

ANALISIS PENGELOMPOKAN TITIK BENCANA DI INDONESIA PERIODE 2023 MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC) BERDASARKAN DATA BADAN NASIONAL PENANGGULANGAN BENCANA (BNPB)

Indah Nur Seha¹, Ria Ester², and Jupron³

¹Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417
e-mail: ¹indahnurseha772@gmail.com

²Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417
e-mail: ²dosen02665@unpam.ac.id

³Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia, 15417
e-mail: ³dosen02664@unpam.ac.id

Abstract

Disasters pose significant challenges for communities worldwide, including Indonesia, which is vulnerable to both natural and human-made disasters. This study analyzes the clustering of disaster points in Indonesia for the year 2023 using the Agglomerative Hierarchical Clustering (AHC) method with data from the National Disaster Management Agency (BNPB). The data encompasses various types of disasters, such as earthquakes, hydrometeorological events, and volcanic eruptions, as well as variables related to casualties and infrastructure damage. The AHC method involves data preprocessing, calculating distances using Euclidean distance, and applying average linkage within AHC. The results identify four clusters: C1 (2497 members), C2 (1 members), C3 (1 member), and C4 (1 member), with a Silhouette Coefficient value of 0.983. This research is expected to provide valuable information for designing more efficient disaster management strategies.

Keywords: Disasters, Disaster Data, Clustering, Agglomerative Hierarchical Clustering (AHC), Indonesia.

Abstrak

Bencana merupakan tantangan signifikan bagi masyarakat di seluruh dunia, termasuk Indonesia yang rentan terhadap bencana alam dan buatan manusia. Penelitian ini menganalisis pengelompokan titik bencana di Indonesia periode 2023 menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) dengan data dari Badan Nasional Penanggulangan Bencana (BNPB). Data mencakup berbagai jenis bencana, seperti gempa bumi, hidrometeorologi, dan erupsi gunung berapi, serta variabel terkait korban dan kerusakan infrastruktur. Metode AHC dilakukan melalui *preprocessing data*, perhitungan jarak dengan *Euclidean distance*, dan penerapan *average linkage*. Hasil penelitian mengidentifikasi 4 *cluster*, yaitu C1 (2497 anggota), C2 (1 anggota), C3 (1 anggota), dan C4 (1 anggota) dengan hasil nilai *Sillhouette Coefficient* 0,983. Penelitian ini diharapkan dapat memberikan informasi untuk merancang strategi penanggulangan bencana yang lebih efisien.

Kata Kunci: Bencana, Data Bencana, Pengelompokan, *Agglomerative Hierarchical Clustering* (AHC), Indonesia.

1. PENDAHULUAN

Indonesia terletak di area pertemuan lempeng tektonik aktif, jalur gunung berapi, dan berada dalam iklim tropis. Indonesia memiliki banyak daerah yang rentan terhadap bencana alam. Indonesia memiliki dua musim utama, yaitu kemarau dan hujan, berbagai bencana pun terjadi sesuai dengan perubahan musim. Saat musim kemarau, potensi kekeringan, kebakaran hutan, dan kebakaran lahan meningkat. Sementara itu, musim hujan membawa risiko banjir, longsor, dan puting beliung.

Masyarakat Indonesia umumnya kurang memerhatikan ancaman bencana sebelum terjadi. Akibatnya, bencana alam dapat membawa dampak berupa kerugian moral, material, gangguan psikis, dan jatuhnya korban jiwa, yang berujung pada melemahnya mental masyarakat. Keterlibatan masyarakat dalam mencegah bencana alam merupakan bagian dari hak dan kewajibannya. Hal ini bertujuan untuk membantu meningkatkan efektivitas kegiatan penanggulangan bencana.

Untuk menangani berbagai masalah yang ada, pemerintah memerlukan data yang akurat dan sesuai dalam menangani bencana alam di seluruh provinsi di Indonesia. Namun, hingga kini, data bencana masih belum terintegrasi dengan baik, dan standar yang berbeda antar Kementerian atau Lembaga menyulitkan pembagian serta pelaporan data.

Berdasarkan uraian yang telah dijelaskan, penelitian ini menerapkan metode *Agglomerative Hierarchical Clustering* (AHC). Metode ini merupakan teknik *clustering* yang mengelompokkan data ke dalam hierarki kelompok dengan pendekatan *bottom-up*, yang digunakan untuk mengevaluasi efisiensi AHC. Kelebihan dari metode *Agglomerative Hierarchical Clustering* (AHC) yaitu tidak memerlukan jumlah *cluster* awal yang ditentukan sebelumnya, kemudian kelemahan pada metode AHC yaitu dapat menjadi *sensitive* terhadap *outlier* dalam data, yang dapat mempengaruhi pembentukan *cluster* yang optimal. Metode ini dipilih untuk mengidentifikasi pola kelompok yang memiliki kesamaan atribut dalam mengelompokkan titik bencana berdasarkan karakteristik tertentu.

2. PENELITIAN YANG TERKAIT

Berikut adalah jurnal yang terkait dengan Analisis Pengelompokan Titik Bencana di Indonesia Periode 2023 Menggunakan Metode *Agglomerative Hierarchical Clustering* (AHC) Berdasarkan Data Badan Nasional Penanggulangan Bencana (BNPB):

a. Pada penelitiannya dengan judul “Analisis Clustering Donor Darah dengan Metode *Agglomerative Hierarchical Clustering*”. Penelitian

ini bertujuan untuk melakukan sebuah pengelompokan yang diperoleh dari Unit Donor Darah PMI Purwakarta. Dengan permasalahan yang cukup tinggi dan jumlah pendonor darah yang minim, maka penelitian ini untuk menghasilkan sebuah informasi tentang jumlah pendonor, golongan serta ketersediaan darah pada tiap kecamatan yang berada di kabupaten Purwakarta dengan menggunakan *Agglomerative Hierarchical Clustering*. [1]

b. Pada penelitiannya dengan judul “Pengelompokan Kabupaten atau Kota Di Sulawesi Berdasarkan Indikator Pendidikan Menggunakan Analisis Klaster *Average Linkage* dan *Median Linkage*”. Penelitian ini bertujuan untuk mengelompokkan kabupaten atau kota di Sulawesi berdasarkan indikator pendidikan mengetahui pemerataan pendidikan atau karakteristik tingkat pendidikan maka dilakukan pengelompokan yang menggunakan analisis klaster. Hasil dari penelitian ini adalah menunjukkan bahwa *median linkage* memperoleh nilai rasio simpangan baku sebesar 0,061 lebih kecil dibandingkan dengan nilai rasio simpangan baku *average linkage* yaitu 0,078. Metode yang mempunyai rasio terkecil merupakan metode dengan kinerja terbaik. [2]

c. Pada penelitiannya dengan judul “Perbandingan Analisis *Agglomerative Hierarchical Clustering* Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat”. Penelitian ini bertujuan untuk mengetahui gambaran pada sektor pendidikan di provinsi Jawa Barat 2021 yang masih terdapat indikasi belum tercapainya pemerataan pendidikan khususnya pada jenjang sekolah menengah atas hingga perguruan tinggi serta terpusatnya jumlah sekolah dan guru di klaster tertentu. Hasil perbandingan nilai koefisien korelasi *cophenetic* dari keempat metode pada analisis AHC, metode *Ward* adalah metode terbaik dengan koefisien korelasi *cophenetic* terbesar (0,952). [3]

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode kualitatif. Metode kualitatif merupakan metode penelitian yang menghasilkan data deskriptif, baik dalam bentuk kata-kata tertulis, lisan, maupun pengamatan perilaku. Ciri khas pendekatan ini adalah pengumpulan data yang bersumber langsung dari kondisi alami, bersifat deskriptif, dan lebih memprioritaskan proses penelitian dari pada hasil akhirnya.

Metode Pengumpulan data dalam penelitian ini adalah dengan Studi Lapangan (*Field Research*), peneliti mengunjungi kantor BNPB untuk memperoleh

data yang diperlukan yang berhubungan dengan judul yang diangkat, dengan melakukan teknik-teknik pengumpulan data sebagai berikut:

a. Studi Pustaka

Studi Pustaka ini dilakukan untuk mengumpulkan informasi yang berbentuk *literature* tertulis atau buku sebagai landasan teori dalam penyusunan penulisan ini sehingga tidak terlibat langsung dengan subjek atau objek penelitian.

b. Observasi

Dilakukan di daerah Jl. Pramuka No.38 11, Kec Matraman, Kota Jakarta Timur dengan cara pengamatan untuk mengetahui informasi mengenai Data Bencana di Indonesia Periode 2023.

c. Wawancara

Dalam hal ini penulis melakukan wawancara untuk melengkapi data yang diperoleh selama observasi. Kemudian melakukan wawancara dengan pemangku kepentingan agar dapat memberikan informasi tambahan yang diperlukan agar data lebih lengkap dan jelas.

d. Sampling

Teknik sampling adalah teknik yang digunakan dalam pengambilan data-data yang berguna untuk mempermudah penelitian yang dilakukan, Adapun data yang digunakan adalah data Bencana di Indonesia Periode 2023 pada bulan Januari sampai Desember 2023.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data yang telah tercatat oleh Badan Nasional Penanggulangan Bencana (BNPB) pada periode 2023 bencana di Indonesia dengan total data yang diperoleh 5000 data dengan menggunakan atribut Meninggal dan Hilang, Luka-Luka, Mengungsi dan Terdampak, Rumah Rusak Berat, Rumah Rusak Sedang, Rumah Rusak Ringan, Rumah Terendam, Satuan Pendidikan, Rumah Ibadat, Fasilitas Layanan Kesehatan, Kantor, Jembatan, Lahan Terdampak. Pada penelitian ini akan menggunakan 2500 data untuk mengimplementasikan metode penelitian ini. Proses analisis *clustering* ini akan dijelaskan dari tahap *preprocessing*, perhitungan jarak, pembentukan *dendrogram*, dan evaluasi *clustering*.

a. Preprocessing Data

Tahap awal yang dilakukan yaitu melakukan pembersihan data dimana data yang telah didapatkan digabung menjadi satu *dataset*. Pada pembersihan data akan dilakukan yang dimana langkah pertama yaitu menghitung nilai *mean*, varian, dan standar deviasi. Kemudian data di transformasi menggunakan *Z-Score* agar setiap nilai atribut menjadi seimbang.

MEAN	0.0012	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MEAN	0.0012	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
VARIAN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
STANDAR DEVIASI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Gambar 4. 1 Mean, Varian, dan Standar Deviasi Variabel Data

NORMALISASI NILAI	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.1881	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Gambar 4. 2 Normalisasi Data Z-Score

b. Perhitungan Jarak (*Euclidean Distance*)

Menghitung jarak antar objek datanya. Perhitungan jarak ini menggunakan *Euclidean Distance*. Perhitungan jarak dilakukan untuk mengetahui kedekatan antar objek datanya. Data hasil normalisasi *Z-Score* selanjutnya akan dilakukan perhitungan jarak dengan *Euclidean Distance*.

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
0.00000000	1.0781179	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
1.0781179	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

Gambar 4. 3 Perhitungan Jarak Euclidean Distance

Dapat dilihat pada gambar diatas bahwa menghitung jarak ke dirinya sendiri akan selalu bernilai 0 sesuai dengan kondisi (iv). Sedangkan jika menghitung jarak ke terdekatnya itu akan menghasilkan nilai seperti pada gambar diatas bahwa 1 bertemu dengan V2 dia akan menghasilkan nilai 1,873. Nilainya sama dengan V2 bertemu 1, hal ini sesuai dengan kondisi (i) sehingga jarak antar 2 kota akan selalu bernilai sama. Pada gambar hasil perhitungan jarak menggunakan *Euclidean Distance* diatas hanya beberapa data yang dimasukkan sebagai acuan bagaimana *Euclidean Distance* diterapkan pada perhitungan jarak.

c. Pembentukan *Dendrogram*

Pembentukan *dendrogram* dari 2500 Kabupaten atau Kota yang diambil dalam penelitian ini dari *Agglomerative Hierarchical Clustering* (AHC) dengan menggunakan metode *Average Linkage*.



Gambar 4. 4 Hasil dendrogram dengan Average Linkage

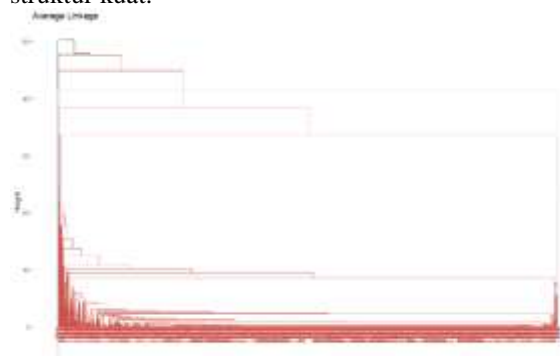
d. Evaluasi *Clustering*

Untuk menganalisis hasil dari proses *Agglomerative Hierarchical Clustering* (AHC) dengan *Average Linkage* maka diperlukan evaluasi *clustering* untuk mengetahui metode dan jumlah *cluster* dengan struktur terbaik berdasarkan *silhouette coefficient* nya. Pada evaluasi *clustering* ini jumlah *cluster* paling sedikit adalah 4 untuk mengetahui karakteristik *cluster* nya berdasarkan dari nilai *Z-Score*. Pada perhitungan *silhouette coefficient* ini adalah menghitung nilai *silhouette coefficient* dari metode *average linkage* dengan jumlah *cluster* sebanyak 4 *cluster*.

Tabel 4. 1 Nilai Silhouette Coefficient AHC

Jumlah Cluster	Average Linkage
4	0,983

Agglomerative Hierarchical Clustering (AHC) dengan *Average Linkage* hasil *cluster* terbaiknya adalah 4 *cluster* dengan nilai *silhouette coefficient* nya adalah 0,983 sehingga termasuk dalam *cluster* yang memiliki struktur kuat.



Gambar 4. 5 Hasil Visualisasi Dendrogram Berdasarkan Cluster nya

Pada *cluster* 1 menunjukkan karakteristik pada variabel “Mengungsi dan Terdampak” dan “Rumah Terendam” yang memiliki nilai tertinggi dibanding dengan *cluster* lainnya, hal ini menunjukkan bahwa *cluster* 1 mengalami dampak besar dari bencana di Indonesia periode 2023 yang mengakibatkan banyak orang untuk mengungsi dan pada *cluster* 1 ini terdapat Lahan Terdampak dengan nilai yang diperoleh 20,635. Jadi tidak hanya kerusakan infrastruktur yang signifikan namun lahan pun ikut terdampak dengan nilai yang cukup tinggi dari *cluster* lainnya.

Pada *cluster* 2 menunjukkan karakteristik pada variabel “Luka-Luka” yang memiliki nilai tertinggi pada *cluster* 2. Hal ini menunjukkan bahwa *cluster* 2 adanya korban luka-luka yang diakibatkan oleh bencana yang sangat signifikan.

Pada *cluster* 3 menunjukkan karakteristik pada variabel “Mengungsi dan Terdampak” dan “Rumah

Rusak Berat” yang memiliki nilai tertinggi pada *cluster* 3. Hal ini menunjukkan bahwa *cluster* 3 mengalami kerusakan infrastruktur yang signifikan akibat bencana, namun pada *cluster* 3 variabel “Meninggal dan Hilang” yang menunjukkan tidak hanya terjadi pada kerusakan infrastruktur namun adanya korban jiwa yang diakibatkan oleh bencana.

Pada *cluster* 4 menunjukkan karakteristik pada variabel “Mengungsi dan Terdampak” dan “Rumah Rusak Ringan” yang memiliki nilai tertinggi pada *cluster* 4. Pada *cluster* 4 menunjukkan mengalami kerusakan terhadap infrastruktur yang signifikan akibat bencana, namun pada *cluster* 4 ini variabel “Luka-Luka” nilai yang diperoleh juga cukup tinggi. Dalam hal ini *cluster* 4 mengalami kerusakan pada rumah dan dampak yang cukup besar dengan banyak korban luka-luka yang diakibatkan oleh bencana.

5. KESIMPULAN

Berdasarkan uraian dari bab-bab yang sudah dibahas sebelumnya dan pengujian yang telah dilakukan terhadap data Bencana di Indonesia Periode 2023 menggunakan *Agglomerative Hierarchical Clustering* (AHC) dengan metode *Average Linkage* maka dapat diambil kesimpulan sebagai berikut:

- Didapatkan bahwa *Average Linkage* hasil terbaiknya adalah 4 *cluster* dengan nilai *silhouette Coefficient* nya adalah 0,983 dengan struktur yang kuat.
- Metode *Average Linkage* adalah metode terbaik yang digunakan dalam penelitian ini karena memiliki hasil nilai *Silhouette Coefficient* dengan struktur yang kuat. Dengan karakteristik pada Kabupaten atau Kota *cluster* 1 yang perlu diperhatikan karena jumlah mengungsi dan terdampak paling tinggi serta jumlah yang tinggi terhadap lahan yang terdampak yang artinya pada *cluster* 1 terdampak kerusakan infrastruktur yang sangat signifikan dan perlu diperhatikan akibat bencana.

DAFTAR PUSTAKA

- Iyan Yulianti, D., Iman Hermanto, T., & Defriani, M., “Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering,” 2023.
- Oktavianty, E., Lilies Handayani, dan, Studi Statistika, P., Matematika, J., Mipa, F., Tadulako, U., Soekarno Hatta Km, J., & Palu Sulawesi Tengah, T., “Pengelompokan Kabupaten/Kota Di Sulawesi Berdasarkan Indikator Pendidikan Menggunakan Analisis Klaster Average Linkage Dan Median Linkage,” *Natural Science: Journal of Science and Technology*, 8(3), 197-191, 2019, <https://doi.org/10.22487/25411969.2019.V8.I3.14960>

- [3] Khoirunnisa', A., Wibowo, F. A. S., & Kismiantini, K, "Perbandingan Analisis Agglomerative Hierarchical Clustering Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat," 2023.
- [4] Anggraini, T., & Agustian, D, "Peran Badan Penanggulangan Bencana Daerah Dalam Upaya Pencegahan Bencana Kebakaran Hutan Dan Lahan (Karhutla) Di Kabupaten Musi Banyuasin," 2021.
- [5] Arya, E., Safitri, Y., & Riyadhno, F. A, "Pemantauan Dan Mitigasi Tingkat Potensi Bencana Kekeringan Di Kota Dumai," 2021.
- [6] Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," 2019.
- [7] Fatiatun, F., Firdaus, F., Jumini, S., & Adi, N. P, "Analisis Bencana Tanah Longsor Serta Mitigasinya," 2019.
- [8] Imasdiani, I., Purnamasari, I., & Amijaya, F. D. T, "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward," 2022.
- [9] Mujibulloh, M., Martanto, M., & Hayati, U, "Clustering Produk Ekspor Indonesia Berdasarkan Tingkat Permintaan Menggunakan Metode K-Means Tahun 2020-2022," 2023.