

CLUSTERING ZONASI DAERAH RAWAN BENCANA ALAM DI PROVINSI SUMATERA BARAT MENGGUNAKAN ALGORITMA K-MEANS DAN LIBRARY GEOPANDAS

David Aryo Wicaksono ^{1*}, Yeremia Alfa Susetyo ²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Email: 672019307@student.uksw.edu ^{1*}, yeremia.alfa@uksw.edu ²

Histori Artikel:

Dikirim 10 Februari 2023; *Diterima dalam bentuk revisi* 5 Maret 2023; *Diterima* 10 April 2023; *Diterbitkan* 10 Mei 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Indonesia merupakan salah satu negara yang sering terkena bencana alam karena terletak pada cincin asia pasifik. Provinsi Sumatera Barat adalah salah satu provinsi di indonesia yang sering terjadi bencana alam termasuk gempa bumi. Menurut BPS Sumatera Barat dari tahun 2018-2021, tercatat 1.148 kasus gempa dan 1.503 kasus banjir. Penelitian ini bertujuan untuk mengelompokkan daerah rawan bencana alam Provinsi Sumatera Barat menggunakan algoritma K-Means dan library GeoPandas. Setelah melakukan olah data, didapatkan nilai k optimal, didapatkan 3 cluster yaitu cluster 1 tidak rawan bencana alam, cluster 2 rawan bencana alam dan cluster 3 sangat rawan bencana alam. Dari penelitian ini didapatkan cluster 1 memiliki 7 anggota wilayah, cluster 2 memiliki 11 anggota wilayah dan cluster 3 memiliki 1 anggota wilayah.

Kata Kunci: Clustering; K-Means; GeoPandas.

Abstract

Indonesia is one of the countries that are often affected by natural disasters because it is located on the Asia Pacific Ring of Fire. West Sumatra province is one of the most frequently hit by natural disasters including earthquakes. According to the Central Bureau of Statistics of West Sumatra from 2018-2021, there were 1,148 earthquake cases and 1,503 flood cases. The objective of this study is to reveal areas prone to natural disasters in West Sumatra province using K-Means algorithm and GeoPandas visualization. After doing data processing for the optimal K value, 3 clusters were obtained, namely, cluster 1 is not prone to natural disasters, cluster 2 is prone to natural disasters and cluster 3 is very prone to natural disasters. From this research, it is found that cluster 1 has 7 members of the region, cluster 2 has 11 members of the region, and cluster 3 has 1 member of the region.

Keyword: Clustering; K-Means; GeoPandas.

1. Pendahuluan

Bencana alam merupakan suatu peristiwa yang berdampak bagi populasi manusia. Indonesia terletak pada cincin Asia Pasifik, dimana adanya aktivitas tektonik di wilayah tersebut. Hal ini menyebabkan Indonesia harus menghadapi resiko dari aktivitas tersebut yang tidak lain adalah bencana alam [1]. Bencana alam bisa terjadi karena faktor alam dan manusia. Gempa terjadi dikarenakan adanya aktivitas tektonik dari bumi itu sendiri [2]. Tingginya curah hujan menyebabkan meluapnya air sungai, ada juga karena lebih rendahnya permukaan tanah daripada permukaan air yang bisa menyebabkan banjir [3]. Dari bencana alam tersebut menyebabkan manusia menerima dampaknya.

Provinsi Sumatera Barat berada di zona subdiksi lempeng Indo-Australia ke bawah lempeng Eurasia yang memberikan kontribusi tektonik di laut maupun daratan [4]. Menurut data dari bps tahun 2018 – 2021 tercatat data bencana alam banjir 1.503 kasus dan gempa 1.148 kasus di Provinsi Sumatera Barat [5]. Angka ini menunjukkan bahwa Provinsi Sumatera Barat menjadi wilayah rawan bencana alam dan juga bisa dijadikan lokasi penelitian karena banyaknya kasus bencana alam, sehingga masyarakat dan pemerintah mendapat informasi daerah kabupaten atau kota di Provinsi Sumatera Barat yang rawan bencana alam agar dapat mempersiapkan diri untuk menghadapi akibat dari bencana alam itu sendiri.

Berdasarkan latar belakang tersebut, maka akan dilakukan penelitian yang berjudul Clustering Zonasi Daerah Rawan Bencana Alam di Provinsi Sumatera Barat Menggunakan Algoritma K-Means dan Visualisasi GeoPandas. Penelitian ini bisa memberikan informasi kepada masyarakat maupun pemerintah tentang wilayah rawan bencana sehingga masyarakat atau pemerintah bisa mempersiapkan diri untuk menghadapi resiko bencana alam. Bahkan penelitian ini bisa digunakan pemerintah untuk upaya antisipasi wilayah bencana alam.

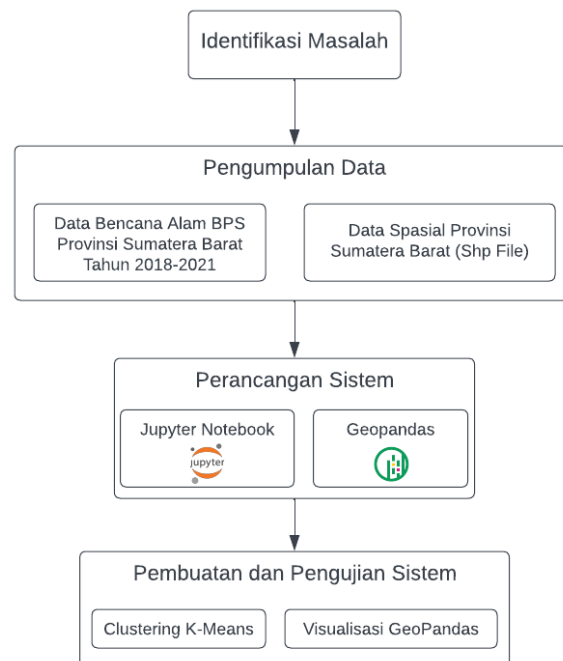
Beberapa penelitian yang sudah dilakukan sebelumnya, seperti pada penelitian yang berjudul Analisa Cluster Menggunakan K-Means dan Fuzzy C-Means Dalam Pengelompokkan Provinsi Menurut Data Intensitas Bencana Alam di Indonesia Tahun 2017-2021, peneliti menggunakan 2 metode yaitu K-Means dan Fuzzy C-Means. Untuk data yang digunakan berasal dari BNPB dari tahun 2017-2021 dan diakses melalui website resmi BNPB. Setelah dilakukan proses menggunakan K-Means dan hasil optimasinya menggunakan elbow method, didapatkan 2 klaster. Klaster 1 memiliki jumlah anggota sebanyak 31 provinsi dan pemetaannya berwarna hijau. Sedangkan cluster 2 memiliki jumlah anggota sebanyak 3 provinsi dan pemetaannya berwarna merah. Lalu untuk algoritma Fuzzy C-Means yang juga menggunakan elbow method untuk hasil optimasinya, didapatkan anggota klaster 1 sebanyak 32 provinsi dan pemetaannya berwarna hijau. Untuk klaster 2 memiliki jumlah anggota sebanyak 2 provinsi dan pemetaannya berwarna merah. Pada akhir penelitian dilakukan perbandingan nilai rata-rata Silhouette Coefficient antara 2 algoritma ini dan didapatkan nilai Fuzzy C-Means memiliki nilai rata-rata 0.8403194 lebih tinggi dari pada K-Means yaitu 0.8031255 [6].

Penelitian yang kedua berjudul “Analisa Spasial Clustering Zonasi Rawan Bencana Tanah Longsor Wilayah Bogor Selatan Berbasis WebGIS”, menggunakan K-Means untuk mengelompokkan wilayah yang terkena longsor disertai dengan visualisasi WebGIS. Data yang digunakan berasal dari BPBD Kabupaten Bogor dari Januari 2018 – Maret 2018. Peneliti menggunakan R Studio untuk melakukan clustering dan setelah itu divisualisasikan dalam bentuk WebGIS. Pada penelitian ini di dapatkan wilayah rawan bencana tanah longsor sebanyak 5 klaster, klaster 5 memiliki 6 titik, klaster 2 dan klaster 3 memiliki 5 titik, dan klaster 1 dan 4 sebanyak 1 titik [7]. Penelitian yang ketiga berjudul “Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenaga Kerja Menggunakan Algoritma K-Means”, pada penelitian ini menggunakan data UMKM, penduduk miskin, dan pencari kerja di Kabupaten Purwakarta. Peneliti menggunakan algoritma K-Means untuk melakukan proses clustering dan diikuti dengan pemetaan ke peta menggunakan GeoPandas. Setelah dilakukan proses clustering, didapatkan 3 klaster. Klaster 1 memiliki jumlah anggota 3 kecamatan, klaster 2 memiliki jumlah anggota 6 kecamatan, dan klaster 3 memiliki jumlah anggota 8 kecamatan. Untuk pemetaan klaster 1 menggunakan warna cream, untuk klaster 2 warna hijau muda, dan klaster 3 menggunakan warna hijau tua [8].

Penelitian dengan topik clustering zonasi daerah rawan bencana alam di Provinsi Sumatera Barat menggunakan algoritma K-Means dan library GeoPandas menjadi sangat penting untuk dilakukan. Penelitian ini akan memberikan manfaat yang besar bagi pemerintah dan masyarakat dalam upaya pengurangan risiko bencana alam di wilayah tersebut. Dengan melakukan clustering zonasi, pemerintah dapat menentukan daerah yang rentan terhadap bencana alam dan mengambil langkah-langkah preventif yang tepat dan efektif untuk mengurangi risiko bencana alam. Selain itu, penelitian ini juga akan memberikan informasi yang berguna bagi masyarakat untuk meningkatkan kewaspadaan dan persiapan menghadapi bencana alam. Dengan menggunakan algoritma K-Means dan library GeoPandas, penelitian ini dapat memberikan hasil yang akurat dan representatif untuk pengelompokan daerah rawan bencana alam di Provinsi Sumatera Barat. Oleh karena itu, penelitian ini sangat penting untuk dilakukan guna meningkatkan kesiapsiagaan dan pengurangan risiko bencana alam di wilayah tersebut.

2. Metode Penelitian

Clustering merupakan pengelompokan objek berdasarkan kemiripan ataupun keunikan atribut masing-masing [9]. Data bencana alam di Provinsi Sumatera Barat merupakan data yang akan dikelompokkan ke dalam masing-masing cluster. Tantangan dari penelitian ini adalah bagaimana cara melakukan klusterisasi terhadap data yang digunakan dan visualisasinya ditampilkan menggunakan geopandas. Gambar 1 menjelaskan tahapan yang akan dilakukan selama penelitian.



Gambar 1. Metode Penelitian

Berdasarkan alur dari metodologi penelitian diatas, tahap identifikasi permasalahan dilakukan pemahaman masalah untuk memahami fenomena alam yang terjadi terkait bencana alam di Provinsi Sumatera Barat. Jenis bencana yang diamati yaitu banjir dan gempa. Tahap berikutnya yaitu melakukan proses pengumpulan data bencana alam bulanan di Provinsi Sumatera Barat pada tahun 2018-2021. Data tersebut diperoleh dari website Badan Pusat Statistik terkait jumlah bencana alam yang terjadi di Provinsi Sumatera Barat. Selain itu ada juga data spasial perbatasan administrasi kabupaten atau kota di Provinsi Sumatera Barat yang didapatkan dari website Indonesia Geospasial. Dalam data tersebut berisikan 22 baris dan 28 kolom yang merupakan identitas dari tiap daerah.

Selanjutnya ada perancangan sistem, pada tahap ini dataset disimpan dalam bentuk txt. Selanjutnya, clustering akan dilakukan menggunakan bahasa R, dikarenakan ukuran file yang disimpan lebih kecil, bersifat open-source dan tools yang disediakan oleh bahasa ini terintegrasi untuk analisis data [10]. Setelah itu dimasukkan pada aplikasi R studio untuk mendapatkan hasil clustering. Setelah itu divisualisasikan menggunakan GeoPandas dengan tools Jupyter Notebook. Berikutnya, tahap pembuatan dan pengujian sistem dilakukan pengelompokan atau klusterisasi dan memvisualisasikannya ke GeoPandas. Algoritma clustering yang digunakan yaitu K-Means karena mampu mengelompokkan data cukup besar dengan efisien dan waktu yang dibutuhkan singkat [11]. Clustering ini diawali dengan menentukan nilai k paling optimal menggunakan metode silhouette dan metode elbow. Metode elbow bekerja dengan melihat titik siku yang terbentuk dari hasil membandingkan nilai persentase dari setiap jumlah cluster. Rumus perhitungan dari metode elbow dirumuskan pada persamaan (1).

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - c_k|^2 \quad (1)$$

Keterangan :

K = cluster ke-c

X_i = jarak data objek ke-i

C_k = pusat cluster ke-i

Selanjutnya untuk nilai silhouette ditentukan dengan beberapa perhitungan sebagai berikut :

- 1) Menghitung average distance objek ke-o dengan semua objek dalam satu cluster menggunakan persamaan 2.

$$a(o) = \frac{\sum_{o' \in C_k, o \neq o'} \text{dist}(o, o')}{|C_i - 1|} \quad (2)$$

Keterangan :

o : Objek ke-o pada cluster i

o' : Objek lain pada cluster i selain objek ke-o

C_i : Cluster i

|C_i| : Jumlah objek pada cluster i

dist(o, o') : Jarak data o dengan data o'

- 2) Menghitung average distance objek ke-o dengan semua objek dari cluster lain menggunakan persamaan (3)

$$b(i) = \min_{C_j: 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} \text{dist}(o, o')}{|C_j - 1|} \right\} \quad (3)$$

Keterangan :

o : Objek ke-o pada cluster j

o' : Objek lain pada cluster j selain objek ke-o

C_j : Cluster j

|C_j| : Jumlah objek pada cluster j

dist(o, o') : Jarak data o dengan data o'

- 3) Pada objek k memiliki nilai silhouette coefficient dengan memasukkan nilai a(o) dan b(o) pada persamaan (4) :

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (4)$$

Keterangan :

S_i^j : Silhouette index data ke-i dalam satu cluster

a_i^j : Rata-rata jarak data ke-i terhadap semua data dalam satu cluster

b_i^j : Rata-rata jarak data ke-i terhadap semua data yang tidak dalam

Setelah mendapatkan nilai k optimal dari perhitungan metode elbow dan silhouette, dilakukan pengelompokan menggunakan algoritma K-Means. Berikut proses pengolahan dataset yang telah dipersiapkan sebelumnya :

- 1) Memasukkan nilai k optimal yang dijadikan sebagai titik centroid, selanjutnya akan ditentukan centeroid cluster untuk iterasi berikutnya dengan persamaan 5

$$v = \frac{\sum_{i=1}^n X_i}{n}; i = 1, 2, 3 \dots n \quad (5)$$

Keterangan :

V : Titik pusat kelompok

x_i : Objek data ke = i

n : Jumlah objek yang menjadi anggota cluster

- 2) Selanjutnya dilakukan perhitungan jarak setiap objek ke titik centeroid menggunakan Ecludian distance pada persamaan (6). Hal ini dilakukan agar setiap objek yang memiliki karakteristik yang sama berada di cluster yang sama [12].

$$D(x_2, x_1) = \|x_2, x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (6)$$

Keterangan :

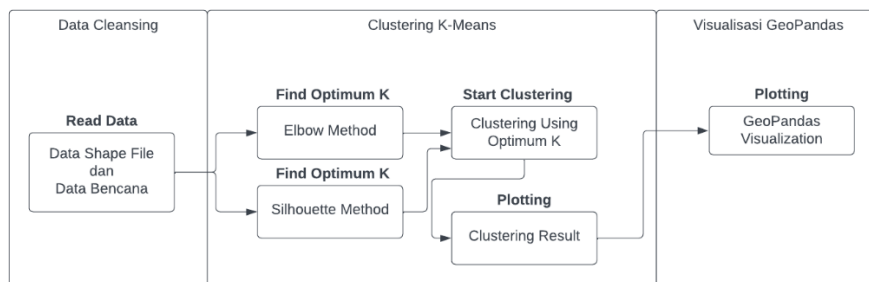
D : jarak kuadrat Euclidean antara objek ke x2 dan x1

P : jumlah variabel cluster

X_{2j} : data ke-2 pada variabel ke-j

X_{1j} : data ke-1 pada variabel ke-j

Berikutnya, Library Geopandas merupakan library Python yang digunakan untuk mengolah dan memanipulasi data geospasial, seperti shapefile (shp), GeoJson dan lainnya. Data geospasial sendiri merupakan data geografis yang memiliki titik koordinat dari sebuah objek tertentu [13]. Geopandas juga merupakan plotting package yang berintegrasi dengan Matplotlib yang berfungsi membaca file dalam bentuk format GIS [14]. Proses visualisasi ini diawali dengan mengambil hasil klasterisasi dari proses clustering sebelumnya. Setelah itu, library ini akan menampilkan peta dengan warna yang berbeda berdasarkan hasil clustering sebelumnya.



Gambar 2. Arsitektur Penelitian

Gambar 2 merupakan alur penelitian yang dimulai dari memasukkan data yang sudah kita dapatkan, setelah itu proses clustering diawali dengan pencarian nilai optimum menggunakan dua metode yaitu elbow dan silhouette method. Setelah didapatkan nilai k optimal, maka akan dilakukan clustering. Selanjutnya, plotting hasil clustering ditampilkan dalam bentuk gambar dan plotting berikutnya menggunakan geopandas. Untuk mempermudah pemahaman terhadap penelitian ini, dengan cara memahami deskripsi algoritma pemrograman komputer yang digunakan dengan bahasa manusia merupakan fungsi pseudocode [15], berikut pseudocode yang ditunjukkan pada tabel 1.

Tabel 1. Pseudocode Alur Penelitian

Input: k : Jumlah kluster optimum, inisiasi kluster yang didapat dari metode <i>elbow</i> dan <i>silhouette</i> d : {d ₁ , d ₂ , d ₃ ,...,d _n } dataset berisi n objek
Output : c : {c ₁ , c ₂ , ..., c _k } Nilai <i>centroid</i> L : {l(e) e = 1, 2, ..., n} anggota kluster
Method: 1) Inisiasi <i>centroid</i> awal menggunakan persamaan 5 dengan nilai yang sudah didapatkan sebelumnya 2) Membuat partisi sejumlah k dari segmentasi yang akan dibentuk 3) Memasukkan anggota d kedalam persamaan 6 4) Hitung nilai <i>euclidean distance</i> antar anggota d dengan pusat kluster menggunakan persamaan 6 5) Ulangi langkah 3 – 4 hingga anggota dataset sudah sampai ke-n 6) STOP 7) Hasil pengelompokan data.

3. Hasil dan Pembahasan

Penelitian ini diawali dengan pengambilan data melalui website resmi BPS dari tahun 2018-2021. Setelah mendapatkan data per tahun, dilakukan proses penjumlahan data dan didapatkan seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Data Total Bencana Tahun 2018-2021

No	Daerah	Banjir	Gempa
1	Solok	76	59
2	Pesisir Selatan	227	504
3	Kep. Mentawai	44	64
4	Dharmasraya	61	87
5	Solok	22	0
6	Sawahlunto	4	0
7	Pariaman	11	0
8	Kab. Solok	60	24
9	Padang Panjang	5	0
10	Sijunjung	96	5
11	Tanah Datar	3	16
12	Bukit Tinggi	40	48
13	Payakumbuh	8	11
14	Agam	20	49
15	Lima Puluh Koto	57	6
16	Pasaman barat	89	43
17	Pasaman	80	16
18	Padang	78	30
19	Padang Pariaman	72	186

Data berikutnya merupakan shapefile (shp) yang didapatkan dari indonesia-geospasial. Dalam data tersebut terdapat banyak atribut atau kolom yang berisikan identitas dari tiap-tiap wilayah kabupaten atau kota seperti yang tunjukan pada Tabel 3.

Tabel 3. Data Shp Provinsi Sumatera Barat

No	kdppum	namobj	remark	wadmkk	wiadkk	wadmpr
1	13	Solok Selatan	None		None	Solok	Sumatera Barat
2	13	Pesisir Selatan	None		Pesisir Selatan	None	Sumatera Barat
3	13	Kepulauan Mentawai	None		Kepulauan Mentawai	Padang Pariaman	Sumatera Barat
4	13	Dharmasraya	None		Dharmasraya	Sawah Lunto	Sumatera Barat
5	13	Kota Solok	None		Kota Solok	None	Sumatera Barat
6	13	Kota Sawahlunto	None		Kota Sawahlunto	None	Sumatera Barat
7	13	Kota Pariaman	None		Kota Pariaman	Pariaman	Sumatera Barat
8	13	Solok	None		Solok	None	Sumatera Barat
9	13	Kota Padang Panjang	None		Kota Padang Panjang	None	Sumatera Barat
10	13	Sijunjung	None		Sijunjung	None	Sumatera Barat
11	13	Tanah Datar	None		Tanah Datar	None	Sumatera Barat
12	13	Kota Bukittinggi	None		Kota Bukittinggi	None	Sumatera Barat
13	13	Kota Payakumbuh	None		Kota Payakumbuh	None	Sumatera Barat
14	13	Agam	None		Agam	None	Sumatera Barat
15	13	Lima Puluh Koto	None		Lima Puluh Koto	Labuhanbatu	Sumatera Barat
16	13	Pasaman Barat	None		Pasaman Barat	Pasaman	Sumatera Barat
17	13	Pasaman	None		Pasaman	None	Sumatera Barat
18	13	Provinsi Sumatera Barat	None		None	None	Sumatera Barat
19	13	Provinsi Sumatera Barat	None		None	None	Sumatera Barat
20	13	Kota Padang	None		Kota Padang	None	Sumatera Barat
21	13	Padang Pariaman	None		Padang Pariaman	None	Sumatera Barat
22	13	Provinsi Sumatera Barat	None		None	None	Sumatera Barat

Data tersebut akan di-cleaning data yang bertujuan untuk mengambil kolom yang diperlukan untuk identitas kabupaten atau wilayah dan mengubah nama kolom. Dengan kata lain cleaning data adalah tahap menghapus kolom yang tidak digunakan agar efisien dan optimal dalam penggunaan memori saat pemrosesan data.

Kode Program 1 *Cleaning Data*

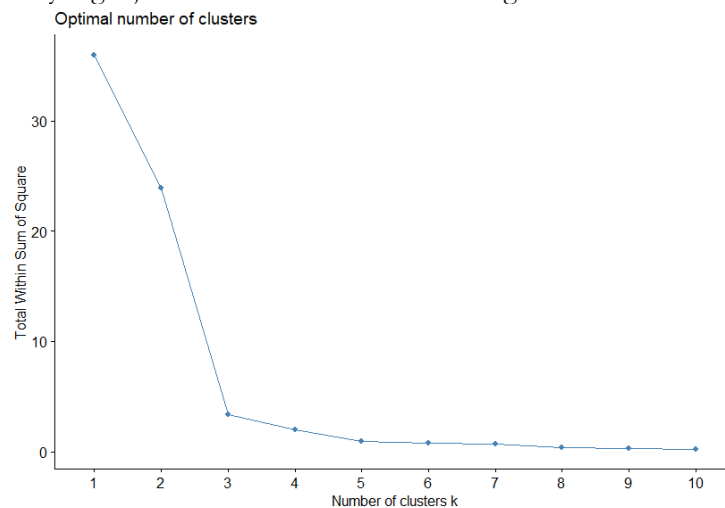
```
prov = prov[["wadmkk", "wadmpr", "shape_leng", "shape_area", "geometry"]]  
prov = prov.drop([17, 18, 21])  
prov = prov.reset_index()  
bencana = bencana.rename(columns={"wadmkk": "Wilayah", "wadmpr": "Provinsi"})
```

Pada Kode Program 1 diatas, baris 1 merupakan perintah untuk mengambil kolom yang diperlukan. Setelah itu, drop digunakan untuk menghapus baris, karena ada data duplikat pada index 17, 28, 21. Pada baris ketiga dilakukan reset_index agar index dari tiap tiap baris menjadi berurutan kembali setelah dilakukan perintah drop. Sedangkan pada baris keempat merupakan perintah untuk me-rename nama kolom agar mudah dimengerti. Data yang digunakan merupakan data total jumlah bencana alam dari tahun 2018 – 2021 yaitu gempa dan banjir. Proses clustering menggunakan algoritma K-Means akan dilakukan di aplikasi R studio menggunakan data tersebut.

Kode Program 2 *Scaling and Clustering*

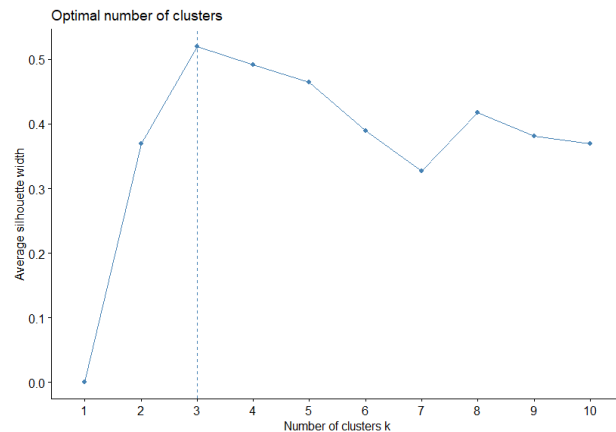
```
baru<-scale(dataku)  
fviz_nbclust(baru, kmeans, method = "wss")  
fviz_nbclust(baru, kmeans, method = "silhouette")  
final = kmeans(baru, 3)
```

Pada Kode Program 2 baris pertama, digunakan untuk standarisasi nilai data yang telah didapatkan sebelumnya agar jarak antar nilai tidak terlalu beragam.



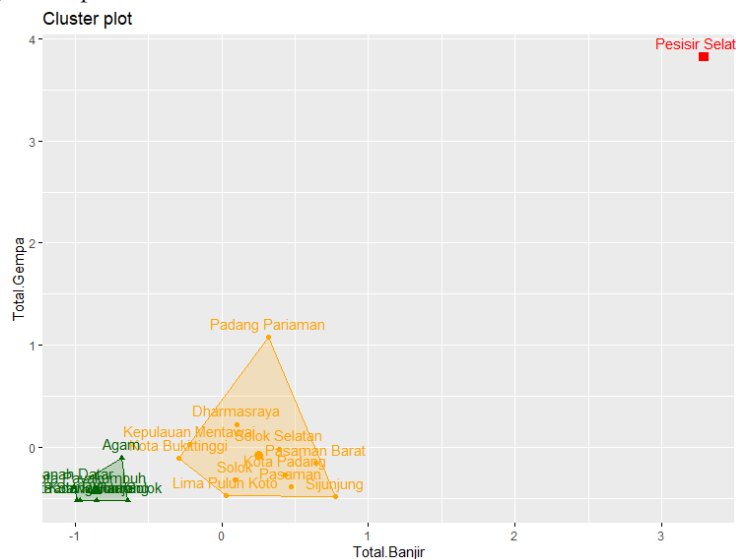
Gambar 3. Metode Elbow

Setelah itu dilakukan proses perhitungan nilai k paling optimal menggunakan metode Elbow pada baris Kode Program 2 baris kedua dan mengeluarkan nilai optimal seperti yang ditunjukkan pada Gambar 3. Nilai k optimal yang didapatkan dari metode ini adalah 3, berdasarkan sudut yang dibentuk menyerupai elbow nilai ke 3.



Gambar 4. Metode Silhouette

Selanjutnya pada Kode Program 2 baris ketiga, merupakan perintah untuk mencari nilai k paling optimal menggunakan metode silhouette dan mengeluarkan gambar grafik seperti yang ditunjukkan pada Gambar 4. Perhitungan ini menghasilkan nilai k paling optimal yaitu 3 berdasarkan garis putus-putus yang ditunjukkan pada Gambar 4.



Gambar 5. Hasil Plotting Clustering

Dari kedua metode tersebut, didapatkan nilai k paling optimal yaitu 3. Pada Kode Program 2 baris terakhir dilakukan proses clustering menggunakan nilai k optimal dan mengeluarkan hasil plotting berdasarkan hasil clustering seperti yang ditunjukkan pada Gambar 5. Keterangan dari ketiga cluster tersebut yaitu cluster 1 tidak rawan bencana, cluster 2 rawan bencana dan cluster 3 sangat rawan bencana. Hasil clustering tersebut akan di-concat dengan data shp. Pada tahap ini digunakan menggabungkan 2 dataframe menjadi 1 dataframe untuk mempermudah proses visualisasinya dan hasil dari concat data ditunjukkan pada Gambar 7. Berdasarkan hasil clustering menggunakan algoritma k-means, cluster memiliki 7 anggota, cluster 2 memiliki 12 anggota, dan cluster 3 memiliki 1 anggota.

Tabel 4. Hasil Concat Data

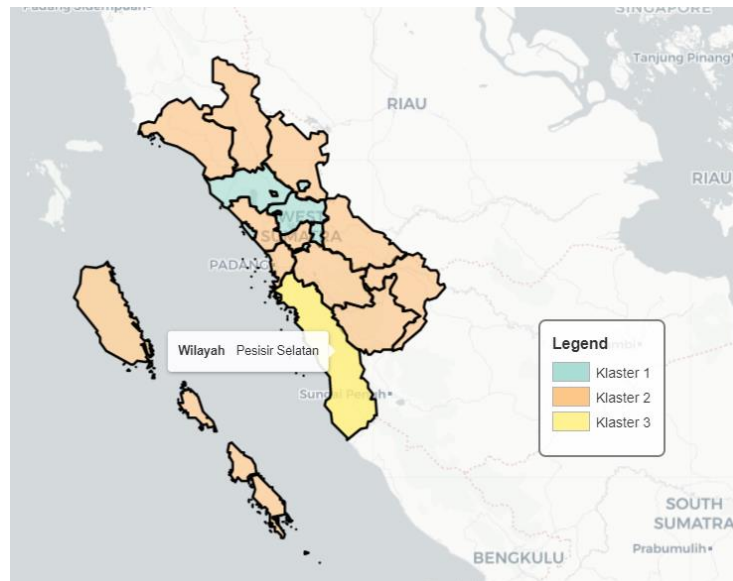
No	Wilayah	Total Banjir	Total Gempa	Cluster
1	Solok Selatan	76	59	2
2	Pesisir Selatan	227	504	3
3	Kepulauan Mentawai	44	64	2
4	Dharmasraya	61	87	2
5	Kota Solok	22	0	1
6	Kota Sawahlunto	4	0	1
7	Kota Pariaman	11	0	1
8	Solok	60	24	2
9	Kota Padang Panjang	5	0	1
10	Sijunjung	96	5	2
11	Tanah Datar	3	16	1
12	Kota Bukittinggi	40	48	2
13	Kota Payakumbuh	8	11	1
14	Agam	20	49	1
15	Lima Puluh Koto	57	6	2
16	Pasaman Barat	89	43	2
17	Pasaman	80	16	2
18	Kota Padang	78	30	2
19	Padang Pariaman	72	186	2

Setelah mendapatkan hasil clustering, dilakukan visualisasi menggunakan GeoPandas dalam bentuk peta dengan Kode Program 3.

Kode Program 3 Visualisasi Geopandas

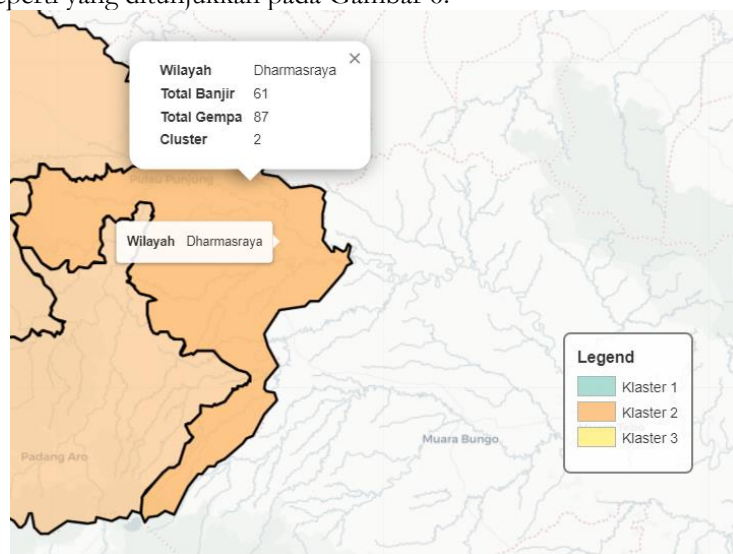
```
import geopandas as gpd
m = bencana.explore( column="Cluster",
    tooltip="Wilayah",
    popup=True,
    tiles="CartoDB positron",
    cmap="Set3",
    style_kwds=dict(color="white"))
```

Pada Kode Program 3 baris kesatu dilakukan import library GeoPandas. Selanjutnya proses visualisasi ke dalam bentuk peta dengan menggunakan GeoPandas. Atribut pada baris pertama yaitu column, yang menjadi acuan untuk pewarnaan peta, tooltip berfungsi ketika kursor meng-hover peta, maka akan keluar nama wilayah. Sedangkan untuk atribut lainnya berfungsi untuk menentukan tipe peta. Hasil dari visualisasi ke dalam peta ditunjukkan pada Gambar 8.



Gambar 6. Visualisasi *GeoPandas*

Visualisasi peta akan mengeluarkan identitas wilayah ketika pointer dari mouse diarahkan ke wilayah di peta seperti yang ditunjukkan pada Gambar 6.



Gambar 7. Keterangan Visualisasi *GeoPandas*

Ketika suatu objek di-klik, maka akan mengeluarkan informasi dari objek tersebut seperti yang ditunjukkan pada Gambar 7.

4. Kesimpulan

Hasil penelitian ini dapat menunjukkan wilayah Provinsi Sumatera Barat memiliki wilayah rawan bencana alam dan dapat digunakan sebagai informasi awal untuk pemerintah mempersiapkan diri menanggulangi dampak dari bencana alam di masing-masing daerah pemerintahan. Dari analisis penelitian ini, kategori yang bisa dimasukkan dalam kluster 1 yaitu wilayah yang mengalami bencana alam yang cenderung sedikit sehingga kluster 1 merupakan wilayah aman atau tidak rawan bencana. Sedangkan kluster 2, yaitu daerah rawan bencana karena wilayah ini mengalami bencana alam cukup

banyak. Serta klaster 3 merupakan wilayah yang sering terjadi bencana alam sehingga klaster 3 ini disebut wilayah sangat rawan bencana alam. Wilayah Kab. Pesisir Selatan merupakan wilayah sering terjadi bencana alam dalam 4 tahun terakhir yaitu 227 kasus banjir dan 504 kasus gempa bumi dan merupakan satu-satunya anggota dari klaster 3. Sedangkan anggota wilayah cluster 2 yaitu memiliki 12 anggota wilayah dan anggota wilayah cluster 1 memiliki 7 anggota wilayah.

5. Daftar Pustaka

- [1] Sadewo, M.G., Windarto, A.P. and Wanto, A., 2018. Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2(1). DOI: <http://dx.doi.org/10.30865/komik.v2i1.943>.
- [2] Rais, L. N. 2021. Analisis Bencana Gempa Bumi Dan Mitigasi Bencana Di Daerah Kertasari. *Jurnal Samudra Geografi*, 4(2), 14-19. DOI: <https://doi.org/10.33059/jsg.v4i2.3773>.
- [3] Findayani, A., 2018. Kesiap siagaan masyarakat dalam penanggulangan banjir di Kota Semarang. *Jurnal Geografi: Media Informasi Pengembangan Dan Profesi Kegeografian*, 12(1), pp.102-114.
- [4] Metrikasari, R. and Choiruddin, A., 2021. Pemodelan Risiko Gempa Bumi di Pulau Sumatera Menggunakan Model Inhomogeneous Neyman-Scott Cox Process. *Jurnal Sains dan Seni ITS*, 9(2), pp.D102-D107. DOI: <https://doi.org/10.12962/j23373520.v9i2.52318>.
- [5] BPS Provinsi Sumatera Barat. <https://sumbar.bps.go.id/>.
- [6] Firdaus, H., 2022. Analisa Cluster Menggunakan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Provinsi Menurut Data Intesitas Bencana Alam Di Indonesia Tahun 2017-2021. *MATHunesa: Jurnal Ilmiah Matematika*, 10(1), pp.50-60. DOI: <https://doi.org/10.26740/mathunesa.v10n1.p50-60>.
- [7] Suwardi, A.M., Susetyo, B. and Hermawan, E., 2019, October. Analisa Spasial Clustering Zonasi Rawan Bencana Tanah Longsor Wilayah Bogor Selatan Berbasis WebGIS. In *Seminar Nasional Teknologi Informasi* (Vol. 2, pp. 182-190).
- [8] Kurniadewi, H., Hakim, R.A., Jajuli, M. and Jaman, J.H., 2022. Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenaga Kerja Menggunakan Algoritma K-Means. *Journal of Applied Informatics and Computing*, 6(2), pp.113-119. DOI: <https://doi.org/10.30871/jaic.v6i2.4227>.
- [9] Murdiaty, M., Angela, A. and Sylvia, C., 2020. Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means. *Jurnal Media Informatika Budidarma*, 4(3), pp.744-752. DOI: <http://dx.doi.org/10.30865/mib.v4i3.2213>.
- [10] Sihombing, R.E., Rachmatin, D. and Dahlan, J.A., 2019. Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. *Jurnal EurekaMatika*, 7(1), pp.58-79.



- [11] Herald, H.Y., Aprilia, N.C. and Pratiwi, H., 2019. Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode K-Means. *Indonesian Journal of Applied Statistics*, 2(2), pp.137-144. DOI: <https://doi.org/10.13057/ijas.v2i2.34911>.
- [12] Dewi, D.A.I.C. and Pramita, D.A.K., 2019. Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 9(3), pp.102-109. DOI: <https://doi.org/10.31940/matrix.v9i3.1662>.
- [13] Geopandas. <https://geopandas.org/en/stable/>
- [14] Yim, A., Chung, C. and Yu, A., 2018. *Matplotlib for Python Developers: Effective techniques for data visualization with Python*. Packt Publishing Ltd.
- [15] Fuadi, W., Meliala, S. and Fariadi, D., 2021. Rancang Bangun Aplikasi Game Pembelajaran Pseudocode Dasar. *JET (Journal of Electrical Technology)*, 6(1), pp.31-34.