

## PENGKLASTERAN DAERAH DI JAWA TIMUR BERDASARKAN CURAH HUJAN

Elok Rizqi Auliya

Jurusan Matematika, FMIPA, Universitas Negeri Surabaya

e-mail : elok.17030214010@mhs.unesa.ac.id

## Abstrak

Curah hujan di beberapa provinsi di Indonesia memiliki intensitas yang cukup tinggi, termasuk di Provinsi Jawa Timur. Intensitas curah hujan yang tinggi dan terjadi secara terus menerus akan menimbulkan berbagai bencana hidrometeorologi. Sehingga diperlukan mitigasi bencana untuk mengurangi dampak yang diakibatkan. Salah satu upaya yang dapat dilakukan yaitu dengan mengklaster daerah yang memiliki karakteristik curah hujan yang hampir sama. Oleh karena itu, dalam artikel ini akan dilakukan pengklasteran daerah berdasarkan curah hujannya di Jawa Timur menggunakan *single linkage*, *complete linkage*, dan *average linkage*. Dari pengklasteran dengan menggunakan tiga metode tersebut, diperoleh dendrogram untuk masing-masing metode. Kemudian akan dibentuk pengklasteran sebanyak 2 klaster, 3 klaster, 4 klaster, dan 5 klaster. Dari masing-masing klaster yang dibentuk dan metode *clustering* yang digunakan, dilakukan *cluster validation* untuk menentukan metode terbaik dengan menggunakan nilai *Dunn Index* yang terbesar. Nilai *Dunn Index* yang terbesar yaitu sebesar 1.120301 dengan 3 klaster baik dengan metode *single linkage*, *complete linkage*, maupun *average linkage*.

**Kata kunci:** Curah Hujan, Pengklasteran, *Cluster Validation*

## Abstract

Rainfall in several provinces in Indonesia has a high enough intensity, including in East Java Province. The high intensity of rainfall that occurs continuously will cause various hydrometeorological disasters. So that disaster mitigation is needed to reduce the resulting impacts. One effort that can be done is to cluster areas that have almost the same rainfall characteristics. Therefore, in this article, we will cluster the areas based on rainfall in East Java using single linkage, complete linkage, and average linkage. From the clustering using these three methods, a dendrogram is obtained for each method. Then 2 clusters will be formed, 3 clusters, 4 clusters, and 5 clusters. From each cluster formed and the clustering method used, cluster validation is carried out to determine the best method by using the largest Dunn Index value. The largest Dunn Index value is 1.120301 with 3 clusters using single linkage, complete linkage, and average linkage methods.

**Keywords :** Rainfall, Clustering, Cluster Validation

## 1. PENDAHULUAN

Curah hujan saat ini tidak menentu, bahkan intensitasnya semakin tinggi. Banyak faktor yang mempengaruhi intensitas curah hujan, beberapa diantaranya yaitu ketinggian dari suatu daerah, arah angin, dan suhu.

Hujan dengan intensitas tinggi yang terjadi secara terus menerus dapat menyebabkan berbagai bencana hidrometeorologi seperti banjir, banjir bandang, dan tanah longsor. Menurut Badan Meteorologi Klimatologi dan Geofisika (BMKG), ada 15 provinsi yang berstatus waspada terhadap bencana salah satunya yaitu Provinsi Jawa Timur.

Menurut Gubernur Jawa Timur, secara geografis dan geologis Provinsi Jawa Timur sangat rentan terhadap bencana alam. Ada 22 daerah yang rawan terjadinya bencana hidrometeorologi. Potensi terjadinya bencana banjir disebabkan karena luapan sungai brantas, sungai

welang, dan sungai kemuning. Sedangkan untuk daerah yang rawan terjadinya longsor salah satunya yaitu Jombang. Oleh karena itu, diperlukan adanya mitigasi bencana sebagai salah satu upaya yang dapat dilakukan untuk meminimalisir dampak yang ditimbulkan. Salah satu upaya yang dapat dilakukan yaitu dengan melakukan pengklasteran daerah berdasarkan curah hujannya. Hal ini dikarenakan setiap daerah memiliki curah hujan yang berbeda-beda, sehingga perlu dilakukan pengklasteran daerah yang karakteristik curah hujannya hampir sama. Diharapkan dengan adanya pengklasteran ini, daerah yang berpotensi dilanda bencana dapat meningkatkan antisipasi dan mitigasi bencananya sehingga dampaknya akan berkurang. Berdasarkan hal tersebut, peneliti tertarik untuk mengklaster daerah berdasarkan curah hujannya di Jawa Timur.

Pada artikel ini, pengklasteran daerah berdasarkan curah hujannya di Jawa Timur menggunakan metode *clustering hierarki* yaitu *single linkage*, *complete linkage*,

dan *average linkage*. *Cluster validation* yang digunakan untuk mengetahui metode *clustering* terbaik yang cocok dalam mengklaster daerah berdasarkan curah hujan dapat menggunakan *dunn index*. Nilai *dunn index* yang terbesar yang menunjukkan metode terbaik.

Penelitian sebelumnya yang terkait dengan pengklasteran curah hujan telah dilakukan oleh (Machiwal, et al., 2017) menggunakan *Principal Component Analysis* (PCA) dan *Hierarchical Cluster Analysis* (HCA) di wilayah India, (H & Thomas, 2016) menggunakan *Ward's Method* yang mengklaster negara bagian Kerala berdasarkan zona homogen curah hujan bulanan.

Sedangkan penelitian sebelumnya yang terkait dengan pengklasteran dengan *clustering hierarki* yaitu penelitian yang telah dilakukan oleh (Onwukwe & Ezeorah, 2009) menggunakan *single linkage clustering* dalam menganalisis produk domestik bruto, (Astuti & Untari, 2017) menggunakan *average linkage clustering* dalam mengklaster kabupaten sidoarjo berdasarkan geografi, pendidikan, pertanian, dan industry, (Nafisah & Chandra, 2017) menggunakan *average linkage* untuk mengklaster kabupaten/kota di Provinsi Jawa Timur berdasarkan faktor-faktor kemiskinan.

Berdasarkan latar belakang tersebut, pada penelitian ini akan dilakukan pengklasteran daerah berdasarkan curah hujannya di Jawa Timur menggunakan metode *single linkage*, *complete linkage*, dan *average linkage*. Diharapkan dengan adanya penelitian ini, dapat digunakan sebagai acuan pemerintah dalam meningkatkan mitigasi bencana terutama di daerah yang rawan terjadi bencana hidrometeorologi.

## 2. KAJIAN TEORI

### Clustering

Pengklasteran data atau *clustering* merupakan proses pengidentifikasian klaster berdasarkan beberapa ukuran kesamaan (Rokach & Maimon, 2005). Algoritma *clustering* pada umumnya dibagi menjadi 2 yaitu *clustering hierarki* dan *clustering non hierarki*. *Clustering hierarki* mengklaster data ke dalam klaster yang ditentukan di akhir. Metode *clustering hierarki* dapat dibagi menjadi dua yaitu *Divisive* dan *Agglomerative* (Rani & Rohil, 2013). *Divisive* menganggap semua data dalam satu klaster dan memisahkan data yang berbeda ke dalam klaster lain. Sedangkan *agglomerative* menganggap semua data satu klaster dan menggabungkan data yang memiliki kemiripan ke dalam satu klaster. Beberapa contoh metode *clustering hierarki* yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward's method*. *Clustering non hierarki* mengklaster data ke dalam  $k$  klaster yang telah ditentukan di awal. Contoh metode *clustering non hierarki* yaitu metode *K-Means*.

### Single Linkage

*Single Linkage* merupakan salah satu metode *agglomerative hierarchical clustering*. Pada metode *single linkage*, pengklasteran berdasarkan jarak yang paling dekat antar klaster. Jarak antar data akan dihitung dengan *euclidean distance* dengan rumus sebagai berikut.

$$d_{ik} = \sqrt{\sum_{i=1}^n (x_i - x_k)^2} \quad (1)$$

dengan  $x_i$  merupakan data ke- $i$  dan  $x_k$  merupakan data ke- $k$ . Jarak terdekat dari dua klaster didefinisikan sebagai jarak minimum antar dua titik di dua klaster (Mohbey & Thakur, 2013). Langkah pertama yang dilakukan yaitu menentukan nilai terkecil dalam matriks jarak. Kemudian menggabungkan menjadi satu klaster. Misalkan untuk mencari jarak antara klaster (UV) dengan klaster W dengan *single linkage*, dapat ditentukan dengan rumus sebagai berikut.

$$d_{(UV)W} = \min\{d_{UV}, d_{VW}\} \quad (2)$$

Hasil pengklasteran dapat dilihat pada dendrogram yang dihasilkan.

Algoritma *single linkage* yaitu sebagai berikut.

1. Menentukan jarak antar data menggunakan persamaan (1)
2. Menentukan jarak yang paling minimum pada matriks jarak
3. Menggabungkan 2 data yang jaraknya paling minimum menjadi satu klaster dan jaraknya diambil yang paling minimum diantara 2 data tersebut menggunakan persamaan (2).
4. Melakukan langkah kedua sampai menjadi 2 klaster
5. Menggambar dendrogram dengan kombinasi titik dan ketinggian yang telah ditentukan.

### Complete Linkage

*Complete Linkage* merupakan salah satu metode *agglomerative hierarchical clustering*. Pengklasteran dengan metode *complete linkage* berdasarkan jarak terjauh antar data, sehingga dapat dipastikan semua data dalam satu klaster memiliki jarak yang jauh satu sama lain. Jarak antar data menggunakan *Euclidean distance* sesuai persamaan (1) dan membentuk matriks jarak. Yang dilakukan pertama yaitu menentukan nilai terkecil dalam matriks jarak. Kemudian menggabungkan data menjadi satu klaster. Untuk mencari jarak antara klaster menggunakan *complete linkage* yaitu misalkan jarak antara klaster (UV) dengan klaster W, dapat ditentukan dengan rumus sebagai berikut.

$$d_{(UV)W} = \max\{d_{UV}, d_{VW}\} \quad (3)$$

Algoritma *complete linkage* yaitu sebagai berikut.

1. Menentukan jarak antar data dengan persamaan (1)

- Menentukan jarak yang paling minimum pada matriks jarak
- Menggabungkan 2 data yang jaraknya paling minimum menjadi satu kluster dan jaraknya dipilih yang paling maksimum diantara 2 data tersebut menggunakan persamaan (3)
- Melakukan langkah kedua sampai menjadi 2 kluster
- Menggambar dendrogram berdasarkan titik dan ketinggian yang telah diperoleh.

**Average Linkage**

*Average Linkage* termasuk ke dalam metode *agglomerative hierarchical clustering*. Pengklasteran dengan *average linkage* berdasarkan jarak rata-rata antar kluster. Jarak antar data dapat menggunakan *euclidean distance* pada persamaan (1). Pertama yang dilakukan yaitu menentukan nilai terkecil dalam matriks jarak. Kemudian menggabungkan data yang memiliki jarak terkecil menjadi satu kluster. Misalkan untuk mencari jarak antara kluster (UV) dengan kluster W menggunakan *average linkage*, dapat ditentukan dengan rumus sebagai berikut.

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)}N_W} \tag{4}$$

Algoritma *complete linkage* yaitu sebagai berikut.

- Menentukan jarak antar data dengan persamaan (1)
- Menentukan jarak yang paling minimum pada matriks jarak
- Menggabungkan 2 data yang jaraknya paling minimum menjadi satu kluster dan jaraknya merupakan jarak rata-rata antar 2 data tersebut menggunakan persamaan (4)
- Melakukan langkah kedua sampai menjadi 2 kluster
- Menggambar dendrogram berdasarkan titik dan ketinggian yang telah diperoleh.

**Dunn Index**

*Dunn Index* merupakan salah satu metode untuk *cluster validation* yang bertujuan untuk mengevaluasi hasil pengklasteran. Semakin tinggi nilai *Dunn Index*, maka semakin baik metode pengklasteran tersebut. Nilai *Dunn Index* dapat diperoleh dengan rumus sebagai berikut.

$$Dunn\ Index = \frac{Dist_{min}}{Dist_{max}} \tag{5}$$

dengan  $Dist_{min}$  merupakan jarak minimum antara dua data dalam kluster yang berbeda dan  $Dist_{max}$  merupakan jarak maksimum antara dua data pada kluster yang sama.

**3. METODE (untuk penelitian terapan)**

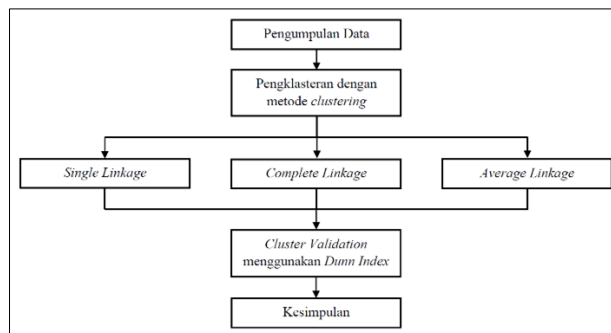
**Data Penelitian**

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diperoleh dari website resmi Pusat Database BMKG yang dapat diakses melalui [www.dataonline.bmkg.go.id](http://www.dataonline.bmkg.go.id). Data yang digunakan yaitu data curah hujan di Jawa Timur setiap harinya dari bulan

Januari 2016 sampai Desember 2020. Data diambil dari 11 stasiun pengamatan BMKG yang ada di Jawa Timur yaitu Stasiun Klimatologi Malang, Stasiun Geofisika Malang, Stasiun Meteorologi Banyuwangi, Stasiun Geofisika Pasuruan, Stasiun Meteorologi Juanda, Stasiun Geofisika Nganjuk, Stasiun Meteorologi Tuban, Stasiun Meteorologi Sangkapura, Stasiun Meteorologi Kalianget, Stasiun Meteorologi Perak 1, dan Stasiun Meteorologi Maritim Tanjung Perak.

**Diagram Alir**

Diagram alir yang digunakan dalam penelitian ini yaitu sebagai berikut.



Gambar 1. Diagram Alir Penelitian

Berdasarkan Gambar 1, langkah yang pertama dilakukan yaitu melakukan pengumpulan data dari 11 stasiun pengamatan BMKG di Jawa Timur. Kemudian melakukan pengklasteran dengan 3 metode *clustering*, yaitu *Single Linkage*, *Complete Linkage*, dan *Average Linkage*. Setelah didapatkan dendrogram dari masing-masing metode, kemudian melakukan pembentukan kluster sebanyak 2 kluster, 3 kluster, 4 kluster, dan 5 kluster. Hasil dari masing-masing kluster dan metode pengklasteran yang digunakan, kemudian dilakukan *cluster validation* menggunakan *Dunn Index*. Jika nilai *Dunn Index* semakin besar, maka semakin baik metode tersebut.

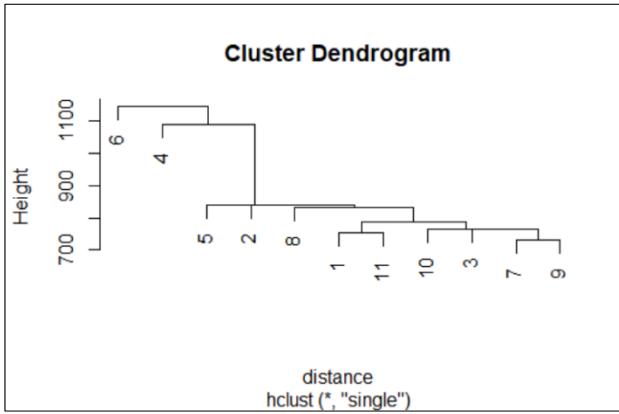
**4. PEMBAHASAN**

Dalam bagian ini akan dijelaskan cara pengklasteran daerah di Jawa Timur berdasarkan curah hujannya. Metode *clustering* yang digunakan yaitu *Single Linkage*, *Complete Linkage*, dan *Average Linkage*. Sedangkan untuk *validation cluster* menggunakan *Dunn Index*.

**1. Single Linkage Clustering**

Langkah pertama yang dilakukan pada metode *single linkage clustering* yaitu dengan menentukan jarak antar data menggunakan persamaan (1) dan dihasilkan matriks jarak dengan ukuran 11 x 11. Kemudian diambil nilai terkecil dari matriks jarak yang telah ditentukan.

Dengan bantuan Rstudio, dendrogram dari pengklasteran dengan *single linkage clustering* yaitu sebagai berikut.



Gambar 2. Dendrogram dari *Single Linkage*

Pengklasteran menggunakan *single linkage*, jumlah anggota setiap kluster yaitu sebagai berikut.

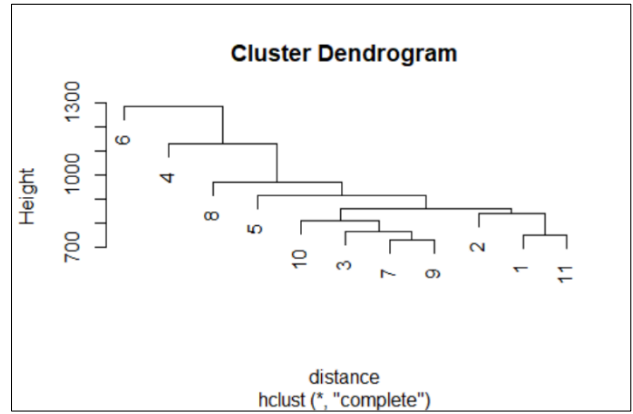
Tabel 1. Jumlah Anggota perkluster dengan *Single Linkage*

|   | Kluster   | Jumlah Anggota | Persentase |
|---|-----------|----------------|------------|
| 2 | Klaster 1 | 10             | 90.91%     |
|   | Klaster 2 | 1              | 9.09%      |
| 3 | Klaster 1 | 9              | 81.82%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
| 4 | Klaster 1 | 8              | 72.73%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
|   | Klaster 4 | 1              | 9.09%      |
| 5 | Klaster 1 | 7              | 63.64%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
|   | Klaster 4 | 1              | 9.09%      |
|   | Klaster 5 | 1              | 9.09%      |

## 2. Complete Linkage Clustering

Langkah pertama yang dilakukan pada pengklasteran dengan metode *complete linkage* sama dengan pengklasteran menggunakan metode *single linkage* yaitu dengan menentukan jarak antar data menggunakan persamaan (1) dan dari masing-masing jarak tersebut akan membentuk matriks jarak. Kemudian mengambil nilai terkecil dari matriks jarak tersebut.

Dengan bantuan RStudio, dendrogram yang dihasilkan dari pengklasteran dengan *single linkage clustering* yaitu sebagai berikut.



Gambar 3. Dendrogram dari *Complete Linkage*

Dari dendrogram yang terbentuk, jumlah anggota setiap kluster hasil pengklasteran menggunakan *complete linkage* yaitu sebagai berikut.

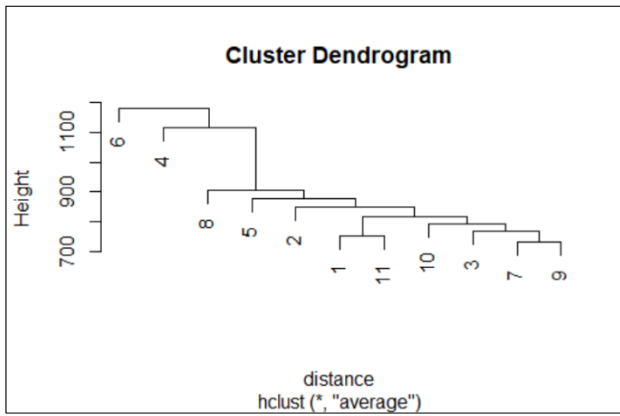
Tabel 2. Jumlah Anggota perkluster dengan *complete Linkage*

|   | Kluster   | Jumlah Anggota | Persentase |
|---|-----------|----------------|------------|
| 2 | Klaster 1 | 10             | 90.91%     |
|   | Klaster 2 | 1              | 9.09%      |
| 3 | Klaster 1 | 9              | 81.82%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
| 4 | Klaster 1 | 8              | 72.73%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
|   | Klaster 4 | 1              | 9.09%      |
| 5 | Klaster 1 | 7              | 63.64%     |
|   | Klaster 2 | 1              | 9.09%      |
|   | Klaster 3 | 1              | 9.09%      |
|   | Klaster 4 | 1              | 9.09%      |
|   | Klaster 5 | 1              | 9.09%      |

## 3. Average Linkage Clustering

Pada pengklasteran dengan metode *average linkage*, langkah pertama yang dilakukan yaitu dengan menentukan jarak antar data menggunakan *Euclidean distance* berdasarkan persamaan (1) dan dari jarak tersebut membentuk matriks jarak. Kemudian nilai dalam matriks jarak tersebut, diambil nilai terkecil.

Dendrogram dari pengklasteran dengan *average linkage clustering* dengan bantuan Rstudio yaitu sebagai berikut.



Gambar 3. Dendrogram dari Average Linkage

Berdasarkan dari gambar 3, jumlah anggota setiap kluster menggunakan pengklasteran dengan metode average linkage yaitu sebagai berikut.

Tabel 4. Jumlah Anggota perkluster dengan Average Linkage

|   | Kluster   | Jumlah Anggota | Persentase |
|---|-----------|----------------|------------|
| 2 | Kluster 1 | 10             | 90.91%     |
|   | Kluster 2 | 1              | 9.09%      |
| 3 | Kluster 1 | 9              | 81.82%     |
|   | Kluster 2 | 1              | 9.09%      |
|   | Kluster 3 | 1              | 9.09%      |
| 4 | Kluster 1 | 8              | 72.73%     |
|   | Kluster 2 | 1              | 9.09%      |
|   | Kluster 3 | 1              | 9.09%      |
|   | Kluster 4 | 1              | 9.09%      |
| 5 | Kluster 1 | 7              | 63.64%     |
|   | Kluster 2 | 1              | 9.09%      |
|   | Kluster 3 | 1              | 9.09%      |
|   | Kluster 4 | 1              | 9.09%      |
|   | Kluster 5 | 1              | 9.09%      |

**4. Cluster Validation dengan Dunn Index**

Setelah dilakukan pengklasteran dengan metode single linkage, complete linkage, dan average linkage, kemudian dilakukan cluster validation menggunakan Dunn Index berdasarkan persamaan (5). Dengan bantuan Rstudio, nilai Dunn Index masing-masing kluster yang terbentuk dari setiap metode clustering hierarki yaitu sebagai berikut.

Tabel 5. Nilai Dunn Index

| Kluster | Single Linkage | Complete Linkage | Average Linkage |
|---------|----------------|------------------|-----------------|
| 2       | 1.008178       | 1.008178         | 1.008178        |
| 3       | 1.120301       | 1.120301         | 1.120301        |
| 4       | 0.890091       | 0.9062188        | 0.9062188       |
| 5       | 0.8892958      | 0.9628253        | 0.9628253       |

Berdasarkan tabel 5, nilai dunn index tertinggi yaitu sebesar 1.120301 dengan 3 kluster baik dengan single linkage, complete linkage, maupun average linkage. Sehingga metode terbaik untuk pengklasteran daerah di

Jawa Timur berdasarkan curah hujannya yaitu dengan 3 kluster menggunakan metode baik single linkage, complete linkage, maupun average linkage.

**5. PENUTUP**

**Simpulan**

Metode terbaik untuk pengklasteran daerah di Jawa Timur berdasarkan curah hujannya berdasarkan nilai Dunn Index yaitu dengan 3 kluster baik dengan single linkage, complete linkage, maupun average linkage sebesar 1.120301. sehingga kluster daerah di Jawa Timur yang terbentuk berdasarkan curah hujannya yaitu kluster pertama meliputi Stasiun Klimatologi Malang, Stasiun Geofisika Malang, Stasiun Meteorologi Banyuwangi, Stasiun Meteorologi Juanda, Stasiun Meteorologi Tuban, Stasiun Meteorologi Sangkapura, Stasiun Meteorologi Kalianget, Stasiun Meteorologi Perak 1, dan Stasiun Meteorologi Maritim Tanjung Perak. Sedangkan kluster kedua meliputi Stasiun Geofisika Pasuruan, dan kluster ketiga meliputi Stasiun Geofisika Nganjuk.

**Saran**

Dalam penelitian selanjutnya, pengklasteran dapat menggunakan metode clustering lainnya baik hierarki maupun non hierarki yang diharapkan dapat mengklaster dengan lebih baik lagi.

**DAFTAR PUSTAKA**

Astuti, C. C., & Untari, R. S. (2017). Applied Hierarchical Cluster Analysis with Average Linkage Algoritm. *CAUCHY –Jurnal Matematika Murni dan Aplikasi*, 5(1), 1-7.

H, S. S., & Thomas, R. (2016). Regionalization of Rainfall in Kerala State. *International Conference on Emerging Trends in Engineering, Science and Technology*, 15-22.

Machiwal, D., Kumar, S., Meena, H. M., Santra, P., Singh, R. K., & Singh, D. V. (2017). Clustering of rainfall stations and distinguishing influential factors using PCA and HCA techniques over the western dryregion of India. *Meteorological Applications*, 300-311.

Mohbey, K. K., & Thakur, G. S. (2013). An Experimental Survey on Single Linkage Clustering. *International Journal of Computer Applications*, 76(17), 6-10.

Nafisah, Q., & Chandra, N. E. (2017). Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur. *Zeta – Math Journal*, 3(2), 31-36.

Onwukwe, C. E., & Ezeorah, J. N. (2009). Application of Single-Linkage Clustering Method in the Analysis of Growth Rate of Gross Domestic Product (GDP) at

1990 Constant Basic Prices (Million Naira). *Global Journal of Mathematical Sciences*, 8(2), 83-93.

Rani, Y., & Rohil, D. H. (2013). A Study of Hierarchical Clustering Algorithm. *International Journal of Information and Computation Technology*, 3(11), 1225-1232.

Rokach, L., & Maimon, O. (2005). Clustering Methods. Dalam O. Maimon, & L. Rokach (Penyunt.), *Data Mining and Knowledge Discovery Handbook* (hal. 321-352). Boston: Springer.