

## Perbandingan Analisis *K-means* dan *Hierarchical Clustering* dalam Mengelompokkan Kecamatan di Kabupaten Grobogan Berdasarkan Jumlah Titik Kejadian Bencana Alam

Sakti Wijayanti<sup>1,\*</sup>, Ayundyah Kesumawati<sup>2</sup>

<sup>1,2</sup> Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14.5, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55584, Indonesia.

\*Corresponding author: [sakti.wijayanti@students.uii.ac.id](mailto:sakti.wijayanti@students.uii.ac.id)



P-ISSN: 2986-4178

E-ISSN: 2988-4004

### Riwayat Artikel

Dikirim: 14 November 2024

Direvisi: 03 Februari 2025

Diterima: 21 Februari 2025

### ABSTRAK

Kabupaten Grobogan adalah salah satu kabupaten di Jawa Tengah yang rentan terhadap bencana alam. Tingginya frekuensi bencana alam di Kabupaten Grobogan seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang menjadi masalah yang serius bagi masyarakat. Penelitian ini bertujuan untuk melakukan perbandingan keefektivitasan antara metode *k-means* dan *hierarchical clustering* dalam mengelompokkan kecamatan berdasarkan data jumlah titik kejadian bencana alam dan menentukan metode mana yang dinilai lebih baik dalam menganalisis data bencana. Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari website Sistem Informasi Bencana Alam (SIGANA) Badan Penanggulangan Bencana Daerah Kabupaten Grobogan. Hasil analisis *k-means* dan *hierarchical clustering* adalah terbentuknya 3 *cluster* optimum dengan keterangan 14 kecamatan kategori rendah, 4 kecamatan kategori sedang dan 1 kecamatan kategori tinggi. Berdasarkan validasi *cluster internal measure*, kedua metode tersebut memperoleh nilai *connectivity*, *dunn index* dan *silhouette* yang sama yaitu nilai *connectivity* sebesar 7.3115, nilai *dunn index* sebesar 0.8991, dan nilai *silhouette* sebesar 0.5932. Sedangkan dalam Selanjutnya dalam validasi *cluster* dengan *stability measure* metode *k-means* menghasilkan nilai APN sebesar 0.0962, AD sebesar 13.6823, ADM sebesar 4.1335, dan FOM sebesar 7.4761. Sedangkan, metode *hierarchical clustering* menghasilkan nilai APN sebesar 0.0391, AD sebesar 13.4789, ADM sebesar 3.5167, dan FOM sebesar 7.1598. Secara keseluruhan, dapat disimpulkan bahwa metode yang dinilai lebih baik untuk mengelompokkan data jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 adalah metode *hierarchical clustering*.

**Kata Kunci:** Bencana Alam, Grobogan, *K-Means*, *Hierarchical Clustering*.

### ABSTRACT

*Grobogan Regency is one of the districts in Central Java that is vulnerable to natural disasters. The high frequency of natural disasters in Grobogan Regency, such as floods, droughts, fires, landslides and strong winds, is a serious problem for the community. This research aims to compare the effectiveness between the *k-means* and *hierarchical clustering* methods in grouping sub-districts based on data on the number of natural disaster occurrence points and determine which method is considered better in analyzing disaster data. The data used in*

*this research is secondary data sourced from the Natural Disaster Information System (SIGANA) website of the Grobogan Regency Regional Disaster Management Agency. The results of the k-means and hierarchical clustering analysis were the formation of 3 optimum clusters with information about 14 sub-districts in the low category, 4 sub-districts in the medium category and 1 sub-district in the high category. Based on cluster internal measure validation, the two methods obtained the same connectivity, dunn index and silhouette values, namely a connectivity value of 7.3115, a dunn index value of 0.8991, and a silhouette value of 0.5932. Meanwhile, further validation of the cluster using the stability measure k-means method produces an APN value of 0.0962, AD of 13.6823, ADM of 4.1335, and FOM of 7.4761. Meanwhile, the hierarchical clustering method produces an APN value of 0.0391, AD of 13.4789, ADM of 3.5167, and FOM of 7.1598. Overall, it can be concluded that the method that is considered better for grouping data on the number of natural disaster occurrence points in Grobogan Regency in 2023 is the hierarchical clustering method.*

**Keyword:** *Grobogan, Hierarchical Clustering, K-Means, Natural Disasters.*

## 1. Pendahuluan

Menurut Undang-undang Nomor 24 Tahun 2007 tentang Penanggulangan Bencana, menyebutkan bahwa definisi “bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan oleh faktor alam dan/atau faktor manusia sehingga mengakibatkan timbulnya korban jiwa, kerusakan lingkungan, kerugian harta benda dan dampak psikologis”. Definisi tentang bencana tersebut menyebutkan bahwa bencana disebabkan oleh beberapa faktor, yaitu faktor alam, non alam, dan manusia. Undang-Undang Nomor 24 Tahun 2007 mendefinisikan juga mengenai definisi “bencana alam adalah bencana yang diakibatkan oleh peristiwa atau serangkaian peristiwa yang disebabkan oleh alam antara lain berupa gempa bumi, tsunami, gunung meletus, banjir, kekeringan, angin topan dan tanah longsor”. Selanjutnya, “Bencana non alam adalah bencana yang diakibatkan oleh peristiwa atau serangkaian peristiwa non alam yang antara lain berupa gagal teknologi, gagal modernisasi, epidemic dan wabah penyakit”. Sedangkan, “bencana sosial adalah bencana yang diakibatkan oleh peristiwa atau serangkaian peristiwa yang diakibatkan oleh manusia yang meliputi konflik sosial antar kelompok dan antar komunitas masyarakat serta teror” [1].

Letjen TNI Suharyanto, S.Sos., M.M., Kepala Pelaksan Badan Nasional Penanggulangan Bencana (BNPB), menyampaikan dalam konferensi pers Kaleidoskop Bencana 2023 dan *Outlook Bencana 2024* bahwa tahun 2023 Badan Nasional Penanggulangan Bencana mencatat lebih dari 4.940 kejadian bencana alam terjadi Indonesia. Bencana alam yang paling sering terjadi adalah bencana hidrometeorologi, seperti bencana banjir 1.170 kali, kekeringan 168 kali, cuaca ekstrem sebanyak 1.115 kali, tanah longsor sebanyak 579 kali, kebakaran hutan dan lahan (karhutla) menyumbang 1.802 kali, gelombang pasang dan abrasi juga menyumbang 31 kali, erupsi gunung api sebanyak 4 kali dan gempa bumi sebanyak 31 kali. Beliau menegaskan bahwa dalam menghadapi potensi bencana alam setiap tahunnya, Badan Nasional Penanggulangan Bencana akan terus bekerja sama dengan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) [2].

Letak Kabupaten Grobogan berada di Provinsi Jawa Tengah dan secara administratif daerah tersebut terbagi menjadi 19 kecamatan dan 280 desa/kelurahan dengan luas wilayah 1.975,86 km<sup>2</sup>. Berdasarkan Sensus Penduduk Indonesia Tahun 2020, jumlah penduduk Kabupaten Grobogan mencapai 1.453.526 jiwa dengan kepadatan penduduk sekitar 719 jiwa/km<sup>2</sup>. Secara letak geografis, Letak geografisnya terletak di antara 110° 15' BT – 111° 25' BT dan 7° LS – 7°30' LS. Berdasarkan geografis dan reliefnya, Kabupaten Grobogan adalah daerah yang perekonomiannya sangat bergantung pada sektor pertanian dan cenderung mengalami kesulitan mendapatkan air bersih [3].

Seperti daerah lain, Kabupaten Grobogan rentan terhadap bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang. Pada tahun 2023, tercatat sebanyak 101 titik kejadian banjir, 341 titik kejadian kekeringan, 154 titik kejadian kebakaran, 20 titik kejadian tanah longsor, dan 41 titik kejadian angin kencang yang terjadi di berbagai wilayah Kabupaten Grobogan. Berdasarkan data bencana yang bersumber dari situs *website* Badan Penanggulangan Bencana Daerah Kabupaten Grobogan tahun 2023, pada bulan Januari terjadi sebuah bencana banjir yang disebabkan oleh curah hujan yang begitu tinggi. Selanjutnya, pada bulan agustus terjadi bencana kekeringan akibat rendahnya curah hujan mengakibatkan air sumur mengering dan kesulitan memenuhi kebutuhan air. Kekeringan melanda di beberapa kecamatan di Kabupaten Grobogan, salah satunya adalah Kecamatan Kradenan yang berdampak pada 114 KK [4].

Terdapat penelitian terdahulu yang dilakukan oleh Natalis & Nataliani (2022) dengan judul “Pemanfaatan *K-means Clustering* dan *Analytic Hierarchy Process* terhadap Penilaian Prestasi Kerja Pegawai”. Penelitian tersebut bertujuan untuk untuk menggambarkan tingkat kinerja pegawai (baik, cukup atau kurang). Dari hasil pengolahan data, ditentukan tiga *cluster* yaitu *cluster* baik, cukup, dan kurang, sehingga diperoleh hasil terdapat enam pegawai pada *cluster* baik, sembilan pegawai pada *cluster* cukup, dan lima pegawai pada *cluster* kurang. Selanjutnya, pada *cluster* baik akan dicari untuk siapa pegawai dengan nilai terbaik, maka dilakukan seleksi dengan metode AHP untuk menentukan yang terbaik pada *cluster* baik. Berdasarkan hasil dari proses AHP, didapatkan bahwa terdapat dua pegawai yang mempunyai nilai terbaik, sehingga dua pegawai tersebut dinilai layak untuk menjadi yang terbaik dari *cluster* baik. Hasil ini dapat menjadi acuan ke depannya, sebagai contoh pegawai dengan kinerja baik mendapatkan kenaikan jabatan atau pangkat, sementara pegawai dengan kinerja kurang mendapatkan pelatihan. Selain itu hasil ini dapat menjadi tolak ukur pada instansi pemerintahan Kabupaten Bengkayang, terutama pegawai Dinas Pendidikan dan Kebudayaan Kabupaten Bengkayang [5].

Penelitian lain dilakukan oleh Sibrani dkk (2024) yang berjudul “Penggunaan *K-Means* dan *Hierarchical Clustering Single Linkage* dalam Pengelompokan Stok Obat” dengan tujuan untuk meningkatkan kualitas perawatan dengan memastikan pasokan obat yang dibutuhkan tersedia. Hal ini juga membantu rumah sakit untuk memahami kebutuhan obat-obatan secara lebih baik, terutama di era digital saat ini. Berdasarkan hasil pengujian dan pembahasan *hierarchical clustering* dengan metode *single linkage* dapat memberikan hasil kualitas *clustering* yang lebih baik dengan *Silhouette Score* yang lebih tinggi jika dibandingkan dengan *k-means clustering*. *Hierarchical clustering single linkage* dengan tiga *cluster* menghasilkan *Silhouette Score* sebesar 0.976, yang menunjukkan kualitas *clustering* yang sangat baik. Sementara itu, *k-means clustering* juga menunjukkan performa yang cukup baik dengan *Silhouette Score* sebesar 0.954 [6].

*K-means clustering* adalah metode pengelompokan data yang membagi data menjadi beberapa kelompok berdasarkan kedekatannya dengan titik pusat atau centroid. Keunggulan metode ini adalah cepat dan efisien, sehingga cocok untuk *dataset* besar, seperti data kecamatan di Kabupaten Grobogan. *K-means* juga memungkinkan kita untuk

menentukan jumlah *cluster* terlebih dahulu, yang membantu mengelompokkan kecamatan berdasarkan tingkat kerentanannya terhadap bencana alam, seperti banjir atau gempa. Sementara itu, *hierarchical clustering* adalah metode yang mengelompokkan data secara bertahap dengan membangun struktur bertingkat. Keunggulan dari metode ini adalah fleksibilitasnya, karena tidak memerlukan penentuan jumlah *cluster* di awal. Selain itu, hasilnya dapat digambarkan dalam bentuk dendrogram, yang memperlihatkan hubungan antar kecamatan, sehingga memudahkan analisis pola kerentanannya terhadap bencana alam. Kedua metode ini memiliki keunggulannya masing-masing, di mana *k-means* efektif untuk efisiensi dan pengelompokan yang jelas, sementara *hierarchical clustering* cocok untuk analisis yang lebih mendalam dan fleksibel.

Oleh karena itu, dalam penelitian ini dilakukan analisis *cluster* menggunakan metode *k-means* dan *hierarchical clustering* untuk mengelompokkan kecamatan di Kabupaten Grobogan berdasarkan jumlah titik kejadian bencana alam tahun 2023. Metode *k-means* dan *hierarchical clustering* merupakan salah satu “*unsupervised machine learning algorithms*” yang sangat sederhana dan sering digunakan oleh peneliti. Pada penelitian berjudul “*An Analysis of Natural Disaster Data by Using K-means and K-medoids Algorithm of Data Mining Techniques*” oleh Prihandoko dkk dijelaskan bahwa metode *K-means* dinilai lebih efektif dan efisien dibandingkan metode *K-medoids* karena metode *K-medoids* kurang efektif digunakan apabila banyak data yang mengandung angka 0. Dengan adanya penelitian ini, diharapkan dapat mengetahui kecamatan mana saja yang merupakan kecamatan dengan rawan bencana alam yang nantinya dapat membantu penanganan operasional dan mengoptimalkan mitigasi penanggulangan bencana alam di Kabupaten Grobogan.

## 2. Metodologi Penelitian

### 2.1. Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *website* Sistem Informasi Grobogan Aman Bencana (SIGANA) Badan Penanggulangan Bencana Daerah Kabupaten Grobogan. Dalam penelitian ini terdapat lima variabel yaitu bencana banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang yang akan ditampilkan dan dijelaskan secara rinci dalam **Tabel 1** sebagai berikut.

**Tabel 1** Keterangan Variabel dan Simbol

Variabel	Simbol	Pengertian	Keterangan
Banjir	$X_1$	Peristiwa atau keadaan di mana terendamnya suatu daerah atau daratan karena volume air yang meningkat	Jumlah titik kejadian bencana alam berdasarkan kecamatan di Kabupaten Grobogan yang terdaftar dalam data banjir tahun 2023 dengan satuan kejadian
Kekeringan	$X_2$	Ketersediaan air yang jauh di bawah kebutuhan air untuk kebutuhan hidup, pertanian, kegiatan ekonomi dan lingkungan	Jumlah titik kejadian bencana alam berdasarkan kecamatan di Kabupaten Grobogan yang terdaftar dalam data kekeringan tahun 2023 dengan satuan kejadian
Kebakaran	$X_3$	Situasi di mana bangunan pada suatu tempat seperti rumah/pemukiman, pabrik, pasar, gedung dan lain-lain dilanda api yang menimbulkan korban dan/atau kerugian	Jumlah titik kejadian bencana alam berdasarkan kecamatan di Kabupaten Grobogan yang terdaftar dalam data kebakaran tahun 2023 dengan satuan kejadian

Variabel	Simbol	Pengertian	Keterangan
Tanah Longsor	$X_4$	Salah satu jenis gerakan massa tanah atau batuan, ataupun percampuran keduanya, menuruni atau keluar lereng akibat terganggunya kestabilan tanah atau batuan penyusun lereng	Jumlah titik kejadian bencana alam berdasarkan kecamatan di Kabupaten Grobogan yang terdaftar dalam data tanah longsor tahun 2023 dengan satuan kejadian
Angin Kencang	$X_5$	Angin yang datang secara tiba-tiba, mempunyai pusat, bergerak melingkar menyerupai spiral dengan kecepatan 40-50 km/jam hingga menyentuh permukaan bumi dan akan hilang dalam waktu singkat (3-5 menit)	Jumlah titik kejadian bencana alam berdasarkan kecamatan di Kabupaten Grobogan yang terdaftar dalam data angin kencang tahun 2023 dengan satuan kejadian

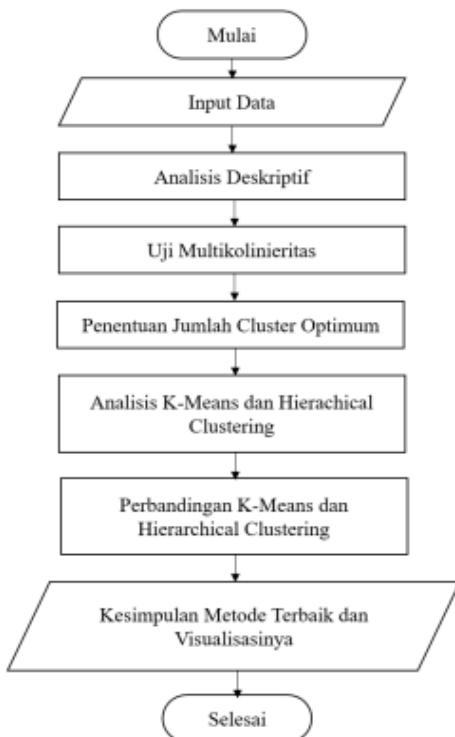
## 2.2. Tahapan Penelitian

Dalam penelitian analisis *cluster* ini digunakan dua metode yaitu *k-means* dan *hierarchical clustering* untuk mengelompokkan kecamatan di Kabupaten Grobogan berdasarkan jumlah kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang tahun 2023. Pengolahan data bencana alam tersebut dilakukan menggunakan bantuan *software R*.

Tahapan-tahapan yang analisis *cluster* untuk mengelompokkan kecamatan di Kabupaten Grobogan berdasarkan jumlah kejadian bencana alam menggunakan metode *k-means* dan *hierarchical clustering* sebagai berikut:

1. Pertama, dimulai dengan memasukkan data jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang.
2. Melakukan analisis deskriptif yang meliputi statistika deskriptif untuk mengetahui gambaran umum atau ringkasan data tersebut.
3. Melakukan uji multikolinearitas untuk mengetahui hubungan/korelasi antara variable satu dengan variabel lain.
4. Menentukan *cluster* optimum (*k*) metode *k-means* menggunakan metode Elbow dan metode *hierarcgical clustering* menggunakan hasil dendogram yang terbentuk.
5. Kemudian, melakukan analisis *cluster k-means* maupun *hierarchical clustering* untuk mengetahui kriteria setiap *cluster* yang terbentuk (*k*).
6. Selanjutnya, melakukan perbandingan hasil *cluster* kedua metode tersebut untuk mengetahui perbedaan metode satu dengan metode lain.
7. Setelah itu, menentukan metode *clustering* terbaik dengan menggunakan validasi *cluster internal measure* dan *stability measure* dan dilanjutkan dengan membuat visualisasi hasil metode terbaik dalam bentuk peta sederhana.

Berikut merupakan *flowchart* atau diagram alir yang memuat tahapan-tahapan analisis dalam penelitian..



**Gambar 1** Diagram Alir Penelitian

### 2.3. Analisis Deskriptif

Analisis deskriptif adalah teknik penelitian statistik yang bertujuan untuk meringkas, menggambarkan, dan menganalisis data secara numerik atau grafis. Metode ini berfokus pada penyajian data secara terstruktur dan mudah dipahami, sehingga memungkinkan kita untuk mengidentifikasi pola, tren, dan karakteristik dari data yang dianalisis. Selain itu, analisis deskriptif juga memungkinkan perbandingan data dari berbagai kelompok atau variabel yang berbeda. Pendekatan yang umum digunakan analisis deskriptif mencakup statistik deskriptif, grafik, dan tabel. Melalui analisis deskriptif, kita dapat memperoleh pemahaman yang lebih kompetitif dan mendalam tentang data yang tersedia, serta dapat mengambil keputusan yang lebih informasional [7]

Analisis deskriptif dalam konteks penelitian ini dimulai dengan pengumpulan data kejadian bencana (banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang) di setiap kecamatan di Kabupaten Grobogan. Data yang sudah terkumpul kemudian dibersihkan untuk menghilangkan nilai yang hilang. Selanjutnya, dihitung statistik deskriptif seperti rata-rata (*mean*) kejadian bencana, nilai tengah (*median*), dan sebaran data (standar deviasi) untuk memberikan gambaran tentang frekuensi dan variasi bencana di berbagai kecamatan. Visualisasi data, seperti grafik batang atau peta, digunakan untuk mempermudah pemahaman dan menunjukkan kecamatan yang paling sering terdampak bencana. Hasil analisis ini kemudian dapat diinterpretasikan, misalnya untuk mengidentifikasi kecamatan yang memiliki tingkat kerentanannya lebih tinggi terhadap bencana tertentu, yang menjadi dasar untuk perencanaan dan pengambilan keputusan dalam mitigasi bencana di daerah tersebut.

## 2.4. Analisis Cluster

*Clustering* adalah proses pembentukan kelompok di mana setiap anggota dalam setiap partisi memiliki sifat yang serupa ketentuan tertentu. Analisis *cluster* adalah teknik analisis data yang digunakan untuk melakukan pengelompokan terhadap objek atau individu dalam beberapa klaster atau kelompok berdasarkan sifat yang berbeda, sehingga individu atau objek yang tergabung dalam kelompok memiliki sifat yang relative sama atau homogen. Hal tersebut dilakukan dengan tujuan adalah untuk mengelompokan objek atau individu tersebut ke dalam beberapa kluster atau kelompok [8].

Secara umum, *cluster* dibagi menjadi dua metode, yaitu metode *partitional clustering* dan *hierarchical clustering*. Dalam *hierarchical clustering*, data akan dikelompokkan menggunakan dendogram hirarki, di mana kelompok-kelompok terdekat secara bertahap digabungkan atau seluruh data dibagi ke dalam *cluster*. Contoh metode *hierarchical clustering* meliputi *single linkage*, *average linkage*, *complete linkage*, *ward's linkage* dan *centroid linkage*. Sementara *partitional clustering*, data tersebut akan dikelompokkan ke dalam beberapa bagian tanpa struktur dendogram hirarki antara satu *cluster* dengan *cluster* lain. Beberapa metode *partitional clustering* lainnya meliputi *k-means*, *mixture modelling* dan *fuzzy c means* [9].

### 2.4.1. K-means Clustering

*K-means clustering* merupakan algoritma “*unsupervised learning*” atau algoritma pembelajaran tanpa pengawalan. Fungsi utama metode *k-means* adalah mengelompokkan data kedalam beberapa *cluster* tanpa memerlukan label kategori. *K-means* adalah metode *clustering non-hierarchy*, di mana algoritma metode ini adalah mengelompokkan berbagai data menjadi satu kelompok dengan karakteristik yang sama dan memiliki karakteristik yang sangat berbeda dengan kelompok lain. *Cluster sampling* adalah metode yang digunakan untuk pengambilan sampel dengan melakukan pemilihan secara acak unit-unit populasi dari klaster yang telah ada. Salah satu permasalahan yang dapat diselesaikan dengan pembelajaran tanpa pengawalan (*unsupervised learning*) adalah metode pendekatan analisis *clustering*. *K-means clustering* merupakan salah satu metode analisis data *data mining* dalam melakukan pemodelan *unsupervised learning* untuk melakukan pengelompokan data ke dalam beberapa partisi. Tujuan utama dari metode *k-means* adalah meminimalkan fungsi fungsi yang telah ditetapkan dalam proses klasterisasi. Hal tersebut dilakukan untuk minimalisasi variasi antar kluster dengan maksimalisasi variasi dalam klaster yang berbeda [10].

### 2.4.2. Hierarchical Clustering

*Hierarchical clustering* adalah metode yang digunakan untuk klasterisasi dengan menciptakan hierarki data dengan cara melakukan pengelompokan berbagai entitas berdasarkan tingkat persamaan tertentu. Tujuan utama metode ini adalah untuk mendapatkan hasil representasi visual yang memperlihatkan hubungan antar entitas. Pada tahun 1963, Robert L. Sibson memperkenalkan sebuah algoritma yang dinamakan “*single linkage*” di mana algoritma ini menggabungkan klaster berdasarkan jarak yang paling dekat antar dua entitas dalam klaster. Kemudian, pada tahun 1967, L.L. Cavalli-Sforza dan A. W. Edwards mengembangkan algoritma yang diberinama algoritma “*complete linkage*” yang menggabungkan klaster berdasarkan jarak paling jauh antara entitas dalam klaster. Pada tahun yang sama, algoritma “*average linkage*” muncul dan diperkenalkan oleh Maurice S. Johnson dengan menggunakan rata-rata jarak antara entitas sebagai cara untuk penggabungan klaster [11].

Proses analisis dengan metode *hierarchical clustering* dimulai dengan melakukan pengelompokan dua atau lebih individu atau objek yang memiliki hubungan terdekat satu

sama lain. Kemudian, objek lain yang mempunyai jarak terdekat kedua ditambahkan, sampai selanjutnya. Jarak Euclidean digunakan sebagai metode untuk melakukan pengukuran kedekatan dengan menggunakan rumus sebagai berikut.

$$d_{ik} = \sqrt{\sum_{j=1}^p (X_{ij} - X_{kj})^2} \quad (1)$$

Di mana:  $d_{ik}$  = jarak Euclidean,  $X_{ij}$  = objek ke-i dalam variabel ke-j,  $X_{kj}$  = objek ke-k dalam variabel ke-j, dan  $p$  = jumlah variabel penelitian.

Secara umum, *hierarchical clustering* dilakukan dengan menggunakan dua metode, yaitu *agglomerative nesting* dan *divisive analysis* yang akan dijelaskan sebagai berikut.

### 1. *Agglomerative Nesting (AGNES)*

*Agglomerative clustering* atau yang sering disebut dengan “*bottom-up clustering*,” dimulai dengan menjadikan setiap objek sebagai individu dalam kelompok. Selanjutnya, beberapa objek tersebut yang dinilai hampir identik akan digabungkan dengan cara berpasangan-pasangan menjadi satu *cluster* [12]. *Agglomerative nesting* dapat dilakukan dengan beberapa metode, yaitu:

#### a) *Single Linkage (Nearest Neighbor)*

*Single Linkage* berfungsi untuk mengukur jarak antar klaster sebagai jarak minimum antara pasangan titik data dari dua klaster.

#### b) *Complete Linkage (Furthest Neighbor)*

*Complete linkage* berfungsi untuk mengukur jarak antara dua klaster sebagai jarak maksimum antara pasangan titik data dari dua klaster.

#### c) *Average Linkage (Centroid Groups Methods)*

*Average linkage* berfungsi untuk mengukur jarak antar klaster sebagai rata-rata jarak antara seluruh pasangan titik data dari dua klaster. Sedangkan, *centroid* adalah rata-rata jarak antar objek dalam klaster. Dalam metode ini, proses pembentukan ulang *centroid* dilakukan setiap klaster baru sampai terbentuk sebuah klaster tetap.

#### d) *Centroid Linkage*

*Centroid linkage* berfungsi untuk mengukur jarak antara dua klaster sebagai jarak antara *centroid* (titik pusat rata-rata) dari dua klaster.

#### e) *Ward's Linkage*

*Ward's linkage* berfungsi untuk mengukur jarak antara dua klaster berdasarkan peningkatan jumlah total varian dalam klaster setelah penggabungan dua *cluster*.

### 2. *Divisive Analysis (DANA)*

*Divisive clustering*, atau yang dikenal dengan “*top-down clustering*,” di mana proses ini dimulai dengan mengelompokkan seluruh objek dalam satu kelompok besar. Kemudian, kelompok besar tersebut akan dipecah berdasarkan perbedaan signifikan menjadi beberapa kelompok antara objek-objeknya yang lebih kecil. Proses ini akan tetap berlanjut sampai setiap objek atau individu tergabung dalam klaster yang berbeda [13].

## 2.5. *Validasi Cluster*

Validasi *cluster* adalah tahapan di mana peneliti memastikan bahwa hasil klasterisasi dapat merepresentatif data secara umum dan mampu digeneralisasi untuk objek-objek lain. Diharapkan bahwa representasi ini akan tetap stabil selama beberapa periode waktu. Proses validasi ini bertujuan untuk memeriksa kestabilan hasil analisis klaster yang diperoleh dari proses analisis data. Menurut Fauziyyah & Sholikhah, terdapat dua jenis pengukuran yang sering digunakan untuk menentukan kualitas kluster yang baik, yaitu pengukuran *internal measures* dan *stability measures* [14].

### 2.5.1. Internal Measure (Ukuran Internal)

*Internal Measure* adalah ukuran yang digunakan untuk mengevaluasi kualitas klaster yang dihasilkan dari suatu metode klastering. Tujuan dari internal measure adalah untuk menilai seberapa baik objek-objek dalam klaster yang sama berkumpul bersama dan seberapa jauh klaster tersebut terpisah dari klaster atau kelompok lainnya [15]. Berikut ini merupakan jenis dari validasi *internal measure* beserta rumusnya.

#### 1. Connectivity Index

*Connectivity index* berfungsi untuk mengukur jarak kedekatan setiap objek dalam satu klaster dengan objek-objek lain dalam klaster tersebut. Rumus *connectivity index* adalah sebagai berikut.

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)} \quad (2)$$

di mana:  $N$  = jumlah kluster dan  $nni(j)$  = pengamatan untuk tetangga terdekat (*nearest neighbor*) i ke-j dan L.

#### 2. Silhouette Index

*Silhouette index* berfungsi untuk mengukur seberapa cocok setiap objek dengan klasternya sendiri dibandingkan dengan klaster tetangga terdekatnya. Rumus *silhouette index* adalah sebagai berikut.

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \quad (3)$$

di mana:  $a(x)$  = rata-rata jarak antara objek dalam kluster dan  $b(x)$  = jarak rata-rata antara objek dalam kluster tetangga terdekat.

#### 3. Dunn Index

*Dunn index* berfungsi untuk mengukur rasio antara jarak terpendek antara dua klaster yang berbeda dengan jarak paling jauh di antara objek yang berbeda dalam klaster yang sama. Rumus *dunn index* adalah sebagai berikut.

$$D = \frac{d_{min}}{d_{max}} \quad (4)$$

di mana:  $d_{min}$  = jarak terdekat antara objek dalam klaster lain dan  $d_{max}$  = jarak paling besar dalam masing-masing klaster.

### 2.5.2. Stability Measure (Ukuran Stabilitas)

*Stability Measure* adalah ukuran yang digunakan untuk mengukur seberapa konsisten hasil klastering ketika data diulang/ketika sampel data ditambah atau diubah sedikit. Semakin konsisten hasil klasteringnya, semakin baik algoritma *clustering* tersebut. Rumus dari validasi *stability measure* adalah sebagai berikut [15].

#### 1. Average Proportion of Non-overlap (APN)

*Average Proportion of Non-overlap* berfungsi untuk mengukur proporsi rata-rata jumlah observasi pada klaster yang sama antara data lengkap dengan data yang telah dihilangkan satu variabelnya. Misalkan  $C^{i,0}$  mewakili *cluster* yang berisi observasi  $i$  menggunakan *clustering* asli (berdasarkan data lengkap), dan  $C^{i,1}$  mewakili *cluster* yang berisi observasi  $i$  di mana *clustering* didasarkan pada *dataset* dengan satu variabel dihilangkan. Kemudian dengan jumlah klaster (k), maka ukuran APN didefinisikan sebagai berikut.

$$APN(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M \left(1 - \frac{n(C^{i,1} \cap C^{i,0})}{n(C^{i,0})}\right) \quad (5)$$

Interval nilai dari APN [0,1], dengan nilai mendekati 0 akan sesuai dengan hasil *clustering* yang sangat konsisten.

2. *Average Distance* (AD)

*Average Distance* untuk menghitung rata-rata jarak antar observasi dalam klaster yang sama pada data lengkap dan data yang telah dilakukan eliminasi satu variabelnya. AD didefinisikan sebagai berikut.

$$AD(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M \frac{1}{n(C^{i,0})n(C^{i,1})} \left[ \sum_{i \in C^{i,0}, j \in C^{i,1}} dist(i, j) \right] \quad (6)$$

*Average Distance* memiliki nilai antara 0 sampai tak hingga, dan nilai yang semakin kecil akan semakin baik.

3. *Average Distance Between Means* (ADM)

*Average Distance Between Means* berfungsi untuk menghitung jarak rata-rata antara pusat klaster dalam klaster berdasarkan data lengkap dan data yang telah dihilangkan satu variabelnya. ADM dapat didefinisikan sebagai berikut.

$$ADM(K) = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M dist(\bar{x}_{C^{i,1}}, \bar{x}_{C^{i,0}}) \quad (7)$$

di mana:  $\bar{x}_{C^{i,0}}$  = rata-rata dari observasi pada *cluster* yang berisi observasi  $i$  ketika *clustering* didasarkan pada data lengkap, dan  $\bar{x}_{C^{i,1}}$  didefinisikan serupa. ADM memiliki nilai antara nol sampai dengan tak hingga, dan semakin kecil nilai akan semakin baik.

4. *Figure of Merit* (FOM)

*Figure of Merit* untuk mengukur rata-rata varians dalam klaster pada observasi dengan data yang telah dieliminasi satu variabel, di mana *clustering* didasarkan pada sampel yang tersisa. FOM didefinisikan sebagai berikut.

$$FOM(I, K) = \sqrt{\frac{1}{MN} \sum_{i=1}^N \sum_{l \in C_k(i)} dist(x_{i,l}, \bar{x}_{C_k(l)})} \quad (8)$$

di mana:  $x_{i,l}$  = nilai ke- $i$  observasi di kolom ke- $l$  *cluster*  $C_k(l)$  dan  $\bar{x}_{C_k(l)}$  = rata-rata *cluster*  $C_k(l)$ .

Saat ini, satu-satunya jarak yang tersedia untuk FOM adalah Euclidean. FOM dikalikan dengan faktor penyesuaian  $\sqrt{\frac{N}{N-R}}$ , untuk mengurangi kecenderungan menurun seiring dengan bertambahnya jumlah *cluster*. Skor akhir adalah rata-rata dari semua kolom yang dihapus, dan memiliki nilai antara nol sampai tak hingga, dengan nilai yang lebih kecil berarti kinerja yang lebih baik.

### 3. Hasil dan Pembahasan

#### 3.1. Analisis Deskriptif

Analisis deskriptif adalah suatu tahap eksplorasi data yang sangat diperlukan dalam penelitian untuk mengetahui informasi atau gambaran umum dari setiap variabel dengan

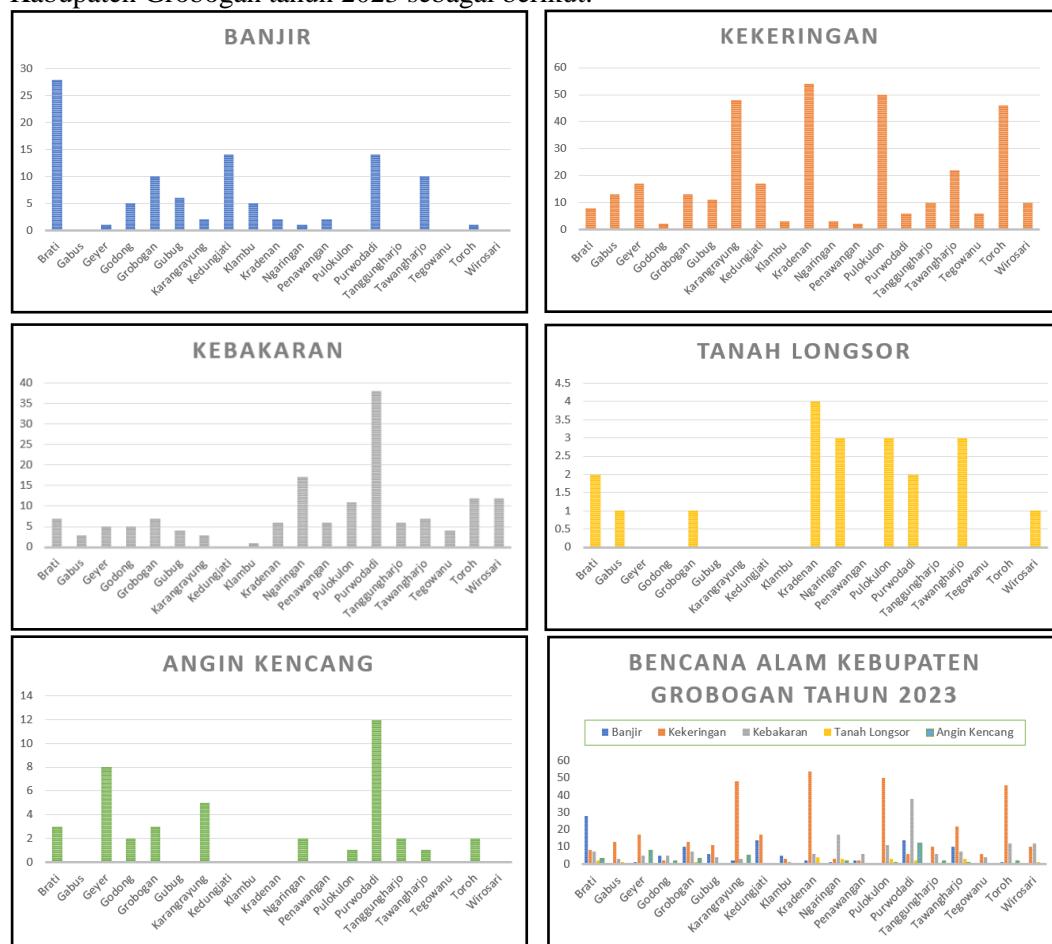
tujuan data yang disajikan menjadi mudah dipahami oleh pembaca. Berikut merupakan hasil analisis deskriptif setiap variabel yang akan disajikan dalam **Tabel 2** berikut.

**Tabel 2** Keterangan Variabel dan Simbol

Variabel	Uraian	Percentase
Banjir	101	15%
Kekeringan	341	52%
Kebakaran	154	24%
Tanah Longsor	20	3%
Angin Kencang	41	6%
<b>Total</b>	<b>556</b>	<b>100%</b>

Berdasarkan **Tabel 2**, dapat dilihat persentase bencana alam di Kabupaten Grobogan tahun 2023. Diketahui bahwa terdapat 556 titik kejadian bencana alam yang terjadi di Kabupaten Grobogan tahun 2023, di mana sekitar 15% atau sebanyak 101 titik adalah kejadian banjir, sekitar 52% atau sebanyak 341 titik adalah kejadian kekeringan, sekitar 24% atau sebanyak 154 titik adalah kejadian kebakaran, sekitar 3% atau sebanyak 20 titik kejadian tanah longsor, dan sekitar 6% atau sebanyak 41 titik kejadian angin kencang.

Selanjutnya, akan ditampilkan grafik dari data jumlah titik kejadian bencana alam Kabupaten Grobogan tahun 2023 sebagai berikut.



**Gambar 2** Grafik Bencana Alam di Kabupaten Grobogan Tahun 2023

Berdasarkan **Gambar 2**, dapat dilihat bahwa tercatat sebanyak 556 titik kejadian bencana alam yang terjadi di Kabupaten Grobogan tahun 2023. Di mana 556 titik kejadian bencana alam tersebut terdiri dari 101 titik kejadian bencana alam banjir, 341 titik kejadian

bencana alam kekeringan, 154 titik kejadian bencana alam kebakaran, 20 titik kejadian bencana alam tanah longsor, dan 41 titik kejadian bencana alam angin kencang. Secara keseluruhan, dapat disimpulkan bahwa bencana alam yang paling sering terjadi di Kabupaten Grobogan tahun 2023 adalah kekeringan dengan titik kejadian kekeringan terbanyak terjadi di Kecamatan Kradenan dengan jumlah 54 titik kejadian bencana alam. Selanjutnya, bencana alam yang terjadi di hampir seluruh daerah di Kabupaten Grobogan adalah bencana kekeringan dan kebakaran. Kecamatan dengan titik kejadian bencana alam terbanyak adalah Kecamatan Purwodadi dengan total kejadian sebanyak 72 titik kejadian. Hal tersebut terjadi karena Kecamatan Purwodadi sering mengalami bencana alam, apalagi masalah banjir karena letak Kecamatan Purwodadi dekat dengan Sungai Lusi dan Sungai Serang yang sering meluap karena tidak mampu menampung debit air. Sedangkan, kecamatan dengan titik kejadian bencana alam terendah adalah Kecamatan Brati dengan total kejadian sebanyak 9 titik kejadian.

Setelah itu, akan ditampilkan tabel yang memuat ringkasan data titik kejadian bencana alam Kabupaten Grobogan tahun 2023 sebagai berikut.

**Tabel 3** Ringkasan Bencana Alam di Kabupaten Grobogan Tahun 2023

Statistika Deskriptif	Variabel				
	Banjir	Kekeringan	Kebakaran	Tanah Longsor	Angin Kencang
Min	0	2	0	0	0
Median	2	11	6	0	1
Mean	5.316	17.95	8.105	1.053	2.158
Max	28	54	38	4	12

Berdasarkan **Tabel 3**, dapat dilihat ringkasan data dari setiap variabel jumlah titik kejadian bencana alam. Di mana diketahui bahwa terdapat lima variabel yang digunakan yaitu jumlah titik kejadian bencana seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin kencang. Analisis deskriptif setiap variable penelitian akan dijelaskan secara rinci sebagai berikut.

- *Min* menunjukkan jumlah titik kejadian minimum setiap variabel jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 yaitu sebanyak 0 titik kejadian untuk banjir, kebakaran, tanah longsor, dan angin kencang. Sedangkan, untuk kekeringan terdapat sebanyak 2 titik kejadian bencana alam.
- Median menunjukkan nilai tengah setiap variabel jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 yaitu sebesar 2 untuk titik kejadian banjir, 11 untuk titik kejadian kekeringan, 6 untuk titik kejadian kebakaran, 0 untuk titik kejadian tanah longsor, dan 1 untuk titik kejadian angin kencang.
- *Mean* menunjukkan nilai rata-rata setiap variabel jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 yaitu sebesar 5.316 untuk titik kejadian banjir, 17.95 untuk titik kejadian kekeringan, 8.105 untuk titik kejadian kebakaran, 1.053 untuk titik kejadian tanah longsor, dan 2.158 untuk titik kejadian angin kencang.
- *Max* menunjukkan jumlah titik kejadian bencana alam maksimum dari setiap variabel jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 yaitu sebesar 28 titik kejadian untuk banjir, 54 kejadian untuk titik kejadian kekeringan, 38 titik kejadian untuk kebakaran, 4 kejadian untuk titik kejadian untuk tanah longsor, dan 12 titik kejadian untuk angin kencang.

### 3.2. Uji Multikolinearitas

Uji multikolinearitas dilakukan dengan tujuan untuk mengetahui ada tidaknya korelasi antar variable independent/bebas dalam suatu model regresi. Tujuan pengujian

multikolinearitas adalah untuk menentukan apakah model regresi menemukan korelasi antara variabel independent/bebas bebas (Ghozali, 2016). Untuk memeriksa apakah multikolinearitas ada dalam model regresi, periksa nilai toleransi dan *Variance Inflation Factor* (VIF). Nilai toleransi mengukur variabilitas dalam variabel independen yang dipilih yang tidak dapat dijelaskan oleh variabel independen lainnya. Sehingga, nilai nilai toleransi yang rendah setara dengan nilai VIF yang tinggi. Hal ini karena  $VIF = 1/\text{nilai toleransi}$  yang menunjukkan kolinearitas tinggi. Nilai *cut off* yang digunakan adalah nilai toleransi 0.10 atau nilai VIF lebih besar dari 10. Pengujian tersebut dapat dilakukan dengan melihat nilai toleransi dan *Variance Inflation Factor* (VIF) dari model regresi [16]. Kriteria pengujian multikolinearitas adalah sebagai berikut:

1. Apabila nilai  $VIF < 10$  atau nilai toleransi  $> 0.01$ , maka dinyatakan tidak terjadi multikolinearitas.
2. Apabila Nilai  $VIF > 10$  atau nilai toleransi  $< 0.01$ , maka dinyatakan terjadi multikolinearitas.
3. Apabila koefisien korelasi masing-masing variabel independen/bebas  $> 0.8$ , maka terjadi multikolinearitas. Tetapi, apabila koefisien korelasi masing-masing variabel independen/bebas  $< 0.8$  maka tidak terjadi multikolinearitas.

Nilai korelasi/hubungan antara variabel penelitian tersebut akan disajikan dalam **Tabel 4** sebagai berikut.

**Tabel 4** Korelasi Antar Variabel Penelitian

Variabel	Banjir	Kekeringan	Kebakaran	Tanah Longsor	Angin Kencang
Banjir	1	-0.2404	0.1544	0.1629	0.2574
Kekeringan	-0.2404	1	-0.0819	0.3401	-0.0306
Kebakaran	0.1544	-0.0819	1	0.3879	0.6709
Tanah Longsor	0.1629	0.3401	0.3879	1	0.0238
Angin Kencang	0.2574	-0.0306	0.6709	0.0238	1

Berdasarkan **Tabel 4**, dapat dilihat bahwa nilai korelasi/hubungan antara masing-masing variabel di mana nilai korelasi variabel banjir dengan kekeringan sebesar -0.2404, korelasi variabel banjir dengan kebakaran sebesar 0.1544, nilai korelasi variabel banjir dengan tanah longsor sebesar 0.1629, nilai korelasi variabel banjir dengan angin kencang sebesar 0.2574, korelasi variabel kekeringan dengan kebakaran sebesar -0.0819, nilai korelasi variabel kekeringan dengan tanah longsor sebesar 0.3401, korelasi variabel kekeringan dengan angin kencang sebesar -0.0306, nilai korelasi variabel kebakaran dengan tanah longsor sebesar 0.3879, nilai korelasi variabel kebakaran dengan tanah longsor sebesar 0.1544, dan korelasi variabel tanah longsor dengan angin kencang sebesar 0.0238. Hal tersebut dapat dilihat bahwa nilai korelasi/hubungan antara masing-masing variabel  $< 0.8$ . Sehingga, secara keseluruhan dapat diambil kesimpulan bahwa tidak terjadi multikolinearitas.

### 3.3. Penentuan Jumlah *Cluster* Optimum

Dalam penelitian ini, akan dilanjutkan dengan menentukan banyak *cluster* optimum ( $k$ ) menggunakan dua metode yang berbeda: *k-means* dan *hierarchical clustering*. Algoritma dalam menentukan banyak *cluster* optimum ( $k$ ) menggunakan metode *k-means* dan *hierarchical clustering* adalah sebagai berikut.

#### 3.3.1. *Cluster* Optimum *K-means clustering*

Dalam proses menganalisis *K-means clustering* menggunakan data yang telah distandardisasi. Penentuan jumlah *cluster* optimum dalam *k-means* menggunakan pendekatan metode Elbow. Menurut Bholowalia & Kumar (2014) algoritma metode Elbow dalam menentukan jumlah *cluster* dalam metode *k-means* adalah sebagai berikut [17].

1. Mulai
2. Menginisiasi nilai  $k$  awal
3. Menaikkan nilai  $k$
4. Menghitung nilai jumlah kesalahan kuadrat total pada setiap nilai  $k$
5. Mengamati penurunan yang signifikan dalam jumlah kesalahan kuadrat total dalam setiap nilai  $k$
6. Menentukan nilai  $k$  yang membentuk siku pada grafik
7. Selesai.

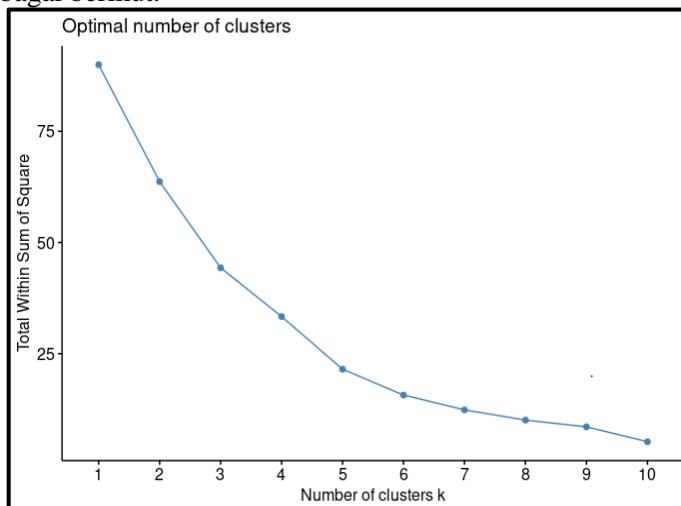
Berdasarkan algoritma tersebut, untuk menentukan jumlah *cluster* optimum ( $k$ ) pada data jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 akan disajikan table **Tabel 5** yang memuat nilai *sum of square error* adalah sebagai berikut.

**Tabel 5** Nilai *Sum of Square Error*

Jumlah <i>Cluster</i> ( $K$ )	<i>Sum of Square Error</i>	Selisih
$K = 2$	64.1%	64.1%
$K = 3$	78.3%	<b>14.2%</b>
$K = 4$	86.8%	8.5%
$K = 5$	89.4%	2.6%

Berdasarkan **Tabel 5**, dapat dilihat bahwa nilai *sum of square error* mengalami penurunan secara drastis atau optimum terjadi pada saat jumlah *cluster* ( $k$ ) = 3 yaitu mengalami penurunan sebesar 14,2%.

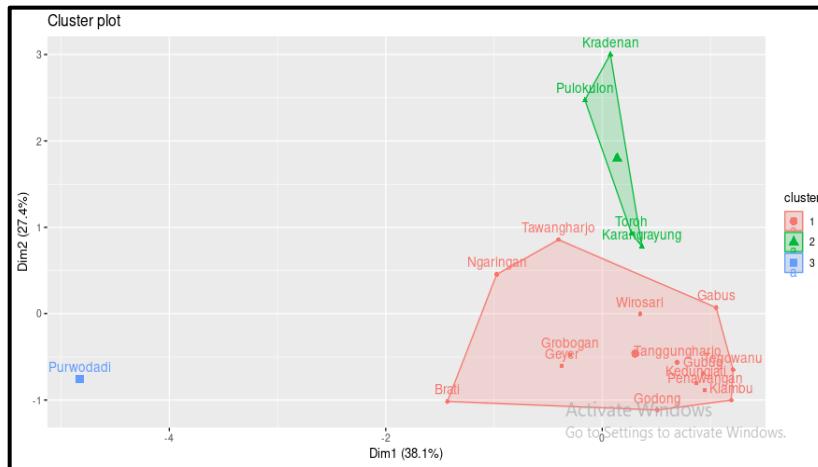
Selanjutnya, akan ditampilkan grafik Elbow dalam menentukan jumlah *cluster* optimum ( $k$ ) sebagai berikut.



**Gambar 3** Metode Elbow

Berdasarkan **Gambar 3**, dapat dilihat bahwa penurunan nilai  $k$  secara drastis terjadi pada  $k = 1$  menuju  $k = 2$  namun mengalami landai atau membentuk siku pada  $k = 3$ . Menurut (Bholowalia & Kumar, 2014) apabila dalam grafik *cluster* pertama dan *cluster* kedua mengalami penurunan drastis yang begitu besar, maka *cluster* tersebut dianggap sebagai *cluster* yang optimal. Sehingga, secara keseluruhan dapat disimpulkan bahwa berdasarkan nilai *sum of square error* dan grafik Elbow didapatkan jumlah *cluster* optimum ( $k$ ) yaitu sebanyak 3 *cluster*.

Setelah mendapatkan *cluster* optimum yaitu sebanyak 3 *cluster* dengan menggunakan metode *k-means*, akan dilanjutkan dengan melakukan membuat *plot* dengan pengelompokan data sebagai berikut.



Gambar 4 Hasil Pengelompokan *K-means Clustering*

Berdasarkan **Gambar 4**, dapat dilihat hasil pengelompokan menggunakan metode *k-means clustering* dengan jumlah *cluster* optimum (*k*) sebanyak 3 *cluster* beserta anggotanya. Terlihat bahwa *cluster* 1 memiliki anggota sebanyak 14 kecamatan, *cluster* 2 memiliki anggota sebanyak 4 kecamatan, dan *cluster* 3 memiliki anggota sebanyak 1 kecamatan. Hal tersebut terjadi karena Kecamatan Purwodadi sering mengalami bencana alam, apalagi masalah banjir karena letak Kecamatan Purwodadi dekat dengan Sungai Lusi dan Sungai Serang yang sering meluap karena tidak mampu menampung debit air.

### 3.3.2. Cluster Optimum *Hierarchical clustering*

Penentuan jumlah *cluster* optimum dalam *hierarchical clustering* menggunakan metode *agglomerative clustering*. Dalam metode ini akan dilakukan perbandingan nilai koefisien korelasi *cophenetich* dengan ketentuan nilai berkisar antara nilai -1 sampai 1, di mana apabila nilai korelasi *cophenetich* atau *correlation cophenetich* mendekati 1 maka proses pengelompokan yang dihasilkan dapat dikatakan cukup baik.

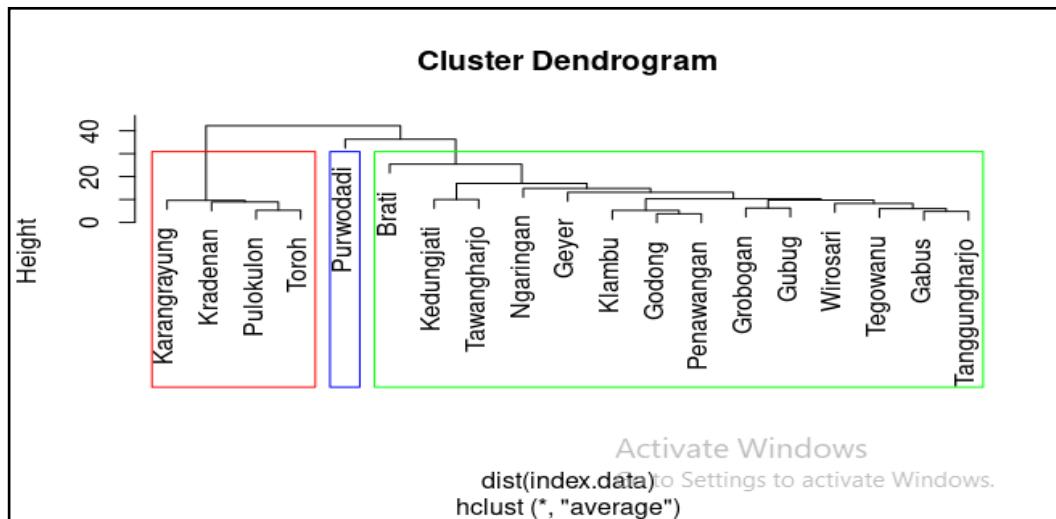
Dalam hal ini, akan dilakukan perbandingan nilai koefisien *correlation cophenetich* menggunakan metode *single linkage*, *complete linkage* dan *average linkage* akan disajikan dalam tabel berikut.

Tabel 6 Perbandingan Nilai Koefisien *Correlation Cophenetich*

Metode	Correlation Cophenetich
<i>Single Linkage</i>	0.9303105
<i>Average Linkage</i>	0.9419865
<i>Complete Linkage</i>	0.9120342

Berdasarkan **Tabel 6**, dapat dilihat bahwa nilai koefisien korelasi *cophenetich* tertinggi atau yang mendekati 1 adalah metode *average linkage*. Sehingga dapat disimpulkan bahwa dengan merujuk pada nilai korelasi *cophenetich*, metode pendekatan *average linkage* merupakan metode yang cukup baik digunakan untuk mengelompokkan data jumlah titik kejadian bencana alam Kabupaten Grobogan tahun 2023 dibandingkan dengan metode *single linkage* dan *complete linkage*.

Setelah mendapatkan metode terbaik, berdasarkan metode *average linkage* akan dilanjutkan dengan menentukan jumlah *cluster* (*k*) optimum. Fadliana & Rozi, 2016 menyatakan bahwa “untuk menentukan banyaknya *cluster* selain menggunakan selisih koefisien jarak penggabungan terbesar, juga dapat menggunakan dendrogram, yakni dengan cara memotong garis yang memiliki selisih terpanjang” [18]. Selanjutnya, akan ditampilkan dendrogram beserta pengelompokan data sebagai berikut.



Gambar 5 Hasil Pengelompokan *Hierarchical clustering*

Berdasarkan **Gambar 5**, dapat dilihat bahwa dendogram tersebut terpotong membentuk 3 *cluster* beserta kelompok anggotanya. Sehingga, dapat disimpulkan bahwa dengan metode *average linkage hierarchical clustering* jumlah *cluster* optimum (*k*) yang terbentuk dalam mengelompokkan kecamatan di Kabupaten Grobogan berdasarkan jumlah titik kejadian bencana alam tahun 2023 adalah sebanyak 3 *cluster*.

### 3.4. Analisis Cluster

Setelah didapatkan sebanyak 3 *cluster* optimum baik metode *k-means* maupun *hierarchical clustering*. Kemudian, akan dilanjutkan dengan melakukan analisis pada masing-masing metode sebagai berikut.

#### 3.4.1. Analisis *K-means Clustering*

Diketahui bahwa jumlah *cluster* optimum (*k*) yang terbentuk dengan menggunakan metode *k-means clustering* berdasarkan nilai *sum of square error* dan grafik Elbow adalah 3 *cluster*. Hasil analisis dari *cluster* yang terbentuk akan disajikan dalam bentuk **Tabel 7** sebagai berikut.

Tabel 7 Analisis *K-means Clustering*

Cluster	Rata-rata						Kategori
	Banjir	Kekeringan	Kebakaran	Tanah Longsor	Angin Kencang		
1	5.8571	9.7857	6	0.7857	1.5		Rendah
2	1.25	49.5	8	1.75	2.0		Sedang
3	14.0	6.0	38	2.0	12.0		Tinggi

Berdasarkan **Tabel 7**, dapat diketahui bahwa rata-rata jumlah titik kejadian bencana alam Kabupaten Grobogan tahun 2023 pada setiap *cluster* dikategorikan berdasarkan nilai tersebut. *Cluster* dengan nilai rata-rata terendah diberi keterangan “rendah”, *cluster* dengan nilai rata-rata sedang diberi keterangan “sedang”, dan *cluster* dengan nilai dengan rata-rata tertinggi diberi keterangan “tinggi”. Interpretasi *cluster* beserta anggota masing-masing *cluster* dengan menggunakan metode *k-means* adalah sebagai berikut:

1. *Cluster 1*: *Cluster* ini dikategorikan sebagai “rendah” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebuk terendah yakni masing-masing 5.8571, 9.7857, 6, 0.8757 dan 1.5. Terdapat 14 kecamatan dalam *cluster* ini, termasuk Kecamatan Brati,

- Geyer, Gabus, Godong, Gubug, Grobogan, Kedungjati, Klambu, Ngaringan, Penawangan, Tawangharjo, Tanggunharjo, Tegowanu, dan Wirosari.
2. *Cluster 2:* *Cluster* ini dikategorikan sebagai “sedang” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebut sedang yakni masing-masing 1.25, 49.5, 8, 1.75 dan 2. Terdapat 4 kecamatan dalam *cluster* ini, termasuk Kecamatan Karangrayung, Kradenan, Pulokulon, dan Toroh.
  3. *Cluster 3:* *Cluster* ini dikategorikan sebagai “tinggi” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebut tertinggi yakni masing-masing 14, 6, 38, 2, dan 12. Hanya terdapat 1 kecamatan dalam *cluster* ini, yaitu Kecamatan Purwodadi.

### 3.4.2. Analisis *Hierarchical clustering*

Setelah itu, dilanjutkan dengan menganalisis metode *hierarchical clustering*. Sudah diketahui bahwa jumlah *cluster* yang terbentuk dengan metode *hierarchical clustering* berdasarkan dendogram adalah 3 *cluster*. Hasil analisis dari *cluster* yang terbentuk akan disajikan dalam bentuk tabel seperti berikut.

**Tabel 8** Analisis *Hierarchical clustering*

Cluster	Rata-rata					Kategori
	Banjir	Kekeringan	Kebakaran	Tanah Longsor	Angin Kencang	
1	5.8571	9.7857	6	0.7857	1.5	Rendah
2	1.25	49.5	8	1.75	2.0	Sedang
3	14.0	6.0	38	2.0	12.0	Tinggi

Berdasarkan **Tabel 8**, diketahui bahwa rata-rata jumlah titik kejadian bencana alam masing-masing *cluster* dan diberi kategori berdasarkan nilai tersebut. Kemudian, dilanjutkan dengan melakukan interpretasi *cluster* beserta anggota masing-masing *cluster* dengan metode *hierarchical clustering* adalah sebagai berikut.

1. *Cluster 1:* *Cluster* ini dikategorikan sebagai “rendah” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebut terendah yakni masing-masing 5.8571, 9.7857, 6, 0.8757 dan 1.5. Terdapat 14 kecamatan dalam *cluster* ini, termasuk Kecamatan Brati, Geyer, Gabus, Godong, Gubug, Grobogan, Kedungjati, Klambu, Ngaringan, Penawangan, Tawangharjo, Tanggunharjo, Tegowanu, dan Wirosari.
2. *Cluster 2:* *Cluster* ini dikategorikan sebagai “sedang” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebut sedang yakni masing-masing 1.25, 49.5, 8, 1.75 dan 2. Terdapat 4 kecamatan dalam *cluster* ini, termasuk Kecamatan Karangrayung, Kradenan, Pulokulon, dan Toroh.
3. *Cluster 3:* *Cluster* ini dikategorikan sebagai “tinggi” karena karena memiliki nilai rata-rata jumlah titik kejadian bencana alam seperti banjir, kekeringan, kebakaran, tanah longsor, dan angin rebut tertinggi yakni masing-masing 14, 6, 38, 2, dan 12. Hanya terdapat 1 kecamatan dalam *cluster* ini, yaitu Kecamatan Purwodadi.

### 3.5. Perbandingan *K-Means* dan *Hierarchical Clustering*

Setelah itu, akan dilanjutkan dengan membuat perbandingan metode *k-means* dengan *hierarchical clustering* yang dapat dilihat dari jumlah *cluster* optimum (*k*) yang terbentuk dan anggota dari masing-masing *cluster*. Jumlah *cluster* optimum (*k*) yang terbentuk baik menggunakan metode *k-means* maupun *hierarchical clustering* sama yaitu sebanyak 3

*cluster*. Berikut merupakan tabel yang memuat perbandingan anggota *cluster* dari metode *k-means* maupun *hierarchical clustering*.

**Tabel 9** Perbandingan Anggota *K-means* dan *Hierarchical Clustering*

Cluster	Metode		Kategori
	<i>K-means clustering</i>	<i>Hierarchical clustering</i>	
1	Brati, Geyer, Gabus, Godong, Gubug, Grobogan, Kedungjati, Klambu, Ngaringan, Penawangan, Tawangharjo, Tegowanu, Tanggungharjo, Wiroso	Brati, Geyer, Gabus, Godong, Gubug, Grobogan, Kedungjati, Klambu, Ngaringan, Penawangan, Tawangharjo, Tegowanu, Tanggungharjo, Wiroso	Ringan
	Karangrayung, Kradenan, Pulokulon, Toroh	Karangrayung, Kradenan, Pulokulon, Toroh	
	Purwodadi	Purwodadi	
2			Sedang
3			Tinggi

Berdasarkan **Tabel 9**, dapat dilihat bahwa *cluster* yang terbentuk dan anggota klaster metode *k-means* identik dengan metode *hierarchical clustering*. Kedua metode tersebut memiliki jumlah *cluster* sebanyak 3 *cluster* dan anggota *cluster* yang terbentuk dalam kategori rendah sebanyak 14 kecamatan, kategori sedang sebanyak 4 kecamatan, dan kategori tinggi sebanyak 1 kecamatan. Oleh karena itu, untuk mendapatkan metode *cluster* terbaik akan dilanjutkan dengan melakukan validasi *cluster*.

### 3.6. Validasi *Cluster*

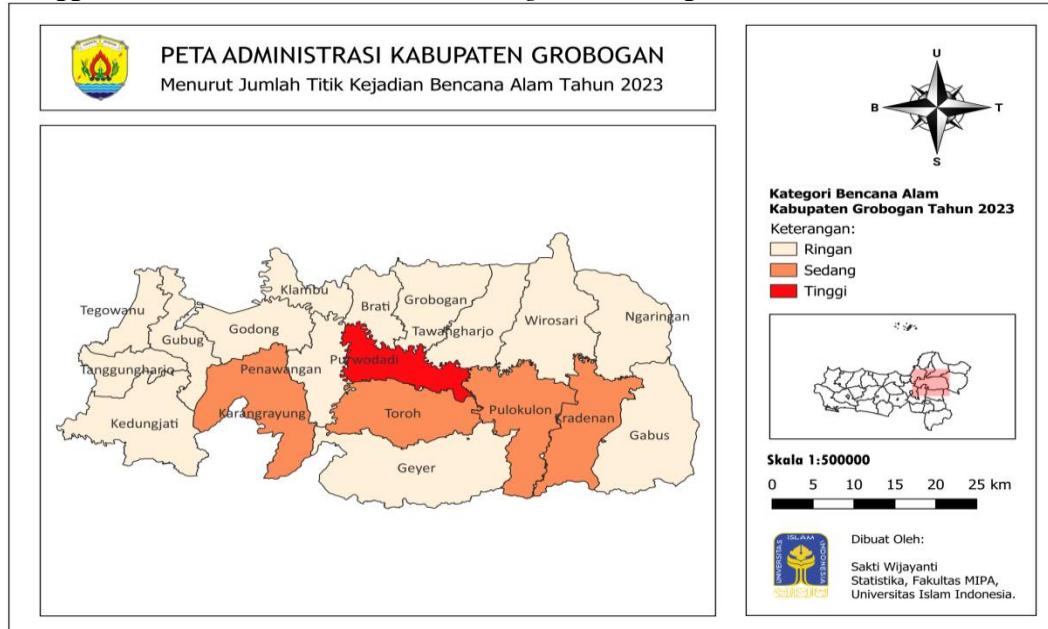
Diketahui bahwa jumlah *cluster* yang terbentuk dan anggota masing-masing *cluster* dengan metode *k-mean* identik dengan metode *hierarchical clustering*. Oleh karena itu, untuk mendapatkan metode *cluster* terbaik akan dilanjutkan dengan melakukan validasi *cluster* dengan *internal* dan *stability measure* dalam metode *k-means* dan *hierarchical clustering* yang akan disajikan dalam bentuk tabel seperti berikut.

**Tabel 10** Validasi *Cluster*

Komponen	Internal Measure		Stability Measure	
	<i>K-Means</i>	<i>Hierarchical</i>	<i>K-Means</i>	<i>Hierarchical</i>
Connectivity	7.3115	7.3115	APN	0.0962
Dunn Index	0.8991	0.8991	AD	13.6823
Silhouette	0.5932	0.5932	ADM	4.1335
			FOM	7.4761
				7.1598

Berdasarkan **Tabel 10**, dapat dilihat bahwa dengan menggunakan metode *k-means* maupun *hierarchical clustering* didapatkan nilai *connectivity*, *dunn index*, dan *silhouette* yang sama yaitu nilai *connectivity* sebesar 7.3115, nilai *dunn index* sebesar 0.8991, dan nilai *silhouette* sebesar 0.5932. Selanjutnya dalam validasi *cluster* dengan *stability measure* metode *k-means* menghasilkan nilai APN sebesar 0.0962, AD sebesar 13.6823, ADM sebesar 4.1335, dan FOM sebesar 7.4761. Sedangkan, metode *Hierarchical clustering* menghasilkan nilai APN sebesar 0.0391, AD sebesar 13.4789, ADM sebesar 3.5167, dan FOM sebesar 7.1598. Dalam hal ini, dapat dilihat bahwa *hierarchical clustering* menghasilkan nilai APN, AD, ADM, dan FOM yang lebih kecil dibandingkan dengan metode *k-means clustering*. Secara keseluruhan, dapat disimpulkan bahwa metode yang dinilai lebih baik untuk mengelompokkan data jumlah titik kejadian bencana alam di Kabupaten Grobogan tahun 2023 adalah metode *hierarchical clustering*.

Setelah diketahui bahwa metode terbaik yang digunakan adalah metode *hierarchical clustering*, akan dilanjutkan dengan membuat visualisasi dari hasil *clustering* dengan menyajikan data statistik dalam bentuk peta sederhana. Visualisasi hasil pengelompokan menggunakan metode *hierarchical clustering* adalah sebagai berikut.



Gambar 6 Visulasi Pengelompokan Metode *Hierarchical Clustering*

Berdasarkan **Gambar 6**, dapat dilihat visualisasi hasil pengelompokan metode *hierarchical clustering* di mana warna merah muda menunjukkan *cluster* 1 dengan kategori rendah (rata-rata titik kejadian bencana alam terendah), warna jingga menunjukkan *cluster* 2 dengan kategori sedang (rata-rata titik kejadian bencana alam sedang), dan warna merah menunjukkan *cluster* 3 dengan kategori tinggi (rata-rata titik kejadian bencana alam tertinggi)

#### 4. Kesimpulan

Berdasarkan hasil penelitian tentang analisis *cluster* untuk mengelompokkan kecamatan di Kabupaten Grobogan menurut jumlah titik kejadian bencana alam tahun 2023 menggunakan metode *k-means* dan *hierarchical clustering*, dapat disimpulkan sebagai berikut:

1. Gambaran umum dari data "Jumlah Titik Kejadian Bencana Alam Kabupaten Grobogan Tahun 2023" adalah terdapat sebanyak 556 titik kejadian bencana alam Kabupaten Grobogan tahun 2023, di mana kerjadian tersebut terdiri dari 101 titik kejadian bencana alam banjir, 341 titik kejadian bencana alam kekeringan, 154 titik kejadian bencana alam kebakaran, 20 titik kejadian bencana alam tanah longsor, dan 41 titik kejadian bencana alam angin kencang. Secara keseluruhan, dapat disimpulkan bahwa bencana alam hampir terjadi di setiap kecamatan di Kabupaten Grobogan. Bencana alam sering terjadi di Kabupaten Grobogan tahun 2023 adalah bencana kekeringan dengan titik kejadian kekeringan terbanyak terjadi di Kecamatan Kradenan dengan jumlah 54 titik kejadian bencana alam. Selanjutnya, bencana alam yang terjadi di hampir seluruh kecamatan di Kabupaten Grobogan adalah bencana kekeringan dan kebakaran. Kecamatan dengan titik kejadian bencana alam terbanyak adalah Kecamatan Purwodadi dengan total kejadian sebanyak 72 titik kejadian. Hal ini terjadi karena Kecamatan Purwodadi sering mengalami

bencana alam, apalagi masalah banjir karena letak Kecamatan Purwodadi dekat dengan Sungai Lusi dan Sungai Serang Sedangkan, kecamatan dengan titik kejadian bencana alam terendah adalah Kecamatan Brati dengan total kejadian sebanyak 9 titik kejadian.

2. Jumlah *cluster* optimum ( $k$ ) yang terbentuk baik menggunakan metode *k-means clustering* maupun *hierarchical clustering* adalah 3 *cluster*. Selanjutnya, setiap *cluster* yang terbentuk diberi kategori rendah, sedang, dan tinggi dengan keterangan kategori berdasarkan nilai rata-rata jumlah titik kejadian bencana alam setiap *cluster*. Kedua metode tersebut memiliki jumlah *cluster* optimum ( $k$ ) sebanyak 3 *cluster* dan dengan menggunakan metode *k-means* maupun *hierarchical clustering* anggota *cluster* yang terbentuk dalam kategori rendah sebanyak 14 kecamatan, kategori sedang sebanyak 4 kecamatan, dan kategori tinggi sebanyak 1 kecamatan.
3. Hasil validasi *cluster internal measure* menggunakan metode *k-means* maupun *hierarchical clustering* didapatkan nilai *connectivity*, *dunn index*, dan *silhouette* yang sama yaitu nilai *connectivity* sebesar 7.3115, nilai *dunn index* sebesar 0.8991, dan nilai *silhouette* sebesar 0.5932. Selanjutnya dalam validasi *cluster* dengan *stability measure* metode *k-means* menghasilkan nilai APN sebesar 0.0962, AD sebesar 13.6823, ADM sebesar 4.1335, dan FOM sebesar 7.4761. Sedangkan, metode *hierarchical clustering* menghasilkan nilai APN sebesar 0.0391, AD sebesar 13.4789, ADM sebesar 3.5167, dan FOM sebesar 7.1598. Dalam hal ini, dapat dilihat bahwa *hierarchical clustering* menghasilkan nilai APN, AD, ADM, dan FOM yang lebih kecil dibandingkan dengan metode *k-means*. Sehingga, secara keseluruhan, dapat disimpulkan bahwa metode yang dinilai lebih baik untuk mengelompokkan data jumlah titik kejadian bencana alam banjir, kekeringan, kebakaran, tanah longsor dan angin kencang di Kabupaten Grobogan tahun 2023 adalah metode *hierarchical clustering*.

## 5. Datar Pustaka

- [1] BNPB, "Definisi Bencana," 2018. [Online]. Available: <https://bnpb.go.id/definisi-bencana>.
- [2] P. Utama, 12 January 2024. [Online]. Available: <https://news.detik.com/foto-news/d-7137805/sepanjang-2023-ada-4-940-bencana-alam-di-indonesia/3>.
- [3] S. Langit, PERATURAN BUPATI GROBOGAN NOMOR 30 TAHUN 2021, Grobogan, 2021.
- [4] Kemenkes, "Pusat Krisis Kesehatan," 30 August 2023. [Online]. Available: <https://pusatkrisis.kemkes.go.id/Kekeringan-di-GROBOGAN-JAWA-TENGAH-30-08-2023-15>.
- [5] A. Natalis and Y. Nataliani, "Pemanfaatan K-Means Clustering dan Analytic Hierarchy Process terhadap Penilaian Prestasi Kerja Pegawai," *Jurnal Teknik Informatika dan Sistem Informasi*, 2022.
- [6] M. A. J. A. Sibrani, I. G. S. M. Diyasa and Sugiarto, "Penggunaan K-Means dan Hierarchical Clustering Single Linkage dalam Pengelompokan Stok Obat," *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 2024.

- [7] Geografi, 07 October 2023. [Online]. Available: <https://geografi.id/jelaskan/pengertian-analisis-deskriptif/>.
- [8] A. W. Talakua, Z. A. Leleury and A. W. Talutta, "Analisis Cluster dengan Menggunakan Meode K-Means Untuk Mengelompokkan Kabupaten di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *Jurnal Ilmu Matematika dan Terapan*, vol. Volume 11 Nomor 2, pp. Hal. 119 -128, 2017.
- [9] E. Irwansyah, "CLUSTERING," 09 March 2023. [Online]. Available: [socs.binus.ac.id:https://socs.binus.ac.id/2017/03/09/clustering/](https://socs.binus.ac.id/2017/03/09/clustering/).
- [10] B. Orleans and E. P. Putra, "Clustering Algoritma (K-Means)," 31 January 2022. [Online]. Available: <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/>.
- [11] Kantinit, "Hierarchical Clustering Adalah: Jenis dan Cara Kerja," 19 Mei 2023. [Online]. Available: [https://kantinit.com/kecerdasan-buatan/hierarchical-clustering-adalah-jenis-dan-cara-kerja/#google\\_vignette](https://kantinit.com/kecerdasan-buatan/hierarchical-clustering-adalah-jenis-dan-cara-kerja/#google_vignette).
- [12] M. Hajar, Y. N. Nasution and S. Prangga, "Penerapan Metode Agglomerative Nesting (Agnes) Pada Pengelompokan Wilayah Desa atau Kelurahan Di Kabupaten Kutai Kartanegara (Studi Kasus: Data Potensi Desa (PODES) Tahun 2018)," *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, 2022.
- [13] J. Han, M. Kamber and J. Pei, Data Mining Third Edition: Concepts and Techniques, Waltham: Morgan Kaufmann, 2012.
- [14] N. A. Fauziyyah and I. Sholikhah, "Introduction to Hierarchical Clustering," 11 February 2021. [Online]. Available: <https://rpubs.com/inayatus/hierarchical-clustering>.
- [15] G. Brock, V. Pihur, S. Datta and S. Datta, "Journal of Statistical Software," vol. Volume 25, 008.
- [16] I. Ghazali, Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23, Edisi 8, Badan Penerbit Universitas Diponegoro, 2016.
- [17] P. Bholowalia and A. Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *International Journal of Computer Applications (0975 – 8887)*, vol. Volume 105, 2014.
- [18] A. Fadliana and F. Rozi, "Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi," 2015.