DIANA(Divisive analysis)

Sayyid Rafif Abdul Hayullah

Import Data

```
data <- read_excel("~/ANALISIS MULTIVARIAT/data uas anmul.xlsx")
head(data)
```

```
## # A tibble: 6 x 6
##   Provinsi           IPM `Umur Harapan Hidup` `Harapan Lama Sekolah`
##   <chr>            <dbl>                <dbl>                  <dbl>
## 1 ACEH              74.0                 70.4                   14.4
## 2 SUMATERA UTARA    74.0                 70.3                   13.5
## 3 SUMATERA BARAT    74.5                 70.3                   14.3
## 4 RIAU              74.8                 72.5                   13.4
## 5 JAMBI             73.4                 72.0                   13.1
## 6 SUMATERA SELATAN  72.3                 70.9                   12.6
## # i 2 more variables: `Rata-rata Lama Sekolah` <dbl>,
## #   `Pengeluaran per Kapita` <dbl>
```

dataset memiliki 38 baris dengan 6 variabel

Cek Data Kosong (Missing Values)

```
print(colSums(is.na(data)))
```

```
##               Provinsi                    IPM     Umur Harapan Hidup
##                      0                      0                      0
##   Harapan Lama Sekolah Rata-rata Lama Sekolah Pengeluaran per Kapita
##                      0                      0                      0
```

Semua variabel tidak memiliki missing values

Cek Data Duplikat

```
cat("\nJumlah baris duplikat:", sum(duplicated(data)))
```

```
##
## Jumlah baris duplikat: 0
```

Tidak terdapat baris yang duplikat di dataset

Pilih Hanya Variabel Numerik

```
data_num <- data[, c("IPM","Umur Harapan Hidup","Harapan Lama Sekolah","Rata-rata Lama Sekolah","Pengelu
```

Standardisasi Data

```
data_scaled <- scale(data_num)
head(data_scaled)
```

```
##              IPM Umur Harapan Hidup Harapan Lama Sekolah Rata-rata Lama Sekolah
## [1,]  0.31873916        -0.02829145           1.15520328               0.6504944
## [2,]  0.31679750        -0.08950570           0.27431248               0.8866871
## [3,]  0.40805562        -0.07037624           1.06711420               0.4876029
## [4,]  0.46630548         0.76749372           0.20579876               0.4794583
## [5,]  0.20223944         0.58385099          -0.06825616               0.0477958
## [6,] -0.01716838         0.15917717          -0.55763993              -0.2209752
##      Pengeluaran per Kapita
## [1,]             -0.31702435
## [2,]             -0.05357270
## [3,]              0.05115846
## [4,]              0.10758339
## [5,]              0.01178279
## [6,]              0.17172108
```

Semua variabel distandarisasi agar setiap variabel memiliki kontribusi setara dan tidak ada variabel yang mendominasi

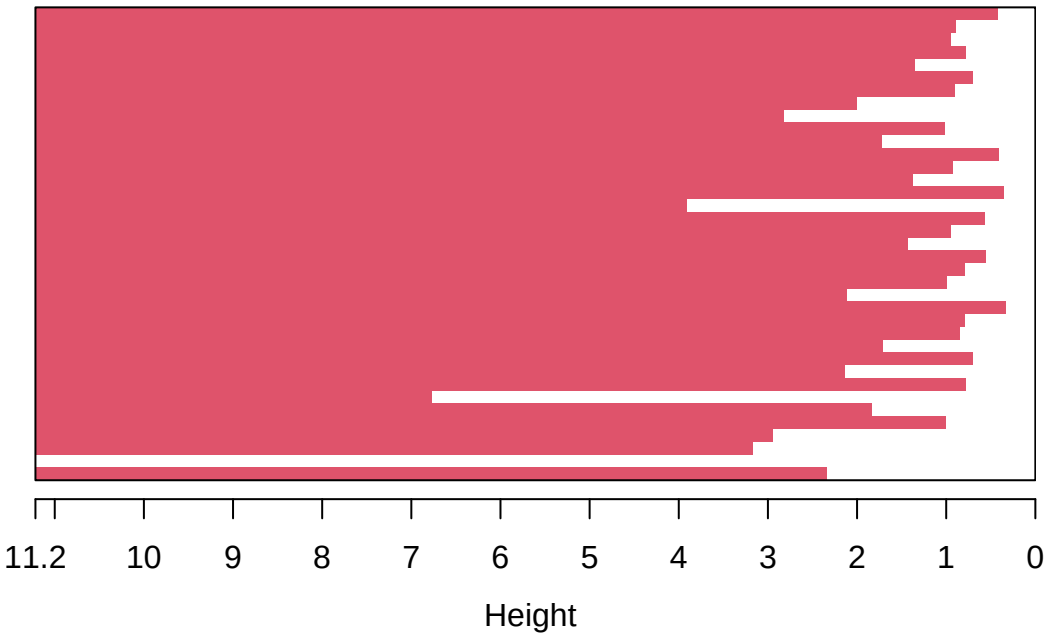
DIANA (Divisive analysis)

```
diana_result <- diana(data_scaled)
```

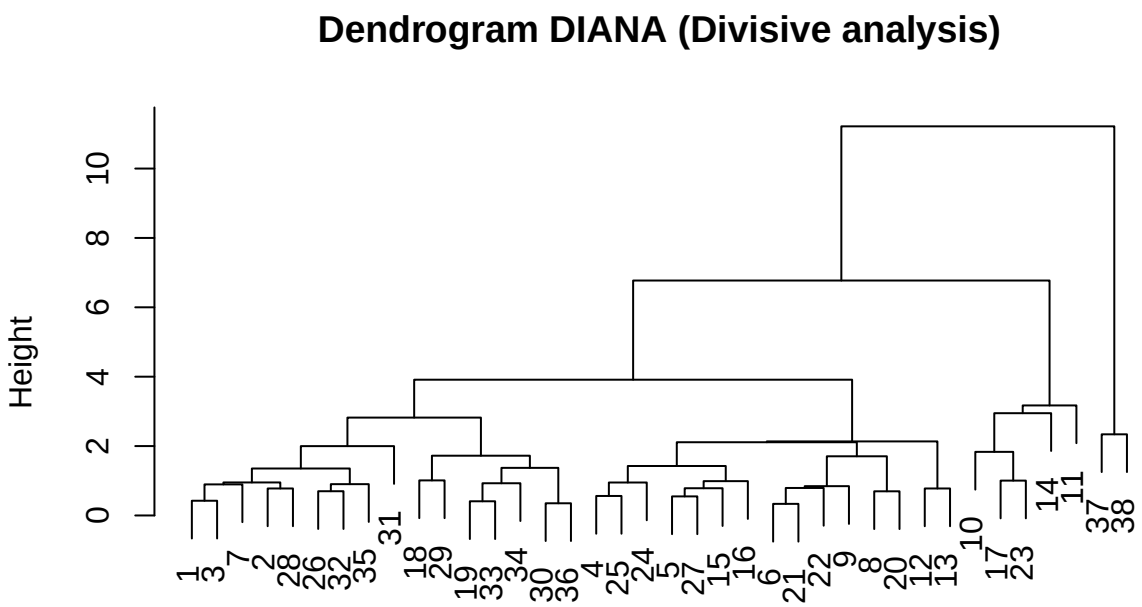
Plot Dendrogram

```
plot(diana_result, main="Dendrogram DIANA (Divisive analysis)")
```

Dendrogram DIANA (Divisive analysis)



Divisive Coefficient = 0.91



Pemotongan Cluster

```
clusters <- cutree(as.hclust(diana_result), k=3)
data$Cluster <- clusters
head(data)
```

```
##           Provinsi   IPM Umur Harapan Hidup Harapan Lama Sekolah
## 1             ACEH 74.03              70.44                14.39
## 2   SUMATERA UTARA 74.02              70.28                13.49
## 3   SUMATERA BARAT 74.49              70.33                14.30
## 4             RIAU 74.79              72.52                13.42
## 5            JAMBI 73.43              72.04                13.14
## 6 SUMATERA SELATAN 72.30              70.93                12.64
##   Rata-rata Lama Sekolah Pengeluaran per Kapita Cluster
## 1                   9.64                   10811       1
## 2                   9.93                   11460       1
## 3                   9.44                   11718       1
## 4                   9.43                   11857       1
## 5                   8.90                   11621       1
## 6                   8.57                   12015       1
```

```r
cat("\n=== PROFIL CLUSTER (Mean Tiap Variabel per Cluster) ===\n")
```

```
##
## === PROFIL CLUSTER (Mean Tiap Variabel per Cluster) ===
```

```r
cluster_profile <- aggregate(data_num, by = list(Cluster = clusters), FUN = mean)
print(cluster_profile)
```

```
##   Cluster      IPM Umur Harapan Hidup Harapan Lama Sekolah
## 1       1 72.20645           70.29323             13.29806
## 2       2 79.83800           73.73000             14.02600
## 3       3 56.58500           65.89500              9.80000
##   Rata-rata Lama Sekolah Pengeluaran per Kapita
## 1               8.844194               11205.77
## 2              10.294000               15920.00
## 3               5.165000                6758.00
```

CLUSTER VALIDATION (Silhouette Score)

```r
cat("\n=== CLUSTER VALIDATION (Silhouette Score, k = 3) ===\n")
```

```
##
## === CLUSTER VALIDATION (Silhouette Score, k = 3) ===
```

```r
dist_eu <- dist(data_scaled, method = "euclidean")
sil_eu <- silhouette(clusters, dist_eu)
cat("\nSilhouete(Euclidean):", mean(sil_eu[, 3]), "\n")
```

```
##
## Silhouete(Euclidean): 0.4640621
```

```r
dist_ma <- dist(data_scaled, method = "manhattan")
sil_ma <- silhouette(clusters, dist_ma)
cat("Silhouette(Manhattan):", mean(sil_ma[, 3]), "\n")
```

```
## Silhouette(Manhattan): 0.4862954
```

```r
dist_ca <- dist(data_scaled, method = "canberra")
sil_ca <- silhouette(clusters, dist_ca)
cat("Silhouette(Canberra):", mean(sil_ca[, 3]), "\n")
```

```
## Silhouette(Canberra): 0.06328778
```

KESIMPULAN

Berdasarkan nilai Silhouette Score metrik Euclidean memperoleh nilai 0.46 dan metrik Manhattan memperoleh nilai 0.48, yang keduanya berada dalam rentang 0.26–0.50 sehingga termasuk kategori Weak clustering. Artinya, struktur cluster yang terbentuk cukup terlihat, namun pemisahan antar cluster belum benar-benar kuat. Sementara itu, metrik Canberra menghasilkan nilai 0.06 yang berada pada rentang 0.00–0.25, sehingga termasuk kategori Bad clustering, menunjukkan bahwa pemisahan cluster dengan Canberra tidak layak digunakan karena objek antar cluster sangat berdekatan dan tidak membentuk struktur yang jelas. Secara keseluruhan, DIANA dengan k = 3 masih dapat diterima untuk Euclidean dan Manhattan meskipun kualitasnya lemah, tetapi Canberra tidak cocok digunakan untuk data IPM dan komponennya.

SIMPAN HASIL FINAL DALAM CSV

```r
write.csv(data, "hasil_cluster_diana.csv", row.names = FALSE)
cat("\nFile 'hasil_cluster_diana.csv' berhasil disimpan.\n")
```

```
##
## File 'hasil_cluster_diana.csv' berhasil disimpan.
```