
Algoritma *K-Means* untuk Mengelompokkan Daerah Rawan Bencana di Sumatera Selatan

Gerry Setapati¹, Dian Hafidh Zulfikar²

¹Prodi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Raden Fatah Palembang

²Prodi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Raden Intan Lampung

¹setapatigerry@gmail.com , ²dianhafidhzulfikar_uin@radenintan.ac.id

ABSTRACT

Indonesia is located in a region where multiple tectonic plates converge, making many areas highly prone to natural disasters, including South Sumatra Province. This study aims to classify disaster-prone areas in South Sumatra using data mining techniques through a clustering method. The algorithm employed for clustering is *K-Means*, a non-hierarchical clustering method that groups data based on similarity. Disaster data with similar characteristics are grouped into the same cluster, while those with different characteristics are placed in separate clusters. The study's results categorize disaster-prone areas into three levels: low-risk, medium-risk, and high-risk regions. These findings are expected to serve as valuable additional information for the government in devising disaster management strategies in South Sumatra. By providing a clearer understanding of disaster risk levels across regions, this study supports more targeted mitigation efforts, enhanced preparedness, and effective response measures. This research highlights the importance of leveraging technology, such as data mining, to facilitate efficient, data-driven disaster risk analysis and management.

Keywords: *Natural Disasters, Data Mining, K-Means, Clustering, Disaster-Prone*

ABSTRAK

Indonesia merupakan negara yang terletak di wilayah pertemuan berbagai lempeng tektonik, sehingga banyak daerahnya memiliki tingkat kerawanan tinggi dari bencana alam, termasuk di Provinsi Sumatera Selatan. Penelitian ini berupaya untuk mengelompokkan wilayah-wilayah rawan bencana di Sumatera Selatan dengan menerapkan teknik data mining menggunakan metode *clustering*. Metode yang digunakan dalam pembentukan klaster adalah *K-Means*, sebuah metode *clustering non-hierarkis* yang mampu mengelompokkan data berdasarkan tingkat kesamaan. Data bencana yang memiliki karakteristik serupa akan dikelompokkan dalam satu klaster, sedangkan data dengan karakteristik yang berbeda akan dimasukkan ke klaster lain. Hasil penelitian ini menghasilkan pengelompokan daerah rawan bencana ke dalam tiga kategori, yaitu daerah dengan tingkat kerawanan rendah, sedang, dan tinggi. Temuan ini dapat menjadi informasi tambahan yang berguna bagi pemerintah dalam upaya penanganan bencana di Sumatera Selatan.

Kata kunci: *Bencana Alama, Data Mining, K-Means, Clustering, Rawan Bencana*

1. PENDAHULUAN

Indonesia memiliki kerentanan tinggi terhadap bencana alam karena letaknya yang berada di pertemuan lempeng tektonik[1]. Kompleksitas ini menyebabkan sebagian besar wilayah Indonesia rentan terhadap berbagai bencana alam[8]. Provinsi Sumatera Selatan adalah salah satu wilayah yang rentan terhadap bencana, terutama karena kombinasi faktor geologi dan aktivitas manusia. Selain bencana akibat faktor geologi, aktivitas manusia seperti penggundulan hutan dan perubahan tata guna lahan juga memperburuk risiko banjir dan tanah longsor di wilayah ini.

Di bawah kondisi geologi yang berlaku, sangat penting untuk memperoleh data komprehensif yang berkaitan dengan kerentanan bencana di wilayah tertentu, terutama di provinsi Sumatera Selatan. Informasi ini dapat dimanifestasikan dalam kategorisasi daerah yang rentan terhadap bencana, sehingga memfasilitasi pengambilan keputusan berdasarkan informasi dalam kerangka manajemen bencana. Pendekatan yang layak untuk mencapai klasifikasi wilayah rawan bencana melibatkan penerapan metodologi penambangan data atau istilahnya data mining[2].

Data mining merupakan suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, tetapi dapat dipahami dan bermanfaat dari database yang besar dan digunakan untuk membuat keputusan bisnis penting [3]. Data mining dimanfaatkan untuk membuat keputusan bisnis dengan mengekstraksi data yang sangat besar dari data warehouse, database, atau repositori [4]. Data mining dikategorikan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi berdasarkan polanya [11]. Data mining adalah inti dari proses *Knowledge Discovery in Database (KDD)*, yang melibatkan algoritma untuk mengeksplorasi data, mengembangkan model, dan menemukan pola yang baru, valid, berguna, dan mudah dipahami dari kumpulan data yang besar dan kompleks[12].

Metode *K-Means* adalah salah satu dari banyak algoritma yang digunakan untuk melakukan data mining. Ini digunakan untuk mengelompokkan data ke dalam kelompok, di mana kelompok data dengan karakteristik yang sama digabungkan dengan kelompok yang berbeda [14].

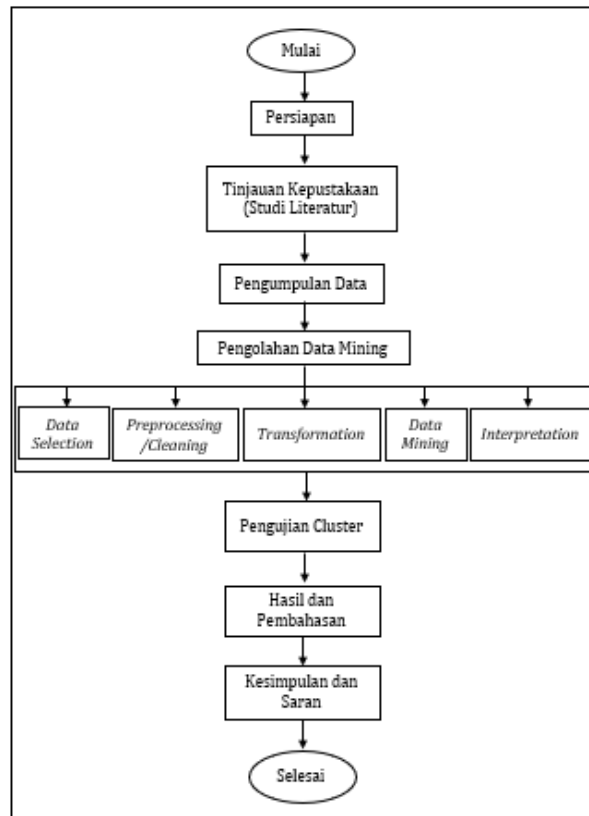
Berdasarkan temuan tersebut, peneliti menyatakan minat yang besar dalam pengelompokan wilayah yang rentan terhadap bencana di Provinsi Sumatera Selatan, memanfaatkan teknik penambangan data bersama algoritma *K-Means*. Oleh karena itu penelitian ini bertujuan untuk mengelompokkan wilayah di Sumatera Selatan berdasarkan tingkat kerawanan bencana menggunakan algoritma *K-Means*, sehingga dapat membantu pemerintah dalam strategi mitigasi bencana.

2. METODE PENELITIAN

Penyelidikan ilmiah ini menggunakan teknik analisis pengelompokan menggunakan algoritma *K-Means*. Aspek mendasar dari metodologi pengelompokan ini melibatkan pengurangan fungsi objektif yang diatur dalam kerangka pengelompokan, yang pada dasarnya bertujuan untuk meminimalisir varians dalam cluster sambil memaksimalkan varians di seluruh cluster yang berbeda.

Sebaliknya, algoritma *K-Means* mewakili pendekatan *non-hierarkis* untuk pengelompokan data yang berusaha untuk mengkategorikan data yang diyakini berada dalam satu atau lebih cluster. Algoritma ini secara sistematis mengatur data ke dalam golongan cluster tunggal sampai entitas yang menunjukkan karakteristik analog berasimilasi ke dalam cluster yang sama, sedangkan yang memiliki karakteristik berbeda dialokasikan ke *cluster* alternatif. Tahapan Penelitian menggambarkan prosedur metodologis yang akan dilaksanakan dan memberikan eksposisi komprehensif dari proses penelitian, yang dirujuk seperti Gambar 1 berikut.

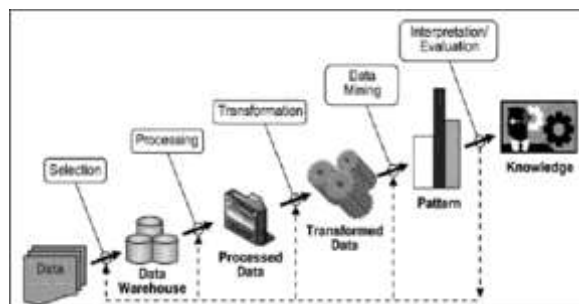
- a. Persiapan: Ini adalah fase menyusun rencana penelitian. antara lain, menemukan lokasi penelitian dan membuat proposal penelitian.
- b. Tinjauan kepustakaan: Dalam kajian pustaka, peristiwa bencana alam dan yang berkaitan diteliti.
- c. Pengumpulan Data: Menggunakan wawancara, observasi, dokumentasi, dan penelitian pustaka.
- d. Pengolahan Data Mining: Tahap *Knowledge Discovery in Database (KDD)* di mana data yang dikumpulkan dipelajari.
- e. Pengujian cluster: Tahap ini melakukan pengujian cluster menggunakan metode *Silhouette Coefficient*.
- f. Hasil dan Pembahasan: Tahapan ini membahas hasil dari algoritma *K-Means* data mining.
- g. Kesimpulan: membuat kesimpulan dari penelitian dan membuat rekomendasi untuk bidang pemerintah Provinsi Sumatera Selatan dan penelitian berikutnya.



Gambar 1. Tahapan Penelitian

2.1. Knowledge Discovery in Database (KDD)

Proses menemukan informasi berguna dan pola-pola yang ada dalam data yang terkandung dalam basis data yang besar, yang sebelumnya tidak diketahui dan mungkin bermanfaat, disebut *Knowledge Discovery in Database (KDD)*. Data mining adalah salah satu dari banyak proses *KDD iterative* [9].



Gambar 2. Tahapan Dalam KDD

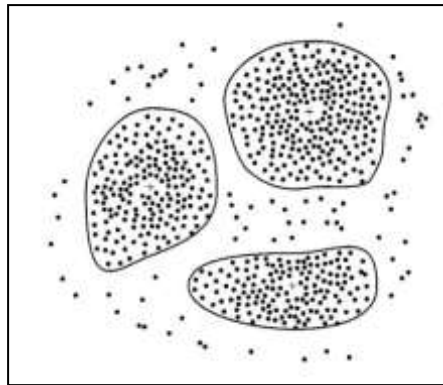
Alut proses KDD secara umum terdiri dari:

1. *Proses Data Selection*
Tahap ini dilakukan pemilihan himpunan data, pembuatan himpunan data target, atau penekanan pada subset variabel, atau sampel data, yang akan digunakan untuk melakukan penemuan. Semua hasil yang dipilih disimpan ke dalam suatu berkas yang berbeda dari basis data operasional.
2. *Preprocessing/Cleaning*
Persiapan dan pembersihan untuk menghilangkan data yang tidak konsisten, duplikasi data, koreksi kesalahan, dan memperkayanya dengan data eksternal yang relevan.
3. *Transformation*

- Tahap ini menggabungkan atau mengubah data dalam bentuk yang lebih sesuai untuk melakukan proses data mining.
4. *Data Mining*
Proses menemukan pola atau informasi yang menarik dalam data tertentu dengan menggunakan metode, teknik yang sesuai dengan tujuan umum proses *KDD*.
 5. *Interpretation/Evaluation*
Interpretasi dan Evaluasi Pola informasi dari proses data mining harus ditampilkan dengan cara yang mudah dipahami oleh pihak yang berkepentingan. Ini adalah tahap dalam proses *KDD* yang disebut interpretasi. Pada proses ini mencakup memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2. Clustering

Misalnya, jarak euclidean adalah jarak terdekat dari centroid dalam data jika kita menggunakan rumus jarak euclidean. Menurut guchi, mengenali masalah atau membuat keputusan tidak selalu menghasilkan hasil yang sama, tetapi cenderung menghasilkan hasil yang sama. Ini disebabkan oleh penggunaan clustering yang luas [4].



Gambar 3. Gambaran *Clustering*

Proses pembagian sejumlah dataset ke dalam kelompok berdasarkan kemiripan berbagai nilai atribut digambarkan dalam Gambar 3. Belajar tanpa pengawasan adalah istilah untuk ini. Tujuannya adalah untuk mendapatkan cluster berkualitas tinggi untuk meningkatkan kecepatan komputasi. Sekumpulan objek data dalam cluster yang sama yang terpisah dan mirip satu sama lain disebut cluster [4].

2.3. Metode *K-Means*

Kumpulan data dengan atribut yang berbeda dikelompokkan ke dalam kumpulan yang berbeda dan kumpulan data dengan atribut yang sama dipartisi ke dalam kumpulan yang sama melalui metode clustering yang dikenal sebagai *K-Means*. Berikut adalah algoritma utama yang digunakan untuk teknik ini [1].

1. Menentukan jumlah cluster (dibagi ke dalam cluster awal K).
2. Alokasikan data ke dalam cluster.
3. Lakukan proses perhitungan dari daftar item, menandai item mana yang harus dimasukkan ke dalam cluster berdasarkan pusat (mean) terdekat. Hitung centroid atau rata-rata dari data yang ada di masing-masing cluster.
4. Kemudian alokasikan masing-masing data ke *centroid* atau rata-rata terdekat. Kemudian kembali ke langkah 3 jika *cluster* data masih berpindah atau nilai *centroid* berubah.

Untuk menemukan pusat terdekat, Anda dapat menggunakan ruang jarak geometris, yang merupakan ruang jarak yang digunakan untuk menghitung jarak antara data dan *centroid*. Persamaan ruang jarak (*distance space*) geometris dapat ditemukan dalam Persamaan 1 dan 2 [15].

$$d(x, y) = \left[\sum_{i=0}^p |x_i - y_i|^2 \right]^{\frac{1}{2}} \quad (1)$$

Objek jarak diformulasikan pada $d(x, y)$, dimana p merupakan dimensi data, x_i merupakan koordinat objek x pada dimensi i , dan y_i merupakan koordinat objek y pada dimensi i .

2.4. Normalization

Strategi linier digunakan untuk mengubah data di rentang nilai sebelumnya ke rentang nilai baru dalam *normalisasi min-max*. Ini menghasilkan nilai perbandingan seimbang antara data sebelum dan sesudah proses, sehingga data menjadi seimbang antara 0 dan 1. Metode ini adalah salah satu metode untuk mengubah data yang kompleks tanpa menghilangkan isi, membuat olahan menjadi lebih mudah [20]. Ini adalah persamaan normalisasi Min-Max:

$$v' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

penjelasan :

v' = Data atribut yang akan dinormalisasi.

X_{min} = Atribut nilai terkecil.

X_{max} = Atribut nilai tertinggi.

2.5. Silhouette Coefficient

Dimana kualitas dan kekuatan sebuah cluster, serta seberapa baik suatu objek ditempatkan di dalamnya, dapat dinilai melalui koefisien silinder [19]. Metode ini menggabungkan teknik *cohesion* dan *separation*. Metode penghitungan koefisien *Silhouette* adalah sebagai berikut [6]:

1. Gunakan Persamaan 3 untuk menghitung jarak data rata-rata. Kemudian, misalkan i terhadap semua data lain dalam satu cluster sebagai berikut.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Yang mana:

$a(i)$ = beda rata-rata objek (i) ke semua objek lain pada A
 $d(i, j)$ = jarak antara data j dengan i
 A = Cluster

2. Jarak data i dengan semua data di cluster lain dan temukan nilai terkecil dengan menggunakan Persamaan 4

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (4)$$

Yang mana:

$d(i, C)$ = Menunjukkan Perbedaan rata-rata objek (i) ke semua objek lain pada C
 C = Merupakan cluster lain selain cluster A atau cluster C tidak sama dengan cluster A .

3. Nilai $d(i, C)$ dihitung untuk setiap C , Persamaan 5 digunakan untuk mendapatkan nilai terkecil.

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Pada Cluster B yang mencapai minimum (yaitu, $d(i, B) = b(i)$) disebut tetangga dari objek (i). Ini adalah cluster terbaik kedua untuk objek (i).

4. Kemudian Nilai *Silhouette Coefficient* dikalkulasi seperti pada Persamaan 6.

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \quad (6)$$

Kriteria Penilaian subjektif pengukuran pengelompokan berdasarkan *Silhouette Coefficient* menurut *Kauffman* dan *Roesseeuw* [10], terlihat pada Tabel 1.

Tabel 1. Nilai *Silhouette Coefficient* dan *Rousseuw*

Nilai <i>Silhouette Coefficient</i>	Struktur
$0.7 < SC \leq 1$	Kuat
$0.5 < SC \leq 0.7$	Sedang
$0.25 < SC \leq 0.5$	Lemah
$SC \leq 0.25$	Tidak terstruktur

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berasal dari BPBD Provinsi Sumatera Selatan, berupa data bencana Sumatera Selatan tahun 2014-2019.

3.1. Data Selection

Bencana alam dan non-alam adalah dua kategori yang termasuk dalam data ini. Data alam hanya digunakan dalam penelitian ini, dan data bencana non-alam tidak digunakan. Kabupaten atau Kota, Jenis dan Jumlah Kerusakan, Korban Terdampak (KT), dan Korban Meninggal (KM) adalah atribut yang digunakan dalam penelitian ini. Tabel 2 menunjukkan ini:

Keterangan:

KT : Korban Terdampak

KM : Korban Meninggal

Tabel 2. Data Selection

Data ke-	Kabupaten/Kota	Jenis Bencana	Jenis dan Jumlah Kerusakan	KT	KM
1	OKU Timur	Banjir	Sawah Terendam 2758,75 Ha Rumah Terendam 599 Unit	0	0
2	Muara Enim	Banjir	Rumah Terendam 1003 Unit	3762	0
3	Muara Enim	Banjir	Rumah Terendam 584 Unit	2699	0
4	Ogan Ilir	Banjir	Rumah Terendam 1627 Unit	8139	0
5	Ogan Ilir	Banjir	Rumah Terendam 1716 Unit	8593	0
6	Ogan Ilir	Banjir	Rumah Terendam 404 Unit	2039	0
7	Ogan Ilir	Banjir	Rumah Terendam 314 Unit	1574	0
8	Ogan Ilir	Banjir	Rumah Terendam 1033 Unit	5169	0
9	Ogan Ilir	Banjir	Rumah Terendam 82 Unit	409	0
10	Ogan Ilir	Banjir	Rumah Terendam 77 Unit	381	0
...
621	Lahat	Banjir Bandang	Rumah Rusak Berat, 13 Unit Rumah Rusak Berat, 10 Unit Rumah Rusak Ringan, 53 Unit	226	0

3.2. Preprocessing/Cleaning

Pembersihan data adalah proses mendeteksi, memperbaiki, atau menghapus data yang tidak sesuai, *missing value*, dan *redundant* dengan mengatur kembali data yang ada pada tabel. Pada tahap ini, pembersihan data diselesaikan secara manual, yaitu dengan memeriksa setiap data secara individual dan menghapus data yang tidak relevan, kurang nilai, dan berlebihan. Studi ini menemukan bahwa dua daftar memiliki data kosong, jadi data harus dihapus. Dengan demikian, jumlah dataset dari 621 menjadi 619.

3.3. Transformation

Proses ini ada tiga prosedur yaitu, *Attribute Construction*, *Agregasi* dan *Normalisasi*. Tahapan *Transformation* adalah tahap mengubah data menjadi bentuk yang sesuai untuk diproses dalam data mining.

1. *Attribute Construction*

Ada fitur baru yang membantu pemrosesan data dalam proses ini. Kabupaten/Kota, Korban Terdampak (KT), Korban Meninggal (KM), Rumah Rusak Ringan (RRR), Rumah Rusak Sedang (RRS), Rumah Rusak Berat (RRB), dan Rumah Terendam (RT) adalah empat atribut baru yang dibuat berdasarkan jenis dan jumlah kerusakan. Tabel 3 menunjukkan hasil dari penambahan atribut.

Keterangan:

KT : Merupakan Korban Terdampak

KM : Merupakan Korban Meninggal

RRR : Merupakan Rumah Rusak Ringan

RRB : Merupakan Rumah Rusak Berat

RT : Merupakan Rumah terendam

Tabel 3. *Attribute Construction*

Data ke-	Kabupaten/Kota	KT	KM	RRR	RRS	RRB	RT
1	OKU Timur	0	0	0	0	0	600
2	Muara Enim	3762	0	0	0	0	1003
3	Muara Enim	2700	0	0	0	0	584
4	Ogan Ilir	8140	0	0	0	0	1626
5	Ogan Ilir	8594	0	0	0	0	1717
6	Ogan Ilir	2040	0	0	0	0	406
7	Ogan Ilir	1573	0	0	0	0	314
8	Ogan Ilir	5170	0	0	0	0	1033
9	Ogan Ilir	410	0	0	0	0	81
10	Ogan Ilir	385	0	0	0	0	76
...
619	Lahat	228	0	53	0	23	0

2. *Agregasi*

Pada tahapan ini akan dilakukan pengkombinasian dua atau lebih objek ke dalam sebuah objek tunggal. Pada prosedur agregasi diselesaikan penjumlahan terhadap setiap atribut yang ada berdasarkan nama Kabupaten/Kota. Terlihat pada Tabel 4.

KT : Korban Terdampak

KM : Korban Meninggal

RRR : Rumah Rusak Ringan

RRB : Rumah Rusak Berat

RT : Rumah terendam

Tabel 4. *Agregasi*

No.	Kabupaten/Kota	KT	KM	RRR	RRS	RRB	RT
1.	Banyuasin	13759	0	267	5	193	2331
2.	Empat Lawang	12446	2	79	0	50	1992
3.	Lahat	2070	0	53	5	78	391
4.	Lubuk Linggau	551	0	32	1	24	50
5.	Muara Enim	10124	3	464	0	16	2341
6.	Musi Banyuasin	23929	1	107	35	56	5037
7.	Musi Rawas	26223	2	328	58	229	5167
8.	Musi Rawas Utara	214189	0	83	18	11	24713
9.	Ogan Ilir	40832	0	18	60	53	7642

10.	OKI	24850	1	20	0	55	5067
11.	OKU	14368	0	231	0	5	3979
12.	OKU Selatan	8922	8	122	23	82	334
13.	OKU Timur	2783	0	118	18	38	1181
14.	Pagar Alam	319	0	3	0	11	74
15.	Palembang	203	1	93	0	7	100
16.	PALI	7686	0	59	7	21	1863
17.	Prabumulih	11631	0	0	0	0	2421

3. Normalisasi

Pada prosedur ini diselesaikan menggunakan metode normalisasi *min-max* dengan skala 0 sampai 1.

KT : Korban Terdampak
 KM : Korban Meninggal
 RRR : Rumah Rusak Ringan
 RRB : Rumah Rusak Berat
 RT : Rumah terendam

Tabel 5. Normalisasi

No.	Kabupaten/Kota	KT	KM	RRR	RRS	RRB	RT
1.	Banyuasin	0,062	0	0,575	0,083	0,843	0,092
2.	Empat Lawang	0,056	0,250	0,170	0	0,218	0,079
3.	Lahat	0,008	0	0,114	0,083	0,341	0,014
4.	Lubuk Linggau	0,002	0	0,069	0,017	0,105	0
5.	Muara Enim	0,046	0,375	1	0	0,070	0,093
6.	Musi Banyuasin	0,111	0,125	0,231	0,583	0,245	0,202
7.	Musi Rawas	0,122	0,250	0,707	0,967	1	0,207
8.	Musi Rawas Utara	1	0	0,179	0,300	0,048	1
9.	Ogan Ilir	0,190	0	0,039	1	0,231	0,308
10.	OKI	0,115	0,125	0,043	0	0,240	0,203
11.	OKU	0,066	0	0,498	0	0,022	0,159
12.	OKU Selatan	0,041	1	0,263	0,383	0,358	0,012
13.	OKU Timur	0,012	0	0,254	0,300	0,166	0,046
14.	Pagar Alam	0,001	0	0,006	0	0,048	0,001
15.	Palembang	0	0,125	0,200	0	0,031	0,002
16.	PALI	0,035	0	0,127	0,117	0,092	0,074
17.	Prabumulih	0,053	0	0	0	0	0,096

3.4. Data Mining

Pada perhitungan data mining dilakukan sebanyak iterasi 2 kali. Untuk hasil perhitungan data mining dapat dilihat pada Tabel 6 berikut.

Keterangan:

JC0 : Jarak Centroid 0 ke data
 JC1 : Jarak Centroid 1 ke data
 JC2 : Jarak Centroid 2 ke data
 C : Cluster

Tabel 6. Hasil Hitung Data Mining

Kab/Kot	JC0	JC1	JC2	C
Banyuasin	0,818	0,439	1,592	C1
Empat Lawang	0,236	0,807	1,384	C0
Lahat	0,263	0,799	1,445	C0
Lubuk Linggau	0,260	0,969	1,445	C0
Muara Enim	0,906	0,729	1,623	C1
Musi Banyuasin	0,427	0,709	1,250	C0

Musi Rawas	1,282	0,724	1,758	C1
Musi Rawas Utara	1,326	1,528	0	C2
Ogan Ilir	0,868	1,094	1,295	C0
OKI	0,267	0,899	1,255	C0
OKU	0,440	0,784	1,330	C0
OKU Selatan	0,927	0,985	1,733	C0
OKU Timur	0,203	0,732	1,379	C0
Pagar Alam	0,313	1,051	1,454	C0
Palembang	0,259	0,914	1,451	C0
PALI	0,166	0,896	1,351	C0
Prabumulih	0,320	1,074	1,354	C0

3.5. Interpretation

Corak atau data yang diperoleh dari proses data mining harus ditampilkan dalam bentuk yang mudah dipahami pada tahapan awal prosedur.

- i) Ada 13 kabupaten/kota termasuk daerah rawan bencana rendah (C0) yaitu, Lubuk Linggau, Ogan Ilir, OKU, Pagar Alam, OKU Selatan, OKI, Lahat, OKU Timur, Palembang, Empat Lawang, Prabumulih dan PALI, Musi Banyuasin.
- ii) Ada tiga (3) kabupaten/kota termasuk daerah rawan bencana sedang (C1) yaitu, Banyuasin, Muara Enim, dan Musi Rawas.
- iii) Ada satu (1) kabupaten/kota termasuk daerah rawan bencana tinggi (C2) yaitu Musi Rawas Utara.

3.6. Pengujian Cluster

Silhouette Coefficient adalah pengujian cluster yang digunakan. Hasil pengujian cluster dengan menggunakan *Silhouette Coefficient* disajikan pada Tabel 7 berikut.

Tabel 7. Nilai *Silhouette Coefficient*

No.	Kabupaten/Kota	a(i)	b(i)	s(i)
1.	Empat Lawang	0,427	1,001	0,573
2.	Lahat	0,412	1,081	0,618
3.	Lubuk Linggau	0,433	0,997	0,565
4.	Musi Banyuasin	0,392	0,997	0,607
5.	Ogan Ilir	0,581	0,957	0,392
6.	OKI	0,922	1,259	0,266
7.	OKU	0,437	1,081	0,595
8.	OKU Selatan	0,554	0,960	0,422
9.	OKU Timur	0,984	1,175	0,163
10.	Pagar Alam	0,428	0,960	0,553
11.	Palembang	0,423	1,207	0,649
12.	PALI	0,369	1,079	0,657
13.	Prabumulih	0,435	1,227	0,645
14.	Banyuasin	0,637	0,919	0,305
15.	Muara Enim	0,783	0,994	0,211
16.	Musi Rawas	0,778	1,346	0,421
17.	Musi Rawas Utara	0	1,395	1

Dari hasil terakhir dari *silhouette coefficient* yakni dengan hitung rerata nilai s(i) segala data yaitu 0,509.

4. KESIMPULAN

Penelitian ini menyelesaikan permodelan dengan algoritma *K-Means* menggunakan data yang dikaji berdasarkan *Knowledge Discovery in Database* (KDD). Data bencana alam di Provinsi Sumatera Selatan dibagi menjadi tiga cluster: *Cluster 0* (C0) menunjukkan daerah rawan bencana rendah, *Cluster 1* (C1) menunjukkan daerah rawan bencana sedang, dan *Cluster 2* (C2) menunjukkan daerah rawan bencana tinggi.

Dari 17 kabupaten/kota di Provinsi Sumatera Selatan, 13 di antaranya termasuk dalam kategori daerah rawan bencana rendah (C0), yaitu Lubuk Linggau, Ogan Ilir, OKU, Pagar Alam, OKU Selatan, OKI, Lahat, OKU Timur, Palembang, Empat Lawang, Prabumulih, dan PALI, Musi Banyuasin. Ada juga 3 kabupaten/kota yang termasuk dalam kategori daerah rawan bencana sedang (C1), yaitu Banyuasin, Muara Enim, dan Musi Rawas Setelah menggunakan hasil pengelompokan algoritma K-Means, metode *Silhouette Coefficient* digunakan untuk menguji *cluster*. Hasil pengujian menunjukkan nilai 0,509, yang menunjukkan bahwa struktur *cluster* yang dibentuk memiliki struktur sedang.

DAFTAR RUJUKAN

- [1] Ariyanto, D. (2022). Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut. *Jurnal Sistim Informasi Dan Teknologi*. <https://doi.org/10.37034/jsisfotek.v4i1.117>
- [2] Connolly, T., dan Begg, C. 2010. *Database Systems: A Practical Approach to Design, Implementation, and Management* (5th ed.). United States: Pearson.
- [3] Didi, Usmadi. (2023). Maximum entropy application in predicting the vulnerability of land and forest fires in South Sumatra Province, Indonesia. *IOP conference series*, 1183 (1):012105-012105. doi: 10.1088/1755-1315/1183/1/012105
- [4] Han, Jiawei., Kamber, Micheline., Pei, Jian. 2012. *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kauffmann.
- [5] Handoyo, R., Mangkudjaja, R., & Nasution, S. M. 2014. Perbandingan metode clustering menggunakan metode Single Linkage dan K-means pada Pengelompokan Dokumen. *Jurnal Sifo Mikroskil*, 15(2), 73-82.
- [6] L. Kauffman dan P. J. Rousseeuw. 1990. *Finding groups in data: An introduction to cluster analysis*. USA: Wiley Series in Probability and Statistics.
- [7] Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [8] Maimon, Oded dan Lior Rokach. 2010. *Data Mining and Knowledge Discovery Handbook*. New York: Springer. 53
- [9] Mediana, Desfita., Taufika, Ophiyandri., Benny, Hidayat., Jendrius, Jendrius. (2022). Gender Mainstreaming in Post Disaster Reconstruction in West Sumatera. doi: 10.4108/eai.30-8-2021.2316299
- [10] Murdiaty, M., Angela, A., & Sylvia, C. (2020). Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means. <https://doi.org/10.30865/MIB.V4I3.2213>
- [11] Oded, Maimon., Lior, Rokach. (2005). Introduction to Knowledge Discovery in Databases. 1-17. doi: 10.1007/0-387-25465-X_1
- [12] Oscar, Johan Ong. 2013. Implementasi Algoritma K-Means Clustering untuk menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri*. 12(1), 10-20.
- [13] Rita, Nasmirayanti., Rafki, Imani., Meldia, Fitri., Jihan, Melasari., Maiyozzi, Chairi., Nadra, Arsyad. (2022). Literasi kesiapsiagaan gempa di sumatera barat: sebuah review. *Construction And Material Journal*, 4(1):39-48. doi: 10.32722/cmj.v4i1.4527
- [14] Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEEAccess*, 8, 80716–80727. doi:10.1109/access.2020.2988796
- [15] Singh, S., & Singh, P. (2020). Speaker specific feature based clustering and its applications in language independent forensic speaker recognition. *International Journal of Electrical and Computer Engineering (IJECE)*, 10 (4) ,3508. doi:10.11591/ijece.v10i4.pp3508-3518

-
- [16] Suroso, D. S., Kombaitan, B., & Setiawan, B. (2013). Exploring the use of Risk Assessment Approach for Climate Change Adaptation in Indonesia: Case Study of Flood Risk and Adaptation Assessment in the South Sumatra Province☆. *Procedia Environmental Sciences*. <https://doi.org/10.1016/J.PROENV.2013.02.050>
- [17] Tanzil Furqon, Muhammad, and Lailil Muflikhah. 2016. Clustering the Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application with Noise (Dbscan). *Journal of Enviromental Engineering and Sustainable Technology*. 3(1), 1-8.
- [18] Utomo, D., P & Mesran.(2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada DataSet Penyakit Jantung.” *JURNAL MEDIA INFORMATIKA BUDIDARMA*”,2(4), 437 –444. <https://doi.10.30865/mib.v4i2.2080>
- [19] Witten, Ian H., Frank, Eibe. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Complementary literature None.