**Praktikum 10**

**ANALISIS CLUSTER MENGGUNAKAN R DAN EXCEL**

Dataset :

|  |  |  |  |
| --- | --- | --- | --- |
| Supplier | DAR | DSR | Quality |
| S1 | 96.81 | 73.85 | 100 |
| S2 | 99.64 | 65.79 | 100 |
| S3 | 96.5 | 71.63 | 62.86 |
| S4 | 99.349 | 79.38 | 96.86 |
| S5 | 100 | 88.24 | 100 |
| S6 | 71.4 | 60.7 | 71.4 |
| S7 | 99.827 | 84.54 | 79.43 |
| S8 | 99.9 | 82.98 | 88 |
| S9 | 99.058 | 95.16 | 98 |
| S10 | 98.1 | 90.77 | 81.43 |
| S11 | 99.574 | 79.66 | 95.14 |
| S12 | 99.606 | 80.4 | 71.1 |

Script k-mean clustering

|  |
| --- |
| > data\_analisis\_cluster=read.delim("clipboard")  > data\_latih <-data\_analisis\_cluster[,c(2,3,4)]  > data\_latih\_matrix <- as.matrix(scale(data\_latih))  library(factoextra)  library(NbClust)  > nb\_steven <- NbClust(data\_latih, distance = "euclidean", min.nc = 2,max.nc = 8, method = "complete", index ="all")    km.res=kmeans(data\_latih, 3, nstart=25)  km.res  K-means clustering with 3 clusters of sizes 1, 4, 7  Cluster means:  DAR DSR Quality  1 71.40000 60.70000 71.40000  2 98.50825 81.83500 73.70500  3 99.19014 80.72286 96.85714  Clustering vector:  [1] 3 3 2 3 3 1 2 3 3 2 3 2  Within cluster sum of squares by cluster:  [1] 0.0000 417.3477 662.7461  (between\_SS / total\_SS = 71.5 %)  Available components:  [1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"  [7] "size" "iter" "ifault"  library(dplyr)  fviz\_cluster(km.res, data = data\_latih, geom = "point",stand = FALSE, frame.type = "norm")    fviz\_cluster(km.res, data = data\_latih) |

Script hierarchical clustering

|  |
| --- |
| > data\_analisis\_cluster2=read.delim("clipboard")  > df=scale(data\_analisis\_cluster2[,2:4])  res.dist <- dist(df, method = "euclidean")  res.hc <- hclust(d = res.dist, method = "ward.D2")  library("factoextra")  fviz\_dend(res.hc, cex = 0.5)  fviz\_dend(res.hc, k = 4, cex = 0.5, k\_colors = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"), color\_labels\_by\_k = TRUE, rect = TRUE) |

Lakukan perhitungan menggunakan excel menggunakan metode k-means dan hierarchical clustering. File excel disertakan dalam email.

Perbandingan hasil manual dan R :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Hasil dari R:**  K-means clustering with 3 clusters of sizes 1, 4, 7  Cluster means:  DAR DSR Quality  1 71.40000 60.70000 71.40000  2 98.50825 81.83500 73.70500  3 99.19014 80.72286 96.85714  Clustering vector:  [1] 3 3 2 3 3 1 2 3 3 2 3 2  **Menggunakan excel:**  K-means clustering with 3 clusters of sizes 5, 4, 3   |  |  |  |  | | --- | --- | --- | --- | |  | DAR | DSR | Quality | | 1 | 99.0746 | 77.384 | 98.4 | | 2 | 99.22125 | 88.3625 | 86.715 | | 3 | 89.16867 | 70.91 | 68.45333 | |  |  |  |  |   Clustering vector:  [1] 1 1 3 1 1 3 2 2 2 2 1 3 |

Sumber :

<https://datascienceplus.com/k-means-clustering-in-r/>

<https://www.datanovia.com/en/blog/types-of-clustering-methods-overview-and-quick-start-r-code/>

<https://www.youtube.com/watch?v=mtkWR8sx0NA>