

# Perbandingan Penerapan Metode *Agglomerative* dengan Metode *K-Means* pada Data Curah Hujan di Wilayah Bogor

Hera Khoirunnisa<sup>1, a)</sup>, Budi Nurani Ruchjana<sup>1, b)</sup>, Iin Irianingsih<sup>1, c)</sup> dan Bambang Suhandi<sup>2)</sup>

<sup>1</sup>*Departemen Matematika, Fakultas MIPA, Universitas Padjadjaran*

<sup>2</sup>*Balai Pengelola Observatorium Nasional Kupang*

<sup>a)</sup>*email: koirunnisahera@gmail.com*

<sup>b)</sup>*email: budi.nurani@unpad.ac.id*

<sup>c)</sup>*email: iin.irianingsih@unpad.ac.id*

<sup>2)</sup>*email: bsuhandi63@gmail.com*

## Abstrak

Bogor merupakan salah satu wilayah di Jawa Barat yang dijuluki sebagai kota hujan, karena memiliki curah hujan relatif lebih besar dibandingkan dengan wilayah lain, sehingga perlu dilakukan pengelompokan wilayah berdasarkan tinggi rendahnya curah hujan sebagai acuan pemerintah dalam penanganan bencana. Analisis Statistika Multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristiknya adalah analisis *cluster*. Metode dari analisis *cluster* yang digunakan pada penelitian ini yaitu *Agglomerative* dan *K-Means*. Perbedaan yang signifikan pada kedua metode tersebut terdapat pada proses pembentukan *cluster*. Oleh karena itu, tujuan pada penelitian ini adalah membandingkan metode yang terbaik berdasarkan kerapatan *cluster*. Data yang digunakan dalam penelitian ini adalah data agregat curah hujan bulanan musim basah (Desember-Januari-Februari) dari 24 stasiun pos hujan di wilayah Bogor tahun 2017-2018. Hasil penelitian menunjukkan bahwa curah hujan wilayah Kabupaten Bogor dapat dibagi menjadi 2 *cluster* yaitu *cluster* 1 kategori curah hujan sedang dan *cluster* 2 kategori curah hujan tinggi dengan perbandingan nilai kerapatan *cluster* kedua metode menghasilkan nilai yang sama yaitu sebesar 49,4%, sehingga kedua metode tersebut baik untuk digunakan dalam pembentukan *cluster* curah hujan di wilayah Bogor dan bisa dijadikan sebagai rekomendasi bagi instansi terkait penggunaan data curah hujan seperti LAPAN dan BMKG.

*Kata kunci: Curah Hujan, Analisis Cluster, Perbandingan Kerapatan Cluster, Agglomerative, K-Means*

## Abstract

Bogor is one of the regions in West Java, which is dubbed as a city of rain because it has a relatively greater rainfall compared to other regions, so it is necessary to group regions based on high and low rainfall as a reference for the government in handling disasters. Multivariate Statistical Analysis which aims to group objects based on their characteristics is cluster analysis. The method of cluster analysis used in this study

is Agglomerative and K-Means. Significant differences in the two methods are found in the cluster formation process. Therefore, the aim of this study is to compare the best method based on cluster density. The data used in this study is the aggregate data of monthly wet season rainfall (December-January-February) from 24 rain stations in the Bogor region in 2017-2018. The results showed that the rainfall in the Bogor Regency area can be divided into 2 clusters, namely cluster 1, the category of moderate rainfall and cluster 2, the category of high rainfall, with the comparison of the cluster density values of the two methods producing the same value of 49.4%, so both methods good for use in the formation of rainfall clusters in the Bogor region and can be used as recommendations for agencies related to the use of rainfall data such as LAPAN and BMKG.

Keywords: *Rainfall, Cluster Analysis, Cluster Density Comparison, Agglomerative, K-Means*

## Pendahuluan

Bogor seringkali dijuluki sebagai kota hujan karena intensitas curah hujan yang terjadi di daerah tersebut relatif lebih besar dibandingkan dengan daerah lain sehingga perlu diadakannya pengelompokan curah hujan di wilayah Bogor sebagai referensi bagi pemerintah dalam penanggulangan tanggap bencana. Dengan adanya pembentukan *cluster* wilayah Bogor berdasarkan curah hujannya, maka penanggulangan bencana pada daerah yang termasuk *cluster* curah hujan yang tinggi dengan daerah yang termasuk *cluster* curah hujan rendah membutuhkan penanganan yang berbeda. Oleh karena itu objek pada penelitian ini adalah fenomena curah hujan bulanan pada musim basah dengan data sampel yang diambil dari 24 stasiun pos hujan di wilayah Bogor.

Salah satu teknik statistika yang digunakan untuk mengelompokkan objek berdasarkan karakteristiknya adalah analisis *cluster*. Analisis *cluster* terbagi menjadi dua metode yaitu metode hirarki dan metode non-hirarki. Metode hirarki dan metode non-hirarki yang digunakan pada penelitian ini adalah *Agglomerative* dan *K-Means*. Proses pembentukan *cluster* pada kedua metode tersebut sangat berbeda. Metode *Agglomerative* menentukan banyaknya *cluster* setelah semua objek tergabung dalam satu *cluster* utama sedangkan metode *K-Means* menentukan banyaknya *cluster* dilakukan sebagai proses awal. Penelitian sebelumnya menyimpulkan bahwa metode *Agglomerative* merupakan metode yang lebih baik jika dibandingkan dengan metode *K-Means* dalam menentukan *cluster* data mahasiswa pemohon beasiswa [1]. Penelitian lain mengemukakan bahwa metode *K-Means* merupakan metode terbaik dalam menentukan *cluster* penumpang bus Trans Jogja [2]. Oleh karena itu, pada penelitian ini penulis membandingkan metode terbaik yang membentuk *cluster* curah hujan di wilayah Bogor dengan membandingkan kerapatan *cluster*.

## Metode

### 1. Analisis *Cluster*

Analisis *cluster* adalah salah satu teknik analisis statistika multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik dari setiap kelompok tertentu. Hasil dari pengelompokan suatu objek harus menunjukkan kehomogenan yang tinggi di dalam satu *cluster* dan heterogenan yang tinggi antar *cluster*. Metode analisis *cluster* secara umum terbagi menjadi 2 yaitu metode hirarki dan metode non-hirarki [3].

## 2. Metode dalam Analisis Cluster

### a. Metode Agglomerative

Metode *Agglomerative* merupakan metode analisis *cluster* hirarki yang dimulai dengan menentukan masing-masing objek sebagai *cluster* yang berbeda. Dua objek yang memiliki kemiripan digabungkan menjadi *cluster* yang baru. Proses tersebut dilakukan berulang sampai semua *cluster* bergabung dalam satu *cluster*. Hasil akhir dari metode *Agglomerative* ditampilkan dalam bentuk diagram pohon dua dimensi atau disebut *dendogram*. Terdapat beberapa cara perhitungan jarak kesamaan antar objek pada metode *Agglomerative*, pada penelitian ini perhitungan jarak kesamaan antar objek menggunakan metode *Complete Linkage* [3].

*Complete Linkage* adalah algoritma *cluster* hirarki yang mendefinisikan kesamaan objek sebagai jarak maksimum antar objek tunggal dalam satu *cluster* dan objek tunggal lainnya,  $\mathbf{D} = \{d_{ik}\}$  atau dalam arti lain jarak antara dua *cluster*  $U$  dan  $V$  didefinisikan sebagai jarak maksimum diantara titik di  $U$  dan titik di  $V$  dengan rumus:

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (1)$$

dengan:

$d_{UW}$  : Jarak antar objek pada *cluster*  $U$  dan *cluster*  $W$

$d_{VW}$  : Jarak antar objek pada *cluster*  $V$  dan *cluster*  $W$

Pada proses analisis *cluster* dengan metode *Agglomerative* diperlukan pemilihan *cluster* untuk menentukan jumlah *cluster* optimum. Pemilihan jumlah *cluster* optimum dapat digunakan dengan menggunakan kriteria nilai *Pseudo F statistics* dengan rumus:

$$Pseudo\ F\ statistics = \frac{R^2/k - 1}{1 - R^2/N - k} \quad (2)$$

Dan

$$R^2 = \frac{JKTotal - JKDalam\ Kelompok}{JKTotal} \quad (3)$$

$$JKTotal = \sum_{i=1}^n \sum_{p=1}^c \sum_{j=1}^k (x_{ipj} - \bar{x})^2 \quad (4)$$

$$JKDalam\ Kelompok = \sum_{i=1}^n \sum_{p=1}^c \sum_{j=1}^k (x_{ipj} - \bar{x}_j)^2 \quad (5)$$

dengan:

$x_{ipj}$  : objek ke- $i$  pada variabel ke- $p$  *cluster* ke- $j$

$\bar{x}$  : rata-rata seluruh sampel

$\bar{x}_j$  : rata-rata sampel pada *cluster* ke- $j$

Nilai *Pseudo F statistics* tertinggi menunjukkan bahwa jumlah kelompok data telah optimal [4].

### b. Metode K-Means

Istilah *K-Means* digunakan untuk menjelaskan sebuah algoritma yang memiliki *centroid* terdekat di setiap *cluster* [3]. Penentuan *centroid* awal dilakukan secara acak dan ditentukan partisi objek sebanyak  $k$

*cluster*, kemudian menghitung *centroid* ke-*i* variabel ke-*p* berikutnya menggunakan persamaan sebagai berikut:

$$v_{ip} = \frac{\sum_{i=1}^n x_{ip}}{n_{pj}}, i = 1, 2, \dots, n \quad (6)$$

dengan:

$v_{ip}$  : *centroid* ke-*i* variabel ke-*p*

$x_{ip}$  : objek pengamatan ke-*i* variabel ke-*p*

$n_{pj}$  : banyaknya objek pada variabel ke-*p cluster* ke-*j*

### 3. Uji Asumsi Analisis Cluster

#### a. Uji Kecukupan Sampel

Uji kecukupan sampel bertujuan untuk menguji apakah sampel dianggap cukup untuk dianalisis lebih lanjut. Uji kecukupan sampel dapat dilihat dari:

- Indeks *Kaiser Meyer Olkin* (KMO) yang digunakan untuk meneliti apakah data tersebut layak dianalisis atau tidak dengan menggunakan rumus:

$$KMO = \frac{\sum_{p \neq q}^c r_{pq}^2}{\sum_{p \neq q}^c r_{pq}^2 + \sum_{p \neq q}^c b_{pq}^2} \quad (7)$$

dengan:

$r_{pq}$  : koefisien korelasi sederhana antara variabel ke-*p* dan variabel ke-*q*

$b_{pq}$  : koefisien korelasi parsial antara variabel ke-*p* dan variabel ke-*q*.

Sampel dikatakan layak dianalisis jika nilai  $KMO \geq 0,5$  [5].

- Measure of Sampling Adequacy* (MSA) yang digunakan untuk mengetahui apakah variabel memadai untuk dianalisis lebih lanjut. Jika setiap variabel memiliki nilai  $MSA \geq 0,50$  maka variabel tersebut layak untuk dianalisis [5].

#### b. Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji apakah ada korelasi yang tinggi dari variabel yang diikuti sertakan atau tidak. Variabel analisis *cluster* tersebut disebut baik jika antar variabel bebas tidak terjadi multikolinearitas. Salah satu cara untuk mengetahui ada tidaknya multikolinearitas dapat dilihat dari nilai *Variance Inflation Factor* (VIF) dengan rumus sebagai berikut:

$$VIF = \frac{1}{1 - R_p^2} \quad (8)$$

dengan:

$R_p^2$  : koefisien determinasi antara variabel bebas ke-*p* dengan variabel bebas lainnya.

Jika nilai VIF berada pada angka 5 atau 10 maka terindikasi terjadinya multikolinearitas [6].

### 4. Jarak Euclid

Jarak antar objek ke-*i* dan objek ke-*k* disimbolkan dengan  $d_{ik}$  pada variabel ke-*p*. Perhitungan jarak Euclid dari dua objek sebanyak *p* variabel menggunakan persamaan sebagai berikut:

$$d_{ik} = \sqrt{\sum_{p=1}^c |x_{ip} - x_{kp}|^2} \quad (9)$$

dengan:

$d_{ik}$  : jarak Euclid pada objek ke- $i$  dan objek ke- $k$

$x_{ip}$  : data dari objek ke- $i$  variabel ke- $p$

$x_{kp}$  : data dari objek ke- $k$  variabel ke- $p$

Semakin dekat jarak Euclid semakin mirip objek data tersebut [3].

## 5. Kerapatan Cluster

Prinsip dalam membentuk *cluster* yang baik adalah kesamaan antar objek dalam *cluster* yang maksimum dan kesamaan antar objek pada *cluster* berbeda yang minimum. Salah satu metode untuk menentukan *cluster* terbaik adalah dengan menghitung kerapatan *cluster*. *Cluster* yang baik memiliki nilai variansi dalam *cluster* yang minimum dan variansi antar *cluster* yang maksimum [7].

Jika diberikan cluster  $C_j$  sebanyak  $k$  cluster maka untuk mencari nilai variansi *cluster* ke- $j$  ( $s_j^2$ ) menggunakan rumus:

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 ; i = 1, 2, \dots, n \quad (10)$$

karena ( $s_j^2$ ) telah diketahui, untuk mencari variansi dalam *cluster* ( $s_w^2$ ) dapat dihitung menggunakan rumus:

$$s_w^2 = \frac{1}{k} \sum_{j=1}^k s_j^2 \quad (11)$$

Variansi antar *cluster* ( $s_b^2$ ) digunakan untuk mengukur variabilitas dari rata-rata setiap kelompok berdasarkan rata-rata global ( $\bar{x}$ ). Rumus variansi antar *cluster* ( $s_b^2$ ) sebagai berikut:

$$s_b^2 = \frac{1}{k - 1} \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 \quad (12)$$

dengan:

$x_{ji}$  : anggota ke- $i$  setiap *cluster*

$\bar{x}_j$  : rata-rata *cluster* ke- $j$

$n_j$  : banyaknya anggota *cluster* ke- $j$

$k$  : banyaknya *cluster*

$\bar{x}$  : anggota ke- $i$  setiap *cluster*

*Cluster* yang baik memiliki nilai ( $s_w^2$ ) yang minimum dan nilai ( $s_b^2$ ) yang maksimum [7], artinya *cluster* yang baik memiliki nilai minimum *Rasio* dengan rumus:

$$Rasio = \frac{s_w^2}{s_b^2} \quad (13)$$

### Hasil dan Diskusi

Data yang digunakan dalam penelitian ini adalah data agregat curah hujan bulanan pada musim basah periode Desember-Januari-Februari (DJF). Data tersebut diperoleh dari Badan Pusat Statistik Kabupaten Bogor tahun 2018 [8]. Data agregat curah hujan bulanan di wilayah Bogor pada 24 stasiun pos hujan adalah sebagai berikut:

- $X_1$  : data agregat curah hujan bulan Desember tahun 2017  
 $X_2$  : data agregat curah hujan bulan Januari tahun 2018  
 $X_3$  : data agregat curah hujan bulan Februari tahun 2018

Contoh data penelitian disajikan pada Tabel 1 berikut ini.

**Tabel 1.** Contoh Data Penelitian Agregat Curah Hujan Kab. Bogor

Stasiun Pos Hujan	Agregat Curah Hujan Bulanan (mm)		
	Des-17	Jan-18	Feb-18
Leuwi Liang	230	293	178
Dramaga	181	190	364
Empang	190	199	473
Ciawi	225	165	743
Gadog	292	316	832
Dayeuh	325	266	394
Cariu	141	168	351
Klapanunggal	309	283	423
Ciriung	209	144	441
Parung Panjang	101	219	316
Jasinga	298	335	203
Cibalagung	225	281	394
...			

Sumber: BPS Kab. Bogor (2018)

#### 1. Hasil Uji Asumsi Analisis Cluster

Pengolahan data untuk uji asumsi Analisis Cluster dilakukan menggunakan software RStudio menggunakan *library* dengan beberapa *command* diantaranya:

```
>library(cluster)
>library(dendextend)
>library(factoextra)
>library(faraway)
>library(readxl)
>library(stats)
>library(psych)
```

```
#Input Data
setwd("E://HERA//SKRIPSI")
data1<-read_excel("daftar CH Bogor musim basah.xlsx")

#Korelasi Data
data=data1[-1]
data
r=cor(data)
#uji asumsi multikolinearitas
vif(data)
#uji asumsi kecukupan sampel
KMOS(data)
```

Hasil uji yang diperoleh disajikan sebagai berikut:

#### a. Hasil Uji Kecukupan Sampel

Uji kecukupan sampel bertujuan untuk menguji apakah sampel dan variabel yang digunakan cukup untuk dianalisis, dapat dilihat dari nilai indeks KMO dan MSA yang ditampilkan pada Tabel 2.

**Tabel 2.** Hasil uji KMO dan MSA

<i>Kaiser Meyer Olkin (KMO)</i>		0,57494		
<i>Measure of Sampling</i>	$X_1$	$X_2$	$X_3$	
<i>Adequacy (MSA)</i>	0,55	0,56	0,74	

Berdasarkan Tabel 2 dapat dilihat bahwa nilai KMO sebesar 0,57494 dan nilai MSA masing-masing variabel  $> 0,5$ . Karena nilai KMO  $> 0,5$  dan nilai MSA  $> 0,5$  maka sampel dan variabel yang digunakan pada penelitian ini layak untuk dianalisis lebih lanjut.

#### b. Hasil Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk mengetahui ada tidaknya korelasi yang sangat kuat antar variabel penelitian. Untuk mendeteksi adanya multikolinearitas menggunakan nilai *Variance Inflation Factor (VIF)* yang ditampilkan dalam Tabel 3.

**Tabel 3.** Hasil uji nilai VIF

Nilai VIF	$X_1$	$X_2$	$X_3$
	1,797	1,688	1,135

Berdasarkan Tabel 3 dapat dilihat bahwa nilai VIF dari semua variabel kurang dari 5 atau 10, itu artinya tidak terdapat multikolinearitas pada data penelitian.

## 2. Proses Pembentukan Cluster

### a. Pembentukan Cluster Metode Agglomerative

Pembentukan *cluster* metode *Agglomerative* dimulai dengan menghitung jarak antar objek menggunakan persamaan (1) dilanjutkan dengan menggabungkan objek yang memiliki kedekatan jarak. Contoh perhitungan kedekatan jarak pada stasiun pos hujan Leuwi Liang dengan stasiun pos hujan Dramaga sebagai berikut:

$$d_{12} = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + (x_{13} - x_{23})^2}$$

$$d_{12} = \sqrt{(230 - 181)^2 + (293 - 190)^2 + (178 - 364)^2}$$

$$d_{12} = 218,188 \text{ mm}$$

sedangkan untuk perhitungan kedekatan jarak stasiun pos hujan Leuwi Liang dengan stasiun pos hujan Empang sebagai berikut:

$$d_{13} = \sqrt{(x_{11} - x_{31})^2 + (x_{12} - x_{32})^2 + (x_{13} - x_{33})^2}$$

$$d_{13} = \sqrt{(230 - 190)^2 + (293 - 199)^2 + (178 - 473)^2}$$

$$d_{13} = 312,188 \text{ mm.}$$

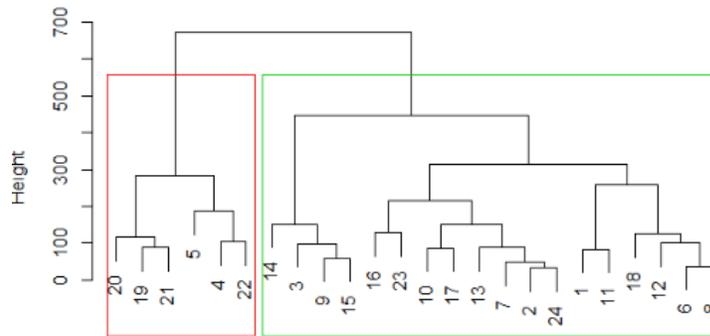
Dari contoh tersebut, dapat dilihat hasil perhitungan kedekatan jarak antara stasiun pos hujan Leuwi Liang dengan stasiun pos hujan Dramaga adalah 218,188 mm sedangkan jarak antara stasiun pos hujan Leuwi Liang dengan stasiun pos hujan Empang adalah 312,188 mm. Hal tersebut menunjukkan bahwa stasiun pos hujan Leuwi Liang memiliki karakteristik yang lebih mirip dengan Stasiun pos hujan Dramaga jika dibandingkan dengan stasiun pos hujan Empang. Perhitungan dilanjutkan sampai semua objek bergabung dalam satu *cluster* utama.

Pemilihan *cluster* optimum menggunakan nilai *Pseudo F statistics* yang ditampilkan pada Tabel 4.

**Tabel 4.** Perbandingan *Pseudo F Statistics*

Banyaknya Cluster	R-Square	Pseudo F Statistics
2	0,14909	3,85459
3	0,24575	3,42111
4	0,27677	2,55129
5	0,27842	1,83273
6	0,28545	1,43816

Berdasarkan Tabel 4 dapat ditunjukkan nilai *Pseudo F statistics* tertinggi adalah 3,85459 dengan jumlah *cluster* yang terbentuk sebanyak 2 *cluster*. Oleh karena itu, *cluster* optimum stasiun pos hujan wilayah Bogor sebanyak 2 *cluster* dengan hasil *dendogram* pada Gambar 1.



**Gambar 1.** Cluster dendrogram

Berdasarkan Gambar 1 klasifikasi pembagian stasiun pos hujan yang masuk ke dalam *cluster* dapat ditampilkan dalam Tabel 5.

**Tabel 5.** Anggota *cluster* yang terbentuk metode *Agglomerative*

Cluster	Stasiun Pos Hujan
Cluster 1	Leuwi Liang, Dramaga, Dayeuh, Cariu, Klapanunggal, Parung Panjang, Jasinga, Cibalagung, Citayam, Beji Depok, Tanjung Rasa, Tunggilis, Kebun Raya, Cimanggu, Empang, Ciriung, Lanud Atang Sanjaya, dan Depok BMKG.
Cluster 2	Ciawi, Gadog, Citeko, Naringgul, Gunung Mas, dan Katulampa.

Berdasarkan Tabel 5 dapat dilihat bahwa *cluster* 1 sebanyak 18 stasiun pos hujan yang memiliki rata-rata agregat curah hujan bulanan sebesar 254,20 mm dengan kategori *cluster* curah hujan sedang. *Cluster* 2 sebanyak 6 stasiun pos hujan yang memiliki rata-rata agregat curah hujan bulanan sebesar 439,94 mm dengan kategori *cluster* curah hujan tinggi.

#### b. Pembentukan *Cluster* Metode *K-Means*

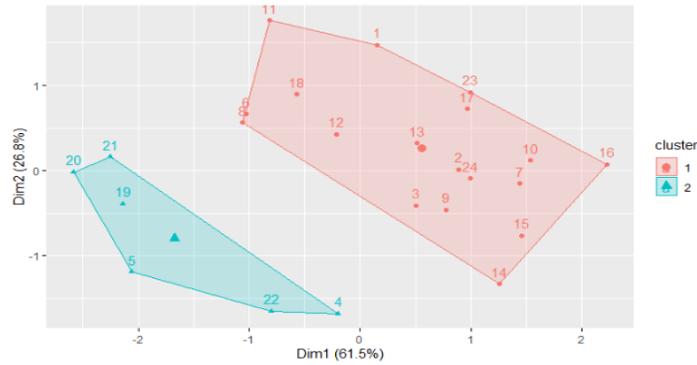
Pada pembentukan *cluster* metode *Agglomerative*, *cluster* telah optimum sebanyak 2 *cluster*. Berdasarkan hal tersebut, pembentukan *cluster* metode *K-Means* sebanyak 2 *cluster*. Inisiasi *centroid K-Means* dilakukan secara acak dengan  $C_1$  adalah *centroid cluster* 1 dan  $C_2$  adalah *centroid cluster* 2. Perhitungan jarak objek terhadap stasiun pos hujan sesuai persamaan (9). Contoh perhitungan jarak stasiun pos hujan Leuwi Liang dengan  $C_1$  berikut:

$$\begin{aligned} \text{Leuwi Liang} &= \sqrt{(x_{11} - 225)^2 + (x_{12} - 177)^2 + (x_{13} - 180)^2} \\ &= \sqrt{(230 - 225)^2 + (293 - 177)^2 + (178 - 180)^2} \\ &= 116,12 \text{ mm.} \end{aligned}$$

Perhitungan jarak stasiun pos hujan Leuwi Liang dengan  $C_2$  sebagai berikut:

$$\begin{aligned} \text{Leuwi Liang} &= \sqrt{(x_{11} - 292)^2 + (x_{12} - 177)^2 + (x_{13} - 180)^2} \\ &= \sqrt{(230 - 292)^2 + (293 - 177)^2 + (178 - 180)^2} \\ &= 657,33 \text{ mm} \end{aligned}$$

Stasiun pos hujan Leuwi Liang memiliki jarak terdekat ke  $C_1$  yang berarti bahwa stasiun pos hujan Leuwi Liang masuk ke *cluster* 1. Demikian seterusnya sampai stasiun pos hujan ke-24 bergabung ke dalam *cluster* sesuai jarak terdekat. Plot *cluster* metode *K-Means* ditampilkan pada Gambar 2.



**Gambar 2.** Plot *cluster* metode *K-Means*

Berdasarkan Gambar 2. klasifikasi pembagian stasiun pos hujan yang masuk ke dalam *cluster* dapat ditampilkan dalam Tabel 6.

**Tabel 6.** Anggota *cluster* yang terbentuk metode *K-Means*

<b>Cluster</b>	<b>Stasiun Pos Hujan</b>
<i>Cluster 1</i>	Leuwi Liang, Dramaga, Dayeuh, Cariu, Klapanunggal, Parung Panjang, Jasinga, Cibalagung, Citayam, Beji Depok, Tanjung Rasa, Tunggilis, Kebun Raya, Cimanggu, Empang, Ciriung, Lanud Atang Sanjaya, dan Depok BMKG.
<i>Cluster 2</i>	Ciawi, Gadag, Citeko, Naringgul, Gunung Mas, dan Katulampa.

Pada Tabel 6 dapat dilihat bahwa *cluster 1* sebanyak 18 stasiun pos hujan yang memiliki rata-rata agregat curah hujan bulanan sebesar 254,20 mm dengan kategori *cluster* curah hujan sedang. *Cluster 2* sebanyak 6 stasiun pos hujan yang memiliki rata-rata agregat curah hujan bulanan sebesar 439,94 mm dengan kategori *cluster* curah hujan tinggi.

Berdasarkan hasil anggota *cluster* yang terbentuk diperoleh bahwa pembentukan *cluster* metode *Agglomerative* dengan metode *K-Means* menghasilkan anggota yang sama pada kategori *cluster* yang sama dengan rata-rata curah hujan bulanan setiap *cluster* yang sama pula. Oleh karena itu, baik metode *Agglomerative* maupun metode *K-Means* menghasilkan 2 *cluster*, *cluster 1* stasiun pos hujan dengan kategori curah hujan sedang dan *cluster 2* stasiun pos hujan dengan kategori curah hujan tinggi.

Sebaran *cluster* stasiun pos metode *Agglomerative* dan metode *K-Means* dapat ditampilkan dalam peta wilayah Bogor pada Gambar 3.



**Gambar 3.** Sebaran *cluster* pada peta wilayah Bogor

Sumber: <https://www.bogorchannel.com> [9]

Keterangan:

- Cluster 1
- Cluster 2

### c. Perbandingan Kerapatan Cluster

Alat perbandingan untuk menentukan *cluster* terbaik adalah perhitungan kerapatan *cluster*. Setelah dilakukan pembentukan *cluster* menggunakan metode *Agglomerative* dan metode *K-Means*, hasil yang diperoleh ternyata anggota *cluster* yang dihasilkan oleh kedua metode tersebut sama. Sehingga kategori *cluster* untuk metode *Agglomerative* dan metode *K-Means* pun sama. Selanjutnya dihitung variansi setiap *cluster* yang ditampilkan pada Tabel 7.

**Tabel 7.** Variansi setiap *cluster*

Cluster	Variansi
Cluster 1	34273,57
Cluster 2	170524,98

Berdasarkan nilai variansi setiap *cluster* pada Tabel 7 dapat dihitung nilai variansi dalam kelompok sebagai berikut:

$$s_w^2 = \frac{1}{2}(34272,57 + 170524,98)$$

$$s_w^2 = 102399,28.$$

Setelah dihitung variansi dalam *cluster*, maka dihitung variansi antar *cluster* ( $s_b^2$ ) dengan terlebih dahulu menghitung rata-rata keseluruhan dari setiap *cluster* sebagai berikut:

$$\bar{x} = \frac{1}{2}(254,20 + 439,94)$$

$$\bar{x} = 347,07$$

maka perhitungan variansi antar *cluster* ( $s_b^2$ ) adalah sebagai berikut:

$$s_b^2 = \frac{1}{2-1}[(254,20 - 347,07) + (439,94 - 347,07)]$$

$$s_b^2 = 206997,74$$

selanjutnya dihitung nilai rasio dari variansi dalam *cluster* ( $s_w^2$ ) dengan variansi antar *cluster* ( $s_b^2$ ) sebagai berikut:

$$Rasio = \frac{s_w^2}{s_b^2} = \frac{102399,28}{206997,74} = 0,4946879$$

Berdasarkan perhitungan hasil bahwa anggota *cluster* untuk kedua metode sama pada masing-masing *cluster*, maka kedua metode tersebut menghasilkan nilai perbandingan *cluster* yang sama yaitu sebesar 49,4% dengan perbedaan yang signifikan terletak pada proses pembentukan awal *cluster*.

### Kesimpulan

Berdasarkan hasil dari pembahasan dapat disimpulkan bahwa pembentukan *cluster* metode *Agglomerative* dan metode *K-Means* pada 24 stasiun pos hujan wilayah Bogor dengan menggunakan data

agregat curah hujan bulanan pada musim basah periode Desember tahun 2017 sampai Februari tahun 2018 terbentuk sebanyak 2 *cluster*, yaitu *cluster* 1 berupa kelompok stasiun pos hujan wilayah Bogor dengan kategori curah hujan sedang dan *cluster* 2 berupa kelompok stasiun pos hujan wilayah Bogor dengan kategori curah hujan tinggi. Setelah dilakukan perbandingan menggunakan kerapatan *cluster*, hasil kerapatan *cluster* metode *Agglomerative* dan metode *K-Means* untuk data curah hujan di wilayah Bogor menghasilkan nilai yang sama. Oleh karena itu, kedua metode dalam analisis *cluster* dapat digunakan untuk menentukan kelompok lokasi-lokasi curah hujan di wilayah Bogor.

### Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Rektor Universitas Padjadjaran yang telah memberikan dukungan dana penelitian melalui *Academic Leadership Grant* tahun 2020 dengan nomor kontrak 1427/UN6.3.1/LT/2020 untuk diseminasi hasil penelitian dosen dan mahasiswa.

### Referensi

- [1] Satria, F. & Aziz, R. A., "Perbandingan Kinerja Metode Ward dan K-Means Dalam Menentukan Cluster Data Mahasiswa Pemohon Beasiswa", *Jurnal TIM Darmajaya*, pp. 12-26, 2016.
- [2] Zahrotun, L., "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode *Clustering K-Means* dan *Agglomerative Hierarchical Clustering (AHC)*", *Jurnal Informatika*, Vol. 9, pp. 1039-1047, 2015.
- [3] Johnson, R. A. & Wichern, D. W., "Applied Multivariate Statistical Analysis", Penerbit Pearson, New Jersey, 2007.
- [4] Putri, M. M. & Fithriasari, K., "Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Metode Kohonen SOM dan K-Means", *Jurnal Sains dan Seni ITS*, Vol. 4 No. 1, pp. 13-18, 2015.
- [5] Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E., "Multivariate Data Analysis", Penerbit Pearson Prentice Hall, New Jersey, 2010.
- [6] James, G., Witten, D., Hastie, T. & Tibshirani, R., "An Introduction to Statistical Learning", Penerbit Springer, New York, 2013.
- [7] Barakbah, A. R. & Arai, K., "Determining Constraint of Moving Variance to Find Global Optimum and Make Automatic Clustering", *Industrial Electronics Seminar (IES)*, pp. 409-413, 2004.
- [8] <http://bogorkab.bps.go.id/> [diakses 01 November 2017].
- [9] <https://www.bogorchannel.com>, [diakses 01 November 2017]