Pengembangan Model Pengelompokan Jenis Bencana Alam di Jawa Barat menggunakan Algoritma K-Means

Panji Adi Pura[1], Martanto[2], Arif Rinaldi Dikananda[3], Dede Rohman[4]
STMIK IKMI CIREBON[1234]
panjiadiepura69@gmail.com[1] , martantomusijo@gmail.com[2] , rinaldi21crb@gmail.com[3],
dederohman17@gmail.com[4]

Abstrak - Salah satu masalah terbesar yang dihadapi masyarakat Jawa Barat adalah bencana alam. Analisis berbasis data diperlukan untuk memahami pola kejadian bencana dan mendukung kebijakan mitigasi yang efektif karena berbagai jenis bencana. Untuk menganalisis data kejadian bencana di Jawa Barat selama periode 2020–2023, penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases (KDD).*Tahapan KDD meliputi *pembuatan dataset*, *preprocessing* untuk normalisasi dan penanganan data hilang, serta *transformasi* guna menentukan atribut utama. Algoritma K-Means digunakan dalam proses data mining untuk mengelompokkan wilayah berdasarkan jenis bencana dan intensitasnya. Tahap terakhir adalah *interpretasi hasil*, yang bertujuan untuk memahami pola distribusi bencana. Hasil klasterisasi menghasilkan lima kluster utama. *Cluster 0* menunjukkan dominasi kejadian banjir dan kebakaran lahan, sering ditemukan di dataran rendah dengan karakteristik lingkungan yang rawan pembakaran liar. *Cluster 1* didominasi oleh kejadian tanah longsor di wilayah perbukitan yang curah hujannya tinggi. *Cluster 2* mencerminkan kombinasi kejadian hujan angin dan kekeringan di daerah pedesaan dengan sumber daya air terbatas. *Cluster 3* menunjukkan kejadian bencana dengan frekuensi rendah dan distribusi yang merata, seringkali terkait dengan daerah urban. Sementara itu, *Cluster 4* memiliki tingkat heterogenitas tertinggi, mencakup berbagai jenis bencana dengan intensitas bervariasi di wilayah pegunungan dan lembah. Kualitas klasterisasi diukur menggunakan *Davies-Bouldin Index (DBI)* sebesar 0.085, mengindikasikan pemisahan kluster yang baik. Selain itu, analisis *Performance Vector* menunjukkan jarak total antar-kluster sebesar 2.311, dengan jarak terbesar pada *Cluster 4* (4.672). penelitian ini diharapkan dapat membantu dalam perencanaan dan alokasi sumber daya yang lebih tepat sasaran untuk mitigasi bencana.
Kata Kunci: K-Means, bencana alam, agregasi, pengurangan bencana, Jawa Barat.

Abstract *- One of the biggest problems faced by the people of West Java is natural disasters. Data-based analysis is necessary to understand disaster occurrence patterns and support effective mitigation policies due to various types of disasters. To analyze disaster occurrence data in West Java during the period 2020–2023, research uses the Knowledge Discovery in Databases (KDD) approach. The KDD stages include dataset creation, preprocessing for normalization and handling missing data, as well as transformation to determine key attributes. The K-Means algorithm is used in the data mining process to cluster regions based on the type and intensity of disasters. The final stage is result interpretation. The clustering results in five main clusters. Cluster 0 shows a dominance of flood and land fire events, often found in lowland areas with environmental characteristics prone to wildfires. Cluster 1 is dominated by landslide events in hilly areas. Cluster 2 reflects a combination of wind and drought events in rural areas with limited water resources. Cluster 3 shows low-frequency disaster events with an even distribution, often related to urban areas. Meanwhile, Cluster 4 has the highest level of heterogeneity, encompassing various types of disasters with varying intensities in mountainous and valley regions. The quality of clustering is measured using the Davies-Bouldin Index (DBI) of 0.085, indicating good cluster separation. Performance Vector analysis shows a total cluster distance of 2.311, with the largest distance in Cluster 4. (4.672). this research is expected to assist in more targeted planning and resource allocation for disaster mitigation.*
Keywords: K-Means, natural disasters, aggregation, disaster reduction, West Java.

I. PENDAHULUAN

Indonesia memiliki tingkat kerawanan bencana yang sangat tinggi akibat kondisi geologis dan geografisnya yang unik. Hal ini menjadikan Indonesia sebagai salah satu negara yang paling rentan terhadap berbagai jenis bencana alam. Salah satu wilayah yang memiliki tingkat kerawanan tinggi adalah Jawa Barat, yang sering mengalami bencana seperti banjir, gempa bumi, tanah longsor, dan puting beliung. Kejadian bencana di wilayah ini tidak hanya menyebabkan kerugian materi yang signifikan tetapi juga ancaman serius terhadap keselamatan jiwa dan kerusakan sosial-ekonomi yang mendalam. Dalam menghadapi kondisi tersebut, mitigasi bencana berbasis data menjadi langkah yang sangat penting untuk meningkatkan kesiapsiagaan, mengurangi risiko, serta memastikan respons cepat terhadap bencana yang terjadi (Rosaliyah & Nurhakim, 2023).

Seiring perkembangan teknologi informasi, algoritma *machine learning* seperti *K-Means* telah terbukti memiliki potensi besar dalam membantu identifikasi pola risiko bencana melalui analisis data yang lebih terstruktur. Algoritma ini memungkinkan pengelompokan wilayah berdasarkan parameter tertentu, seperti jenis bencana, intensitasnya, dan karakteristik wilayah yang relevan. Menurut (Kusnadi & Putri, 2021). menunjukkan bahwa penggunaan K-Means sangat efektif dalam mengidentifikasi daerah rawan bencana di Indonesia. Temuan mereka membuktikan bahwa metode ini dapat mendukung perencanaan mitigasi yang lebih tepat sasaran. (Halik & Septiana, 2022). juga menunjukkan keberhasilan algoritma ini dalam memprediksi wilayah rawan bencana di Jawa Barat, menjadikan K-Means sebagai metode yang relevan dalam penanganan risiko bencana yang lebih efisien.

Lebih lanjut, (Hermanto et al., 2021). menemukan bahwa K-Means dapat digunakan untuk mengelompokkan wilayah dengan risiko serupa, sehingga membantu otoritas lokal dalam menyusun strategi mitigasi yang lebih terarah. Studi yang dilakukan (Ramadhani et al., 2022). memanfaatkan metode clustering untuk memprediksi risiko gempa bumi di Pulau Jawa. Sementara itu, (Reviantika et al., 2020). menggunakan K-Means untuk menganalisis pola gempa di wilayah yang sama. Meskipun hasil penelitian ini cukup menjanjikan, kedua studi tersebut masih memiliki keterbatasan karena belum mempertimbangkan data dari jenis bencana lain, seperti banjir atau tanah longsor, yang sering terjadi secara bersamaan di wilayah Jawa Barat.

Penelitian (Baldah et al., 2023). menggarisbawahi bahwa algoritma K-Means mampu memberikan hasil yang signifikan dalam mengidentifikasi daerah rawan bencana di tingkat provinsi, termasuk di Jawa Barat. Namun, mereka juga mencatat bahwa tantangan utama dalam penelitian ini adalah kurangnya integrasi data historis dan *real-time*, yang dapat mengurangi akurasi dalam prediksi risiko bencana. Keterbatasan ini juga ditekankan oleh (Agustina & Mulyawan, 2023). yang menyoroti perlunya sistem klasifikasi yang lebih komprehensif untuk mengelompokkan wilayah berdasarkan risiko multi-bencana.

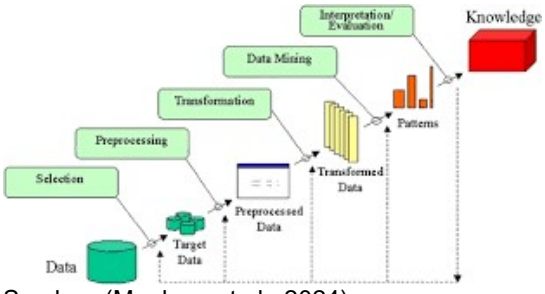
Menurut (Pratama et al., 2022). algoritma K-Means memiliki potensi untuk mengatasi kesenjangan dalam pengelompokan wilayah rawan bencana. Dengan mempertimbangkan berbagai jenis bencana secara bersamaan, K-Means dapat membantu menciptakan gambaran risiko yang lebih holistik. Penemuan ini sejalan dengan temuan (Halik & Septiana, 2022). yang menunjukkan efektivitas algoritma ini dalam memanfaatkan data historis untuk mendukung perencanaan mitigasi. Namun, mereka juga mencatat pentingnya pengembangan model yang lebih kompleks dengan mempertimbangkan data multi-bencana dan faktor lainnya, seperti perubahan iklim atau potensi bencana sekunder.

Penelitian ini bertujuan untuk mengembangkan model clustering berbasis algoritma K-Means guna mengidentifikasi kluster wilayah rawan bencana di Jawa Barat. Model ini akan mencakup analisis terhadap berbagai jenis bencana, seperti banjir, gempa bumi, tanah longsor, dan puting beliung. Dengan pendekatan ini, penelitian diharapkan memberikan kontribusi signifikan bagi manajemen risiko bencana, peningkatan kesiapsiagaan masyarakat, dan penguatan kebijakan mitigasi berbasis data yang lebih efektif. Selain itu, hasil penelitian ini diharapkan menjadi dasar bagi pengembangan teknologi lebih lanjut dalam manajemen bencana, khususnya dengan memanfaatkan *machine learning* untuk analisis risiko secara komprehensif.

II. METODOLOGI PENELITIAN

1. Metode Penelitian

Dalam penelitian ini, metode analisis data Knowledge Discovery Database (KDD) digunakan. Analisis data ini melibatkan langkah-langkah seperti pengumpulan data, pemilihan, transformasi, penambahan informasi yang relevan, pemodelan, evaluasi, dan interpretasi hasil. Metode ini memberikan dasar untuk mengidentifikasi pola dan gagasan yang dapat digunakan dalam proses pengambilan keputusan dan memperoleh pemahaman yang lebih baik tentang konteks data.



Sumber: (Maulana et al., 2024)

Gambar 1. Proses *Knowladge Discovery Database* (KDD)

a) Data Selection

Data bencana alam Kabupaten Cirebon dari lembaga penelitian mencakup informasi tentang Kabupaten, Jenis Bencana, Jumlah Kejadian, dan Tahun Kejadian. Ini adalah data komprehensif yang berasal dari sumber yang dapat dipercaya.

b) Preprocessing

Untuk mendukung analisis K-Means tanpa bias, data dibersihkan dengan memperbaiki data yang hilang, melakukan normalisasi, dan mengubah formatnya.

c) Transformation

Menyediakan data untuk klasterisasi K-Means dengan mengidentifikasi karakteristik terkait dan menggunakan metrik jarak, seperti jarak geometri, untuk membandingkan dan mengelompokkan data berdasarkan kesamaan atau kedekatan antar data.

d) Data Mining

Menggunakan algoritma K-Means untuk mengumpulkan data bencana alam menggunakan analisis statistik dan visualisasi untuk mengevaluasi hasil klasterisasi data bencana alam untuk menemukan pola atau kemiripan antar kelompok. Proses ini mencakup analisis mendalam karakteristik setiap klaster untuk memberikan interpretasi yang signifikan tentang hasil klasterisasi.

e) Knowledge Presentation

Untuk menunjukkan hasil klasterisasi dengan jelas, buat laporan atau visualisasi yang terstruktur dan informatif. Temuan dan interpretasi disampaikan dalam berbagai format, seperti narasi, grafik, peta, atau tabel, sehingga lebih mudah dipahami dan membantu pengambilan keputusan.

2. Sumber Data

Data yang dikumpulkan dari Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Cirebon dikumpulkan melalui metode pengumpulan data sekunder dalam penelitian ini. Data yang digunakan mencakup tahun 2020–2023, memungkinkan analisis menyeluruh terhadap berbagai bencana selama tiga tahun tersebut. Dalam penelitian ini, 156 record peristiwa bencana dikumpulkan. Setiap catatan mengandung informasi lengkap, seperti waktu peristiwa, lokasi, dan data tambahan. Data ini memiliki delapan atribut yang menggambarkan berbagai elemen dan atribut bencana, termasuk jenis bencana dan lokasi bencana. Tujuan dari penelitian ini adalah untuk memberikan gambaran yang lebih mendalam tentang pola kejadian bencana di Kabupaten Cirebon selama periode tersebut melalui pengumpulan data yang terstruktur.

3. Teknik Pengumpulan Data

Data sekunder yang digunakan dalam studi dokumentasi adalah jumlah tingkat bencana alam berdasarkan lokasi kejadian. Data ini diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Cirebon dan berisi jumlah kejadian bencana alam di Kabupaten Cirebon dari tahun 2020 hingga 2023. Data yang mencakup berbagai jenis bencana mencakup 156 record, yang akan dianalisis untuk memahami pola, efek, dan komponen yang berhubungan.

4. Teknik Analisis Data

Untuk melakukan proses clustering, terdapat beberapa tahapan yang harus diikuti, yaitu sebagai berikut:

a) Menentukan berapa banyak kelompok (k) yang akan dibentuk. Langkah pertama adalah menentukan berapa banyak kelompok atau kelompok yang diinginkan, yang diwakili oleh nilai k.

b) Jumlah cluster ini harus disesuaikan dengan tujuan analisis.

c) Inisialisasi centroid secara acak. Setelah menetapkan nilai *k*, langkah selanjutnya adalah menginisialisasi posisi centroid secara acak. Centroid adalah titik pusat dari setiap cluster. Misalnya, jika kita memilih $k = 3$, maka akan ada tiga centroid yang ditandai sebagai *C1, C2*, dan C3, yang awalnya ditempatkan secara acak.

d) Menentukan data point ke centroid terdekat. Pada tahap ini, setiap data point akan dihitung jaraknya ke masing-masing centroid. Pengukuran jarak ini biasanya menggunakan metode Euclidean distance untuk menentukan data point mana yang paling dekat dengan centroid tertentu, sehingga dapat dikelompokkan ke dalam cluster tersebut.

e) Clusterkan setiap data berdasarkan jarak terdekatnya dengan centroid.

f) Optimalisasi dilakukan dengan mengulangi langkah D dan E. Proses

pengelompokan dan perhitungan ulang centroid akan dilakukan lagi dan lagi hingga pergerakan centroid berhenti atau hingga hasil yang diinginkan tercapai. Tujuan dari proses ini adalah untuk meningkatkan ketepatan dan konsistensi pembentukan cluster sesuai dengan standar yang telah ditentukan.

III. HASIL DAN PEMBAHASAN

1. Hasil

a. Selection

Proses ini melibatkan pemilihan dan seleksi data yang relevan tentang peristiwa bencana alam. Kemudian, data ini dikelompokkan menjadi dataset yang akan digunakan dalam penelitian ini dengan menggunakan Rapidminer.

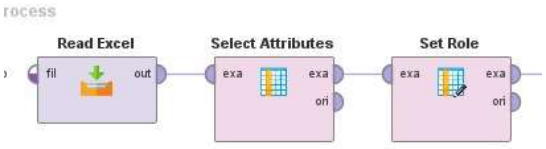
Tabel 1. *Dataset* Bencana Alam 2020-2023

ND	N	B	TL	K	HA	TH
Kec. Arjawinangun	0	4	0	0	3	2020
Kec. Astanajapura	0	4	0	0	3	2020
Kec. Babakan	0	0	0	0	3	2020
Kec. Beber	0	0	0	0	0	2020
Kec. Ciledug	0	2	0	0	0	2020
Kec. Ciwaringin	0	0	0	0	1	2020
Kec. Depok	1	1	1	0	2	2020
Kec. Dukupuntang	0	0	0	0	1	2020
Kec. Gebang	0	3	0	0	4	2020
Kec. Gegesik	0	4	0	0	0	2020
…	…	…	…	…	…	…
Kec. Sedong	1	0	1	85	1	2023
Kec. Sumber	#	0	1	10	6	2023
Kec. Suranenggala	0	1	0	7	2	2023
Kec. Susukan	0	1	0	7	2	2023
Kec. Susukan Lebak	0	0	0	0	0	2023
Kec. Talun	4	1	0	5	1	2023
Kec. Tengahtani	0	7	1	10	0	2023
Kec. Waled	0	10	0	5	0	2023

Keterangan :
ND : Nama Desa
KL : Kebakaran Lahan
B : Banjir
TL : Tanah Longsor
K : Kekeringan
HA : Hujan dan Angin
TH : Tahun

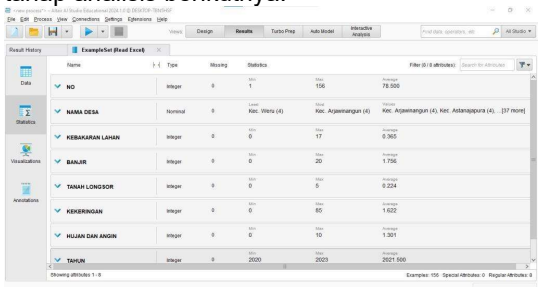
Tabel 1 menunjukkan data yang dipilih dari dataset yang terdiri dari 156 record yang memiliki total 5 atribut. Data ini akan diolah menggunakan tools Rapidminer dalam penelitian ini.

b. Preprocessing



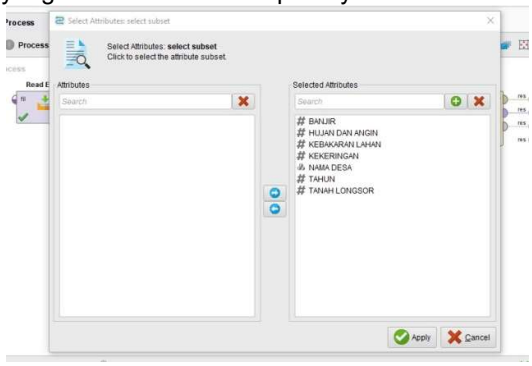
Sumber : Rosaliyah & Nurhakim (2023)
Gambar 2. Operator *Read Excel, Select Attribut dan Set Role*

Gambar 2 menunjukkan antarmuka perangkat lunak menampilkan operator *Read Excel, Select Attribute, dan Set Role* untuk mempermudah impor dataset dari file Excel. Operator ini memungkinkan pengguna memuat data, memilih kolom relevan, dan mengatur peran atribut. Antarmuka dirancang dengan visualisasi jelas agar pengguna dapat memverifikasi dataset sebelum melanjutkan ke tahap analisis berikutnya.



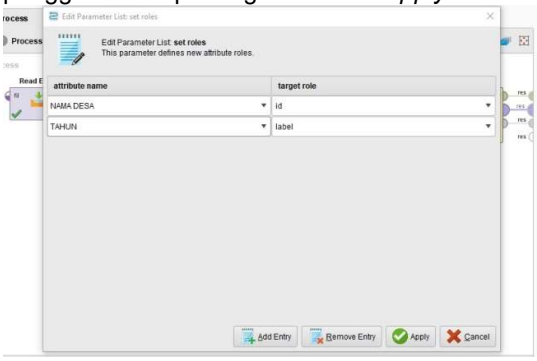
Sumber : Pratama & Yovi (2022)
Gambar 3. *Data Static Missing Value*

Setelah proses impor data selesai, nilai yang hilang dalam dataset ditampilkan secara statistik dalam Gambar 3. Langkah sebelum pengolahan menjadi penting untuk mendukung analisis yang lebih mendalam menggunakan algoritma K-Means. Tahap ini bertujuan mengidentifikasi dan menangani nilai yang hilang (missing value) untuk meningkatkan kualitas data. Dengan membersihkan dataset, analisis dapat dilakukan menggunakan data yang bersih dan lengkap, memastikan hasil yang lebih akurat dan terpercaya.



Sumber : Halik & Septiana (2022)
Gambar 4. Tampilan *Select Attribut*

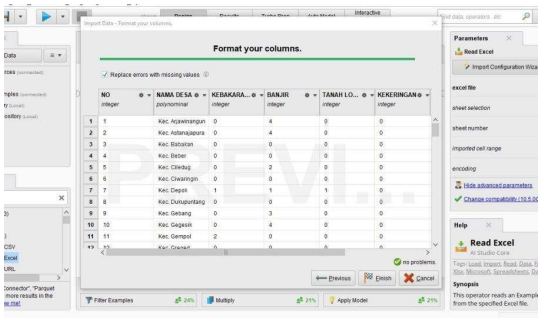
Gambar 4 menunjukkan proses pemilihan kolom data yang relevan untuk analisis. Terdapat dua bagian utama: *Attributes* di sebelah kiri yang berisi semua atribut dalam dataset, dan *Selected Attributes* di sebelah kanan yang menampilkan atribut yang telah dipilih. Dalam contoh, atribut yang dipilih meliputi Banjir, Hujan dan Angin, Kebakaran Lahan, Kekeringan, Nama Desa, Tahun, dan Tanah Longsor**.** Proses ini bertujuan menyaring data agar hanya atribut penting yang digunakan, mempermudah langkah analisis selanjutnya. Untuk menyimpan hasil seleksi, pengguna cukup mengklik tombol *Apply*.



Sumber : Hendrastuty & Nirwana (2024)
Gambar 5. Tampilan *Set Role*

Gambar 5 menunjukkan proses pengaturan peran atribut dalam dataset. Setiap atribut diberikan peran spesifik sesuai fungsinya dalam analisis. Misalnya, atribut Nama Desa diberi peran sebagai *Id* untuk identitas unik, sementara atribut Tahun diberi peran sebagai *Label* untuk variabel target. Pengguna dapat menambah atau menghapus peran atribut menggunakan tombol *Add Entry* dan Remove Entry, lalu menyimpan pengaturan dengan tombol *Apply*. Pengaturan ini memastikan setiap atribut memiliki fungsi yang jelas sebelum proses analisis dilanjutkan.

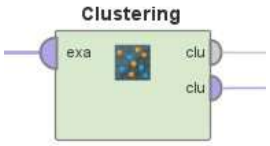
c. Transformation



Sumber : Maulana & Ajay (2024)
Gambar 6. Perubahan Tipe Data

Gambar 6 menunjukkan fase transformasi data, yang menunjukkan proses mengubah jenis data dari polynomial menjadi identifikasi dan perubahan nama atribut dataset. Transformasi ini adalah bagian penting dari pra-pemrosesan data untuk memastikan format data sesuai dengan kebutuhan analisis menggunakan algoritma K-Means. Dengan perubahan ini, data disusun dalam struktur yang lebih cocok, yang meningkatkan efisiensi dan akurasi proses clustering.

d. Data Mining



Sumber : Pratistha & Ragil Nike (2024)
Gambar 7. Operator K-Means

Proses K-Means Clustering ditambahkan ke panel analisis, seperti yang ditunjukkan pada Gambar 7. Pengguna harus menentukan nilai K, yang berkisar antara 2 dan 20. Mereka juga harus menetapkan batas maksimum run sebanyak 10. Selanjutnya, pengguna memilih pengukuran numerik sebagai tipe pengukuran, dan jarak geometri otomatis digunakan sebagai metode pengukuran. Setelah mengatur parameter, garis harus disambungkan ke bagian respiration sistem. Proses clustering dimulai dengan mengklik tombol play. Algoritma

kemudian bekerja sesuai dengan parameter yang telah ditetapkan.

Sumber: Penulis (2024)
Gambar 8. Operator *Performance Distance*

Langkah berikutnya dalam analisis adalah menambahkan operator *performance distance*, seperti yang ditunjukkan pada Gambar 8. Operator ini digunakan untuk mengevaluasi kinerja model clustering yang telah dibangun, memungkinkan pengguna mengukur kualitas hasil dari algoritma *K-Means*. Dengan operator ini, analisis menjadi lebih mendalam, memberikan wawasan tentang efektivitas clustering dan distribusi data dalam setiap kluster.

e. Intrepretation/Evaluation

Row No.	NAMA DESA	cluster	KEBAKARAN...	BANJIR	TANAH LON...	KEKERINGAN	HUJAN DAN ...
1	Kec. Arjawina...	cluster_0	0	4	0	0	3
2	Kec. Astanaja...	cluster_0	0	4	0	0	3
3	Kec. Babakan	cluster_0	0	0	0	0	3
4	Kec. Beber	cluster_0	0	0	0	0	0
5	Kec. Ciledug	cluster_0	0	2	0	0	0
6	Kec. Ciwarin...	cluster_0	0	0	0	0	1
7	Kec. Depok	cluster_0	1	1	1	0	2
8	Kec. Dukupu...	cluster_0	0	0	0	0	1
9	Kec. Gebang	cluster_0	0	3	0	0	4
10	Kec. Gegesik	cluster_0	0	4	0	0	0
11	Kec. Gempol	cluster_0	2	0	0	0	1
12	Kec. Greged	cluster_0	0	0	0	0	2
13	Kec. Gunungj...	cluster_4	1	7	0	0	2
14	Kec. Jamblang	cluster_0	0	0	0	0	3
15	Kec. Kaliwedi	cluster_0	0	0	0	0	0

ExampleSet (156 examples,2 special attributes,5 regular attributes)

Sumber : Baldah & Azizah (2023)
Gambar 9. Hasil *Intrepetation/Evaluation*

Gambar 9 menunjukkan hasil akhir dari proses data mining yang menggunakan algoritma K-Means. Proses ini berhasil menghasilkan lima kluster dari 156 rekord yang ada dalam dataset. Setiap kluster menunjukkan kelompok data yang memiliki kesamaan tertentu, yang memungkinkan pengguna memahami pola dan struktur yang tersembunyi dalam data. Hasil dari analisis ini membantu dalam pengambilan keputusan yang lebih baik karena mereka menjelaskan bagaimana data terdistribusi di antara kluster yang berbeda.

2. Pembahasan

Cluster Model

```
Cluster 0: 137 items
Cluster 1: 1 items
Cluster 2: 1 items
Cluster 3: 2 items
Cluster 4: 15 items
Total number of items: 156
```

Sumber : Rosaliyah & Nurhakim (2023)
Gambar 10. Hasil *Cluster* Model

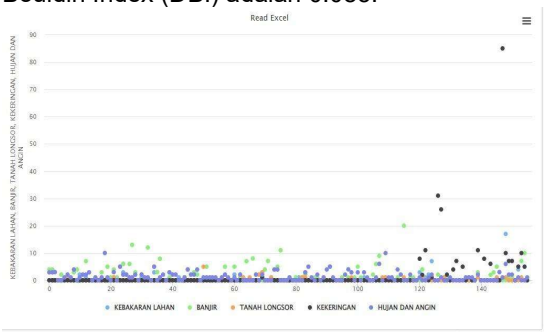
Gambar 10 Menunjukkan hasil *data mining* menggunakan algoritma *K-Means*, menghasilkan 5 *cluster* dari 156 *record* data yang dianalisis. *Cluster* 0 memiliki anggota terbanyak dengan 137 item, sedangkan *Cluster* 1 dan *Cluster* 2 masing-masing hanya memiliki 1 item. *Cluster* 3 berisi 2 item, dan *Cluster* 4 terdiri dari 15 item. Data yang di-*cluster* mencakup rentang tahun 2020 hingga 2023. Pembagian *cluster* ini mencerminkan pola dalam data, yang mungkin disebabkan oleh perbedaan karakteristik antar*record* dalam setiap *cluster*.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 2.311
Avg. within centroid distance_cluster_0: 2.098
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 1.550
Avg. within centroid distance_cluster_4: 4.672
Davies Bouldin: 0.085
```

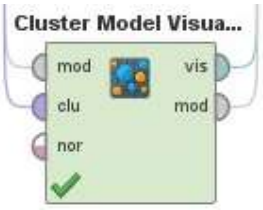
Sumber : Kusnadi & Nahdi (2021)
Gambar 11. Hasil *Perfomance Vector*

Berdasarkan Gambar 11, hasil perhitungan dengan menggunakan *Performance Vector* sebesar 2.311. Untuk Cluster 0, jarak adalah 2.098, Cluster 1 dan Cluster 2 masing-masing 0.000, Cluster 3 memiliki 1.550, dan Cluster 4 memiliki 4.672. Dengan temuan ini, nilai perhitungan Davies-Bouldin Index (DBI) adalah 0.085.



Sumber : Ramadhani & Devy Isya (2022)
Gambar 12 *Scatter Plot* Bencana Alam

Gambar 12 menunjukkan bahwa dari tahun 2020-2023, Bencana alam yang paling umum adalah *Kekeringan* menempati posisi tertinggi dengan 85 kejadian di Kecamatan Sedong. *Banjir* berada di peringkat kedua setelah *Kekeringan*, dengan 20 kejadian di Kecamatan Waled. Selanjutnya, bencana *Hujan dan Angin* tercatat sebanyak 10 kejadian di Kecamatan Kedawung. *Kebakaran Lahan* termasuk kategori sedang, dengan 17 kejadian di Kecamatan Sumber. Sementara itu, bencana dengan jumlah kejadian paling sedikit adalah *Tanah Longsor*, hanya 1 kejadian di Kecamatan Depok.



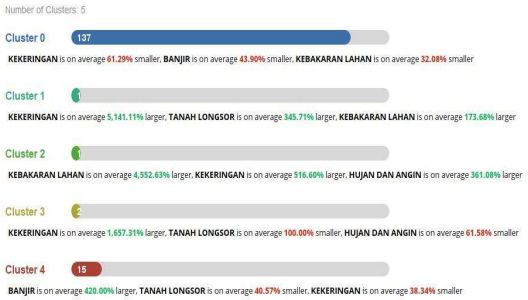
Sumber : Reviantika & Ferin (2020)
Gambar 13. *Cluster Model Visualizer*

Setelah proses clustering selesai, hasil analisis dievaluasi pada Gambar 13 menggunakan Cluster Model Visualizer, yang berperan dalam memvisualisasikan distribusi data di masing-masing cluster.

Cluster	KEBAKARAN LAHAN	BANJIR	TANAH LONGSOR	KEKERINGAN	HUJAN DAN ANGIN
Cluster 0	0.248	0.985	0.226	0.628	1.314
Cluster 1	1	0	1	85	1
Cluster 2	17	0	1	10	6
Cluster 3	0.500	1	0	28.500	0.500
Cluster 4	0.267	9.133	0.133	1	1

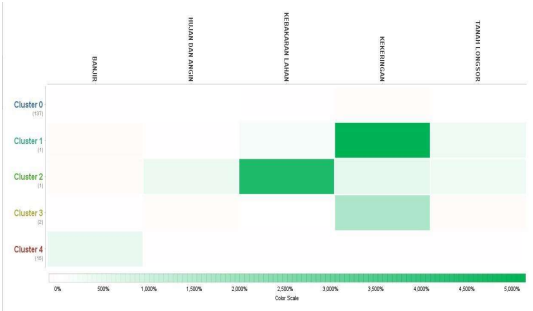
Sumber : Virgo & Ismail (2020)
Gambar 14. *Centroid Table*

Berdasarkan *Centroid Table* yang ditampilkan pada Gambar 4.12, terdapat variasi nilai untuk setiap kategori bencana di masing-masing *cluster*. Pada kategori *Kebakaran Lahan*, nilai yang tercatat adalah 0.248 untuk *Cluster* 0, 1 untuk *Cluster* 1, dan 17 untuk *Cluster* 2. Selanjutnya, *Cluster* 3 memiliki nilai sebesar 0.500, yang sama dengan nilai pada *Cluster* 3 lainnya, sementara *Cluster* 4 mencatat nilai 0.267. Untuk kategori *Banjir*, *Cluster* 0 mencatat nilai 0.985, sedangkan baik *Cluster* 1 maupun *Cluster* 2 memiliki nilai 0. Pada *Cluster* 3, nilainya mencapai 1, sementara *Cluster* 4 menunjukkan angka tertinggi di antara semua *cluster* dengan nilai 9.133. Pada kategori *Tanah Longsor*, *Cluster* 0 memiliki nilai 0.226. Baik *Cluster* 1 maupun *Cluster* 2 masing-masing mencatat nilai sebesar 1. Sementara itu, *Cluster* 3 memiliki nilai 0, dan *Cluster* 4 menunjukkan nilai 0.133. Untuk kategori *Kekeringan*, nilai yang diperoleh cukup bervariasi. *Cluster* 0 memiliki nilai 0.628, diikuti oleh *Cluster* 1 dengan nilai yang jauh lebih tinggi, yaitu 85. *Cluster* 2 mencatat nilai 10, sementara *Cluster* 3 memiliki nilai signifikan sebesar 28.500. Terakhir, *Cluster* 4 memiliki nilai 1. Pada kategori *Hujan dan Angin*, variasi nilai yang terlihat adalah 1.314 untuk *Cluster* 0, dan 1 untuk *Cluster* 1. Kemudian, *Cluster* 2 memiliki nilai 6, sedangkan *Cluster* 3 mencatat angka 0.500. Nilai untuk *Cluster* 4 tercatat sebesar 1.



Sumber : Chusyairi & Ahmad (2019)
Gambar 15. *Record data 5 cluster*

Berdasarkan Gambar 15, penerapan algoritma *K-Means* pada 156 data *record* menghasilkan lima *cluster* dengan distribusi yang bervariasi. *Cluster* 0 mencakup 153 item, menjadikannya *cluster* terbesar, sementara *Cluster* 1 dan *Cluster* 2 masing-masing hanya memiliki 1 item. *Cluster* 3 terdiri dari 2 item, sedangkan *Cluster* 4 memiliki 15 item. Klasterisasi dilakukan dengan mempertimbangkan atribut bencana alam seperti *Banjir*, *Tanah Longsor*, *Hujan dan Angin*, *Kekeringan*, dan *Kebakaran Lahan*. Hasil klasterisasi menunjukkan bahwa kejadian *Kekeringan* paling dominan berada di *Cluster* 0, sedangkan *Cluster* 1 mencatatkan tingkat kejadian paling rendah. Pada kategori *Banjir*, kejadian tertinggi ditemukan di *Cluster* 0, sementara *Cluster* 4 memiliki tingkat kejadian yang lebih rendah. Kejadian *Hujan dan Angin* paling sering terjadi di *Cluster* 3, dengan tingkat kejadian paling rendah tercatat di *Cluster* 2. Untuk *Tanah Longsor*, *Cluster* 4 menunjukkan dominasi tertinggi, diikuti oleh tingkat sedang di *Cluster* 3, sedangkan kejadian paling sedikit tercatat di *Cluster* 1. Terakhir, kejadian *Kebakaran Lahan* paling banyak terjadi di *Cluster* 0, dan yang paling rendah terdapat di *Cluster* 1.



Sumber : Maulana & Danar Dana (2023)
Gambar 16 *Heatmap*

Gambar 16 menampilkan *heatmap* yang menunjukkan distribusi jenis bencana berdasarkan *cluster* hasil algoritma *K-Means*. *Cluster* 0, dengan 137 data titik, menunjukkan

intensitas rendah untuk semua jenis bencana, menandakan wilayahnya relatif tidak rentan terhadap bencana tertentu. Sebaliknya, *Cluster* 1 dan *Cluster* 2, masing-masing hanya memiliki satu data titik, menunjukkan intensitas tinggi pada kategori tertentu, yaitu *Kebakaran Lahan* untuk *Cluster* 1 dan *Kekeringan* untuk *Cluster* 2, menandakan risiko spesifik terhadap bencana tersebut. *Cluster* 3, dengan dua data titik, juga memperlihatkan intensitas tinggi pada *Kebakaran Lahan*, serupa dengan *Cluster* 1. Sementara itu, *Cluster* 4, yang memiliki 15 data titik, menunjukkan intensitas tinggi pada *Kekeringan*, menandakan wilayahnya lebih rentan terhadap bencana tersebut. Pola distribusi ini membantu mengidentifikasi kerentanan bencana di wilayah Jawa Barat, memberikan wawasan tentang variasi risiko di setiap *cluster* dan jenis bencana tertentu.

IV. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *K-Means* efektif untuk mengelompokkan wilayah di Jawa Barat berdasarkan jenis dan intensitas bencana alam seperti banjir, kebakaran lahan, tanah longsor, kekeringan, dan hujan angin. Analisis data dari tahun 2020 hingga 2023 menghasilkan lima *cluster* dengan distribusi bencana yang bervariasi. Berdasarkan perhitungan menggunakan *Performance Vector*, total jarak antar-*cluster* adalah 2.311, yang mencerminkan jarak keseluruhan dalam data yang dikelompokkan. Rincian jarak antar-*cluster* menunjukkan bahwa *Cluster* 0 memiliki jarak 2.098, mengindikasikan dominasi kejadian bencana di wilayah tersebut. Sementara itu, *Cluster* 1 dan *Cluster* 2 memiliki jarak 0.000, yang mencerminkan tingkat homogenitas data yang tinggi. *Cluster* 3 mencatatkan jarak 1.550, menunjukkan tingkat keberagaman sedang, sedangkan *Cluster* 4 memiliki jarak tertinggi sebesar 4.672, merepresentasikan heterogenitas paling tinggi di antara semua *cluster*. Dari analisis ini, diperoleh nilai *Davies-Bouldin Index (DBI)* sebesar 0.085, yang menunjukkan kualitas klasterisasi yang cukup baik, dengan pemisahan antar-*cluster* yang jelas dan tingkat kekompakan internal yang tinggi.

V. REFERENSI

Agustina, M., & Mulyawan. (2023). Implementasi Algoritma K-Means pada Peristiwa Gempa Bumi di Wilayah Jawa Barat. *Jurnal Wahana Informatika (JWI)*, *2*(2), 257–264.

Baldah, A., Duarisah, A. V., & Maulana, R. A. (2023). Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means. *Jurnal Ilmiah Informatika Global*, *14*(2), 31–36. https://doi.org/10.36982/jiig.v14i2.3186

Halik, M. F. Al, & Septiana, L. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering. *Journal of Information System, Applied, Management, Accounting and Research*, *6*(4), 856–870. https://doi.org/10.52362/jisamar.v6i4.939

Hermanto, T. I., Muhyidin, Y., Tinggi, S., Wastukancana, T., Cikopak, J., 53, N., & Barat, J. (2021). Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5Hermanto,*(1), 406.

Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus : Desa Ciomas Bogor). *Jurnal Teknologi Informatika Dan Komputer*, *7*(1), 17–24. https://doi.org/10.37012/jtik.v7i1.498

Maulana, A., Danar Dana, R., & Dienwati Nuris, N. (2024). Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1417–1424. https://doi.org/10.36040/jati.v8i2.9024

Pratama, Y., Hendrawan, H., Rasywir, E., Carenina, B. T., & Anggraini, D. R. (2022). Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1232–1240. https://doi.org/10.47065/bits.v4i3.2549

Ramadhani, D. I., Damayanti, O., Thaushiyah, O., & Kadafi, A. R. (2022). Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam. *JURIKOM (Jurnal Riset Komputer)*, *9*(3), 749. https://doi.org/10.30865/jurikom.v9i3.4326

Reviantika, F., Harahap, C. N., & Azhar, Y. (2020). Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means Analysis Earthquake in Java Island Using Clustering K-Means Algorithm. *Jurnal Dinamika Informatika*, *9*(1), 51–60. https://twitter.com/infobmkg

Rosaliyah, I., & Nurhakim, B. (2023). Clustering Kejadian Bencana Alam di Jawa Barat Berdasarkan Jenis Bencana Menggunakan K-Means. *E-Link: Jurnal Teknik Elektro Dan Informatika*, *18*(1), 10. https://doi.org/10.30587/e-link.v18i1.5318

Agustina, M., & Mulyawan. (2023). Implementasi Algoritma K-Means pada Peristiwa Gempa Bumi di Wilayah Jawa Barat. *Jurnal Wahana Informatika (JWI)*, *2*(2), 257–264.

Baldah, A., Duarisah, A. V., & Maulana, R. A. (2023). Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means. *Jurnal Ilmiah Informatika Global*, *14*(2), 31–36. https://doi.org/10.36982/jiig.v14i2.3186

Halik, M. F. Al, & Septiana, L. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering. *Journal of Information System, Applied, Management, Accounting and Research*, *6*(4), 856–870. https://doi.org/10.52362/jisamar.v6i4.939

Hermanto, T. I., Muhyidin, Y., Tinggi, S., Wastukancana, T., Cikopak, J., 53, N., & Barat, J. (2021). Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5Hermanto,*(1), 406.

Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus : Desa Ciomas Bogor). *Jurnal Teknologi Informatika Dan Komputer*, *7*(1), 17–24. https://doi.org/10.37012/jtik.v7i1.498

Maulana, A., Danar Dana, R., & Dienwati Nuris, N. (2024). Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1417–1424. https://doi.org/10.36040/jati.v8i2.9024

Pratama, Y., Hendrawan, H., Rasywir, E., Carenina, B. T., & Anggraini, D. R. (2022). Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1232–1240. https://doi.org/10.47065/bits.v4i3.2549

Ramadhani, D. I., Damayanti, O., Thaushiyah, O., & Kadafi, A. R. (2022). Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam. *JURIKOM (Jurnal Riset Komputer)*, *9*(3), 749. https://doi.org/10.30865/jurikom.v9i3.4326

Reviantika, F., Harahap, C. N., & Azhar, Y. (2020). Analisis Gempa Bumi Pada Pulau Jawa Menggunakan Clustering Algoritma K-Means Analysis Earthquake in Java Island Using Clustering K-Means Algorithm. *Jurnal Dinamika Informatika*, *9*(1), 51–60. https://twitter.com/infobmkg

Rosaliyah, I., & Nurhakim, B. (2023). Clustering Kejadian Bencana Alam di Jawa Barat Berdasarkan Jenis Bencana Menggunakan K-Means. *E-Link: Jurnal Teknik Elektro Dan Informatika*, *18*(1), 10. https://doi.org/10.30587/e-link.v18i1.5318