

Pengelompokan Algoritma K-Means dan K-Medoid Berdasarkan Lokasi Daerah Rawan Bencana di Indonesia dengan Optimasi Elbow, DBI, dan Silhouette

Dedy Hartama^{1*}, Wanayumini², Irfan Sudahri Damanik¹

¹ Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

² Program Studi Teknik Informatika, Universitas Asahan, Indonesia

Email: dedyhartama@amiktunasbangsa.ac.id, wanayumini@yahoo.co.id, irfansudahri@gmail.com

Email Penulis Korespondensi: dedyhartama@amiktunasbangsa.ac.id

Submitted: 26/08/2024; Accepted: 16/09/2024; Published: 27/09/2024

Abstrak—The study examines the use of K-Means and K-Medoids algorithms in the grouping of disaster area locations in Indonesia, with the aim of identifying patterns and optimizing disaster re-sponse strategies. The data used includes geographical and historical information of various disaster events in Indonesia, such as Aceh Besar, Asahan, Badung, Bangkalan, Bekasi, and others. In the clustering process, optimization techniques such as the Elbow Method, the Davies-Bouldin Index (DBI), and the Silhouette Score are used to determine the optimal number of clusters. Research results show that the K-Means algorithm tends to be more stable in dealing with outliers than K-Means, with the results of the DBI (Davies-Bouldin Index) 0.3737248981 and the cluster 7, resulting in the silhouette score of 0.868728638 and cluster 2, resulting at the elbow 98106477130.371 and cluster 2. The Silhouette Score and Elbow index-es also provide a strong indication that the clustering algorithm used is capable of forming significant and meaningful clusters. The study has made important contributions to the optimization of clustering with three methods used so that it can be the basis for authorities in planning and implementing more effective disaster mitigation policies.

Kata Kunci: Optimization; Data Mining; Clustering; DBI; Silhouette; Elbow

Abstract—Penelitian ini mengkaji penggunaan algoritma K-Means dan K-Medoids dalam pengelompokan lokasi daerah bencana di Indonesia, dengan tujuan untuk mengidentifikasi pola dan mengoptimalkan strategi penanggulangan bencana. Data yang digunakan mencakup informasi geografis dan historis dari berbagai kejadian bencana di Indonesia, seperti Aceh Besar, Asahan, Badung, Bangkalan, Bekasi, dan lainnya. Dalam proses clustering, teknik optimasi seperti Elbow Method, Davies-Bouldin Index (DBI), dan Silhouette Score digunakan untuk menentukan jumlah cluster yang optimal. Hasil penelitian menunjukkan bahwa algoritma K-Means cenderung lebih stabil dalam menangani outlier dibandingkan K-Means, dengan hasil Davies-Bouldin Index (DBI) 0.3737248981 dan cluster 7, hasil Silhouette Score 0.868728638 dan cluster 2, hasil Elbow 98106477130.371 dan cluster 2. Melalui analisis, ditemukan titik optimal yang menunjukkan jumlah cluster terbaik. Indeks Silhouette Score dan Elbow juga memberikan indikasi kuat bahwa algoritma clustering yang digunakan mampu membentuk cluster yang signifikan dan bermakna. Studi ini memberikan kontribusi penting dalam optimasi clustering dengan 3 metode yang digunakan sehingga dapat menjadi dasar bagi pihak berwenang dalam merencanakan dan mengimplementasikan kebijakan mitigasi bencana yang lebih efektif.

Keywords: Optimasi; Data Mining; Clustering; DBI; Silhouette; Elbow;

1. PENDAHULUAN

Indonesia sebagai negara kepulauan yang terletak di wilayah Cincin Api Pasifik, sering kali menghadapi berbagai bencana alam seperti gempa bumi, letusan gunung berapi, tsunami, banjir, dan tanah longsor[1]. Tingginya frekuensi bencana ini menuntut adanya sistem yang efektif untuk mengelompokkan wilayah-wilayah yang rawan bencana guna memudahkan perencanaan dan implementasi strategi mitigasi yang tepat[2]. Pendekatan *clustering* telah banyak digunakan dalam berbagai disiplin ilmu untuk mengelompokkan data berdasarkan karakteristik yang serupa[3], [4], [5]. Algoritma K-Means dan K-Medoids adalah dua metode *clustering* yang populer karena kemampuannya dalam mengelompokkan data dengan jumlah yang besar dan kecepatan komputasinya[6], [7], [8], [9]. Metode pengelompokan data yang dikenal sebagai algoritma K-Means *clustering* lebih sederhana dan mudah digunakan[10]. Proses inti dari algoritma ini adalah membagi atau membagi sejumlah data ke dalam K *cluster* dengan menggunakan jarak rata-rata (Mean) terdekat ke seluruh data *cluster*[11]. Jarak rata-rata ini kemudian dihitung kembali untuk titik data pada iterasi berikutnya. Keunggulan algoritma *clustering* K-Means adalah waktu komputasi yang relatif cepat. Namun, K-Medoids berfungsi sebagai algoritma *clustering* yang memecah sejumlah data ke dalam kelompok K dengan menggunakan data medoids sebagai titik pusat pada iterasi berikutnya[12]. Kedua algoritma ini menggunakan Euclidean Distance untuk menemukan jarak terkecil rata-rata dalam pengelompokkan data mereka. Algoritma K-Medoids menentukan titik data sebagai data representatif. Ini dilakukan untuk mengurangi jarak antara range data yang berbeda. Sementara Algoritma K-Means hanya menghitung jarak rata-rata seluruh data pada setiap *cluster*, K-Medoids dinilai lebih baik[13].

Beberapa studi literatur tentang Penelitian sebelumnya mengenai pengelompokkan analisis *cluster* optimal telah banyak dilakukan sebelumnya. Penelitian yang membandingkan 2 algoritma *clustering* ini salah satunya dilakukan oleh Siti Ramadhani yang diterapkan dalam data mining, disimpulkan bahwa algoritma K-Medoids menghasilkan nilai DBI yang lebih baik (1,56) dibandingkan dengan K-Means (2,79), menunjukkan bahwa K-Medoids lebih unggul dalam hal keakuratan *clustering*. Namun, dalam hal waktu komputasi, K-Means lebih cepat dengan rata-rata waktu 1 detik dibandingkan dengan K-Medoids yang membutuhkan rata-rata 26,7778 detik untuk

mengelompokkan 50 dokumen skripsi[9]. Terdapat Penelitian yang relevan dengan topik optimasi yang diteliti oleh Andre Winarta dengan Judul Penelitian “Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python”. Dengan penerapan Bahasa python dengan topik pengguna narkoba, dengan hasil Penelitian menggunakan algoritma Elbow untuk mendapatkan data cluster optimal dengan metode data kelompok K-Means. Hasilnya menunjukkan bahwa metode ini sangat baik dalam mendapatkan cluster optimal dengan skor 3, SSE 1257.862, dan tes 5.

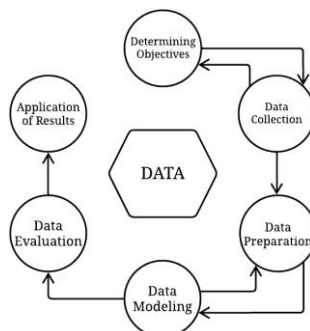
Permasalahan utama yang dihadapi adalah menentukan algoritma yang paling efektif, mengoptimalkan jumlah *cluster*, serta mengevaluasi kualitas *clustering*. Untuk itu, penelitian ini akan menggunakan metode optimasi seperti Elbow, Davies-Bouldin Index (DBI), dan Silhouette Score. Tujuannya adalah mendapatkan pembagian wilayah yang paling representatif, yang dapat mendukung upaya mitigasi dan penanganan bencana secara lebih efektif. Untuk mengatasi permasalahan ini, beberapa teknik optimasi seperti Elbow Method[14], Davies-Bouldin Index (DBI)[15], dan Silhouette Score[16]. Elbow Method membantu dalam menentukan jumlah *cluster* optimal dengan mengidentifikasi titik infleksi pada grafik[17], DBI memberikan ukuran seberapa baik *cluster* terbentuk berdasarkan rasio antara jarak antar *cluster* dan jarak dalam *cluster*[18], sedangkan Silhouette Score mengukur seberapa mirip data dengan *cluster* mereka sendiri dibandingkan dengan *cluster* lainnya[19]. Untuk memastikan hasil *clustering* yang optimal dan dapat diandalkan, ketiga teknik ini akan digunakan. Penelitian ini diharapkan dapat memberikan solusi yang lebih akurat untuk pemetaan wilayah yang rentan terhadap bencana dengan menggabungkan ketiga pendekatan tersebut. Hasilnya akan dapat digunakan sebagai dasar untuk pengambilan keputusan yang lebih baik dalam manajemen bencana. Proses ini meningkatkan pembagian *cluster* dan keakuratan hasil, yang pada akhirnya akan mendorong penanganan bencana yang lebih proaktif dan tepat sasaran.

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dan K-Medoids[20] untuk mengelompokkan lokasi di daerah bencana di Indonesia. Mereka juga akan menggunakan teknik optimasi untuk memastikan kualitas pengelompokan yang dihasilkan. Akibatnya, hasil penelitian diharapkan dapat membantu membangun strategi mitigasi bencana yang lebih efisien dan efektif, serta membantu otoritas memprioritaskan sumber daya untuk daerah yang paling membutuhkan. Penggunaan algoritma K-Means dan K-Medoids dalam analisis kelompok ini akan menghasilkan pemetaan risiko bencana yang lebih akurat, yang akan memungkinkan pengambilan keputusan yang lebih strategis tentang alokasi sumber daya dan upaya penanggulangan bencana jika diimplementasikan dengan benar. Dengan memberi pemerintah dan lembaga terkait pedoman yang lebih jelas, metode ini diharapkan dapat mengurangi dampak bencana secara signifikan.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining adalah proses penemuan pola, hubungan, dan informasi yang bermanfaat dari sejumlah besar data, yang sering kali tersembunyi atau tidak terlihat secara langsung. Proses ini melibatkan penggunaan teknik statistik, matematika, dan kecerdasan buatan untuk menganalisis data dari berbagai sudut pandang dan merangkum informasi tersebut menjadi bentuk yang dapat digunakan[21]. Data mining biasanya diterapkan pada dataset yang sangat besar atau kompleks, seperti database perusahaan, data web, atau data sensor, untuk mengidentifikasi tren atau pola yang dapat membantu dalam pengambilan keputusan[22]. Dalam konteks bisnis, data mining sering digunakan untuk memahami perilaku pelanggan, mengoptimalkan operasi, dan mengidentifikasi peluang pasar baru. Secara teknis, data mining melibatkan berbagai metode seperti klasifikasi, *clustering*, regresi, dan asosiasi. Klasifikasi digunakan untuk mengelompokkan data ke dalam kategori yang telah ditentukan sebelumnya, sementara *clustering* mengidentifikasi kelompok data yang serupa tanpa kategori awal.



Gambar 1. Proses Data Mining

Pada Gambar 1, proses data mining melibatkan beberapa langkah utama yang saling terkait. Pertama, kita menentukan tujuan atau masalah yang ingin dipecahkan, yang akan menjadi panduan seluruh proses. Setelah itu, data yang relevan dikumpulkan dari berbagai sumber, diikuti dengan tahap persiapan data, di mana data dibersihkan dan disiapkan agar siap untuk dianalisis. Pada langkah berikutnya, data tersebut diproses melalui teknik pemodelan

seperti clustering atau klasifikasi, yang bertujuan menemukan pola atau informasi penting. Hasil dari pemodelan ini kemudian dievaluasi untuk memastikan bahwa model yang digunakan efektif dan akurat. Jika hasilnya memuaskan, temuan tersebut dapat diterapkan dalam pengambilan keputusan atau strategi praktis. Proses ini bersifat berulang, sehingga jika diperlukan, langkah-langkah tertentu bisa dilakukan kembali untuk memperbaiki hasil.

2.2 Optimasi

Optimasi adalah proses mencari solusi terbaik dari sekumpulan alternatif yang tersedia, dengan tujuan memaksimalkan atau meminimalkan suatu fungsi tujuan tertentu. Dalam konteks data mining, optimasi digunakan untuk meningkatkan kinerja algoritma dan model, seperti meningkatkan akurasi prediksi, mempercepat waktu pemrosesan, atau mengurangi penggunaan sumber daya[23]. Dalam implementasinya, optimasi yang digunakan berupa *Davies Bouildin-Index*, *Silhouette Coefficient*, dan *Elbow Method*. *Davies Bouildin-Index* menghitung rasio antara jarak rata-rata dari setiap titik di dalam cluster ke pusat clusternya sendiri, Nilai DBI yang lebih rendah menunjukkan clustering yang lebih baik. *Silhouette Coefficient* rata-rata dari semua data point dalam dataset sering digunakan untuk mengevaluasi keseluruhan kualitas clustering. Semakin tinggi nilai rata-rata Silhouette Score, semakin baik hasil clustering. *Elbow Method* menemukan titik di mana penambahan cluster baru tidak memberikan peningkatan yang signifikan dalam hal variasi data yang dijelaskan oleh model[24].

2.3 K-Means

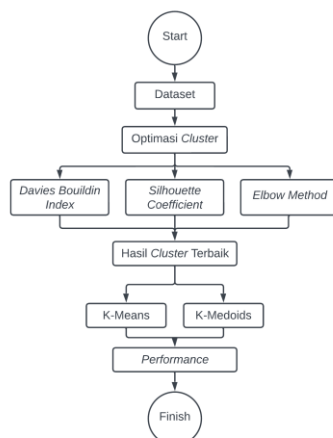
K-Means adalah salah satu algoritma *clustering* yang paling populer dan sering digunakan dalam data mining untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok yang memiliki karakteristik serupa[25]. Algoritma ini bekerja dengan cara membagi data ke dalam *cluster* berdasarkan kedekatan atau kemiripan antar data. Setiap *cluster* diwakili oleh pusat (centroid) yang merupakan rata-rata dari semua data dalam *cluster* tersebut. Proses K-Means dimulai dengan menentukan jumlah *cluster* *K* yang diinginkan, kemudian memilih *K* titik awal secara acak sebagai pusat *cluster*. Selanjutnya, setiap data diassign ke pusat *cluster* terdekat, dan pusat *cluster* kemudian diperbarui dengan menghitung rata-rata posisi dari semua data yang ada di dalamnya[26]. Langkah ini diulang secara iteratif hingga tidak ada perubahan signifikan dalam posisi pusat *cluster* atau data yang berpindah *cluster*.

2.4 K-Medoids

K-Medoids adalah algoritma *clustering* yang mirip dengan K-Means, namun dengan pendekatan yang sedikit berbeda dalam menentukan pusat *cluster*[26]. Alih-alih menggunakan rata-rata dari data dalam *cluster* sebagai pusat (centroid) seperti pada K-Means, K-Medoids menggunakan salah satu data asli dalam *cluster* sebagai pusat (medoid). Medoid adalah data yang memiliki total jarak terendah ke semua data lain dalam *cluster*, sehingga lebih robust terhadap outlier dan data ekstrem dibandingkan dengan K-Means. Proses K-Medoids dimulai dengan pemilihan acak *K* data sebagai medoid awal, kemudian setiap data diassign ke medoid terdekat berdasarkan jarak yang telah ditentukan, seperti jarak Euclidean atau Manhattan. Setelah itu, medoid diperbarui dengan mengevaluasi apakah ada data lain dalam *cluster* yang bisa menjadi medoid baru dengan menurunkan total jarak keseluruhan dalam *cluster*[27].

2.5 Tahapan Penelitian

Rancangan yang dibangun berdasarkan tahapan awal yaitu melakukan processing dataset Lokasi Daerah Bencana yang digunakan dalam penelitian. Hasil dataset tersebut selanjutnya digunakan untuk mencari *cluster* optimal berdasarkan metode *Davies Bouildin-Index*, *Silhouette Coefficient*, dan *Elbow Method*. Dari hasil yang diperoleh, menghasilkan jumlah *cluster* terbaik berdasarkan kriteria nya masing-masing. Proses selanjutnya, implementasi metode K-Means dan K-Medoids dan mencari hasil terbaik dari kedua metode tersebut. Tahapan itu tertulis pada Gambar 2 berikut.



Gambar 2. Rancangan Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Analisis data merupakan proses sistematis untuk menginspeksi, membersihkan, memodelkan, dan menginterpretasi data dengan tujuan mengidentifikasi pola, tren, hubungan, dan wawasan yang mendalam. Proses ini melibatkan berbagai teknik statistik dan komputasional untuk mengubah data mentah menjadi informasi yang berarti, yang kemudian dapat digunakan untuk pengambilan keputusan yang lebih baik. Data yang dianalisis bisa berasal dari berbagai sumber, seperti basis data internal, survei, sensor, atau transaksi bisnis, dan dapat mencakup berbagai jenis data, seperti numerik, kategorikal, teks, atau waktu. Proses data analysis mencakup beberapa tahapan, dimulai dari pengumpulan data, pembersihan data untuk menghilangkan anomali atau noise, transformasi data menjadi format yang dapat dianalisis, hingga analisis eksploratif Teknik-teknik seperti regresi, *clustering*, analisis varians, dan visualisasi data digunakan untuk menemukan korelasi, dan mengidentifikasi insight yang berguna. Hasil analisis dataset terdapat 7 variabel dan 500 record data yang digunakan, diantaranya terdapat variable Meninggal, Hilang, Terluka, Rumah Rumah Rusak, Rumah Terendam, Fasum Rusak, Jumlah Bencana. Hasil dari data analysis membantu processing algoritma yang digunakan untuk mengidentifikasi pola maupun *performance*, dan membuat hasil keputusan yang didasarkan pada fakta dan bukan asumsi.

Tabel 1. Jenis jenis database

Kabupaten	Meninggal	Hilang	Terluka	Rumah Rusak	Rumah Terendam	Fasum Rusak	Jlh Bencana
Aceh Barat	3	0	0	84	10815	22	60
Aceh Barat Daya	0	0	1	80	1045	3	33
Aceh Besar	6	3	0	345	1898	47	125
Aceh Jaya	0	0	0	14	12395	9	60
Aceh Selatan	3	6	1	88	14419	10	91
.....
Wonosobo	10	1	9	602	122	5	74

3.2 Tipe Data

Implementasi atau penetapan tipe data digunakan untuk mempermudah atau dapat dibaca oleh sistem yang digunakan. Sistem maupun software yang digunakan adalah Python, Google Coolab, Orange, dan Rapidminer V10. Tipe data yang digunakan sebagai berikut.

Tabel 2. Tipe Data Lokasi Bencana

Tipe Data	
Variabel	Value
Meninggal	Integer
Hilang	Integer
Terluka	Integer
Rumah Rusak	Integer
Rumah Terendam	Integer
Fasum Rusak	Integer
Jlh Bencana	Integer

3.3 Analisis Python & Google Colab

Berdasarkan tipe data yang digunakan perlu dilakukan analisis Python untuk menentukan optimasi *cluster*. Optimasi digunakan untuk mencari nilai terbaik berdasarkan metode-metode yang diterapkan diantaranya *Davies Bouildin-Index*, *Silhouette Coefficient*, dan *Elbow Method*.

```
def evaluate_clustering(X, max_clusters=10):
    dbi_kmeans = []
    silhouette_kmeans = []
    wcss_kmeans = []

    dbi_kmedoids = []
    silhouette_kmedoids = []
    wcss_kmedoids = []

    for i in range(2, max_clusters + 1):
        # KMeans
        kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
        labels_kmeans = kmeans.fit_predict(X)
        dbi_kmeans.append(davies_bouldin_score(X, labels_kmeans))
        silhouette_kmeans.append(silhouette_score(X, labels_kmeans))
        wcss_kmeans.append(kmeans.inertia_)

        # KMedoids
        kmedoids = KMedoids(n_clusters=i, random_state=0)
        labels_kmedoids = kmedoids.fit_predict(X)
        dbi_kmedoids.append(davies_bouldin_score(X, labels_kmedoids))
        silhouette_kmedoids.append(silhouette_score(X, labels_kmedoids))
        wcss_kmedoids.append(kmedoids.inertia_)

    return dbi_kmeans, silhouette_kmeans, wcss_kmeans, dbi_kmedoids, silhouette_kmedoids, wcss_kmedoids
```

Gambar 2. Penerapan Algoritma K-Means dan K-Medoids



Pada Gambar 2 dilakukan evaluasi *clustering* dengan menentukan Nilai DBI, Silhouette, dan Elbow dari 2 metode yang digunakan. Evaluasi yang dilakukan dengan menentukan rentang *cluster* dari 2 hingga 10 *cluster*. Hasil evaluasi *cluster* dari metode K-Means dan K-Medoids sebagai berikut:

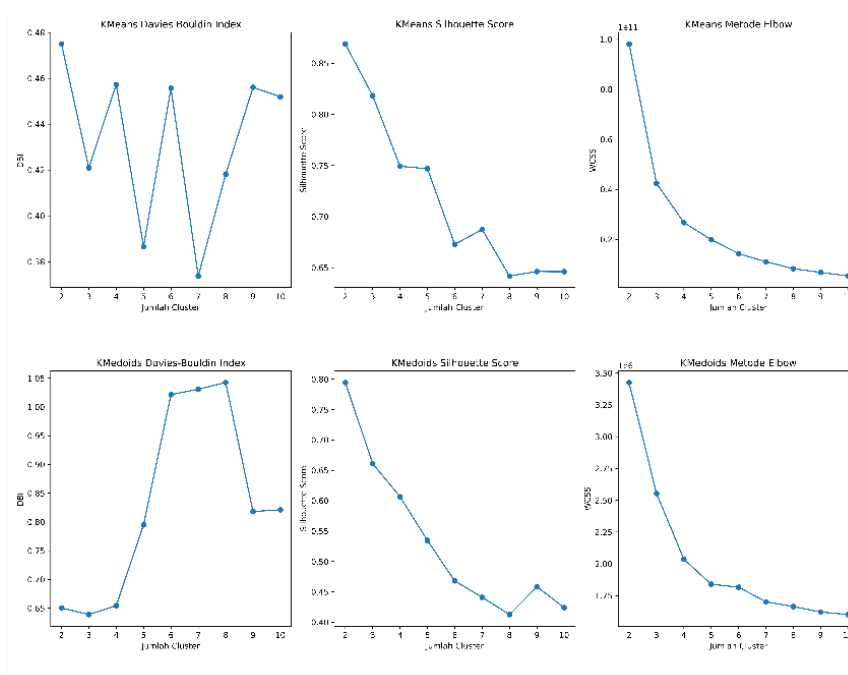
Tabel 3. Hasil Analisis K-Means

K-Means			
Cluster	DBI	SC	Hasil SC
2	0.475	0.869	98106477130
3	0.421	0.818	42407848736
4	0.457	0.749	26744006301
5	0.387	0.747	19901262117
6	0.456	0.673	14287344646
7	0.374	0.687	11057095225
8	0.418	0.642	8289597744
9	0.456	0.646	6834184297
10	0.452	0.646	5371980542

Tabel 4. Hasil Analisis K-Medoids

K-Medoids			
Cluster	DBI	SC	Hasil SC
2	0.650	0.795	3424986
3	0.639	0.661	2554057
4	0.654	0.607	2036935
5	0.795	0.535	1839360
6	1.022	0.468	1816571
7	1.031	0.441	1701456
8	1.042	0.413	1664471
9	0.818	0.458	1621209
10	0.821	0.424	1601421

Hasil plot yang didapatkan dari analisis *cluster* optimal dengan keterangan Gambar sebagai berikut.



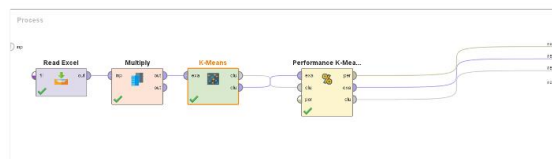
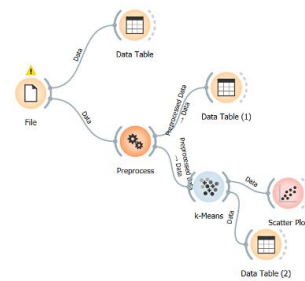
Gambar 3. Hasil *Cluster* Optimal

Dari analisis pada Gambar 3, hasil analisis terbaik dari masing metode dan optimasi dari K-Means dan K-Medoids. Analisis terbaik Davies Bouldin-Index dengan nilai 0.37372490 di 7 Cluster di metode K-Means, Silhouette Coefficient dengan nilai 0.8687286 di 2 Cluster di metode K-Means, dan Elbow Method dengan nilai 98106477130.371 di 2 Cluster di metode K-Means. Hasil dari analisis optimasi cluster mendapatkan metode K-

Means merupakan metode terbaik dalam pengelompokan Lokasi Daerah Bencana dan akan melakukan pengelompokan dengan menggunakan tools Orange an Rapidminer.

3.4 Rancangan Orange dan RapidminerV10

Data dimasukkan ke Microsoft Excel dimulai dari kolom A1, Orange adalah dan Rapidminer perangkat lunak bantuan yang digunakan untuk memproses data. Ketika data dimasukkan ke Orange maupun Rapidminer, tentukan Role Set dengan ketentuan masing-masing variable. Kemudian, menggunakan set data Bencana, algoritma K-Means diterapkan dengan jumlah *cluster* 2 berdasarkan analisis *cluster* optimal. Parameter yang digunakan dalam tools Orange dan Rapidminer sebagai berikut:

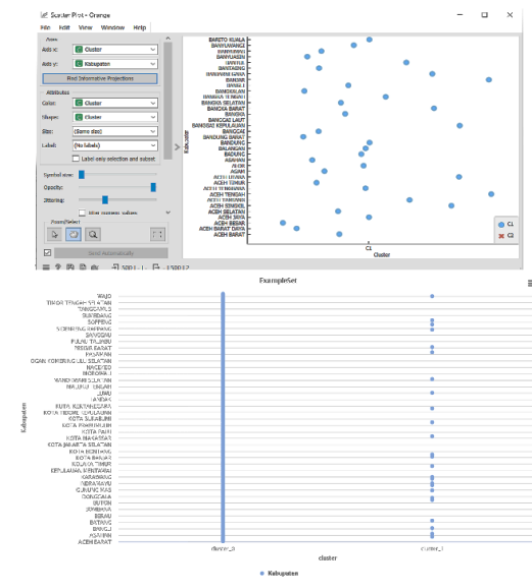


Gambar 4. Design Algoritma K-Means

Tabel 5. Parameter Orange dan Rapidminer

Algoritma	Parameter	Value
K-Means (Orange)	Preprocess	Normalize to interval [0, 1]
	Number of <i>Clusters</i>	2
	Scatter Plot	Axis x : <i>Cluster</i> , Axis y : Kabupaten
	Data Table	-
K-Means (Rapidminer)	Number of <i>Clusters</i>	2
Performance	Main criterion	Avg. within centroid distance

Dari Gambar 4, pada tools orange dilakukan proses *file* berupa insert dataset yang digunakan, dilanjutkan dengan Preprocess untuk transformasi dataset dengan cara interval antara 1 dan 0 sehingga analisis K-Means dapat dilakukan. Untuk membuat data lebih spesifik, nilai dihitung berdasarkan data lokasi daerah bencana. Hasil *cluster* adalah sebagai berikut:



Gambar 5. Hasil *Cluster* K-means



Hasil *cluster* menunjukkan model dengan *cluster_0* : 475 items dan *cluster_1* : 25 items. Performance yang diperoleh dengan parameter Avg. within centroid distance: -196212954.261, dengan jarak rata-rata *cluster_0* sebesar -104094274.744, dan jarak rata-rata *cluster_1* sebesar -1946467865.078. Secara keseluruhan, metode K-Means menunjukkan performa yang konsisten dalam pengelompokan data lokasi daerah bencana di Indonesia, dengan jumlah cluster yang optimal bervariasi tergantung pada metode evaluasi yang digunakan. Meskipun terdapat variasi dalam jumlah cluster yang optimal, hasil ini memberikan wawasan yang berguna untuk pengelompokan daerah bencana yang lebih akurat dan efektif.

4. KESIMPULAN

Studi ini menemukan teknik clustering yang paling efektif untuk mengelompokkan lokasi daerah bencana di Indonesia. Hasilnya menunjukkan bahwa metode K-Means memberikan hasil yang paling optimal setelah penggunaan metode Elbow, Davies-Bouldin Index (DBI), dan Silhouette Coefficient untuk mengoptimalkan hasil. Hasil analisis menunjukkan bahwa metode K-Means pada tujuh cluster menghasilkan nilai DBI terbaik sebesar 0,37372490, yang menunjukkan kejelasan pemisahan cluster yang baik. Selain itu, metode Elbow menghasilkan hasil optimal pada dua cluster dengan nilai 98106477130.371, dan Silhouette Coefficient dengan nilai 0,8687286 menunjukkan kualitas clustering yang tinggi dalam interpretasi dan konsistensi antar item cluster. Metode K-Means terbukti unggul dalam pengelompokan lokasi daerah bencana dibandingkan K-Medoids selama proses optimasi. Hasil clustering dengan alat seperti Orange dan Rapidminer menunjukkan bahwa model K-Means mengelompokkan data menjadi dua cluster: *cluster_0* memiliki 475 item dan *cluster_1* memiliki 25 item. Kinerja model diukur dengan parameter rata-rata dalam jarak centroid sebesar -196212954.261, dengan jarak rata-rata untuk *cluster_0* adalah -104094274.744, dan jarak rata-rata untuk *cluster_1* adalah -1946467865.078. Oleh karena itu, metode K-Means disarankan untuk digunakan dalam pengelompokan data terkait lokasi bencana di Indonesia.

ACKNOWLEDGMENTS

Para peneliti mengucapkan terima kasih kepada Direktorat Riset, Teknologi, dan Pengabdian Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia yang telah mendanai penelitian ini di bawah naungan Skema Penelitian Dasar (Fundamental Research) untuk tahun 2024.

REFERENCES

- [1] S. Nurjanah And E. Mursalin, "Pentingnya Mitigasi Bencana Alam Longsor Lahan: Studi Persepsi Mahasiswa," *Jurnal Basicedu*, Vol. 6, No. 1, Pp. 515–523, Dec. 2021, Doi: 10.31004/basicedu.V6i1.1937.
- [2] K. P. Sinaga And M. S. Yang, "Unsupervised K-Means Clustering Algorithm," *Ieee Access*, Vol. 8, Pp. 80716–80727, 2020, Doi: 10.1109/Access.2020.2988796.
- [3] P. Artioli, A. Maci, And A. Magri, "A Comprehensive Investigation Of Clustering Algorithms For User And Entity Behavior Analytics," *Front Big Data*, Vol. 7, 2024, Doi: 10.3389/fdata.2024.1375818.
- [4] D. Jollyta, S. Efendi, M. Zarlis, And H. Mawengkang, "Analysis Of An Optimal Cluster Approach: A Review Paper," In *Journal Of Physics: Conference Series*, Institute Of Physics, 2023. Doi: 10.1088/1742-6596/2421/1/012015.
- [5] A. A. Jalal And B. H. Ali, "Text Documents Clustering Using Data Mining Techniques," *International Journal Of Electrical And Computer Engineering*, Vol. 11, No. 1, Pp. 664–670, Feb. 2021, Doi: 10.11591/ijece.V11i1.Pp664-670.
- [6] Y. E. Wella, O. Okfalisa, F. Insani, F. Saeed, And A. R. C. Hussin, "Service Quality Dealer Identification: The Optimization Of K-Means Clustering," *Sinergi (Indonesia)*, Vol. 27, No. 3, Pp. 433–442, 2023, Doi: 10.22441/Sinergi.2023.3.014.
- [7] A. A. Aldino, D. Darwis, A. T. Prastowo, And C. Sujana, "Implementation Of K-Means Algorithm For Clustering Corn Planting Feasibility Area In South Lampung Regency," In *Journal Of Physics: Conference Series*, Iop Publishing Ltd, Jan. 2021. Doi: 10.1088/1742-6596/1751/1/012038.
- [8] I. Firman Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, And N. Humaya, "Application Of Data Mining With The K-Means Clustering Method And Davies Bouldin Index For Grouping Imdb Movies," 2022. [Online]. Available: [Http://jurnal.polibatam.ac.id/index.php/jaic](http://jurnal.polibatam.ac.id/index.php/jaic)
- [9] S. Ramadhani, D. Azzahra, And T. Z, "Comparison Of K-Means And K-Medoids Algorithms In Text Mining Based On Davies Bouldin Index Testing For Classification Of Student's Thesis," *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, Vol. 13, No. 1, Pp. 24–33, May 2022, Doi: 10.31849/Digitalzone.V13i1.9292.
- [10] N. Fitrianti Fahrudin, "Mind (Multimedia Artificial Intelligent Networking Database Sequence Clustering In Process Mining For Business Process Analysis Using K-Means," *Journal Mind Journal | Issn*, Vol. 6, No. 1, Pp. 16–30, 2021, Doi: 10.26760/Mindjournal.V6i1.16.
- [11] S. Solikhun, V. Yasin, And Donni Nasution, "Optimization Of The Number Of Clusters Of The K-Means Method In Grouping Egg Production Data In Indonesia," *International Journal Of Artificial Intelligence & Robotics (Ijair)*, Vol. 4, No. 1, Pp. 39–47, Jun. 2022, Doi: 10.25139/Ijair.V4i1.4328.
- [12] Simeftiany Indrilemta Lomo, Endang Darmawan, And Sugiyarto, "Cluster Analysis Of Type Ii Diabetes Mellitus Patients With The Fuzzy C-Means Method," *Annals Of Mathematical Modeling*, Vol. 3, No. 1, Pp. 24–31, Jun. 2023, Doi: 10.33292/Amm.V3i1.28.

- [13] M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, And K. Sadatdiynov, “A Survey Of Data Partitioning And Sampling Methods To Support Big Data Analysis,” Jun. 01, 2020, *Tsinghua University Press*. Doi: 10.26599/Bdma.2019.9020015.
- [14] R. Sammouda And A. El-Zaart, “An Optimized Approach For Prostate Image Segmentation Using K-Means Clustering Algorithm With Elbow Method,” *Comput Intell Neurosci*, Vol. 2021, 2021, Doi: 10.1155/2021/4553832.
- [15] S. E. A. Buananta, M. A. Ahmad, J. Mahmood, And P. Paradise, “Identification Of Evaluation Results In E-Banking Services Transaction For Product Recommendation Using The Birch And Davies Bouldin Index Method,” *Jurnal Infotel*, Vol. 16, No. 2, Pp. 427–440, May 2024, Doi: 10.20895/Infotel.V16i2.1116.
- [16] M. Shutaywi And N. N. Kachouie, “Silhouette Analysis For Performance Evaluation In Machine Learning With Applications To Clustering,” *Entropy*, Vol. 23, No. 6, Jun. 2021, Doi: 10.3390/E23060759.
- [17] G. J. Oyewole And G. A. Thopil, “Data Clustering: Application And Trends,” *Artif Intell Rev*, Vol. 56, No. 7, Pp. 6439–6475, Jul. 2023, Doi: 10.1007/S10462-022-10325-Y.
- [18] S. N. Safitri, Haryono Setiadi, And E. Suryani, “Educational Data Mining Using Cluster Analysis Methods And Decision Trees Based On Log Mining,” *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 6, No. 3, Pp. 448–456, Jul. 2022, Doi: 10.29207/Resti.V6i3.3935.
- [19] A. Tohir, T. Kristianti, M. Indrasari, J. Kristiyono, And U. Supriatna, “Datamining Clustering In Grouping Adolescents And Adults With Information Technology And Computer Skills By Profession,” In *Aip Conference Proceedings*, American Institute Of Physics, Feb. 2024. Doi: 10.1063/5.0183891.
- [20] B. K. Khotimah, F. Agustina, O. R. Puspitarini, A. D. Cahyani, Y. Kustiyahningsih, And D. R. Anamisa, “Hyperparameters And Centroid Improvements In The K-Medoids Method For Grouping Processed Beef Smes,” *Communications In Mathematical Biology And Neuroscience*, Vol. 2024, 2024, Doi: 10.28919/Cmbn/8369.
- [21] C. Anastassopoulou, L. Russo, A. Tsakris, And C. Siettos, “Data-Based Analysis, Modelling And Forecasting Of The Covid-19 Outbreak,” 2020.
- [22] C. W. L. Ho, J. Ali, And K. Caals, “Ensuring Trustworthy Use Of Artificial Intelligence And Big Data Analytics In Health Insurance,” *Bull World Health Organ*, Vol. 98, No. 4, Pp. 263–269, Apr. 2020, Doi: 10.2471/Bl.19.234732.
- [23] D. Riana *Et Al.*, “Identifikasi Citra Pap Smear Repomedunm Dengan Menggunakan K-Means Clustering Dan Glcm,” *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 6, No. 1, Pp. 1–8, Jan. 2022, Doi: 10.29207/Resti.V6i1.3495.
- [24] Anggara Cahya Putra And K. Dwi Hartomo, “Optimalisasi Penyaluran Bantuan Pemerintah Untuk Umkm Menggunakan Metode Fuzzy C-Means,” *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 5, No. 3, Pp. 474–482, Jun. 2021, Doi: 10.29207/Resti.V5i3.2980.
- [25] F. Jingyuan, Q. Xinyi, H. Rongquan, M. Jie, And W. Qingjun, “Gene Expression Profiles For An Immunoscore Model In Bone And Soft Tissue Sarcoma,” *Aging*, Vol. 13, Pp. 13708–13725, May 2021.
- [26] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, And I. H. Sarker, “An Improved K-Means Clustering Algorithm Towards An Efficient Data-Driven Modeling,” *Annals Of Data Science*, 2022, Doi: 10.1007/S40745-022-00428-2.
- [27] A. Bandyopadhyay, K. Deb, A. Chakraborty, A. Das, And R. Bag, “A Neighborhood Impact Driven K-Medoid Clustering And Fuzzy Logic Blended Approach For High Density Impulse Noise Detection And Removal,” *Traitement Du Signal*, Vol. 39, No. 5, Pp. 1737–1749, Oct. 2022, Doi: 10.18280/Ts.390532.