

Nama : Muhammad Fakhri  
NIM : G64170015  
Tipe Soal : Ganjil  
Nomor Meja : 105

1. Buka lah diabet.csv pada R! Lakukan eksplorasi data tersebut! (Lampirkan screenshot kode R serta jelaskan hasilnya)

```
1 library("fpc")
2 library("factoextra")
3 library("mice")
4 library("cluster")
5
6 diabet = read.csv("diabet.csv", sep = ",")
7 diabet1 = read.csv("diabet.csv", sep = ",")
8 str(diabet)
9 summary(diabet)
10
11 md.pattern(diabet)
12 boxplot(diabet)
```

#### Output Kode

```
> str(diabet)
'data.frame': 768 obs. of 9 variables:
 $ preg : int 6 1 8 1 0 5 3 10 2 8 ...
 $ plas : int 148 85 183 89 137 116 78 115 197 125 ...
 $ pres : int 72 66 64 66 40 74 50 0 70 96 ...
 $ skin : int 35 29 0 23 35 0 32 0 45 0 ...
 $ insu : int 0 0 0 94 168 0 88 0 543 0 ...
 $ mass : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
 $ pedi : num 0.627 0.351 0.672 0.167 2.288 ...
 $ age : Factor w/ 53 levels "21","22","23",...: 30 11 12 1 13 10 6 9 33
34 ...
 $ class: Factor w/ 3 levels "", "tested_negative",...: 3 2 3 2 3 2 3 3 3
...
> summary(diabet)
      preg      plas      pres      skin
Min.   :0.000   Min.   :0.0   Min.   :0.00   Min.   :0.00
1st Qu.:1.000   1st Qu.:99.0   1st Qu.:62.00   1st Qu.:0.00
Median :3.000   Median :117.0   Median :72.00   Median :23.00
Mean   :3.845   Mean   :120.9   Mean   :69.11   Mean   :20.54
3rd Qu.:6.000   3rd Qu.:140.2   3rd Qu.:80.00   3rd Qu.:32.00
Max.   :17.000   Max.   :199.0   Max.   :122.00   Max.   :99.00

      insu      mass      pedi      age
Min.   :0.0   Min.   :0.0   Min.   :0.0780   22   :72
1st Qu.:0.0   1st Qu.:27.3   1st Qu.:0.2407   21   :63
Median :30.5   Median :32.0   Median :0.3705   25   :48
Mean   :79.8   Mean   :32.0   Mean   :0.5529   24   :46
3rd Qu.:127.2   3rd Qu.:36.6   3rd Qu.:0.6292   23   :38
```

Ada 8 atribut dengan atribut mass dan pedi memiliki Missing Value

2. Lakukanlah tahapan praproses terhadap data! (Lampirkan screenshot kode R serta jelaskan hasilnya)

- Pengisian missing value.
- Lakukanlah normalisasi pada atribut plas dan insu , menggunakan metode min max.
- Reduksi Dimensi (Pilihlah atribut yang menurut Anda penting atau berpengaruh dan hapus atribut yang tidak digunakan).

```

14 diabet$mass[is.na(diabet$mass)] <- mean (diabet$mass,na.rm= TRUE)
15 diabet$pedi[is.na(diabet$pedi)] <- mean (diabet$pedi,na.rm= TRUE)
16
17
18 for (i in 1:length(diabet$plas)) {
19   diabet$plas[i] = (diabet$plas[i] - min(diabet$plas)) / (max(diabe
20 }
21 |
22 for (i in 1:length(diabet$insu)) {
23   diabet$insu[i] = (diabet$insu[i] - min(diabet$insu)) / (max(diabe
24 }
25
26 diabet$class <- NULL
27 diabet$skin <- NULL
28 diabet$age <-NULL
29

```

Missing value di atribut mass dan pedi semua diganti dengan nilai mean  
 Dan dilakukan normalisasi min max pada plas dan insu  
 Lalu dihilangkan atribut class, skin, dan age

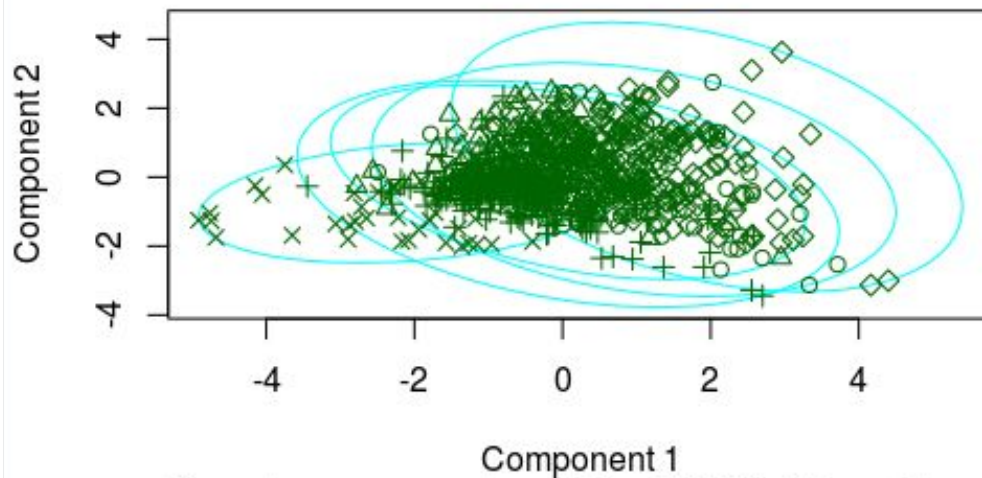
3. Lakukanlah tahapan clustering terhadap data hasil praproses! (Tahapan yang dilakukan pada clustering yaitu clustering menggunakan K-Means (terhadap atribut plas dan insu) dan K-Medoids dengan k=5 serta tampilkan hasil plotnya!)

```

30 kmeans.result <- kmeans(diabet, 5)
31 pamk.result <- pamk(diabet)
32 pam.result <- pam(diabet, 5)
33 plot(pam.result)
34 |

```

**clusplot(pam(x = diabet, k = 5))**



These two components explain 46.02 % of the point variability