang waktu 2019 hingga 2021, Total terdapat 15936 *record* dalam *dataset* tersebut.

3.3. Teknik Pengumpulan Data

Penelitian ini mengandalkan data primer dari instansi yang mencatat kejadian bencana di Jawa Barat dari 2019 hingga 2021. Data melibatkan berbagai jenis bencana dan mencakup 15.936 *records*, yang akan dianalisis untuk memahami pola, dampak, dan faktor-faktor terkait.

3.4. Teknik Analisis Data

Penelitian ini menggunakan analisis data dengan algoritma *K-Means*, sebuah metode yang mengelompokkan data ke dalam beberapa klaster berdasarkan tingkat kesamaan. Langkah-langkahnya adalah sebagai berikut:

1. Menentukan jumlah *cluster* yang diinginkan, disimbolkan sebagai k.
2. Menghasilkan nilai acak untuk titik awal pusat-pusat *cluster (centroid)* sebanyak k.
3. Menghitung jarak antara setiap data input dengan setiap centroid menggunakan rumus jarak *Euclidean* hingga menemukan jarak terdekat dari setiap data ke *centroid* terkait. Berikut adalah persamaan *Euclidean Distance*:

$$d(xi, \mu j) = \sqrt{\Sigma(xi - \mu j)^2} \qquad (1)$$

Dimana :
Xi = data kriteria
μj = *centroid* pada *cluster* ke-j

4. Mengelompokkan setiap data ke dalam *cluster* berdasarkan kedekatannya dengan *centroid* terdekat (jarak terkecil).
5. Memperbarui nilai *centroid* dengan menghitung rata-rata dari setiap *cluster* yang terbentuk dengan menggunakan rumus :

$$\mu j(t+1)^n = \frac{1}{Nsj} \sum_{jeSj} \quad xj \qquad (2)$$

Dimana :
μj(t+1) = *centroid* baru pada iterasi ke-j
(t+1)Nsj = banyak data pada *cluster* Sj

6. Melakukan iterasi dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah. Jika langkah telah terpenuhi, maka nilai pusat *cluster* (μj) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasterisasi data [16].

4. HASIL DAN PEMBAHASAN

4.1. Seleksi Data

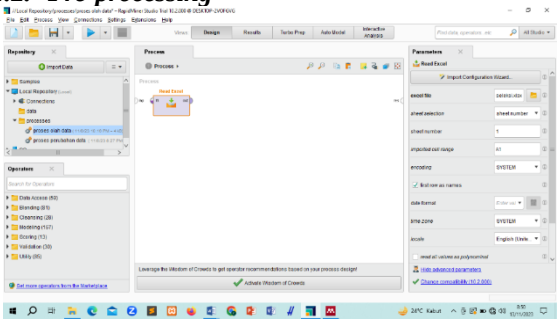
Proses ini melibatkan pemilihan dan seleksi data terkait kejadian Bencana Alam yang relevan. Data tersebut selanjutnya dikelompokkan menjadi *dataset* yang akan digunakan dalam penelitian ini dengan menggunakan RapidMiner.

Tabel 2. Seleksi data bencana alam

Nama Kabupaten	Nama Desa	GB	B	TL	GP	K	Tahun
Kabupaten Bogor	Wanaherang	0	0	0	0	0	2019
Kabupaten Bogor	Bojong Kulur	0	2	0	0	0	2019
Kabupaten Bogor	Ciangsana	0	0	0	0	0	2019
Kabupaten Bogor	Gunung Putri	0	0	1	0	0	2019
Kabupaten Bogor	Bojong Nangka	0	1	0	0	0	2019
Kabupaten Bogor	Tlajung Udik	0	0	0	0	0	2019
Kabupaten Bogor	Cicadas	0	0	0	0	1	2019
Kabupaten Bogor	Cikeas Udik	0	0	0	0	0	2019
Kabupaten Bogor	Nagrak	0	0	0	0	0	2019
….	….	….	….	….	….	….	….
Kota Banjar	Karyamukti	0	0	1	0	1	2021
Kota Banjar	Binangun	0	0	4	0	1	2021
Kota Banjar	Sukamukti	0	0	2	0	1	2021
Kota Banjar	Sinartanjung	0	1	1	0	1	2021
Kota Banjar	Raharja	0	1	0	0	1	2021
Kota Banjar	Mekarharja	0	0	0	0	1	2021
Kota Banjar	Langensari	2	0	0	0	1	2021
Kota Banjar	Rejasari	2	0	0	0	1	2021
Kota Banjar	Waringinsari	2	2	0	0	1	2021
Kota Banjar	Kujangsari	2	0	0	0	1	2021

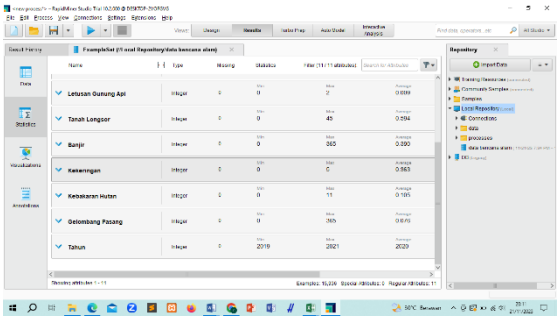
Tabel 2 berisi data terpilih dari *dataset* yang berjumlah 15.936 *record*, dengan total 8 atribut.

4.2. *Pre-processing*



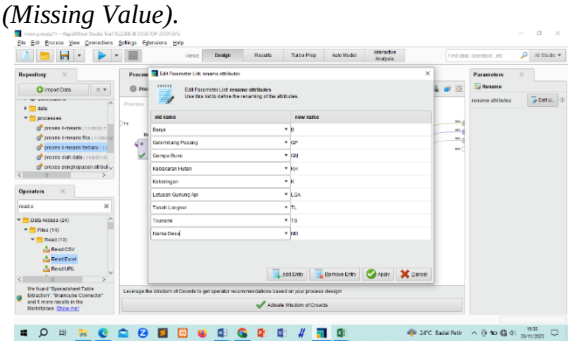
Gambar 2. Penambahan operator *Read Excel*

Gambar 2 menampilkan antarmuka setelah menambahkan operator *Read Excel* untuk mengimpor *dataset* yang akan diproses.



Gambar 3. *Static missing value*

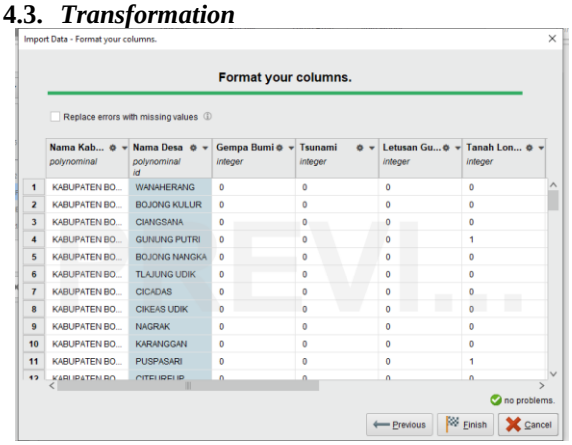
Gambar 3 menunjukkan tampilan *static missing value*. Setelah proses impor data, dilakukan langkah pembersihan atau *pre-processing* data untuk mempermudah analisis menggunakan algoritma *K-Means*. Tujuan dari *pre-processing* ini adalah menghilangkan gangguan atau nilai yang hilang *(Missing Value)*.



Gambar 4. Tampilan *rename* attribute

Jika data tidak mengalami masalah, langkah berikutnya adalah memilih atribut-atribut yang akan digunakan. Setelah pemilihan atribut, langkah selanjutnya adalah mengubah nama atribut seperti yang tertera pada Gambar 4.

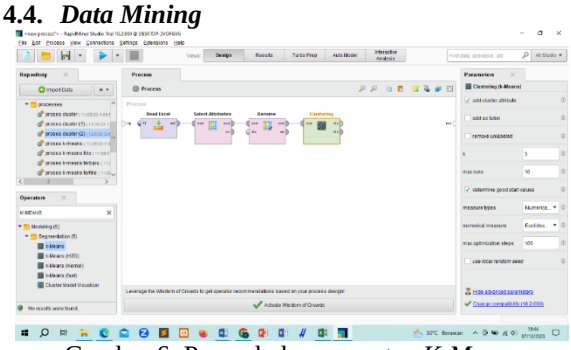
4.3. *Transformation*



Gambar 5. Perubahan tipe data

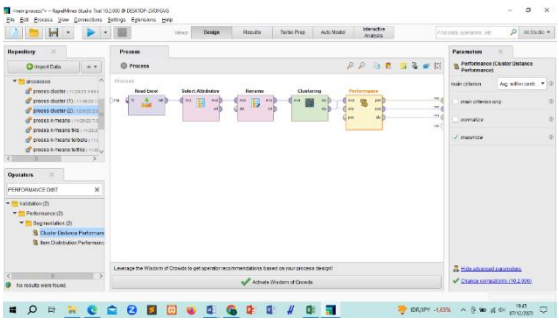
Gambar 5 menggambarkan fase transformasi dilakukan dengan mengubah jenis data dari polinomial menjadi identifikasi, serta modifikasi nama atribut. Tujuan dari pra-pemrosesan ini adalah untuk mempersiapkan data sehingga dapat diolah menggunakan algoritma *K-Means*.

4.4. *Data Mining*



Gambar 6. Penambahan operator *K-Means*

Gambar 6 merupakan tampilan penambahan proses *K-Means Clustering* ke dalam panel. Pada panel parameter, masukkan nilai K dari 2 hingga 20 dengan maksimum *runs* sebanyak 10. Setelah itu, pilih *Numerical Measure* sebagai tipe pengukuran, dan secara otomatis akan muncul *Euclidean Distance*. Selanjutnya, sambungkan garis ke *respiration*. Setelah itu, klik tombol *play*.



Gambar 7. Penambahan operator *Performance Distance*

Langkah berikutnya adalah melengkapi proses dengan menambahkan operator kinerja *performance distance* dengan tampilan seperti pada Gambar 7.

4.5. *Interpretation/Evaluation*

Row No.	ND	cluster	GB	B	TL	GP	K
1	WANAHERA...	cluster_0	0	0	0	0	0
2	BOJONG KU...	cluster_0	0	2	0	0	0
3	CIANGSANA	cluster_0	0	0	0	0	0
4	GUNUNG PU...	cluster_0	0	0	1	0	0
5	BOJONG NA...	cluster_0	0	1	0	0	0
6	TLAJUNG UD...	cluster_0	0	0	0	0	0
7	CICADAS	cluster_0	0	0	0	0	1
8	CIKEAS UDIK	cluster_0	0	0	0	0	0
9	NAGRAK	cluster_0	0	0	0	0	0
10	KARANGGAN	cluster_0	0	0	0	0	0
11	PUSPASARI	cluster_0	0	3	1	0	0
12	CITEUREUP	cluster_0	0	0	0	0	0
13	LEUWINUTUG	cluster_0	0	1	1	0	0
14	TAJUR	cluster_0	0	0	0	0	0
15	SANJA	cluster_0	0	0	0	0	0

ExampleSet (15,936 examples,2 special attributes,5 regular attributes)

Gambar 8. Hasil *cluster* pada *data view* proses *K-Means*

Gambar 8 menampilkan hasil dari proses data *mining* menggunakan algoritma *K-Means*, menghasilkan 20 klaster dari total 15.936 *record*.

Cluster Model

```
Cluster 0: 3119 items
Cluster 1: 1 items
Cluster 2: 3 items
Cluster 3: 1 items
Cluster 4: 254 items
Cluster 5: 11 items
Cluster 6: 2 items
Cluster 7: 7067 items
Cluster 8: 485 items
Cluster 9: 58 items
Cluster 10: 42 items
Cluster 11: 54 items
Cluster 12: 69 items
Cluster 13: 1292 items
Cluster 14: 1931 items
Cluster 15: 151 items
Cluster 16: 641 items
Cluster 17: 49 items
Cluster 18: 702 items
Cluster 19: 4 items
Total number of items: 15936
```

Gambar 9. Tampilan *cluster* model

Dari Gambar 9, terlihat dengan jelas model klaster yang dihasilkan oleh algoritma *K-Means.* Setiap klaster merepresentasikan kelompok data yang memiliki kesamaan tertentu berdasarkan *Euclidean Distance*.

PerformanceVector

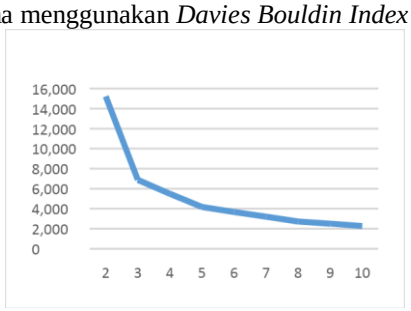
```
PerformanceVector:
Avg. within centroid distance: 0.994
Avg. within centroid distance_cluster_0: 0.617
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 0.000
Avg. within centroid distance_cluster_4: 2.749
Avg. within centroid distance_cluster_5: 73.504
Avg. within centroid distance_cluster_6: 0.000
Avg. within centroid distance_cluster_7: 0.168
Avg. within centroid distance_cluster_8: 2.774
Avg. within centroid distance_cluster_9: 24.621
Avg. within centroid distance_cluster_10: 12.493
Avg. within centroid distance_cluster_11: 32.105
Avg. within centroid distance_cluster_12: 8.120
Avg. within centroid distance_cluster_13: 0.916
Avg. within centroid distance_cluster_14: 0.337
Avg. within centroid distance_cluster_15: 4.476
Avg. within centroid distance_cluster_16: 2.479
Avg. within centroid distance_cluster_17: 6.568
Avg. within centroid distance_cluster_18: 1.329
Avg. within centroid distance_cluster_19: 67.000
Davies Bouldin: 0.765
```

Gambar 10. Tampilan *text view* pada *performance vector*

Dalam Gambar 10, terdapat 20 *cluster* hasil dari proses *data mining* menggunakan algoritma *K-Means.*

4.6. Pembahasan

Hasil analisis menggunakan algoritma *K-Means Clustering* pada data bencana di wilayah Jawa Barat (2019-2021) menunjukkan tiga *cluster* dari 15.936 catatan. Metode *elbow* menetapkan nilai optimal K=3. Setelah menerapkan *K-Means* (K=3, *max runs*=10) dengan *Euclidean Distance*, dilakukan evaluasi performa menggunakan *Davies Bouldin Index* (DBi).



Gambar 11. Grafik *Elbow*

Gambar 11 menampilkan grafik *elbow* hasil percobaan dengan variasi nilai K dari 2 hingga 20 (*max runs* = 10), menunjukkan nilai-nilai *Davies Bouldin Index* (DBi) yang mencerminkan kualitas klasterisasi pada setiap K. Hasil eksperimen menunjukkan DBi terendah adalah 0,004 saat K = 3, mengindikasikan klasterisasi yang lebih baik dibandingkan dengan nilai K lainnya.

Tabel 3. Hasil percobaan dengan nilai parameter *max run / iterations* = 10

K	*K-Means*	
	Avg. Within Centroid Distance	**Dbi**
2	15.223	0.004
3	6.880	0.004
4	5.492	0.576
5	4.175	0.543
6	3.672	0.548
7	3.205	0.741
8	2.715	0.688
9	2.511	0.771
10	2.261	0.605
11	2.093	0.755
12	1.845	0.742
13	1.647	0.642
14	1.539	0.674
15	1.364	0.703
16	1.291	0.669
17	1.226	0.674
18	1.126	0.733
19	1.047	0.745
20	0.994	0.765

Tabel 3 menunjukkan hasil percobaan dengan parameter max run/iterations = 10. Dari tabel tersebut, dapat disimpulkan bahwa nilai optimum untuk jumlah klaster (k) adalah 3, dengan nilai Davies Bouldin Index (DBi) sebesar 0.004 pada algoritma K-Means.

Cluster Model

```
Cluster 0: 15934 items
Cluster 1: 1 items
Cluster 2: 1 items
Total number of items: 15936
```

Gambar 12. Hasil *cluster model*

Gambar 12 menampilkan model klaster dengan k = 3 dan nilai DBi 0.004, menghasilkan 15,934 item pada Cluster 1, 1 item pada Cluster 2, dan 1 item pada Cluster 3, dengan total 15,936 item secara keseluruhan.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 6.880
Avg. within centroid distance_cluster_0: 6.881
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.000
Davies Bouldin: 0.004
```

Gambar 13. Hasil *performance vector*

Berdasarkan Gambar 13, hasil perhitungan dengan menggunakan *Performance Vector* menghasilkan nilai jarak sebesar 6.880. Selanjutnya, nilai jarak untuk *Cluster* 1 adalah 6.881, *Cluster* 2 memiliki jarak 0.000, dan *Cluster* 3 juga memiliki jarak 0.000. Dengan demikian, nilai *Davies Bouldin Index* (DBi) yang dihasilkan adalah 0.004.

Attribute	cluster_0	cluster_1	cluster_2
GB	0.456	0	0
B	0.367	365	2
TL	0.594	0	0
GP	0.053	4	365
K	0.363	1	0

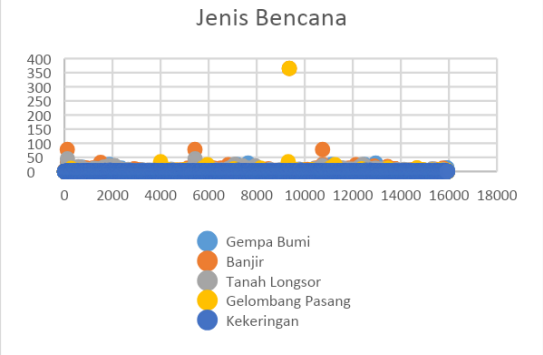
Gambar 14. *Centroid table*

Berdasarkan *Centroid Table* pada Gambar 14, nilai Gempa Bumi untuk *Cluster* 1 adalah 0.456, *Cluster* 2 adalah 0, dan *Cluster* 3 adalah 0. Untuk kategori Banjir, *Cluster* 1 memiliki nilai 0.367, *Cluster* 2 memiliki nilai 365, dan *Cluster* 3 memiliki nilai 2. Pada kategori Tanah Longsor, *Cluster* 1 memiliki nilai 0.594, *Cluster* 2 memiliki nilai 0, dan *Cluster* 3 memiliki nilai 0. Sedangkan untuk Gelombang Pasang, *Cluster* 1 memiliki nilai 0.053, *Cluster* 2 memiliki nilai 4, dan *Cluster* 3 memiliki nilai 365. Untuk kategori Kekeringan, *Cluster* 1 memiliki nilai 0.363, *Cluster* 2 memiliki nilai 1, dan *Cluster* 3 memiliki nilai 0. Selanjutnya, dilakukan interpretasi terhadap klusterisasi dengan K = 3.

Setelah menerapkan algoritma *K-Means* pada 15936 *record* data, terbentuk 3 *cluster* dengan distribusi sebagai berikut: *Cluster* 0 (15934 Desa), *Cluster* 1 (1 Desa), dan *Cluster* 2 (1 Desa).

Gambar 15 Grafik dengan 3 Cluster

Dalam rentang waktu 2019-2021, bencana alam yang paling sering terjadi adalah banjir, mencapai 365 kejadian di daerah Eretan Kulon. Tanah longsor menempati urutan kedua dengan 45 kejadian, terutama di desa Rawa Panjang. Gelombang Pasang merupakan bencana berikutnya dengan 35 kejadian di desa Cemara, sementara Gempa Bumi terjadi sedang dengan 30 kejadian, terutama di desa Bunter.



Gambar 16. *Scatter Plot* bencana alam berdasarkan jenis bencana

Gambar 15 menampilkan *scatter plot* hasil klasterisasi untuk setiap atribut kejadian bencana alam, termasuk Banjir, Tanah Longsor, Gempa Bumi, Kekeringan, dan Gelombang Pasang. *Cluster* 1 mendominasi dalam kejadian Banjir, sedangkan *cluster* 2 memiliki kejadian Banjir yang paling sedikit. Gempa Bumi mendominasi *Cluster* 0, sementara *Cluster* 2 dan *Cluster* 1 memiliki tingkat kejadian yang lebih rendah. Gelombang Pasang paling sering terjadi di *Cluster* 2, sedangkan tingkat kejadiannya yang sedang terdapat di *Cluster* 0, dan yang paling rendah terjadi di *Cluster* 1. Tanah Longsor paling banyak terjadi di *Cluster* 0, sementara tingkat kejadian yang paling rendah terdapat di *Cluster* 1 dan *Cluster* 2. Kekeringan paling sering terjadi di *Cluster* 0, dengan tingkat kejadian yang sedang di *Cluster* 1, dan yang paling rendah di *Cluster* 2.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa terdapat tiga kelompok bencana alam di Jawa Barat dengan tingkat kejadian yang tinggi, sedang, dan rendah. *Cluster* 1 menunjukkan tingkat kejadian sangat tinggi pada bencana banjir, sedangkan *cluster* 0 mendominasi dalam kejadian tanah longsor. Gelombang pasang terkonsentrasi pada *cluster* 2. Evaluasi menggunakan *Cluster Distance Performance*

(CDP) pada *RapidMiner* memberikan nilai *Davies Bouldin Index* sebesar 0.004 pada algoritma *K-Means* dengan K=3. Saran untuk penelitian mengedepankan analisis mendalam terhadap faktor-faktor lain yang mempengaruhi kejadian bencana, seperti geografi, infrastruktur, dan aspek sosial-ekonomi. Validasi hasil dapat dilakukan dengan membandingkan metode klasterisasi lainnya. Kesimpulan ini mendorong perlunya pengarahan kepada masyarakat terkait hasil pengelompokan *K-Means* terhadap bencana alam dengan tingkat kejadian tinggi.

DAFTAR PUSTAKA

[1] M. F. Al Halik and L. Septiana, "Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 4, pp. 856–870, 2022, doi: 10.52362/jisamar.v6i4.939.

[2] Y. Pratama, Hendrawan, E. Rasywir, B. T. Carenina, and D. R. Anggraini, "Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1232–1240, 2022, doi: 10.47065/bits.v4i3.2549.

[3] S. Mariam, F. Handayani, and C. Jualiane, "Penerapan Algoritma Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 231–240, 2023.

[4] O. N. Washilaturrizqi, "Implementasi Algoritma C4 . 5 Untuk Menentukan Penerima," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 373–377, 2023.

[5] O. Nurdiawan, A. Irma Purnamasari, and I. Ali, "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana," *J. Data Sci. dan Inform.*, vol. 1, no. 2, pp. 32–35, 2021.

[6] M. Herviany, S. Putri Delima, T. Nurhidayah, and 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, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.

[7] Murdiaty, A. Angela, and C. Sylvia, "Pengelompokkan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 744, 2020, doi: 10.30865/mib.v4i3.2213.

[8] H. E. D. Moch. Rizki Eko Waluyo, Pramana Yoga Saputra, "Klasterisasi Wilayah Tanah Longsor Berdasarkan Dampak Wilayah Dan Geografis Menggunakan Metode K Means," *Semin. Inform. Apl. POLINEMA 2020*, pp. 226–230, 2020.

[9] N. Fadilah, "Penerapan Metode Algoritma K-Means Untuk Clustering Daerah Rawan Tanah Longsor Di Provinsi Jawa Tengah," *J. BATIRSI*, vol. 6, no. 1, pp. 1–5, 2022, [Online]. Available: https://e-journal.stmik-tegal.ac.id/index.php/batirsi/article/view/31

[10] I. Hidayat, E. Darnila, and Y. Afrillia, "Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma K-Means," *G-Tech J. Teknol. Terap.*, vol. 7, no. 3, pp. 1218–1226, 2023, doi: 10.33379/gtech.v7i3.2880.

[11] M. A. Y. Pratama, A. R. Hidayah, and T. Avini, "Clustering K-Means untuk Analisis Pola Persebaran Bencana Alam di Indonesia," *Juli*, vol. 3, no. 2, pp. 108–114, 2023.

[12] R. R. Khaerullah, N. Suarna, and O. Nurdiawan, "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.

[13] I. Rinjani, S. Anwar, and R. Herdiana, "PENGELOMPOKAN DAERAH BENCANA ALAM MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *J. Ilm. Sist. Inf. DAN ILMU Komput.*, vol. 3, no. 1, pp. 1–14, 2023, [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK558907/

[14] N. S. L. Zulfa and A. Hadiana, "Kajian Data Mining Menggunakan Algoritma K-Means Dan K-Medoids Dalam Strategi Promosi," *J. Fak. Tek. UNISA KUNINGAN*, vol. 2, no. 2, pp. 57–62, 2021.

[15] Firmansyah and O. Nurdiawan, "Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.

[16] M. R. Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.