

Hybrid Hierarchical Clustering dalam Pengelompokan Daerah Rawan Bencana Tanah Longsor di Sulawesi Selatan

Fithriyah Azzahrah¹, Suwardi Annas², Zulkifli Rais³

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: Hybrid Hierarchical Clustering, Landslide, Grouping.

Abstract:

This study aims to describe and classify areas prone to landslides in South Sulawesi. The method used is Hybrid Hierarchical Clustering. The data used is landslide disaster data sourced from the National Disaster Management Agency (BNPB) for 2018-2020 in South Sulawesi. The variables used are the number of landslides, deaths, damaged houses, injured victims, and damaged public facilities. Grouping using the Hybrid Hierarchical Clustering method with mutual clusters using bottom-up and top-down methods. Grouping with bottom-up method produces 2 groups, top-down method produces 2 groups and 1 best mutual cluster. The ratio results in the bottom-up method is 0.84, the top-down method is 1.07 and the mutual cluster is 0.84. The grouping results obtained were 2 groups.

1. Pendahuluan

Analisis *cluster* atau kelompok merupakan salah satu metode dalam analisis statistik multivariat yang digunakan untuk mengelompokkan objek-objek ke dalam suatu kelompok berdasarkan karakteristik yang dimiliki, sehingga objek-objek dalam suatu kelompok memiliki ciri-ciri yang lebih homogen dibandingkan dengan objek dalam kelompok lain. (Azma, 2015). Ada dua jenis metode *clustering*, yaitu Hierarki (*Bottom-up Clustering*) dan Non Hierarki (*Top-down Clustering*). Metode Hierarki terdiri dari *Linkage Method* (*Single Linkage, Complete Linkage, Average Linkage*), *Centroid Linkage* dan *Ward Method*. Sedangkan metode *Non Hierarki* sendiri terdiri dari *K-Means* (Muhidin, 2017). untuk menggunakan kedua metode tersebut digunakan metode Hybrid *hierarchical clustering*.

Hybrid merupakan metode yang menggabungkan keunggulan Hierarki dan Non Hierarki yang akan memberikan hasil terbaik. Sebuah kelompok dipilih berdasarkan kriteria tertentu pada setiap langkah untuk dibagi sehingga menghasilkan jumlah kelompok yang diinginkan (Murugesan & Zhang, 2011). metode *hybrid* pada penelitian ini menggunakan *mutual cluster*. *Mutual cluster* merupakan kelompok yang memakai jarak maksimum antar partner pada kelompok lebih kecil atas jarak minimum menuju titik di kelompok luar (Chipman & Tibshirani, 2006). Metode *hybrid hierarchical clustering* adalah salah satu metode yang cocok digunakan dalam melakukan pengelompokan.

Dengan ini metode *Hybrid hierarchical clustering* dapat diaplikasikan dalam melakukan pengelompokan daerah rawan bencana tanah longsor. 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 (BNPB, 2021). Tanah longsor merupakan salah satu jenis gerakan massa tanah atau batuan, ataupun percampuran keduanya, menuruni atau keluar lereng akibat terganggunya kestabilan tanah atau batuan penyusun lereng (BNPB, 2021)

Berdasarkan latar belakang diatas, maka penulis akan melakukan penelitian yang berjudul “*Hybrid Hierarchical Clustering* dalam Pengelompokan daerah rawan bencana tanah longsor di Sulawesi Selatan”

* Corresponding author.

E-mail address: fithriyah2608@gmail.com



2. Tinjauan Pustaka

2.1. Analisis Cluster

Analisis *cluster* atau kelompok yaitu suatu kelompok dalam teknik multivariat yang terdapat pada metode interdependensi yaitu tidak terdapat satu pun variabel yang diartikan sebagai suatu variabel bebas atau variabel terikat. Tujuan dari analisis *cluster* yaitu melakukan pengelompokan objek berdasarkan variabel yang memiliki karakteristik yang relatif sama (Djafar dkk., 2021).

Adapun langkah Pemrosesan data dari kumpulan data mentah yang dapat dikelompokkan menjadi satu maupun beberapa kelompok adalah sebagai berikut:

- 1) Mengatur ukuran jarak antar objek (similarity).
- 2) Standarisasi data

Untuk melakukan Standarisasi data dapat dilihat pada persamaan berikut:

$$Z_i = \frac{x_i - \bar{x}}{s}$$

Keterangan:

Z_i = standarisasi variabel

x_i = data ke i

\bar{x} = rata-rata keseluruhan data tiap variabel

s = standar deviasi

- 3) Melakukan proses pengelompokan.

2.2. Metode Silhouette Coefficient

Metode *Silhouette Coefficient* merupakan gabungan dari metode *cohesion* dan *separation*. Metode ini digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik atau buruknya suatu objek ditempatkan dalam suatu *cluster* (Dewi & Pramita, 2019). Tahapan perhitungan *Silhouette Coefficient* adalah sebagai berikut (Prima, 2021):

- 1) Menghitung rata-rata jarak dari suatu objek misalkan i dengan semua objek lain yang masih berada dalam satu *cluster*
- 2) Menghitung rata-rata jarak dari objek i tersebut dengan semua objek yang berada di *cluster* lain, dan diambil nilai paling minimumnya.
- 3) Nilai silhouette coefficient

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

Nilai yang didapat dari metode *silhouette coefficient* terletak pada kisaran nilai -1 hingga 1. Jika nilai *silhouette coefficient* mendekati nilai 1, maka semakin baik pengelompokan objeknya dalam satu *cluster*.

2.3. Pengelompokan Bottom-up Clustering

Pengelompokan *bottom-up* merupakan proses pengelompokan yang dilakukan dari pengamatan objek-objek yang memiliki kesamaan sehingga bergabung menjadi suatu kelompok kecil dan dari kelompok-kelompok kecil yang terbentuk akan digabungkan menjadi satu kelompok besar yang berisikan semua kelompok (Nugroho, 2008) Salah satu metode dalam pengelompokan ini adalah *Complete linkage*. *Complete linkage* disebut juga pendekatan tetangga terjauh. (Awaliah, 2018). Metode ini menggunakan jarak yang paling jauh antara dua kelompok yang berbeda B_r dan B_s yang didefinisikan pada Persamaan berikut:

$$h(B_r, B_s) = \max \{d(x_i, x_j) ; x_i \text{ anggota } B_r \text{ dan } x_j \text{ anggota } B_s\}$$

dimana kelompok B_r dan B_s akan digabung jika $h(B_r, B_s)$ adalah jarak yang terbesar.

2.4. Pengelompokan Top-down Clusterin

Metode Non Hierarki (*Top-down clustering*) sering disebut sebagai metode *k-means*. *K-Means* merupakan salah satu metode pengelompokan data kedalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok lain (Hidayat dkk, 2017). Langkah-langkah pengelompokan *top-down clustering* (Mariyani dkk, 2011):

- 1) Mempartisi objek kedalam k kelompok awal
- 2) Menghitung pusat kelompok dapat dilihat pada Persamaan berikut

$$C_{(A)i} = \frac{1}{n} \sum_{j=1}^n y_{ij}$$

dimana:

$C_{(A)i}$	= Pusat kelompok A pada variabel ke-i	A	= 1, 2, ..., k
n_i	= Jumlah objek pada variabel ke-i	y_{ij}	= Nilai dari objek ke-j pada variabel ke-i
I	= 1, 2, ..., p	P	= Banyaknya variabel

- 3) Menghitung jarak setiap objek terhadap pusat kelompok (jarak yang digunakan adalah jarak euclidean)
- 4) Menentukan objek yang lebih dekat dengan pusat kelompok

2.5. Metode Hybrid Hierarchical Clustering

Metode *Hybrid* merupakan metode yang terjadi dari penggabungan dua metode yakni *bottom-up clustering* yaitu metode *cluster* yang diawali dari kelompok kecil menjadi satu kelompok besar dan *top-down clustering* yaitu metode *cluster* yang diawali dari satu kelompok besar yang dipecah menjadi kelompok yang lebih kecil. Metode analisis *hierarchical clustering* terbagi jadi dua yaitu *agglomerative* (penggabungan) dan *divisive* (pemecahan) (Alfira dkk, 2016). Untuk metode *Hybrid hierarchical clustering* ini dilakukan dengan *Mutual cluster*. *Mutual cluster* adalah suatu pengelompokan dimana digunakan jarak paling besar antar bagian pada kelompok terkecil dari jarak yang terpendek pada tiap titik di luar kelompok (Chipman & Tibshirani, 2006). Langkah-langkah atau algoritma dari metode *hybrid hierarchical clustering* dengan *mutual cluster* adalah sebagai berikut:

- 1) Melakukan pengelompokan secara bottom-up.
- 2) Mengidentifikasi mutual cluster.
- 3) Melakukan pengelompokan metode top-down (k-means), dimana tiap mutual cluster tidak boleh terpisah. Hal ini untuk mengubah sementara mutual cluster dari objek dengan pusat mutual cluster tersebut.
- 4) Jika telah melakukan pengelompokan secara top-down, maka dilanjutkan dengan memisahkan tiap mutual cluster dan melakukan pengelompokan top-down tiap mutual cluster

2.6. Penentuan Metode Kelompok Terbaik

Menurut (Laraswati, 2013) Jika nilai s_w minimum dan s_b maksimum atau dalam hal ini metode terbaik menghasilkan nilai rasio simpangan baku minimum s_w terhadap s_b hal ini merupakan pengelompokan yang baik. Dengan rumus seperti pada Persamaan berikut

$$S = \frac{s_w}{s_b} \times 100\%$$

2.7. Bencana Alam Tanah Longsor

Bencana tanah longsor merupakan bencana alam geologi yang dapat menimbulkan korban jiwa dan kerugian material yang sangat besar, seperti terjadinya pendangkalan, terganggunya jalur lalu lintas, rusaknya lahan pertanian, permukiman, jembatan, saluran irigasi dan prasarana fisik lainnya (Faizana dkk, 2015).

Tanah longsor terjadi kerana ada gangguan kestabilan pada tanah atau batuan penyusun lereng. Gangguan kestabilan lereng tersebut dapat dikontrol oleh kondisi morfologi (terutama kemiringan lereng), kondisi batuan atau tanah penyusun lereng, dan kondisi hidrologi atau tata air pada lereng. (Faizana dkk, 2015)

3. Metodologi Penelitian

3.1. Jenis Penelitian

Jenis dari penelitian ini adalah penelitian deskriptif eksploratif yang bertujuan untuk menggambarkan suatu fenomena tanah longsor.

3.2. Jenis dan Sumber Data

Jenis data pada penelitian ini adalah data sekunder yaitu data yang diperoleh dan dikumpulkan dari sumber data informasi yang telah ada. Data yang digunakan pada penelitian ini diperoleh dari publikasi pada *website* Badan Nasional Penanggulangan Bencana 2021 (BNPB, 2021)

3.3. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah:

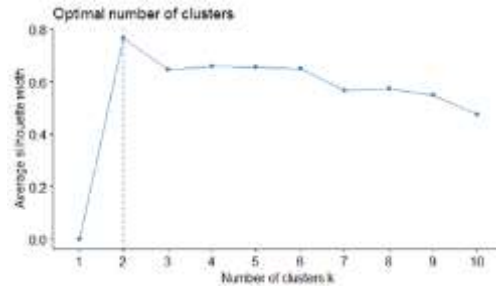
- 1) Melakukan analisis deskriptif untuk mengetahui gambaran suatu data yang akan digunakan.
- 2) Melakukan standarisasi data.
- 3) Pengelompokan tiap kabupaten di Sulawesi Selatan secara *bottom-up clustering* dengan metode *complete linkage*
- 4) Pengelompokan tiap kabupaten di Sulawesi Selatan dengan cara *top-down clustering* dengan *k-means*.
- 5) Pengelompokan kabupaten-kabupaten di Sulawesi Selatan secara *hybrid hierarchical clustering* dengan *mutual cluster*.
- 6) Menentukan metode kelompok terbaik untuk *top-down*, *bottom-up*, dan *mutual cluster* berdasarkan hasil nilai rasio untuk pengelompokan daerah rawan bencana tanah longsor di Sulawesi Selatan.

4. Hasil dan Pembahasan

4.1. Metode Bottom-up Clustering

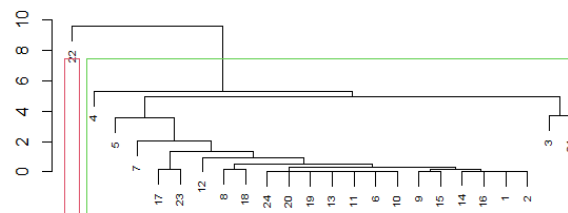
Metode *Bottom-up clustering* yang digunakan dalam melakukan pengelompokan daerah rawan bencana tanah longsor di Sulawesi Selatan, dimana pada metode ini digunakan aturan *complete linkage*.

Pada penelitian ini digunakan metode *Silhouette coefficient* dalam menentukan banyaknya jumlah kelompok yang terbentuk. plot *Silhouette coefficient* dapat dilihat pada Gambar 4.1



Gambar 4.1 Plot *silhouette coefficient complete linkage*

Berdasarkan Gambar 4.1 puncak tertinggi pada plot terdapat pada angka 2, oleh karena itu dapat kita ketahui bahwa jumlah kelompok yang terbentuk adalah sebanyak 2 kelompok.



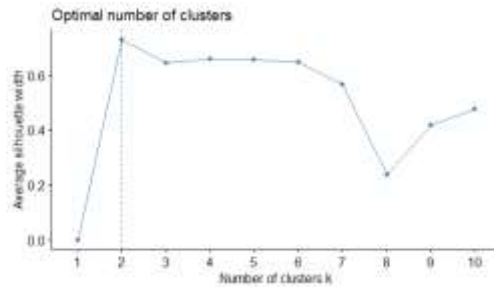
Gambar 4.2 Dendrogram *Bottom-up clustering*

Dendrogram pada Gambar 4.2 merupakan hasil pengelompokan *bottom up clustering* dengan menggunakan metode *complete linkage* yang ditandai dengan *mutual cluster*. Hasil dendrogram tersebut menerangkan bahwa terdapat 1 buah *mutual cluster* yang terbentuk.

4.2. Metode Top-down Clustering

4.2.1. Menentukan banyaknya jumlah kelompok yang akan dibentuk.

Dalam melakukan pengelompokan, terlebih dahulu harus menentukan banyaknya jumlah kelompok, hal ini sangat berpengaruh terhadap hasil pengelompokan yang akan diperoleh. Pada penelitian ini digunakan metode *silhouette coefficient* dalam menentukan banyaknya jumlah kelompok. Berikut Plot *silhouette coefficient*:

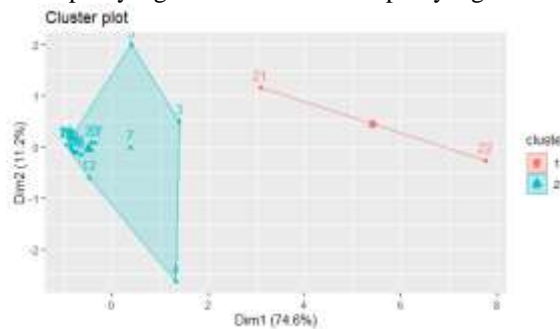


Gambar 4.3 Plot *silhouette coefficient k-means*

Berdasarkan Gambar 4.3 dapat kita ketahui bahwa puncak tertinggi terdapat pada angka 2, maka banyaknya kelompok yang akan dibentuk adalah sebanyak 2 kelompok atau *cluster*.

4.2.2. *Pengelompokan dengan menggunakan metode K-Means.*

K-Means merupakan salah satu metode yang dapat digunakan untuk mengelompokkan daerah rawan bencana tanah longsor di Sulawesi Selatan. Dengan menggunakan metode *silhouette coefficient* untuk menentukan jumlah kelompok yang dibentuk dan kelompok yang terbentuk sebanyak 2 kelompok.

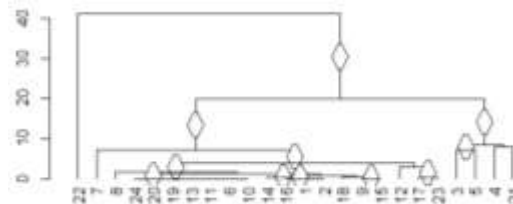


Gambar 4.4 Plot *Top-down clustering*

Pada Gambar 4.4 menunjukkan daerah dengan kode 21 dan 22 atau daerah Tana Toraja dan Toraja Utara merupakan anggota kelompok 1 dan 22 lainnya merupakan anggota kelompok 2.

4.3. *Hybrid Hierarchical Clustering Dengan Metode Mutual Cluster*

Setelah melakukan pengelompokan secara *bottom-up* dan *top-down*, didapatkan hasil yang dapat dilihat pada Gambar 4.5 dimana pada sumbu horizontal merupakan kode pada setiap kabupaten dan sumbu vertikal adalah nilai besaran jarak yang terbentuk pada Gambar 4.5 menunjukkan bahwa terdapat 11 *mutual cluster* yang terbentuk yang ditandai dengan adanya belah ketupat, namun karena masing-masing *cluster* saling berhubungan maka hanya 1 *mutual cluster* yang terbaik



Gambar 4.5 Dendrogram *mutual cluster*

4.4. Menentukan Metode Kelompok Terbaik

Metode terbaik dapat diketahui dengan melihat hasil nilai rasio pada masing-masing metode.

Tabel 4.1 Hasil Nilai Rasio

Metode	Rasio
<i>Bottom-up clustering</i>	0,84
<i>Top-down clustering</i>	1,07
<i>Mutual cluster</i>	0.84

Berdasarkan Tabel 4.1, dapat kita ketahui bahwa kelompok metode terbaik adalah metode *bottom-up* dengan menggunakan aturan *complete linkage* dan *mutual cluster*, karena pada metode tersebut memiliki hasil rasio yang lebih rendah dari metode *top-down clustering*.

Untuk melihat keadaan daerah berdasarkan hasil pengelompokan yang terbentuk, maka digunakan metode terbaik yaitu *mutual cluster*. Dari seluruh objek penelitian sebanyak 24 daerah di Sulawesi Selatan diambil rata-rata dari masing-masing variabel bencana tanah longsor (\bar{X}). Selanjutnya rata-rata masing-masing kelompok dari setiap variabel (\bar{Xc}) sehingga diperoleh deskriptif kelompok setiap variabel yang diberikan pada tabel 4.2.

Tabel 4.2 Deskriptif kelompok *mutual cluster*

Variabel	\bar{X}	\bar{Xc}	
		C_1	C_2
X_1	2,88	-0,19	4,51
X_2	0,59	-0,16	3,61
X_3	0,75	-0,12	2,74
X_4	8,83	-0,13	3,05
X_5	0,5	-0,14	3,29

Setiap variabel pada kelompok diberi tanda jika ($\bar{X} \leq \bar{Xc}$) maka diberi tanda positif (+) sedangkan jika ($\bar{X} \geq \bar{Xc}$) maka diberi tanda (-). uraian karakteristik tiap kelompok dapat dilihat pada Tabel 4.3.

Tabel 4.3 Karakteristik kelompok *mutual cluster*

Cluster	X_1	X_2	X_3	X_4	X_5
1	-	-	-	-	-
2	+	+	+	-	+

Berdasarkan Tabel 4.3 terlihat bahwa daerah yang berada pada kelompok 1 dikelompokkan berdasarkan nilai rata-rata variabel yang rendah. Sehingga daerah yang berada pada kelompok ini merupakan daerah yang tidak rawan akan bencana tanah longsor.

Daerah yang berada pada kelompok 2 merupakan daerah yang memiliki rata-rata yang rendah pada variabel jumlah rumah rusak dan variabel lainnya seperti jumlah kejadian tanah longsor, korban meninggal, korban terluka, dan fasilitas rusak memiliki rata-rata yang tinggi. Oleh karena itu, daerah yang berada pada kelompok 2 dimana anggota kelompoknya adalah Tana Toraja yang merupakan daerah yang rawan terdampak bencana tanah longsor.

5. Kesimpulan

Berdasarkan hasil rasio dari masing-masing metode, diperoleh hasil untuk metode *bottom-up clustering* sebesar 0,84, *top-down clustering* sebesar 1,07, dan *mutual cluster* sebesar 0,84. Dari hasil rasio tersebut dapat diketahui bahwa metode terbaik pada penelitian ini adalah metode *bottom-up clustering* dan *mutual cluster*. Adapun hasil pengelompokan dengan metode *Mutual cluster* adalah 2 kelompok terbentuk dimana pada kelompok 1 terdiri dari 23 anggota kelompok dimana anggotanya adalah Bantaeng, Barru, Bone, Bulukumba, Enrekang, Gowa, Jeneponto, Kepulauan Selayar, Luwu, Luwu Timur, Luwu Utara, Maros, Pangkep, Pinrang, Sidrap, Sinjai, Soppeng, Takalar, Toraja Utara, Wajo, Makassar, Palopo dan Parepare dan kelompok 2 terdiri 1 anggota kelompok yaitu Tana Toraja. Dimana hasil dari daerah yang berada pada kelompok 2 merupakan daerah yang memiliki rata-rata yang rendah pada variabel jumlah rumah rusak dan variabel lainnya seperti jumlah kejadian tanah longsor, korban meninggal, korban terluka, dan fasilitas rusak memiliki rata-rata yang tinggi. maka dari itu, daerah yang berada pada kelompok 2 merupakan daerah yang rawan terdampak bencana tanah longsor, daerah tersebut adalah Tana Toraja.

References

- Alfira, A., Hermin, F., & Wiraningsih, E. D. (2016). *Analisis Hybrid Mutual Clustering menggunakan Jarak Square Euclidean*. 2–3.
- Awaliah, R. (2018). *Analisis Clustering untuk mengelompokkan tingkat kesejahteraan kabupaten/kota berdasarkan sosial ekonomi rumah tangga di wilayah Provinsi Sulawesi Selatan*. 75383.
- Azma, S. (2015). *Perbandingan Analisis Klaster Menggunakan Metode Single Linkage, Complete Linkage, Average Linkage Dan K-Means Untuk Pengelompokan*.
- BNPB. (2021). *Definisi Bencana*. Badan Nasional Penanggulangan Bencana.
- Chipman, H., & Tibshirani, R. (2006). Hybrid hierarchical clustering with applications to microarray data. *Biostatistics*, 7(2), 286–301. <https://doi.org/10.1093/biostatistics/kxj007>
- Dewi, D. A. I. C., & 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), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Djafar, N. M., Wijayanti, L. N., Elprilita, A. R., & Widodo, E. (2021). Pengelompokan Produksi Perkebunan Menurut Kabupaten/Kota Jawa Tengah Tahun 2020 Menggunakan Hierarchical Clustering. *Journal of Mathematics Education and Science*, 4(2), 59–66. <https://doi.org/10.32665/james.v4i2.230>
- Faizana, F., Nugraha, A. L., & Yuwono, B. D. (2015). Pemetan Risiko Bencana Tanah Longsor Kota Semarang. *Jurnal Geodesi Undip*, 4(1), 42.
- Hidayat, R., Wasono, R., & Darsyah, M. Y. (2017). Pengelompokan Kabupaten/Kota Di Jawa Tengah Menggunakan Metode K-Means Dan Fuzzy C-Means. *Prosiding Seminar Nasional & Internasional*, 240–250. <https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/3017/2932>
- Laraswati, T. F. (2013). *Perbandingan kinerja Metode Complete Linkage, Metode Average Linkage, dan Metode K-Means dalam menentukan hasil Analisis Cluster*. 9(2), 62–68.
- Mariyani, D., Purnami, S. W., & Winahju, W. S. (2011). *Penerapan Hybrid Hierarchical Clustering melalui Mutual Cluster dalam Pengelompokan Kabupaten di Jawa Timur Berdasarkan Variabel Sektor Pertanian*. 1–10.
- Muhidin, A. (2017). Analisa Metode Hierarchical Clustering dan K-Mean dengan Model LRFMP pada Segmentasi Pelanggan. *SIGMA, Jurnal Teknologi Pelita Bangsa*, 7(1), 82–83.
- Murugesan, K., & Zhang, J. (2011). *Hybrid Hierarchical Clustering: an Experimental Analysis*. 5. <http://www.cs.uky.edu/~jzhang/pub/MINING/keer1.html>
- Nugroho, S. (2008). *Statistika Multivariat Terapan* (J. Rizal (ed.); Edisi Pert). Unib Press.

Prima, G. R. (2021). *Analisa Perbandingan Nilai K Terbaik Untuk Clustering K-Means Menggunakan Pendekatan Elbow Dan Silhouette Pada Citra Aksara Jawa Oleh : Comparative Analysis Of The Best K-Value For Clustering K-Means Using Elbow And Silhouette Approach On Javanese Script .*