

**PANDUAN PENGGUNAAN
APLIKASI R-GUI: OPTIMALISASI METODE HIERARCHICAL
AGGLOMERATIVE CLUSTERING MENGGUNAKAN CALINSKI-
HARABASZ PSEUDO F-STATISTIC**



**Aplikasi R-GUI: Optimalisasi Metode Hierarchical Agglomerative
Clustering Menggunakan Calinski-Harabasz Pseudo F-Statistic**

TIM INVENTOR:

Ali Mahmudan

Dr. Di Asih I Maruddani, S.Si., M.Si.

Dr. Budi Warsito, S.Si., M.Si.

**DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2021**

Deskripsi Aplikasi

Aplikasi GUI R: Optimalisasi Metode Hierarchical Agglomerative Clustering Menggunakan Calinski-Harabasz Pseudo F-Statistic merupakan aplikasi GUI R yang disusun untuk membantu penelitian mengenai pengelompokan (clustering) dengan Algoritma Agglomerative Hierarchical. Aplikasi ini dilengkapi dengan beberapa metode pengukuran jarak seperti Single Linkage, Complete Linkage, Average Linkage, Centroid, dan Ward yang dapat digunakan sebagai pemilihan metode pengukuran jarak terdekat pada saat melakukan clustering. Aplikasi ini mampu melakukan pemilihan jumlah cluster optimal dengan menggunakan nilai Pseudo F-Statistics. Nilai Pseudo F-Statistics diperoleh dari perhitungan metode Calinski-Harabasz Pseudo F-Statistics. Metode Calinski-Harabasz Pseudo F-Statistics tersebut mampu menemukan jumlah cluster optimal yang merupakan cluster dengan tingkat homogenitas yang tinggi antar anggota dalam cluster tersebut dan heterogenitas yang tinggi antar cluster yang terbentuk.

Langkah Pendahuluan Sebelum Menggunakan Aplikasi

Berikut merupakan langkah pendahuluan yang harus dilakukan sebelum menggunakan aplikasi ini:

1. Memasang program R-GUI (usahakan minimal versi 4.0.0) dan R-Studio.

Agar program GUI dapat berjalan dengan baik maka disarankan menggunakan R-GUI minimal versi 4.0.0 meskipun jika di bawah versi tersebut masih memungkinkan untuk dijalankan, namun akan lebih baik jika menggunakan R-GUI di atas versi 4.0.0. Sedangkan untuk program R-Studio, tidak ada ketentuan terkait minimal versi yang harus terpasang.

2. Persiapan data

Data yang perlu dipersiapkan adalah dua buah data yaitu data indikator *clustering* dan data objek yang ingin diklasterkan. Data variabel indikator harus memiliki satuan yang sama, jika tidak memiliki satuan yang sama maka data harus distandardisasi terlebih dahulu. Semua data yang digunakan haruslah dalam format '.txt'.

3. *Install Packages*

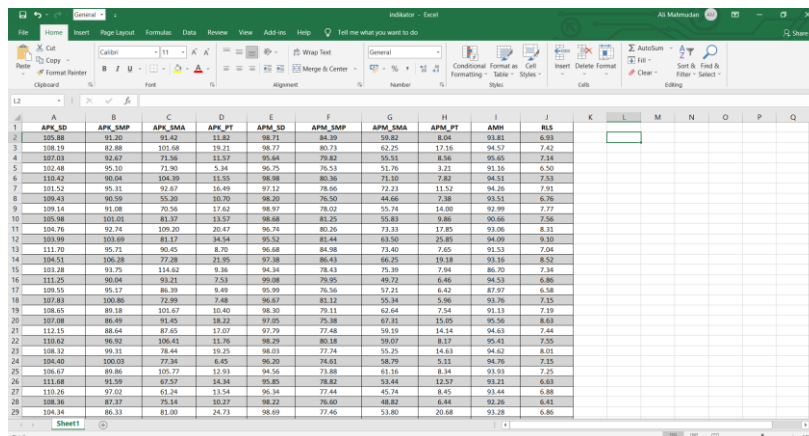
Adapaun beberapa paket yang harus di-*install* terlebih dahulu sebelum menjalankan aplikasi tersebut adalah paket ‘shiny’, ‘shinythemes’, ‘psych’, dan ‘clusterSim’.

Langkah-Langkah Penggunaan Aplikasi

1. Persiapkan data variabel yang akan digunakan sebagai indikator klasterisasi dan data objek yang akan diklasterkan dalam bentuk ‘.txt’. Berikut disajikan ilustrasi cara membentuk data dengan format ‘.txt’:

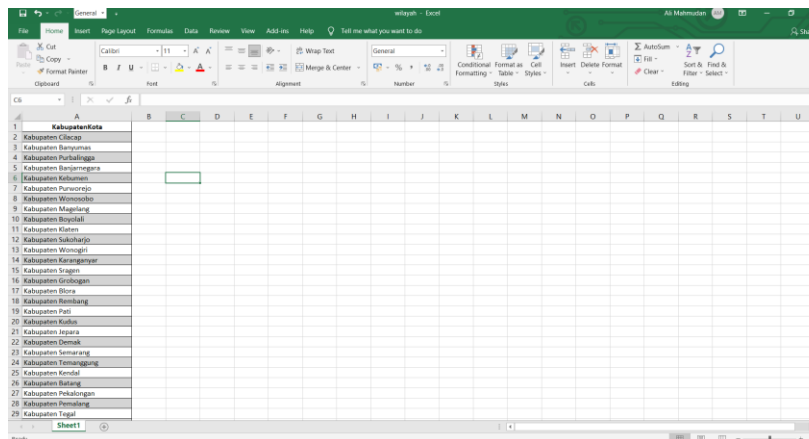
- Input* data menggunakan Ms. Excel.
- Isikan nama masing-masing variabel indikator dan objek di baris pertama.

Variabel indikator



	A	B	C	D	E	F	G	H	I	J
1	APK_SD	APK_SMP	APK_SMA	APK_PT	APM_SD	APM_SMP	APM_SMA	APM_PT	AMN	RLS
2	105.86	91.26	91.42	11.84	98.71	86.39	59.82	8.04	93.88	6.88
3	108.19	82.86	101.68	19.21	98.71	80.73	62.25	17.15	94.57	7.42
4	107.03	92.67	71.56	11.57	95.64	79.82	55.51	8.56	95.45	7.54
5	102.48	95.33	71.90	1.34	96.25	76.53	51.76	2.21	91.26	6.50
6	110.42	90.04	104.39	11.55	98.98	80.36	71.10	7.82	94.51	7.53
7	101.52	95.11	92.67	16.49	97.12	78.66	72.23	11.52	94.26	7.91
8	109.43	90.59	55.20	10.20	98.20	76.50	44.66	7.38	93.51	6.76
9	109.14	91.08	70.56	17.62	98.97	78.02	55.74	14.00	92.99	7.77
10	105.96	101.01	81.37	13.57	98.66	81.25	55.83	9.86	90.86	7.56
11	104.76	92.74	109.20	20.47	96.74	80.36	71.33	17.85	93.09	8.31
12	103.99	101.69	81.17	14.54	95.52	81.44	63.50	25.85	94.09	9.30
13	112.79	95.71	96.45	4.70	96.68	84.98	73.40	7.65	91.53	7.04
14	104.51	106.38	77.28	21.95	97.38	86.43	66.25	19.18	93.35	8.52
15	103.28	93.75	114.62	9.36	94.34	78.43	75.39	7.54	86.70	7.34
16	111.25	90.06	93.21	2.53	90.08	79.95	46.72	6.46	94.53	6.46
17	109.55	95.17	86.39	9.49	95.99	76.56	57.21	6.42	87.07	6.58
18	107.83	100.86	72.99	7.48	96.67	85.52	55.34	5.96	93.76	7.15
19	108.45	89.18	101.67	10.40	98.39	79.11	62.64	7.54	91.13	7.19
20	107.08	86.49	93.45	18.22	97.05	75.38	67.31	15.05	95.56	8.63
21	112.15	88.64	87.65	17.07	97.79	77.48	59.19	14.14	94.45	7.44
22	103.62	96.92	109.41	11.76	98.39	86.38	59.07	8.17	96.61	7.95
23	108.13	99.31	78.44	19.25	98.03	77.74	55.25	14.63	94.62	8.01
24	104.40	100.03	77.34	4.45	96.20	74.61	54.79	5.21	94.26	7.15
25	106.67	89.86	105.77	17.93	94.56	73.88	61.16	8.34	93.53	7.25
26	111.68	91.59	87.57	14.34	95.85	78.82	53.44	12.57	93.25	6.63
27	110.16	97.02	61.34	11.54	96.34	77.44	45.74	8.45	93.44	6.88
28	108.36	87.37	75.14	10.27	98.22	76.80	48.82	6.44	92.26	6.41
29	104.14	86.13	81.00	24.71	98.69	77.46	51.80	20.68	93.28	6.80

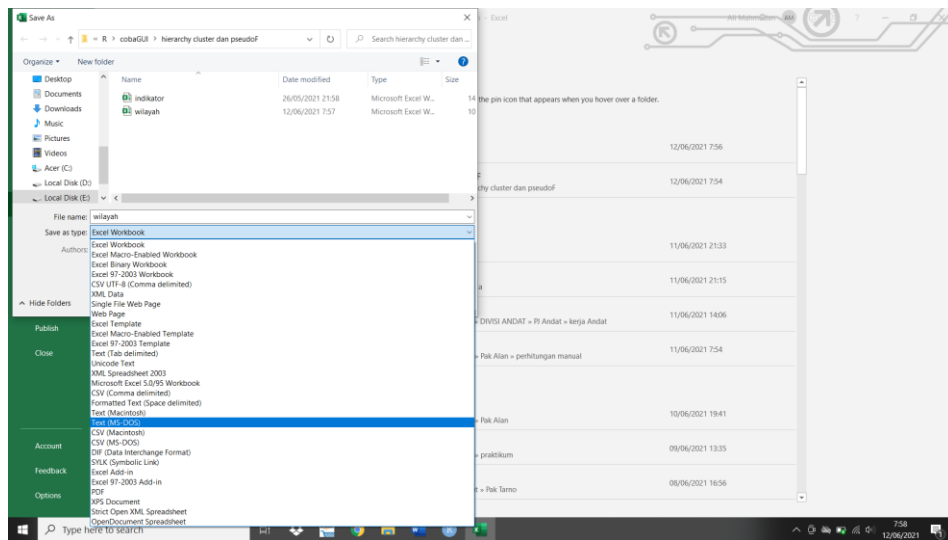
Variabel objek



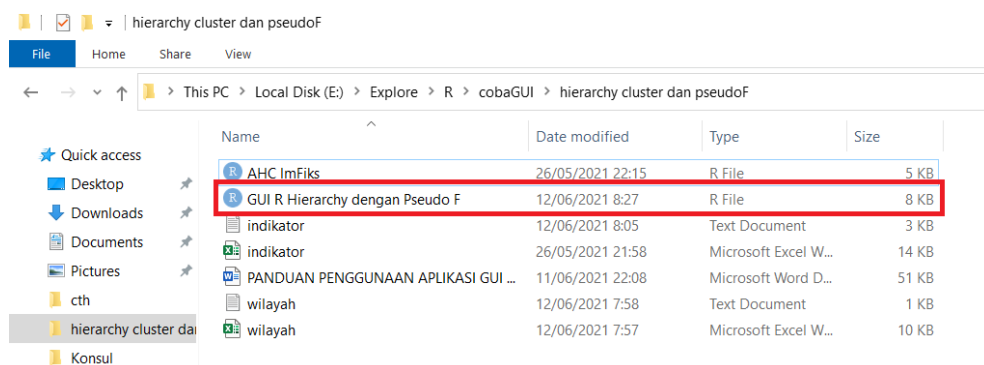
	A	B	C	D	E	F	G	H	I	J
1	KabupatenKuta									
2	KabupatenCibacop									
3	KabupatenBanamara									
4	KabupatenPurhalangan									
5	KabupatenBangoregara									
6	KabupatenKakamei									
7	KabupatenPurworejo									
8	KabupatenWonorejo									
9	KabupatenMagelang									
10	KabupatenBojale									
11	KabupatenKlaten									
12	KabupatenSubaharjo									
13	KabupatenWong									
14	KabupatenKaranganyar									
15	KabupatenKragan									
16	KabupatenGrobogan									
17	KabupatenBlora									
18	KabupatenPonoring									
19	KabupatenPati									
20	KabupatenKudus									
21	KabupatenJepara									
22	KabupatenDemak									
23	KabupatenSemarang									
24	KabupatenTernate									
25	KabupatenKendal									
26	KabupatenBatang									
27	KabupatenPekalongan									
28	KabupatenPembalang									
29	KabupatenTegal									

- Tanda koma dinotasikan dengan ‘.’ (titik).
- Nama tidak perlu menggunakan spasi, jika ingin menggunakan spasi, ganti dengan tanda ‘_’ (*underscore*).

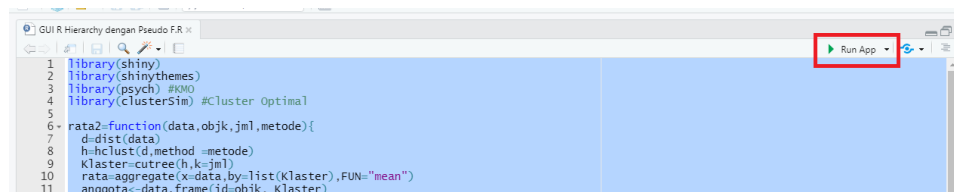
- e. Setelah semua data terinput maka langkah terakhir adalah menyimpan data dalam bentuk ‘.txt’



2. Buka aplikasi R-GUI lalu jalankan aplikasi tersebut.



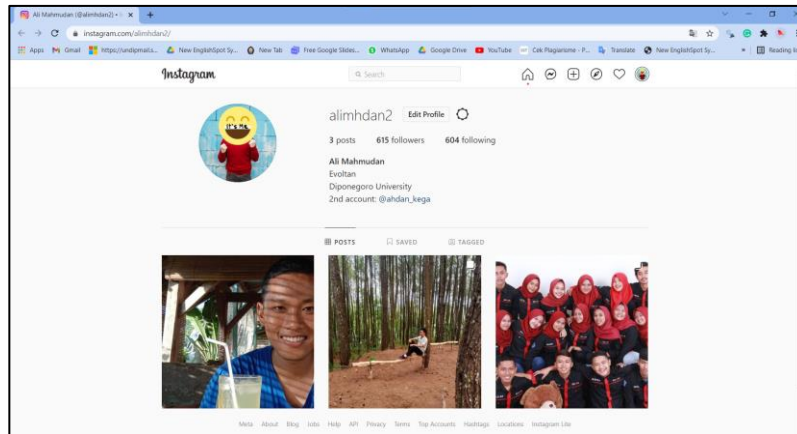
Jalankan aplikasi tersebut dengan cara blok semua *syntax* lalu klik ‘**Run App**’.



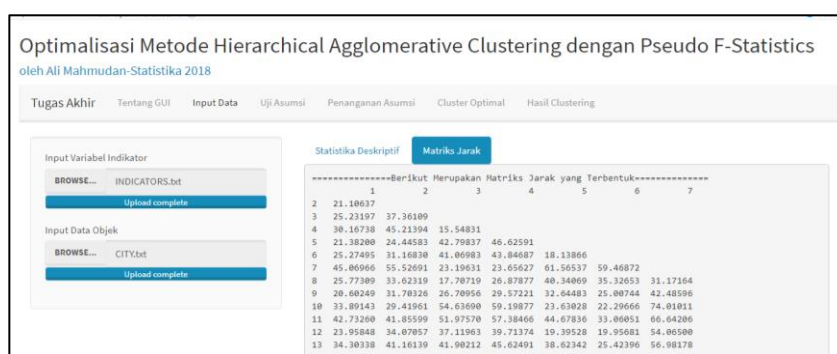
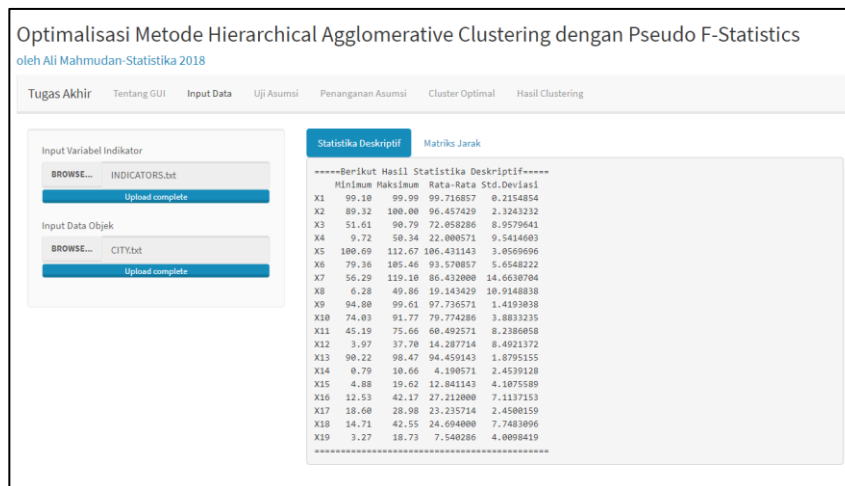
3. Berikut merupakan tampilan awal dari aplikasi R-GUI-nya.



Jika tulisan 'oleh: Ali Mahmudan-Statistika 2018' di klik maka akan menuju ke halaman biodata pembuat GUI (tampilan *profile* Instagram).



4. Lakukan *input* data dengan cara klik '**Browse**' pada panel '**Input Data**'. Setelah data terinput maka akan muncul hasil statistika deskriptif dan matriks jarak seperti pada ilustrasi di bawah ini:



5. Lakukan uji asumsi terhadap data klasterisasi yang digunakan dengan cara klik panel '**Uji Asumsi**', maka akan muncul hasil dari uji uji asumsi seperti ilustrasi di bawah ini:

Sanpel Representatif

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics
oleh Ali Mahmudan-Statistika 2018

Tugas AkhirTentang GUIInput DataUji AsumsiPenanganan AsumsiCluster OptimalHasil Clustering

Representativeness of the SampleNon Multikolinieritas

=====Berikut Merupakan Hasil dari Uji KMO=====
Asumsi Terpenuhi Bila Nilai KMO (Overall MSA) > 0,5

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = ffile)
Overall MSA = 0.57
MSA for each Item =
X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16
0.51 0.50 0.60 0.60 0.70 0.31 0.47 0.75 0.38 0.31 0.53 0.67 0.51 0.39 0.50 0.61
X17 X18 X19
0.16 0.72 0.60
=====

Interpretasi: berdasarkan hasil *output* di atas diperoleh nilai KMO sebesar 0,57 (pada bagian *Overall MSA*) > 0,5 maka asumsi sampel representatif telah terpenuhi.

Non Multikolinieritas

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics
oleh Ali Mahmudan-Statistika 2018

Tugas AkhirTentang GUIInput DataUji AsumsiPenanganan AsumsiCluster OptimalHasil Clustering

Representativeness of the SampleNon Multikolinieritas

=====Berikut Merupakan Hasil dari Variance Inflation Factor=====
Asumsi Terpenuhi Bila Nilai Variance Inflation Factor (VIF) < 10

Nilai Variance Inflation Factor (VIF):
[1] 1.758804 3.745503 19.751998 118.209543 2.604950 8.774381
[7] 49.402655 80.851562 4.860956 4.538150 32.130235 90.006153
[13] 7.404671 15.294684 18.292820 57.337928 10.833450 39.542777
[19] 51.277070
=====

Interpretasi: berdasarkan hasil *output* di atas diperoleh bahwa masih terdapat nilai *Variance Inflation Factor* (VIF) yang > 10 maka asumsi belum terpenuhi dan harus dilakukan penanganan.

6. Jika asumsi belum terpenuhi maka dapat dilakukan penanganan dengan cara klik panel ‘**Penanganan Asumsi**’ lalu akan muncul tampilan sebagai berikut:

Sampel Representatif

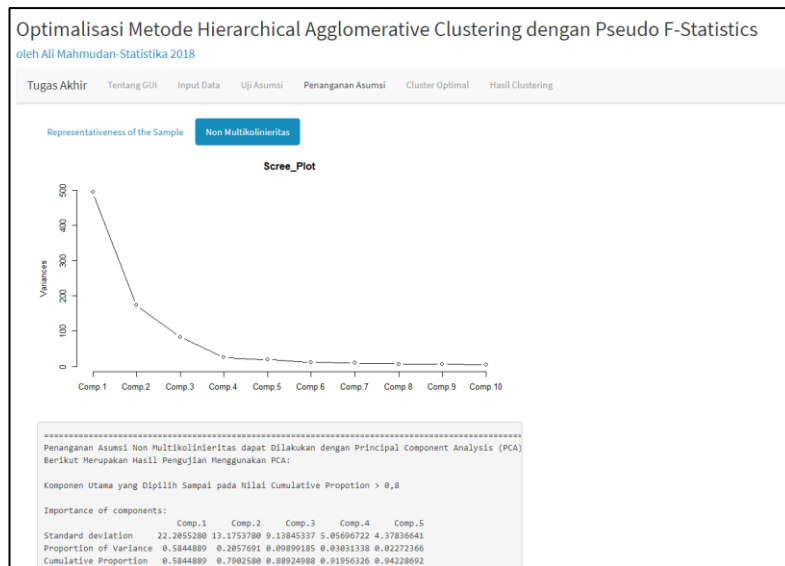
Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics
oleh Ali Mahmudan-Statistika 2018

Tugas AkhirTentang GUIInput DataUji AsumsiPenanganan AsumsiCluster OptimalHasil Clustering

Representativeness of the SampleNon Multikolinieritas

=====
Penanganan Asumsi Sampel Representatif Dapat Dilakukan dengan Manambah Variabel
=====

Non Multikolinieritas



Interpretasi: penanganan asumsi non multikolinieritas dalam kasus ini dilakukan dengan *Principal Component Analysis* (PCA). PCA akan memberikan beberapa komponen yang dapat mewakili seluruh variabel penelitian. Jumlah komponen yang dapat dipilih jika nilai *Cumulative Proportion* > 0,8.

7. Jika asumsi telah terpenuhi maka dapat dilakukan pengecekan jumlah *cluster* optimal dengan cara klik panel '**Cluster Optimal**', lalu pilih metode penentuan jarak terdekat yang diinginkan pada menu '**Metode**' (Metode yang dapat dipilih adalah *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid*, dan *Ward*), isikan jumlah *cluster* maksimal yang ingin dibentuk pada bagian '**Jumlah Maksimal Cluster**', isikan jumlah komponen utama pada bagian '**Jumlah Komponen Utama**' (isikan nilai 0 atau bilangan sembarang apabila tidak menggunakan komponen utama), lalu klik '**OLAH**'.

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics

oleh Ali Mahmudan-Statistika 2018

Tugas Akhir Tentang GUI Input Data Uji Asumsi Penanganan Asumsi **Cluster Optimal** Hasil Clustering

Pilih Metode dan Jumlah Cluster Maksimal
 Metode
 Single Linkage
 Complete Linkage
 Average Linkage
 Centroid
 Ward

Nilai Pseudo-F

OLAH

Pilih Metode dan Jumlah Cluster Maksimal
 Metode
 Single Linkage

Jumlah Maksimal Cluster
 6

Jumlah Komponen Utama
 3

OLAH

Setelah dilakukan pengolahan, maka akan muncul jumlah *cluster* optimal berdasarkan nilai *Pseudo F* tertinggi.

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics

oleh Ali Mahmudan-Statistika 2018

Tugas Akhir Tentang GUI Input Data Uji Asumsi Penanganan Asumsi **Cluster Optimal** Hasil Clustering

Pilih Metode dan Jumlah Cluster Maksimal
 Metode
 Single Linkage
 Jumlah Maksimal Cluster
 6
 Jumlah Komponen Utama
 3

Nilai Pseudo-F

```

=====
No  k-cluster  Pseudo-F
    [,1]      [,2]
[1,]  2 4.922955
[2,]  3 5.175288
[3,]  4 5.413654
[4,]  5 4.667430
[5,]  6 3.942868
=====
Nilai Pseudo-F tertinggi= 5.413654
jumlah cluster optimal yang terbentuk sebanyak 4 cluster
      
```

Nilai Pseudo F Komponen Utama

```

=====
No  k-cluster  Pseudo-F
    [,1]      [,2]
[1,]  2 16.862586
[2,]  3  9.156896
[3,]  4  9.307747
[4,]  5  7.225959
[5,]  6  6.737566
=====
Nilai Pseudo-F tertinggi= 16.86259
jumlah cluster optimal yang terbentuk sebanyak 2 cluster
      
```

Interpretasi: Diperoleh nilai *Pseudo F* tertinggi sebesar 5,1752 maka *cluster* optimal yang terbentuk berdasarkan metode penentuan jarak terdekat *Single Linkage* sejumlah 4 *cluster* (Pilih *output* bagian

bawah ‘Nilai *Pseudo F* Komponen Utama’ apabila menggunakan variabel komponen utama).

8. Lakukan *clustering* berdasarkan hasil jumlah *cluster* optimal yang telah terbentuk pada hasil ‘**Cluster Optimal**’ dengan cara klik panel ‘**Hasil Clustering**’, lalu pilih metode penentuan jarak terdekat yang diinginkan pada menu ‘**Metode**’ (Metode yang dapat dipilih adalah *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid*, dan *Ward*), isikan jumlah *cluster* maksimal yang ingin dibentuk pada bagian ‘**Jumlah Cluster Optimal**’, isikan jumlah komponen utama pada bagian ‘**Jumlah Komponen Utama**’ (isikan nilai 0 atau bilangan sembarang apabila tidak menggunakan komponen utama), lalu klik ‘**CLUSTERING**’.

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics
oleh Ali Mahmudan-Statistika 2018

Tugas Akhir Tentang GUI Input Data Uji Asumsi Penanganan Asumsi Cluster Optimal Hasil Clustering

Clustering sejumlah-k Cluster Optimal

Metode
Ward

Single Linkage
Complete Linkage
Average Linkage
Centroid
Ward

Hasil Clustering Hasil Clustering Komponen Utama

CLUSTERING

Clustering sejumlah-k Cluster Optimal

Metode
Ward

Jumlah Cluster Optimal
4

Jumlah Komponen Utama
3

CLUSTERING

Berikut merupakan hasil *clusterisasi* data yang telah terbentuk:

(Pilih tab panel Hasil Clustering Komponen Utama jika menggunakan variabel komponen utama)

Optimalisasi Metode Hierarchical Agglomerative Clustering dengan Pseudo F-Statistics

oleh Ali Mahmudan-Statistika 2018

Tugas Akhir Tentang GUI Input Data Uji Asumsi Penanganan Asumsi Cluster Optimal Hasil Clustering

Clustering sejumlah-k Cluster Optimal

Metode

Ward

Jumlah Cluster Optimal

4

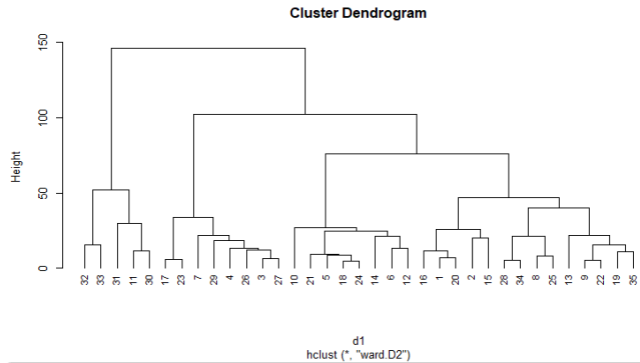
Jumlah Komponen Utama

3

CLUSTERING

Hasil Clustering

Hasil Clustering Komponen Utama



Hasil Clustering (Komponen Utama)

KabupatenKota Cluster

1	Kabupaten Cilacap	1
2	Kabupaten Banyumas	1
3	Kabupaten Purbalingga	2
4	Kabupaten Banjarnegara	2
5	Kabupaten Kebumen	3
6	Kabupaten Purworejo	3
7	Kabupaten Monosobo	2
8	Kabupaten Magelang	1
9	Kabupaten Boyolali	1
10	Kabupaten Klaten	3
11	Kabupaten Sukoharjo	4
12	Kabupaten Wonogiri	3
13	Kabupaten Karanganyar	1
14	Kabupaten Sragen	3
15	Kabupaten Grobogan	1
16	Kabupaten Blora	1
17	Kabupaten Rembang	2
18	Kabupaten Pati	3
19	Kabupaten Kudus	1
20	Kabupaten Jepara	1
21	Kabupaten Demak	3
22	Kabupaten Semarang	1
23	Kabupaten Temanggung	2
24	Kabupaten Kendal	3
25	Kabupaten Batang	1
26	Kabupaten Pekalongan	2
27	Kabupaten Pemalang	2
28	Kabupaten Tegal	1
29	Kabupaten Brebes	2
30	Kota Magelang	4
31	Kota Surakarta	4
32	Kota Salatiga	4
33	Kota Semarang	4
34	Kota Pekalongan	1
35	Kota Tegal	1

Berikut rata-rata tiap cluster-nya:

Group.1	X1	X2	X3	X4	X5	X6	X7
1	99.73000	96.92500	70.07143	21.23571	107.4243	92.27000	83.54786
2	99.76875	93.9875	63.03750	12.86500	106.1550	94.93125	69.97500
3	99.72250	97.2525	78.48250	21.35875	105.9900	94.89500	102.13000
4	99.58800	97.82800	81.77600	39.78600	104.79800	92.91800	95.72200
X8	X9	X10	X11	X12	X13	X14	X15
1	18.55214	98.18214	79.19000	58.39714	14.40643	94.29714	4.176429
2	10.39750	97.24250	79.12500	51.63250	7.55125	94.34875	4.397500
3	14.60625	97.10750	79.75375	68.93625	10.51750	93.83125	5.370000
4	41.90800	98.28600	82.48200	67.02600	30.76600	96.09400	2.012000
X16	X17	X18	X19				
1	27.92714	23.94571	24.21857	6.780			
2	33.72500	22.45750	17.34375	4.500			
3	27.40375	24.07000	23.85000	6.635			
4	14.48200	21.15800	39.13600	15.982			

Selesai