

PANDUAN PENGGUNAAN APLIKASI R-GUI:

**R-SHINY UNTUK PENGELOMPOKAN SUNGAI BERDASARKAN
PARAMETER KUALITAS AIR DENGAN ALGORITMA
*AGGLOMERATIVE HIERARCHICAL CLUSTERING***



TIM INVENTOR:

Ali Mahmudan

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

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

2021

Deskripsi Aplikasi

R-Shiny untuk Pengelompokan Sungai Berdasarkan Parameter Kualitas Air dengan Algoritma *Agglomerative Hierarchical Clustering* merupakan aplikasi GUI R yang disusun untuk membantu penelitian pengelompokan sungai parameter kualitas air menggunakan metode *Agglomerative Hierarchical Clustering*. 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 juga mampu melakukan pemilihan jumlah *cluster* optimal dengan menggunakan nilai *Pseudo F-Statistics*. Nilai *Pseudo F-Statistics* tersebut berasal dari perhitungan sebuah metode yang disebut sebagai metode *Calinski-Harabsz Pseudo F-Statistics*. Metode *Calinski-Harabsz 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 di-*cluster*-kan. 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 *clustering* dan data objek 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

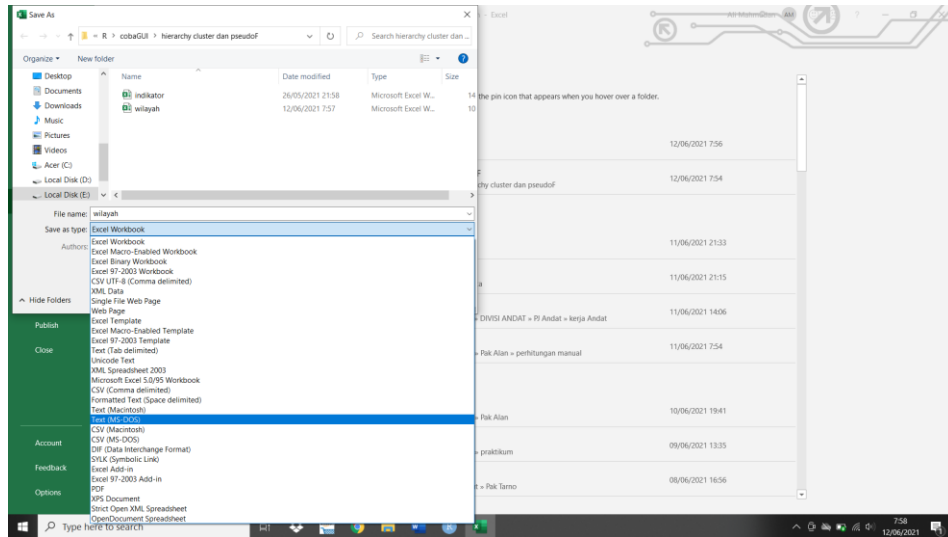
	Temperatur Udara	Temperatur Air	TDS	TSS	DHL	pH	BOD	COD	DO	Total Fosfat	NO ₃ -N	Amonia-NH ₄ -N
1	25.6	26.1	3228	13.1	495	7.51	3.5	15.1	4.1	0.21	3.1	0.21
2	42	31.7	300	43.3	582	7.78	15.2	71.6	3.1	1.25	1.9	0.49
3	32.8	28.7	189	11.5	358	8.22	2	7.6	6.8	0.05	1.9	0.11
4	28.7	27.5	127	148	265	8.11	3.4	14.1	6.2	0.31	2	0.06
5	28.3	24.5	120	73	228	8.32	2	8.8	6.8	0.04	2.1	0.04
6	32	30.6	325	21.5	617	7.14	46.4	211.6	1	0.07	3.6	0.52
7	35	31.2	149	20.5	275	7.93	2	6.8	6.8	0.04	1	0.04
8	33.5	30.7	195	11.8	373	7.75	2	6.5	6.5	0.04	1.2	0.05
9	35.9	28.9	147	136	271	7.74	5	20.2	5.5	0.29	2.8	0.11
10	35.2	28.9	148	130	289	7.7	4.8	27.8	5.4	0.29	2.9	0.23
11	34.4	30.5	295	29.5	563	8.04	2.1	12.2	6.5	0.26	3.2	0.86
12	31.4	26.1	275	5.2	507	8.06	2	4.9	6.2	0.3	2.6	0.08
13	33.2	29.8	280	23	544	7.81	2	14.4	6.2	0.17	2.4	0.08
14	35	26	444	2.5	629	7.81	6.6	19.5	3.1	1.13	0.9	3.76
15	29.2	24.8	135	195	257	7.86	2	21	7.1	0.2	3	0.07
16	35.9	28.2	193	74.3	364	8.29	2.1	18.9	7.1	0.25	2	0.16
17	38.1	30.5	265	39.9	54.2	8.26	2	13.8	7	0.16	2.1	0.1
18	32.3	30.5	231	127	437	7.89	4.5	39.6	4.9	0.28	2.4	0.67
19	29.1	24.6	53.8	4	99.7	8.3	2	3.5	7.8	0.04	0.6	0.03
20	32.1	26.7	52.2	7.8	99.5	8.3	2	4	7.6	0.04	0.6	0.04
21	32.3	27.6	38.9	147.4	73.8	8.22	2	4.2	7.5	0.06	0.9	0.03
22	40.5	26.7	45.5	45.5	86.8	7.56	2	4.6	5.5	1	1.2	0.04
23	30.9	23.4	86.1	54	163.4	7.53	4.8	12.3	7.8	0.3	1.4	0.18
24	34.8	30.6	453	29.1	855	7.23	3.2	13.7	2.7	0.1	1.1	0.6
25	26.5	26.2	460	31.9	848	8.17	2	11.7	6	0.06	3.4	0.06
26	26.3	26.5	333	77.2	626	7.96	2.3	16.8	4.9	0.09	2.2	0.15
27	28	27.5	312	37.9	584	7.82	3.2	13.5	4.8	0.11	2.2	0.3

Varibel objek

Kode
PL1
PL2
UK1
UK2
SM1
SM2
UL1
UL2
GP1
GP2
UK1
UK2
PM1
PM2
PP1
PP2
PP3
PP4
KP1
KP2
KP3
KP4
KP5
KP6
LS1
LS2
LS3
LS4
LSS

- Tanda koma dinotasikan dengan ‘.’ (titik).

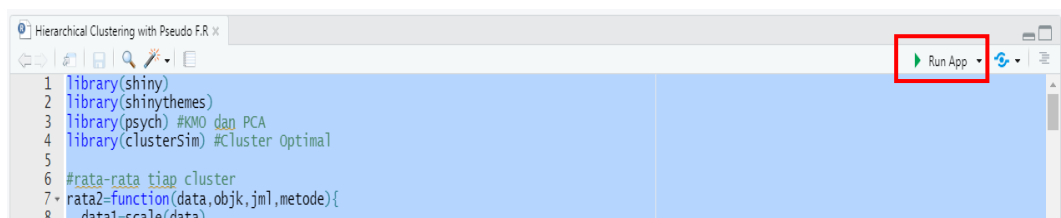
- d. Nama tidak perlu menggunakan spasi, jika ingin menggunakan spasi, ganti dengan tanda ‘_’ (*underscore*).
- e. Setelah semua data terinput maka langkah terakhir adalah menyimpan data dalam bentuk ‘.txt’



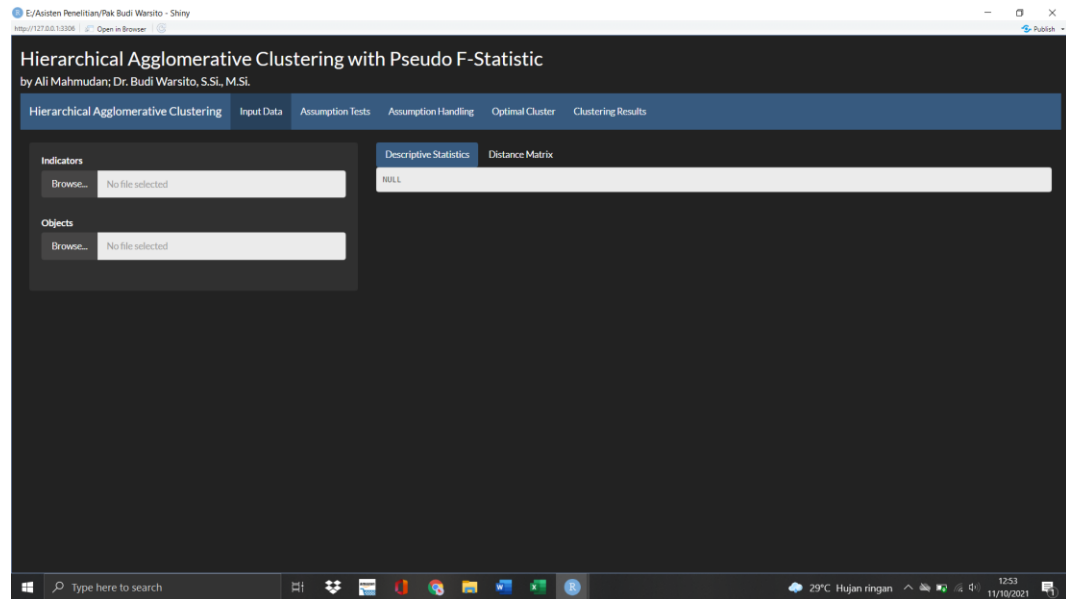
2. Buka aplikasi R-GUI lalu jalankan aplikasi tersebut.

Name	Date modified	Type	Size
pengajuan HKI	11/10/2021 12:44	File folder	
.RData	15/09/2021 14:55	R Workspace	3 KB
.Rhistory	09/10/2021 17:40	RHISTORY File	1 KB
Deskripsi GUI R HAC	15/09/2021 14:55	Microsoft Word D...	663 KB
Hierarchical Clustering with Pseudo F	12/09/2021 17:46	R File	15 KB
indicator	14/09/2021 19:37	Text Document	5 KB

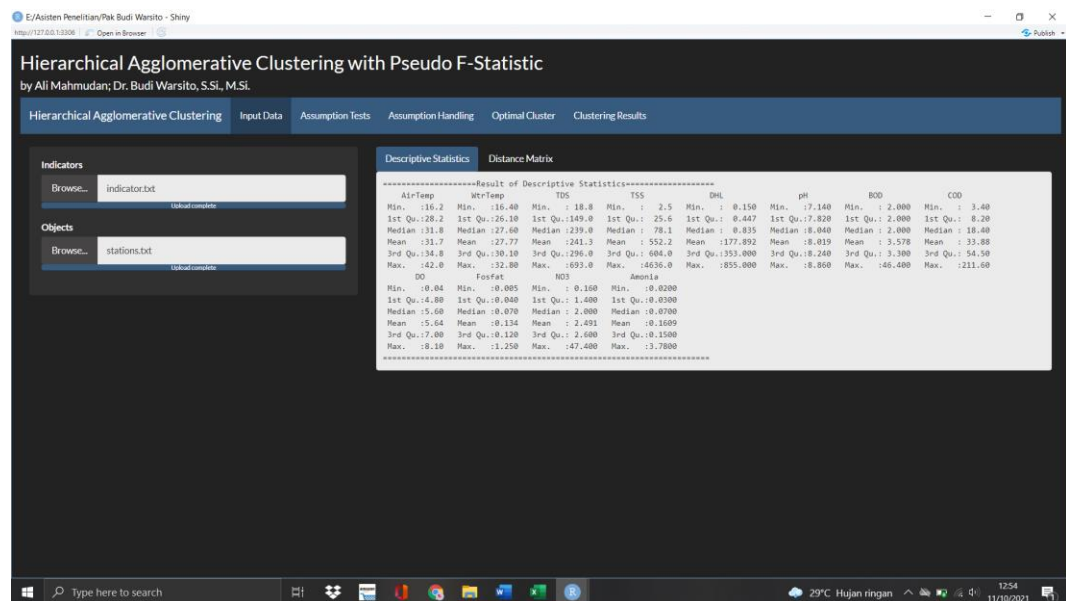
3. Jalankan aplikasi tersebut dengan cara blok semua *syntax* lalu klik ‘**Run App**’.

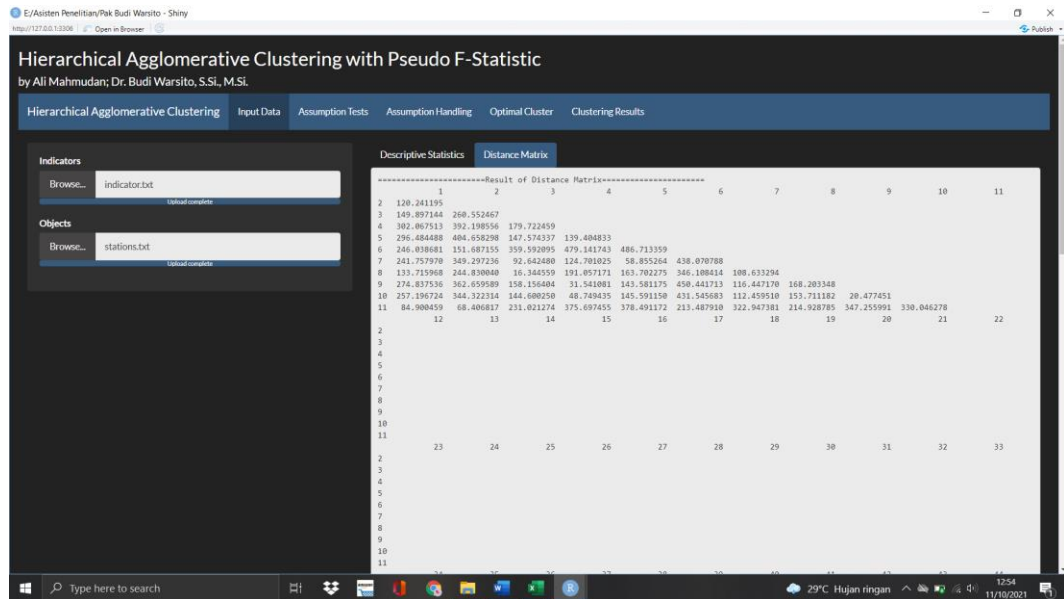


4. Berikut merupakan tampilan awal dari aplikasi R-GUI-nya.



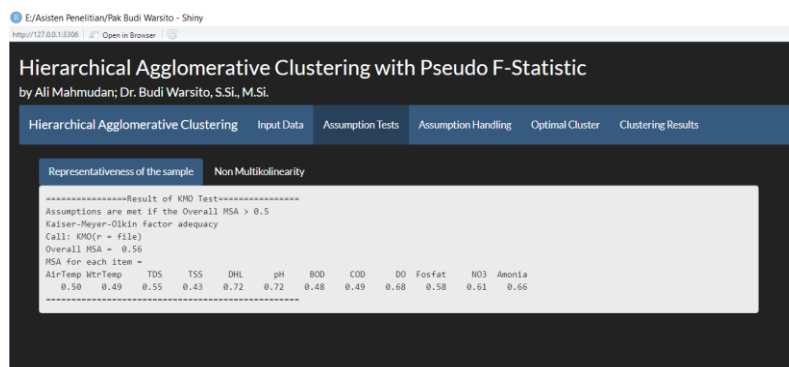
5. Lakukan *input* data dengan cara klik '**Browse**' pada panel '**Input Data**'. Setelah data terinput maka akan muncul hasil statistika deskriptif dan matriks jarak seperti pada ilustrasi di bawah ini:





6. Lakukan uji asumsi terhadap data yang digunakan dengan cara klik panel ‘**Assumption Tests**’, maka akan muncul hasil dari uji uji asumsi seperti ilustrasi di bawah ini:

Sampel Representatif



Interpretasi: berdasarkan hasil *output* di atas diperoleh nilai KMO sebesar 0,56 (pada bagian *Overall MSA*) > 0,5 maka asumsi sampel representatif telah terpenuhi.

Non Multikolinieritas



Interpretasi: berdasarkan hasil *output* di atas diperoleh bahwa nilai $VIF < 10$ pada tiap variabel indikator sehingga asumsi telah terpenuhi.

7. Jika asumsi belum terpenuhi maka dapat dilakukan penanganan dengan cara klik panel ‘**Assumption Handling**’ lalu akan muncul tampilan sebagai berikut:

Sampel Representatif

E:/Asisten Penelitian/Pak Budi Warsito - Shiny
http://127.0.0.1:3306 Open in Browser

Hierarchical Agglomerative Clustering with Pseudo F-Statistic

by Ali Mahmudan; Dr. Budi Warsito, S.Si., M.Si.

Hierarchical Agglomerative Clustering Input Data Assumption Tests Assumption Handling Optimal Cluster Clustering Results

Representativeness of the sample Non Multikolinearity

Handling Representative Sample Assumptions can be Done by Adding or Subtracting Variables

Non Multikolinieritas

Hierarchical Agglomerative Clustering with Pseudo F-Statistic

by Ali Mahmudan; Dr. Budi Warsito, S.Si., M.Si.

Hierarchical Agglomerative Clustering Input Data Assumption Tests Assumption Handling Optimal Cluster Clustering Results

Representativeness of the sample Non Multikolinearity

Handling Non-Multicollinearity Assumptions can be Done with Principal Component Analysis (PCA)
Here are the test results using PCA
Selected Main Components Up to Cumulative Proportion Value ≥ 0.7
Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	937.8541347	244.61214413	123.12765360	2.909633e+01	5.716519e+00	4.286984e+00	3.798600e+00	1.574040e+00
Proportion of Variance	0.9205514	0.06262294	0.01586677	8.860411e-04	3.420115e-05	1.923451e-05	1.510165e-05	2.593038e-06
Cumulative Proportion	0.9205514	0.98317430	0.99904107	9.999271e-01	9.999613e-01	9.999805e-01	9.999956e-01	9.999982e-01

	Comp.9	Comp.10	Comp.11	Comp.12
Standard deviation	1.217452e+00	3.360555e-01	2.505204e-01	1.473771e-01
Proportion of Variance	1.551247e-06	1.181951e-07	6.568458e-08	2.273198e-08
Cumulative Proportion	9.999998e-01	9.999999e-01	1.000000e+00	1.000000e+00

pca

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* $> 0,7$.

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

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

Interpretasi: Diperoleh nilai *Pseudo F* tertinggi sebesar 14,62264 maka *cluster* optimal yang terbentuk berdasarkan metode penentuan jarak terdekat *Average Linkage* sejumlah 5 *cluster* (Pilih *output* bagian bawah ‘Pseudo F of Principal Components’ apabila menggunakan variabel komponen utama).

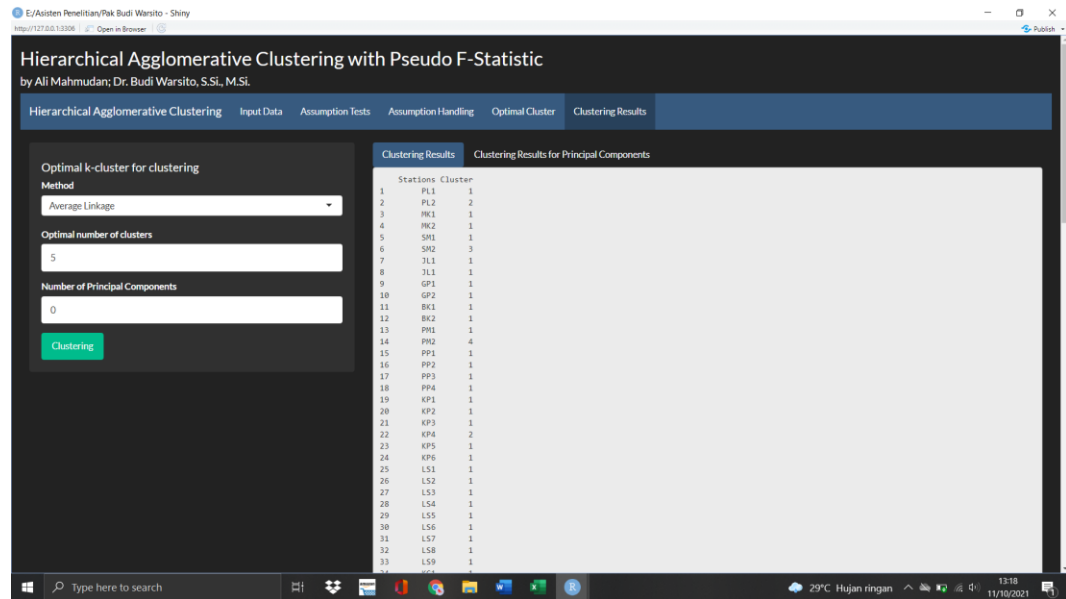
9. Lakukan *clusterisasi* berdasarkan hasil jumlah *cluster* optimal yang telah terbentuk pada hasil ‘**Optimal Cluster**’ dengan cara klik panel ‘**Clustering Results**’, lalu pilih metode penentuan jarak terdekat yang diinginkan pada menu ‘**Method**’ (Metode yang dapat dipilih adalah *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid*, dan *Ward*), isikan jumlah *cluster* maksimal yang ingin dibentuk pada bagian ‘**Maximum Number of Cluster**’, isikan jumlah komponen utama pada bagian ‘**Number of Principal Component**’ (isikan nilai 0 atau bilangan sembarang apabila tidak menggunakan komponen utama), lalu klik ‘**Clustering**’.

The screenshot shows a web application titled "Hierarchical Agglomerative Clustering with Pseudo F-Statistic" by Ali Mahmudan; Dr. Budi Warsito, S.Si., M.Si. The interface has a dark theme with a top navigation bar containing tabs: "Hierarchical Agglomerative Clustering", "Input Data", "Assumption Tests", "Assumption Handling", "Optimal Cluster", and "Clustering Results". The "Clustering Results" tab is active. Below the navigation bar, there are two sub-tabs: "Clustering Results" and "Clustering Results for Principal Components". The "Clustering Results" sub-tab is active. On the left, there is a form titled "Optimal k-cluster for clustering" with the following fields: "Method" (a dropdown menu showing "Average Linkage"), "Optimal number of clusters" (a text input field containing "5"), and "Number of Principal Components" (a text input field containing "0"). Below these fields is a green button labeled "Clustering".

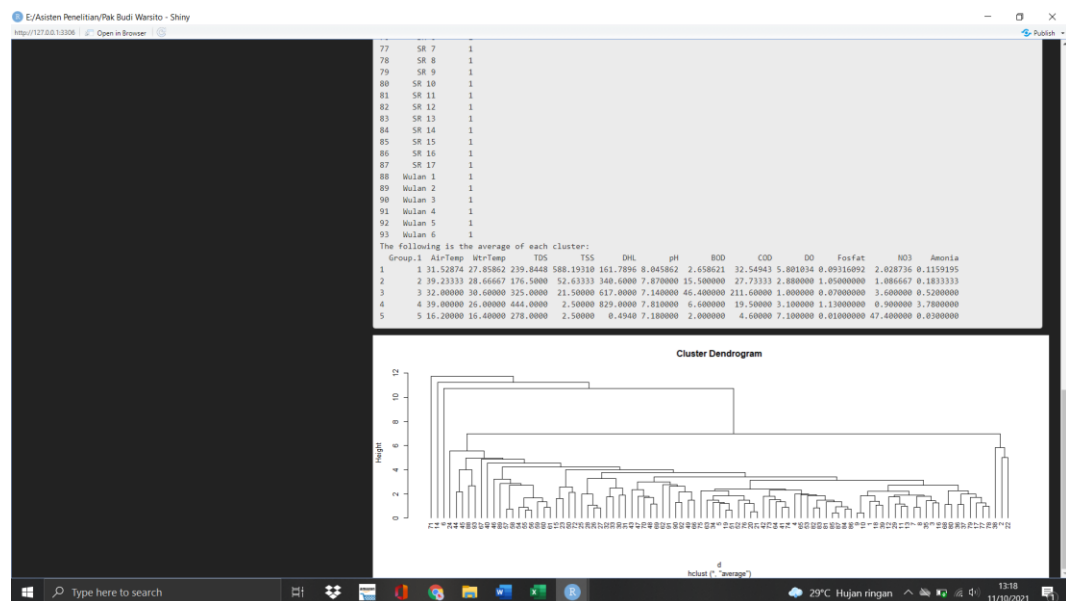
Berikut merupakan hasil *clusterisasi* data yang telah terbentuk:

(Pilih tab panel Clusering Resultsts for Principal Components jika menggunakan variabel komponen utama)

Anggota Tiap Cluster



Rata-Rata Indikator Tiap Cluster dan Dendrogram



Selesai