

PERTEMUAN 1

PENGENALAN LINGKUNGAN KERJA R

TUJUAN PRAKTIKUM

Mahasiswa akan dapat menjelaskan fitur-fitur dalam R dan Weka untuk keperluan penerapan algoritme *data mining*, mencakup fungsi-fungsi sederhana R dan Weka, visualisasi data, *import*, dan *export* data.

TEORI PENUNJANG

Pengenalan R

R adalah lingkungan perangkat lunak tidak berbayar untuk perhitungan statistik maupun pembuatan grafik. R menyediakan berbagai fungsi statistik dan teknik membuat grafik. R dapat dikembangkan atau digunakan lebih lanjut dengan memanfaatkan *package* yang ada. Beberapa keuntungan R adalah :

- a. Perangkat lunak *free* dan *open source*
- b. Banyak *package* yang bisa digunakan dan ada komunitas pengembangnya
- c. Mempunyai kemampuan visual grafik dan keakuratan hasil yang lebih baik daripada perangkat lunak lain (McCullough, B.D. and David A. Heiser, 2008).

1. Fungsi-fungsi pada R

Fungsi dasar pada R:

- a. Visualisasi data : sintax yang digunakan diantaranya : `barplot`, `pie`, `dotchart`, dan `hist`
- b. Manipulasi data : sintax yang digunakan diantaranya : `sample`, `stack`, `unstack`, dan `na.omit`
- c. Antarmuka ke Weka : sintax yang digunakan adalah `RWeka`.

Selain fungsi dasar tersebut, R juga dapat digunakan untuk proses data mining yang lebih kompleks. Masing-masing fungsi pada R direpresentasikan dengan *package*. Berikut beberapa fungsi pada R dan *package* yang digunakan.

- a. *Clustering* : *package* yang digunakan untuk *clustering* diantaranya `fpc`, `cluster`, `pvclust`, `mclust`.
- b. Klasifikasi : *package* yang digunakan untuk klasifikasi diantaranya `rpart`, `tree`, `marginTree`, `party`, `randomForest`, `maptree`.
- c. Asosiasi : *package* yang digunakan untuk asosiasi diantaranya `arules` dan `drn`
- d. *Sequential Pattern* : *package* yang digunakan untuk *sequential pattern* diantaranya `arulesSequences`.
- e. *Time Series* : *package* yang digunakan untuk *time series* diantaranya `timsac`.
- f. Statistik : *package* yang digunakan untuk statistik diantaranya `BaseR` dan `nlme`.

2. Data set dan komponennya

Data set adalah sekumpulan data/*item* yang akan digunakan untuk proses data mining. Komponen-komponen pada data set adalah atribut dan nilai

3. Rcore dan RStudio

Untuk dapat menggunakan R dalam pemrosesan data mining maka dibutuhkan RCore dan RStudio. RCore digunakan untuk memperoleh *package* maupun fungsi-fungsi yang akan digunakan, sedangkan Rstudio digunakan sebagai *interface* ke *user* agar lebih mudah. Sebenarnya RCore sendiri sudah mempunyai GUI yaitu dalam bentuk CLI.

4. Membuka dan menyimpan data R

File yang digunakan pada R adalah file dengan ekstensi `.Rdata`.

a. Membuka data R

Untuk dapat membuka data R dapat dilakukan dengan 2 cara yaitu dengan memakai menu pada *menu bar* atau dengan *command* pada *workspace*. Untuk dapat membuka data R pada *menu bar* klik file >> Open File >> pilih filenya >> klik Open. Atau bisa dapat dengan klik icon >> pilih filenya >> klik Open. Selain itu dapat juga dilakukan pada *workspace* dengan mengetikkan perintah `load()` contoh : `> load("~/R/Latihan/tabel.Rdata")`

b. Menyimpan data R

Untuk dapat menyimpan R data yang telah dikerjakan pada Rstudio maka dapat dilakukan dengan mengetikkan command `save()`. Contoh : `> save(varGab, file="varGab.Rdata")`

5. Import dan Export data to .csv Files

```
> var1 <- 1:5
> var2 <- (1:5) / 10
> var3 <- c("R", "and", "Data Mining", "Examples", "Case Studies")
> a <- data.frame(var1, var2, var3)
> names(a) <- c("VariableInt", "VariableReal", "VariableChar")
> write.csv(a, "data/dummyData.csv", row.names = FALSE)
> #rm(a)
> b <- read.csv("data/dummyData.csv")
> print(b)
```

6. Visualisasi dalam Grafik

a. Pie Chart

```
# Define cars vector with 5 values
cars <- c(1, 3, 6, 4, 9)

# Define some colors ideal for black & white print
colors <- c("white","grey70","grey90","grey50","black")

# Calculate the percentage for each day, rounded to one
# decimal place
car_labels <- round(cars/sum(cars) * 100, 1)

# Concatenate a '%' char after each value
car_labels <- paste(car_labels, "%", sep="")

# Create a pie chart with defined heading and custom colors
# and labels
pie(cars, main="Cars", col=colors, labels=car_labels,
    cex=0.8)

# Create a legend at the right
legend(1.5, 0.5, c("Mon","Tue","Wed","Thu","Fri"), cex=0.8,
    fill=colors)
```

b. Line Chart

```
# Define 2 vectors
cars <- c(1, 3, 6, 4, 9)
trucks <- c(2, 5, 4, 5, 12)

# Calculate range from 0 to max value of cars and trucks
g_range <- range(0, cars, trucks)

# Graph autos using y axis that ranges from 0 to max
# value in cars or trucks vector. Turn off axes and
# annotations (axis labels) so we can specify them ourself
plot(cars, type="o", col="blue", ylim=g_range,
    axes=FALSE, ann=FALSE)

# Make x axis using Mon-Fri labels
axis(1, at=1:5, lab=c("Mon","Tue","Wed","Thu","Fri"))

# Make y axis with horizontal labels that display ticks at
# every 4 marks. 4*0:g_range[2] is equivalent to
# c(0,4,8,12).
axis(2, las=1, at=4*0:g_range[2])

# Create box around plot
box()

# Graph trucks with red dashed line and square points
lines(trucks, type="o", pch=22, lty=2, col="red")

# Create a title with a red, bold/italic font
title(main="Autos", col.main="red", font.main=4)

# Label the x and y axes with dark green text
title(xlab="Days", col.lab=rgb(0,0.5,0))
title(ylab="Total", col.lab=rgb(0,0.5,0))

# Create a legend at (1, g_range[2]) that is slightly
# smaller
# (cex) and uses the same line colors and points used by
# the actual plots
legend(1, g_range[2], c("cars","trucks"), cex=0.8,
    col=c("blue","red"), pch=21:22, lty=1:2);
```

c. Histogram

```
# Read values from tab-delimited autos.dat
autos_data <- read.table("C:/R/autos.dat", header=T,
sep="\t")

# Graph autos with adjacent bars using rainbow colors
barplot(as.matrix(autos_data), main="Autos", ylab= "Total",
beside=TRUE, col=rainbow(5))

# Place the legend at the top-left corner with no frame
# using rainbow colors
legend("topleft", c("Mon", "Tue", "Wed", "Thu", "Fri"),
cex=0.6,
bty="n", fill=rainbow(5));
```

Pengenalan Weka

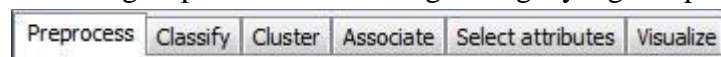


Weka adalah sekumpulan algoritme mesin *learning* untuk melakukan tugas-tugas data mining. Weka dapat digunakan untuk praproses data, klasifikasi, *clustering*, regresi, aturan assosiasi, dan visualisasi.

1. Menu dan Fungsi-fungsi Dasar pada Weka



Pada praktikum kali ini akan digunakan panel explorer. Panel inilah yang akan digunakan untuk mengeksplorasi modul / fungsi-fungsi yang ada pada weka.



Gambar tersebut menunjukkan menu dan tugas-tugas data mining yang dapat dilakukan pada Weka.

- Preprocces : menu ini digunakan untuk melakukan praproses data sebelum digunakan.
- Classify : menu ini digunakan untuk melakukan tugas data mining yaitu klasifikasi
- Cluster : menu ini digunakan untuk melakukan tugas data mining yaitu *clustering*

- d. Associate : menu ini digunakan untuk melakukan tugas data mining yaitu pembentukan aturan assosiasi.
- e. Select attributes : menu ini digunakan untuk memilih atribut yang paling relevan pada data
- f. Visualize : menu ini digunakan untuk memvisualisasikan data yang akan digunakan dalam plot 2D interaktif dari data

2. Membuka, Mengedit, dan Menyimpan Data pada Weka



- a. Untuk membuka data dapat dilakukan dengan klik button “Open file”.
- b. Untuk mengubah data dapat dilakukan dengan
- c. Untuk membatalkan perubahan maka dapat dilakukan dengan klik button “Undo”
- d. Untuk menyimpan data atau hasil kerja yang telah dilakukan dapat dilakukan dengan klik button “Save”.

3. Import and Export .csv Files

- a. Untuk meng-import file .csv maka dilakukan dengan klik button “Open file”.
- b. Untuk meng-export ke file .csv maka dilakukan kemudian memilih ekstensi file penyimpanan .csv

4. Membuat .arff Files

dari File .csv

- a. Buat file.csv seperti data di bawah ini

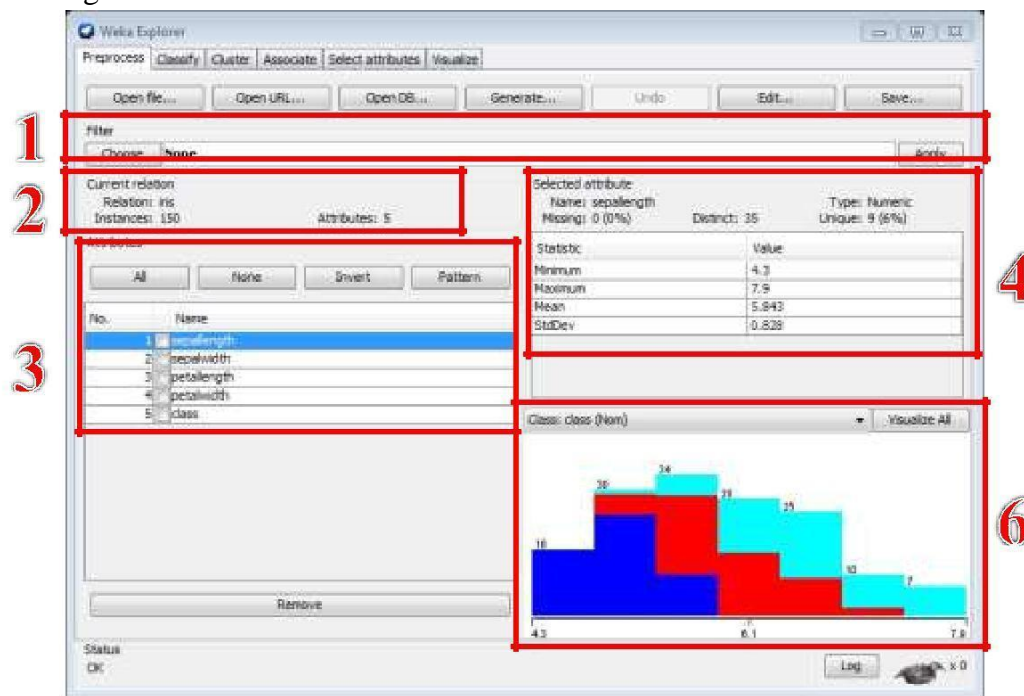
X	Y	Class
2	4	kuadrat
2	3	tambah
3	9	kuadrat
3	4	tambah
7	8	tambah
8	64	kuadrat

- b. Buka file.csv dengan text editor. Tambahkan beberapa kode berikut

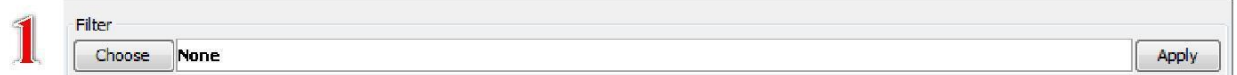
```
@RELATION bilangan
@ATTRIBUTE X NUMERIC
@ATTRIBUTE Y NUMERIC
@ATTRIBUTE Class {kuadrat, tambah} @Data
2,4,kuadrat
2,3,tambah
3,9,kuadrat
3,4,tambah
7,8,tambah
8,64,kuadrat
```

- c. Save file dengan ekstensi .arff

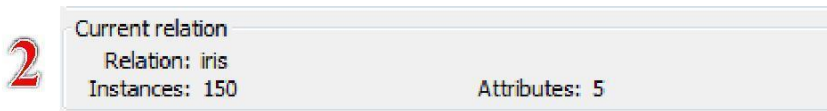
5. Ringkasan Data dan Visualisasi Data



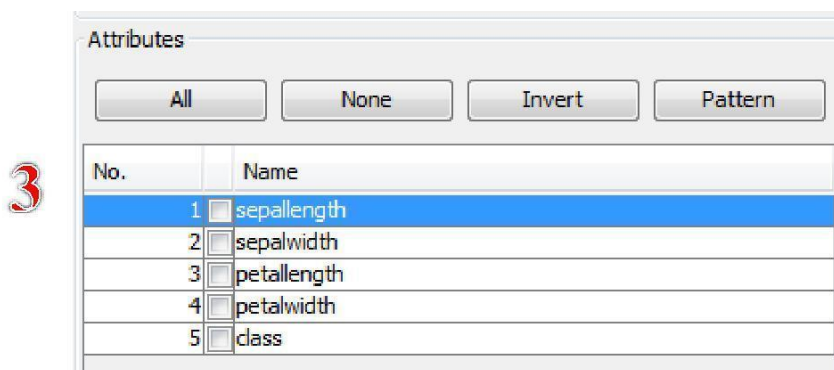
Keterangan :



Tab ini dapat digunakan untuk mengolah data seperti melakukan diskretisasi.



Berisi relasi data, jumlah data, dan jumlah attribut.

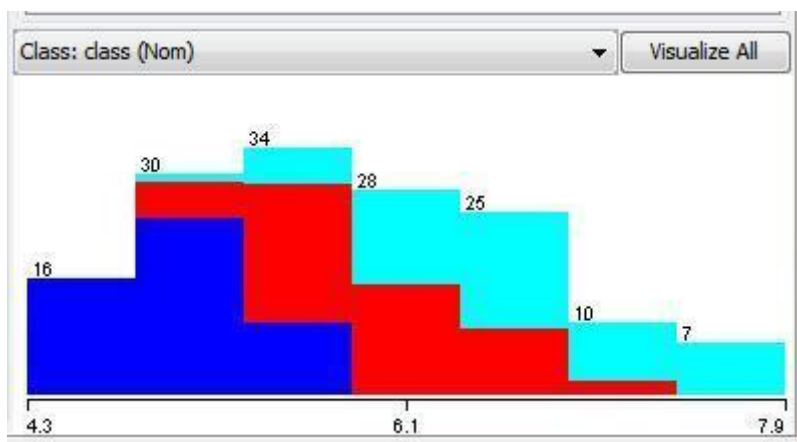


Berisi memilih atribut yang ada.

4

Selected attribute	
Name: sepalength	Type: Numeric
Missing: 0 (0%)	Distinct: 35
	Unique: 9 (6%)
Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Properties data yang terpilih meliputi nama atribut, jumlah data yang hilang, jumlah data yang berbeda, dan tipe data. Selain itu, ditampilkan juga informasi statistik dari data seperti minimum, maximum, mean, standard deviasi dan lain-lain.



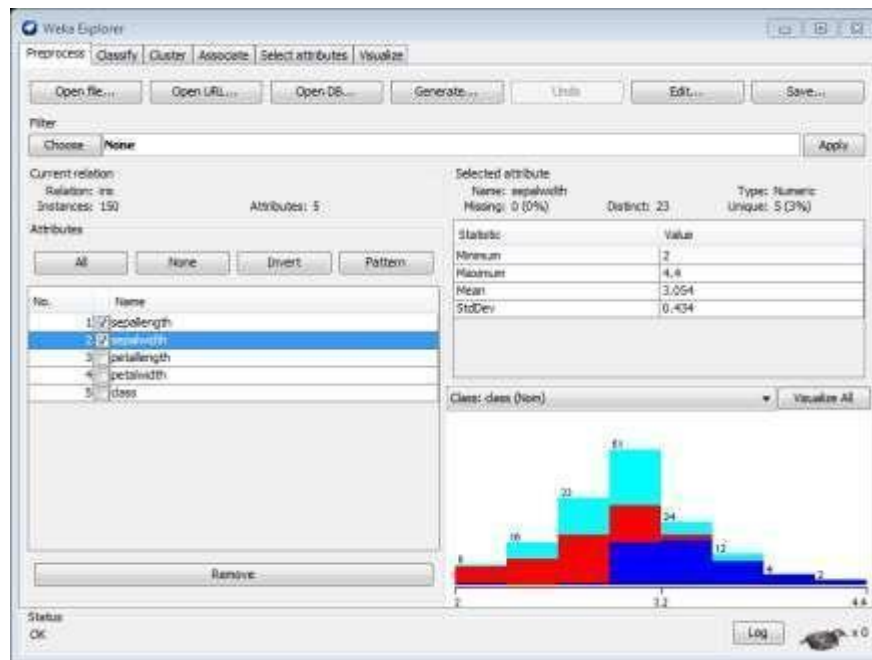
Berisi visualisasi dari data.

6. Contoh Praproses Data dengan Weka

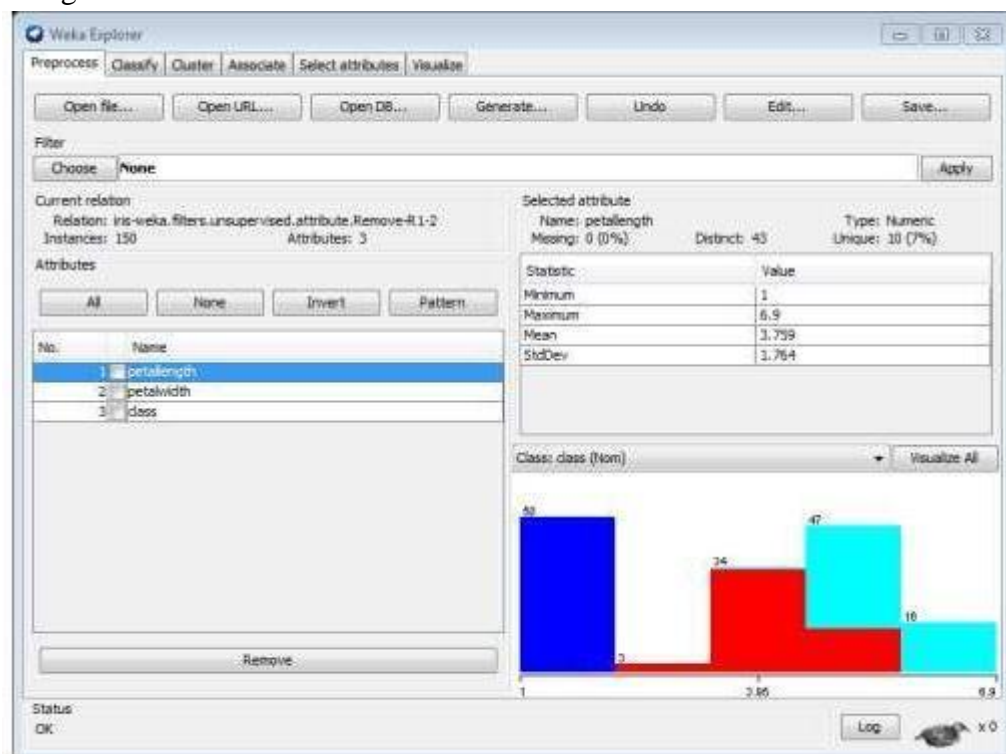
Contoh praproses sederhana adalah pemilihan atribut yang akan digunakan untuk melakukan tugas data mining. Contoh pemilihan atribut yang akan digunakan adalah sebagai berikut.

a. Buka file iris.arff

- b. misal memilih atribut sepallength sama sepalwidth untuk tidak dipakai, klik remove



- c. Maka atribut menjadi hanya petallength dan petalwidth saja yang akan digunakan



LAPORAN PENDAHULUAN

1. Apa yang dimaksud dengan *data mining*?
2. Mengapa perlu dilakukan *data mining*?
3. Apa saja yang dapat dilakukan pada *data mining*?
4. Apa yang anda ketahui tentang R?
5. Apa yang anda ketahui tentang Weka?
6. Tugas *data mining* apa saja yang dapat dilakukan pada Weka?
7. Apakah beda weka dengan R?

MATERI PRAKTIKUM

1. Pengenalan R dan Weka
2. Fungsi-fungsi pada R dan Weka
3. Dataset dan Komponennya
4. Rcore dan RStudio
5. Menyimpan dan Membuka Data R
6. Import and Export to .csv Files
7. Ringkasan dan Visualisasi Data

DAFTAR PUSTAKA

1. Yanchang Zhao. 2012. R and Data Mining: Examples and Case Studies. <http://www.rdatamining.com/docs/RDataMining.pdf>
2. Witten I.H & Frank E. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan-Kauffman. <http://www.cs.waikato.ac.nz/~ml/weka/book.html>. Accessed on 27 January 2013.
3. Bouckaert, R. R.. et al. 2013. *WEKA Manual for Version 3-6-9*. Edition of January 21, 2013. <http://jaist.dl.sourceforge.net/project/weka/documentation/3.6.x/WekaManual-3-6-9.pdf> . Accessed on 27 January 2013.