

Analisis Cluster Non-Hirarki Berbasis GUI R



Disusun Oleh :

Risma Ashali Fauziah

24050119130047

Kelas C

Dosen Pengampu : Dr. Budi Warsito, S.Si, M.Si

I. PENDAHULUAN

1.1 Analisis Cluster

Analisis cluster adalah analisis yang digunakan untuk mengelompokkan objek atau variable ke dalam beberapa kelompok tertentu dimana setiap objek atau variable yang terbentuk memiliki karakteristik yang sama. Ciri-ciri suatu cluster yang baik yaitu memiliki :

1. Homogenitas internal (within cluster) yaitu kesamaan antar anggota dalam suatu cluster.
2. Heterogenitas external (between cluster) yaitu perbedaan antara cluster yang satu dengan yang lain.

1.2 K-Means Cluster

K-means merupakan salah satu metode data clustering yang akan mempartisi data ke dalam bentuk satu atau lebih kelompok. Dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan. Setelah jumlah cluster ditentukan, maka proses cluster dilakukan dengan tanpa mengikuti proses hirarki. Metode ini biasa disebut “K-Means cluster”. K-mean cluster sangat efektif dan efisien jika digunakan untuk mengelompokkan objek yang berjumlah besar.

II. GUI R DENGAN SHINY

- Packages yang digunakan

```
library(shiny)
library(shinythemes)
```

- Syntax UI

```
#membuat UI
myui<-fluidPage(
  theme = shinytheme("united"),
  titlePanel(h1("Analisis Cluster Non Hirarki (K-Means Cluster)")),
  titlePanel(h4(strong("by: Risma Ashali Fauziah"))),
  navbarPage("",
    tabPanel("Data", icon = icon("table"),
      sidebarLayout(
```

```

        sidebarPanel(
            fileInput("data", "Input file di
sini:", accept = ".txt")
        ),
        mainPanel(br(strong("Berikut
merupakan data yang digunakan untuk analisis
cluster")), br(), dataTableOutput("tabel"), h5(strong("Tabel
Statistika
Deskriptif")), verbatimTextOutput("cluster"), br()
    )
)),
    tabPanel("Jarak",
        mainPanel(br("Di bawah ini adalah
visualisasi dan nilai dari jarak antar
data"), h5("Keterangan:"), h5("Hijau = Dekat ; Putih = Sedang
;
Merah =
Jauh"), br(), plotOutput("plot1"), br(), verbatimTextOutput("
cluster1")
    )
),
    tabPanel("Cluster Optimal", icon = icon("bar-
chart-o"),
        sidebarLayout(
            sidebarPanel(
                radioButtons("Metode", "Pilih
Metode:", choices =
c("Elbow" = "wss", "Silhouette" = "silhouette")),
            ),
            mainPanel(plotOutput("plot2"))
        )),
    tabPanel("K-Means Cluster", icon = icon("list-
alt"),
        sidebarLayout(
            sidebarPanel(

```

```

                                textInput("Jml", "Masukkan Jumlah
Cluster"),

                                =
                                icon("redo"), class="btn-success"),
                                ),

                                mainPanel(plotOutput("plot3"), br(), verbatimTextOutput("ce
                                kklaster"), h5(strong("Total
                                Withinss:")), verbatimTextOutput("tw"), h5(strong("Between
                                s:")), verbatimTextOutput("b"), h5(strong("Note:")), h5("Clu
                                ster dinyatakan baik apabila nilai total withinss kecil
                                dan nilai beetweenss besar"), br())
                                ))
                                )
                                )

```

- Syntax Server

```

#membuat server
myserver<-function(input,output,session){
  output$tabel<-renderDataTable({
    datainput=input$data
    if (is.null(datainput)){return()}
    datapakai=read.delim(datainput$datapath,header = T)
    datapakai
  })
  output$cluster<-renderPrint({
    datainput=input$data
    if (is.null(datainput)){return()}
    datapakai=read.delim(datainput$datapath,header = T)
    summary(datapakai)
  })
  output$plot1<-renderPlot({
    datainput=input$data

```

```

    if (is.null(datainput)){return()}
    library(factoextra)
    library(gridExtra)
    datapakai=read.delim(datainput$datapath,header = T)
    distance<-get_dist(datapakai)
    plotfix<-fviz_dist(distance, gradient = list(low =
"green", mid = "white", high = "red"))
    cat("Berikut merupakan jarak korelasi antar data:\n")
    plot(plotfix)
  })
  output$cluster1<-renderPrint({
    datainput=input$data
    if (is.null(datainput)){return()}
    datapakai=read.delim(datainput$datapath,header = T)
    distance<-get_dist(datapakai)
    distance
  })
  output$plot2<-renderPlot({
    datainput=input$data
    if (is.null(datainput)){return()}
    datapakai=read.delim(datainput$datapath,header = T)
    datafix=datapakai[,2:4]
    plotfix2<-fviz_nbclust(datafix, kmeans, method =
input$Metode, print.summary = F)
    plot(plotfix2)
  })
  observeEvent(input$cekbutton,{
    datainput=input$data
    if (is.null(datainput)){return()}
    datapakai=read.delim(datainput$datapath,header = T)
    datafix=datapakai[,2:4]
    jumlah<-input$Jml
    jumlah=as.numeric(jumlah)
    klaster <- kmeans(datafix,centers=jumlah, nstart = 25)

```

```

    TW=klaster$tot.withinss
    B=klaster$betweenss
    p <- fviz_cluster(klaster, geom = "point", data =
datafix)

    output$plot3<-renderPlot(p)
    output$cekkklaster<-renderPrint(klaster)
    output$tw<-renderPrint(TW)
    output$b<-renderPrint(B)
  })
}

```

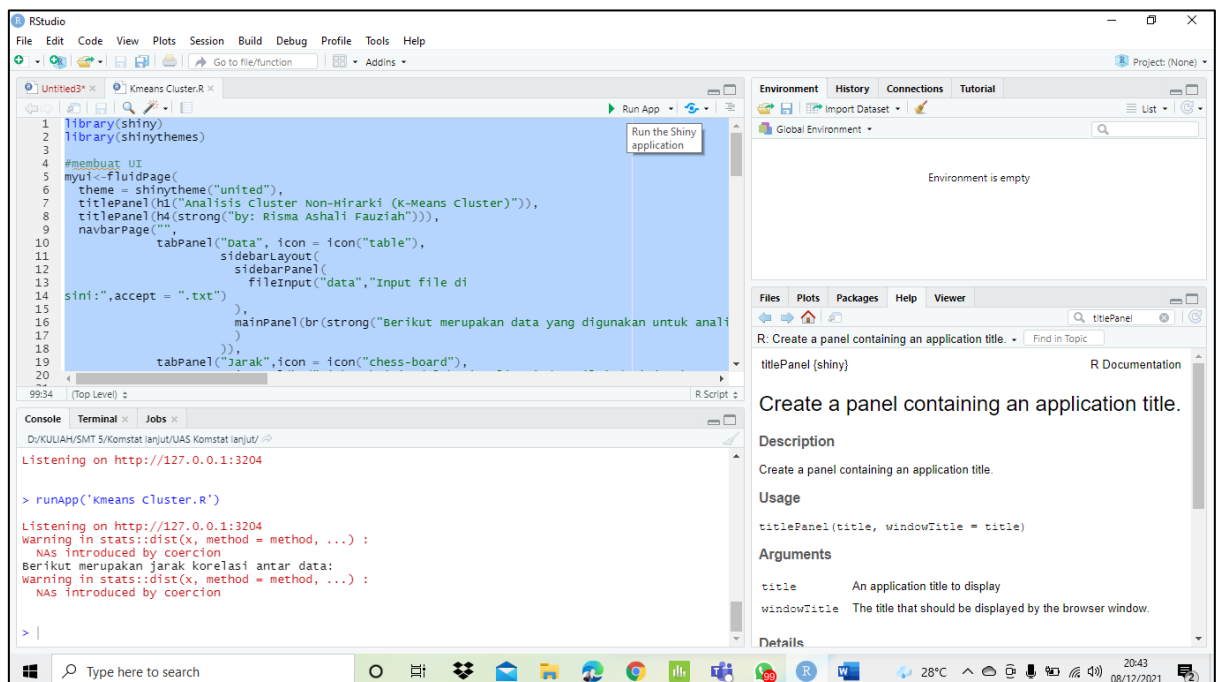
- Jalankan GUI

```
shinyApp(ui=myui,server=myserver)
```

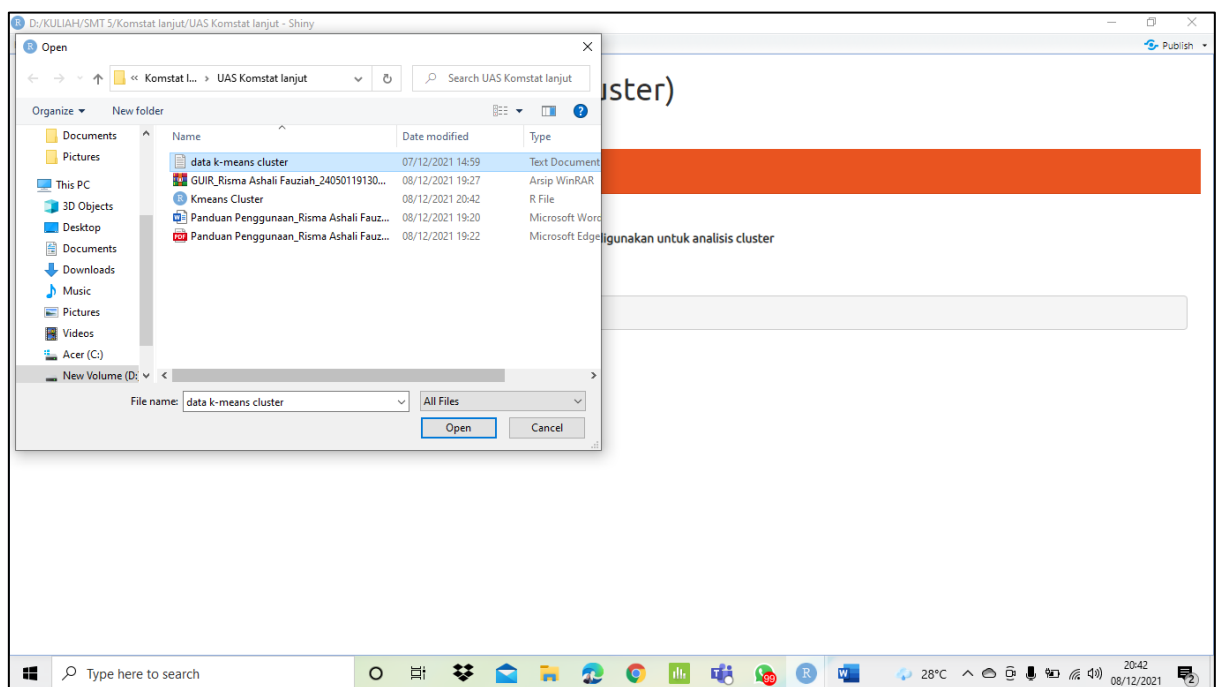
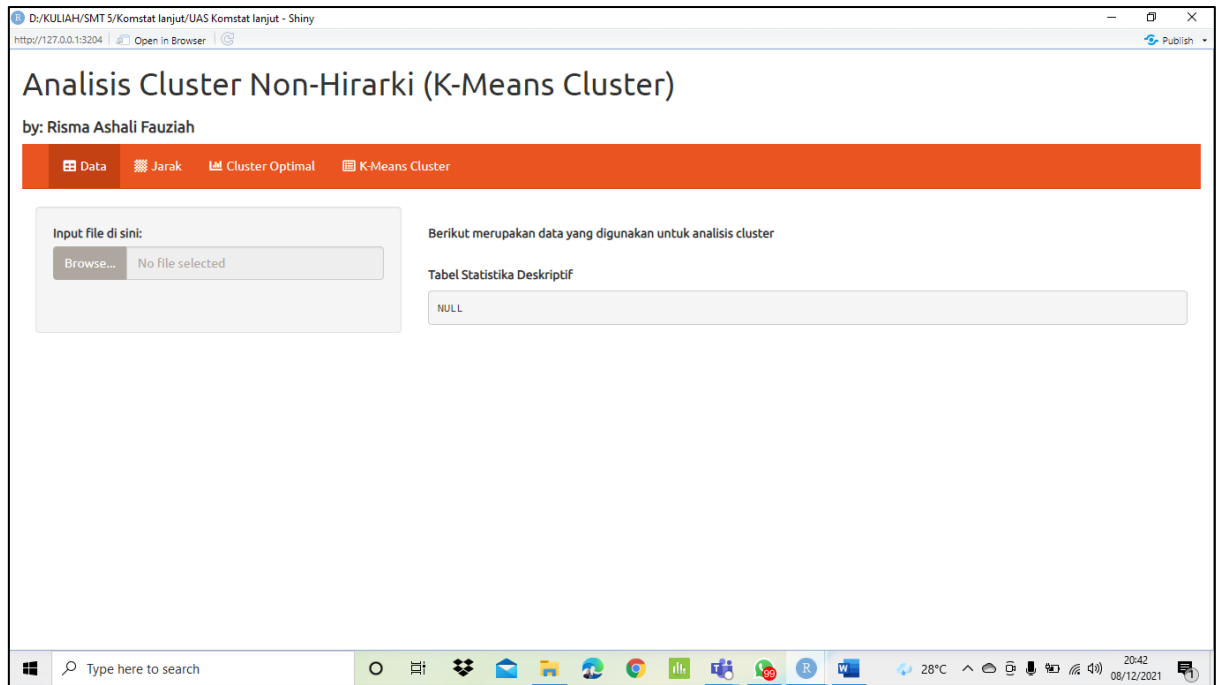
III. PANDUAN PENGGUNAAN GUI R

Panduan ini digunakan untuk menjalankan GUI Analisis Cluster Non-Hirarki (K-Means Cluster).

1. Buka file.R (script) yang memuat syntax dari GUI yang akan digunakan
2. Blok syntax kemudian klik **Run App**



3. Input data dengan cara klik **Browse** kemudian pilih data dengan format .txt yang akan digunakan.



4. Pada tab **Data** akan muncul tampilan data dan tabel statistika deskriptif.

D:/KULIAH/SMT 5/Komstat lanjut/UAS Komstat lanjut - Shiny
http://127.0.0.1:3204 | Open in Browser | Publish

Analisis Cluster Non-Hirarki (K-Means Cluster)

by: Risma Ashali Fauziah

Data Jarak Cluster Optimal K-Means Cluster

Input file di sini:

Browse... data k-means cluster.txt
Upload complete

Berikut merupakan data yang digunakan untuk analisis cluster

Show 25 entries Search:

Kabupaten.Kota	PDRB	TPT	JPM
Kepulauan Selayar	2.5	2.54	17.30
Bulukumba	6.4	3.32	34.70
Bantaeng	3.8	4.30	18.00
Jeneponto	4.8	3.16	55.30
Takalar	4.5	3.16	27.90
Gowa	9.7	3.30	48.98
Sinjai	5.1	0.96	22.90
Maros	10.2	5.77	41.10
Pangkajene dan Kepulauan	12.4	7.54	54.20
Barro	3.5	4.83	16.70

Type here to search 28°C 20:42 08/12/2021

D:/KULIAH/SMT 5/Komstat lanjut/UAS Komstat lanjut - Shiny
http://127.0.0.1:3204 | Open in Browser | Publish

Pinrang	9.0	3.20	30.80
Enrekang	3.4	1.45	28.30
Luwu	6.9	6.70	49.70
Tana Toraja	3.2	3.52	29.60
Luwu Utara	5.7	3.08	43.70
Luwu Timur	13.7	6.58	20.80
Toraja Utara	3.5	3.21	35.00
Kota Makassar	82.7	10.82	64.60
Kota Pare-Pare	3.6	6.81	8.30
Kota Palopo	3.9	9.73	14.80

Kabupaten.Kota PDRB TPT JPM

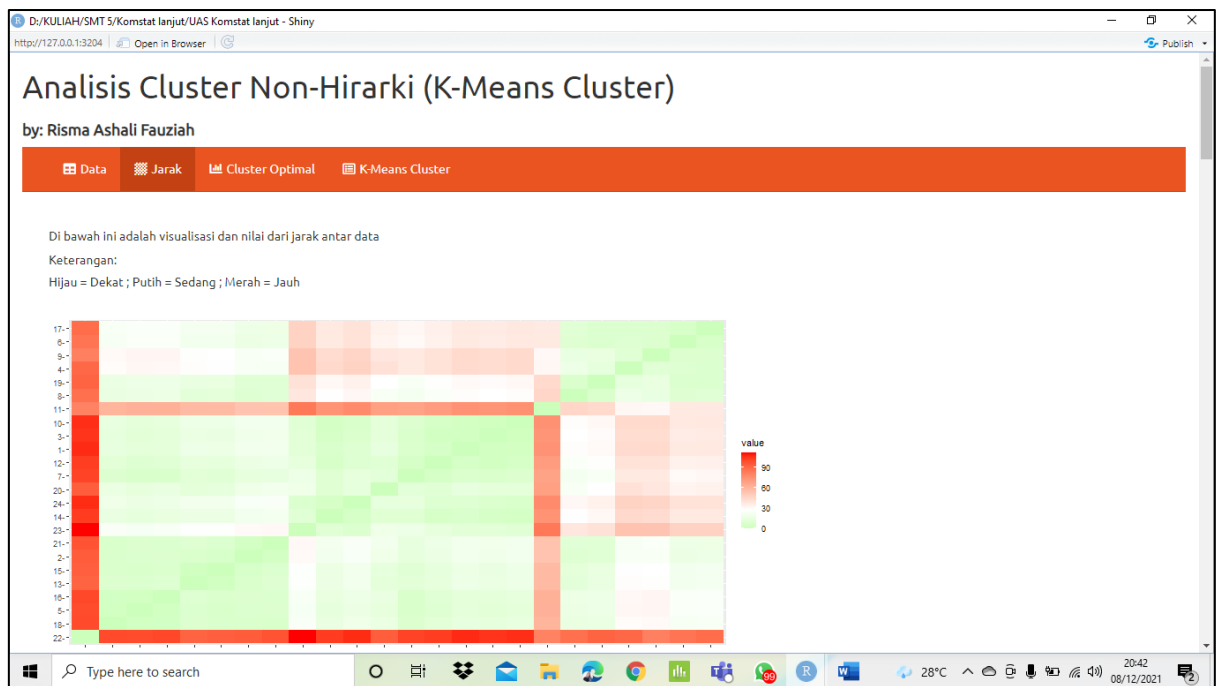
Showing 1 to 24 of 24 entries Previous 1 Next

Tabel Statistika Deskriptif

Kabupaten.Kota	PDRB	TPT	JPM
Length:24	Min. : 2.500	Min. : 0.960	Min. : 8.30
Class :character	1st Qu.: 3.750	1st Qu.: 3.190	1st Qu.:19.43
Mode :character	Median : 5.400	Median : 4.135	Median :30.15
	Mean : 9.775	Mean : 4.747	Mean :33.80
	3rd Qu.: 9.825	3rd Qu.: 6.610	3rd Qu.:45.02
	Max. :82.700	Max. :10.820	Max. :81.00

Type here to search 28°C 20:52 08/12/2021

- Klik tab **Jarak** untuk melihat hasil jarak (*distance*) antara data. Pada bagian ini akan ditampilkan visualisasi jarak beserta nilai dari jarak antar data.

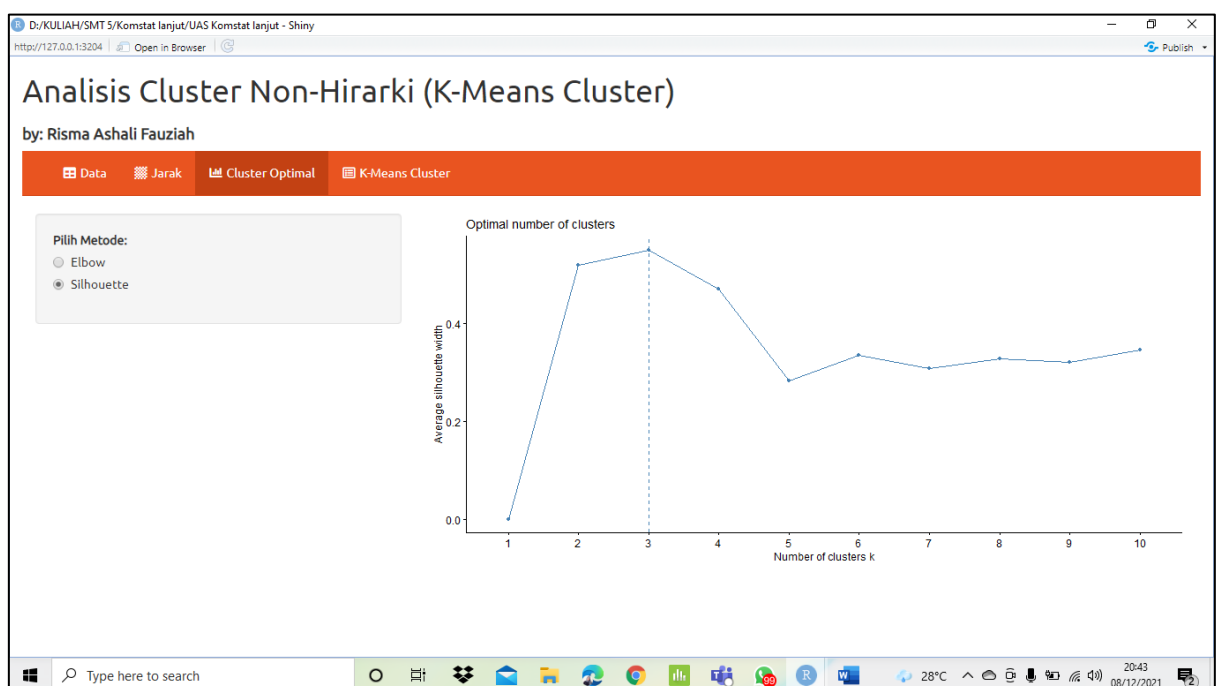
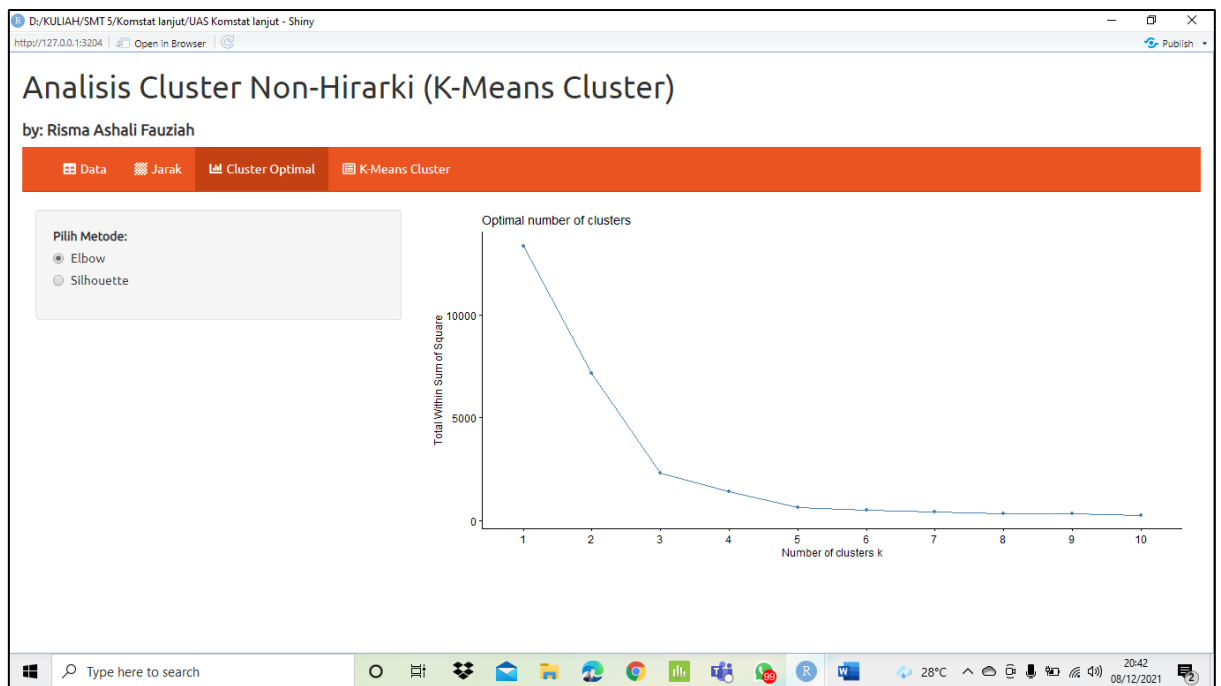


D:/KULIAH/SMT 5/Komstat lanjut/UAS Komstat lanjut - Shiny

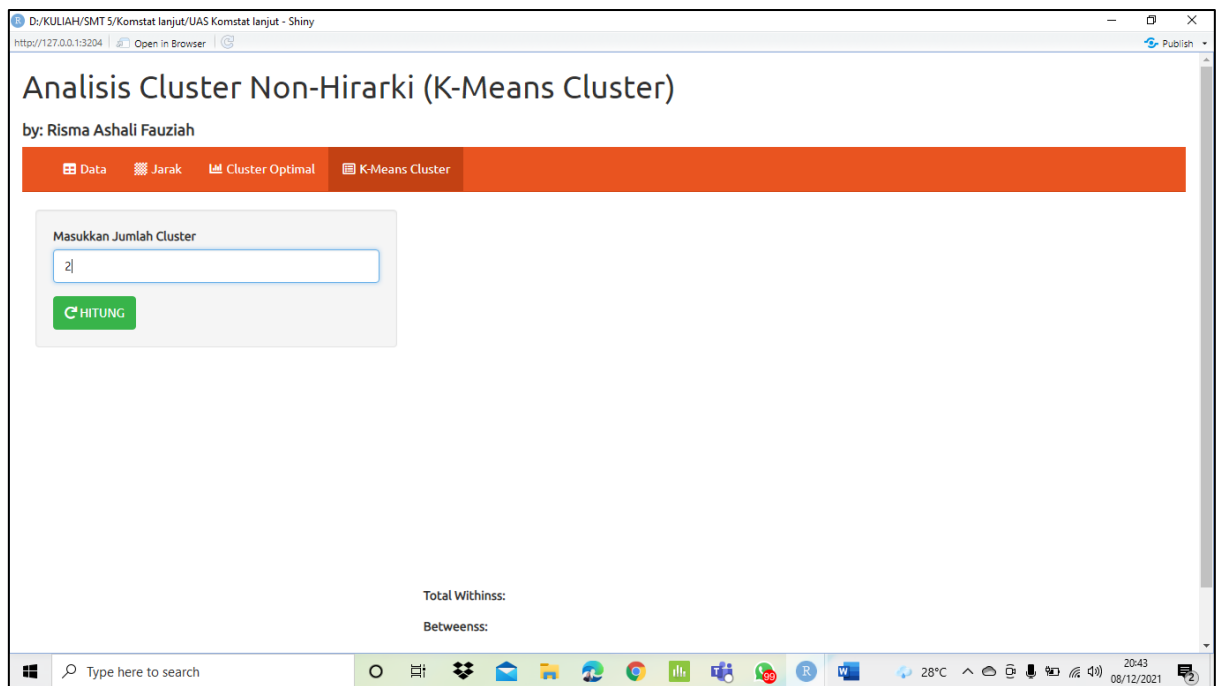
http://127.0.0.1:3204 | Open in Browser

	1	2	3
2	20.609978		
3	2.652697	19.548586	
4	43.964750	23.859187	43.105910
5	12.476346	8.154802	11.535429
6	37.524037	16.923703	36.433874
7	7.359022	13.976151	7.010050
8	29.124169	9.048204	27.730390
9	44.491497	24.057110	43.126134
10	2.967401	21.124712	1.657669
11	74.943557	54.349424	73.846761
12	4.406798	17.193507	2.563565
13	18.071595	6.636515	16.479357
14	6.580993	21.088136	4.265661
15	17.318029	5.414105	16.003750
16	12.806150	8.442504	12.348954
17	38.060051	17.764174	36.882878
18	14.270735	6.956053	13.442639
19	30.713550	10.427374	29.790119
20	14.329997	18.515853	12.168177
21	20.485406	3.368897	19.673268
22	107.937565	95.022453	106.077207
23	11.572548	30.918821	11.571811
24	8.937270	24.313182	7.278727
	4	5	6
2			
3			
4			
5	31.640691		
6	9.235583	25.071258	
7	37.500044	6.345602	30.698578
8	17.799330	16.873731	9.553031
9	10.208127	32.109903	8.367875
10	44.638383	13.126508	37.996128
11	31.874816	62.473599	37.460708
12	40.887261	9.296315	34.041843

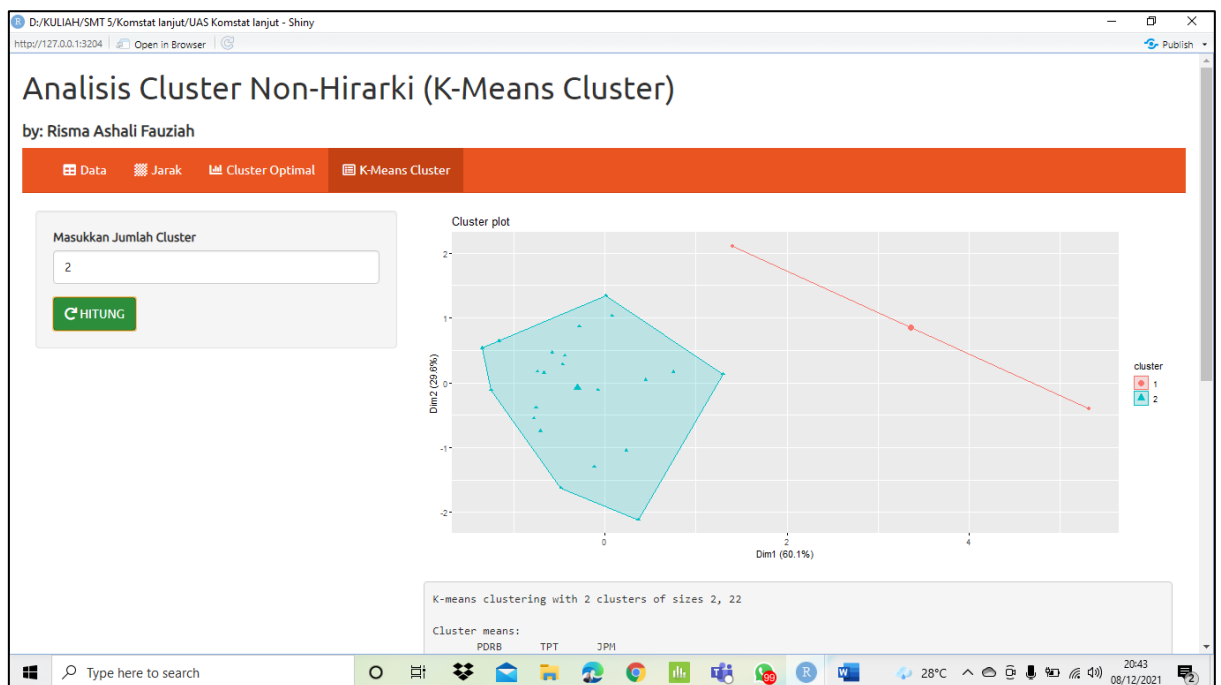
- Klik tab **Cluster Optimal** untuk melihat berapa cluster yang optimal yaitu dengan metode Elbow dan Silhouette.
- Pilih button **Elbow** atau **Silhouette** untuk melihat grafik cluster optimal. Apabila terdapat perbedaan pada kedua metode tersebut dilakukan perbandingan.



8. Klik tab **K-Means Cluster** untuk menganalisis, isi dengan jumlah cluster yang ingin dicek pada bagian **Masukkan Jumlah Cluster**



9. Kemudian klik **HITUNG**. Pada bagian ini akan ditampilkan hasil visualisasi clustering, k-means, total withinss, dan betweenss.



D:/KULIAH/SMT 5/Komstat lanjut/UAS Komstat lanjut - Shiny
http://127.0.0.1:3204 | Open in Browser | Publish

Dim1 (60.1%)

```
K-means clustering with 2 clusters of sizes 2, 22

Cluster means:
      PDRB      TPT      JPH
1 48.750000 7.605000 72.800000
2  6.231818 4.487727 30.249009

Clustering vector:
[1] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2
[21] 2 1 2 2

Within cluster sum of squares by cluster:
[1] 2460.357 4251.213
(between_SS / total_SS = 49.8 %)

Available components:
[1] "cluster"      "centers"
[3] "totss"        "withinss"
[5] "tot.withinss" "betweenss"
[7] "size"         "iter"
[9] "ifault"
```

Total Withinss:

[1] 6711.571

Betweenss:

[1] 6651.504

Note:
Cluster dinyatakan baik apabila nilai total withinss kecil dan nilai betweenss besar

Type here to search | 2043 08/12/2021

10. Lakukan analisis berdasarkan output yang telah muncul.