*Modul Praktikum Data Mining 2016*

# PERTEMUAN 4

**EKSPLORASI DATA MENGGUNAKAN R**



**TUJUAN PRAKTIKUM**

Mahasiswa akan dapat memahami Tipe data, Eksplorasi Data, Statistika ringkasan, Visualisasi menggunakan Weka dan R.



**TEORI PENUNJANG**

# Eksplorasi data

Eksplorasi data dilakukan sebagai langkah awal untuk mengetahui karakteristik dari data. Tahapan ini dilakukan untuk menyeleksi teknik pemrosesan dan analisis data yang sesuai untuk dataset yang kita miliki. Eksplorasi data menggunakan R dimulai dengan memeriksa dimensi data, mencari statistika ringkasan, serta menampilkan berbagai jenis grafik seperti pie chart dan histogram. Dengan menggunakan R kita juga bisa melakukan eksplorasi pada multiple variables. Eksplorasi pada multiple variables dilakukan untuk mengetahui hubungan antara dua variabel atau lebih. Pada R, hal ini dapat dilihat melalui bar chart, boxplot, scatter plot, pairs plot, serta dengan menghitung nilai covariance dan correlation antar variabel.

# Density

# Density/densitas dapat digunakan untuk melihat pola distribusi data/peluang distribusi.

|  |
| --- |
| data(“iris”)plot(density(iris$Sepal.Length)) |

# 

# Berdasarkan plot diatas variabel Sepal Length menyebar normal.

# Pie Chart

|  |
| --- |
| freq <- table(iris$Species)pie(freq) |

# 

# Eksplorasi data pada multiple variables

# Eksplorasi data pada multiple variables dilakukan untuk melihat hubungan antara dua buah variabel atau lebih. Untuk melihat hubungan tersebut bisa dilakukan dengan menghitung nilai covariance dan correlation antara variabel - variabel tersebut.

|  |
| --- |
| > cov(iris$Sepal.Length, iris$Petal.Length)  > cov(iris[,1:4])  > cor(iris$Sepal.Length, iris$Petal.Length)  > cor (iris[,1:4]) > aggregate(Sepal.Length ~ Species, summary, data=iris) |

# Visualisasi data untuk *multiple variables* di R berupa boxplot, scatter plot, 3D scatter plot, pairs plot, heat map, level plot, contour, 3D surface, dan parallel coordinates.

|  |
| --- |
| #boxplot  > boxplot(Sepal.Length~Species, data=iris)  #scatter plot  > with(iris, plot(Sepal.Length, Sepal.Width, col=Species, pch=as.numeric(Species)))  #scatter plot with jitter  > plot(jitter(iris$Sepal.Length, jitter(iris$Sepal.Width))  #matrix of scatter plots (Pairs plot)  > pairs(iris)  #3D scatter plot  > library(scatterplot3d)  > scatterplot3d(iris$Petal.Width, iris$Sepal.Length, iris$Sepal.Width)  #interactive 3D scatter plot  > library(rgl)  > plot3d(iris$Petal.Width, iris$Sepal.Length, iris$Sepal.Width)  #heat map  > distMatrix <- as.matrix(dist(iris[,1:4])  > heatmap(distMatrix)  #level plot  > library(lattice)  > levelplot(Petal.Width~Sepal.Length\*Sepal.Width, iris, cuts=9, col.regions=grey.colors(10)[10:1])  #contour  > filled.contour(volcano, color=terrain.colors, asp=1, plot.axes=contour(volcano, add=T))  #3D surface  > persp(volcano, theta = 25, phi = 30, expand = 0.5, col  = "lightblue")  #Parallel coordinates  > library(MASS)  > parcoord(iris[1:4], col=iris$Species)  #Parallel Coordinates with lattice  > library(lattice)  > parallelplot(~iris[1:4] | Species, data=iris)  #Scatter Plot with ggplot2  > library(ggplot2)  > qplot(Sepal.Length, Sepal.Width, data=iris, facets=Species ~.)  #save as a PDF file  > pdf("myPlot.pdf")  > x <- 1:50  > plot(x, log(x))  > graphics.off()  #Save as a postscript file  > postscript("myPlot2.ps")  > x <- -20:20  > plot(x, x^2)  > graphics.off() |



**MATERI PRAKTIKUM**

1. Penjelasan Eksplorasi Data
2. Eksplorasi dan visualisasi data untuk *individual variables* di R.
3. Eksplorasi dan visualisasi data untuk *multiple variables* di R.
4. Cara membuat dan menyimpan grafik.