

Pengenalan Lingkungan Kerja Weka

Praktikum 2 Data Mining

Materi

- Silakan download Lembar Kerja Praktikum 2 (Alih Jenis) di LMS.

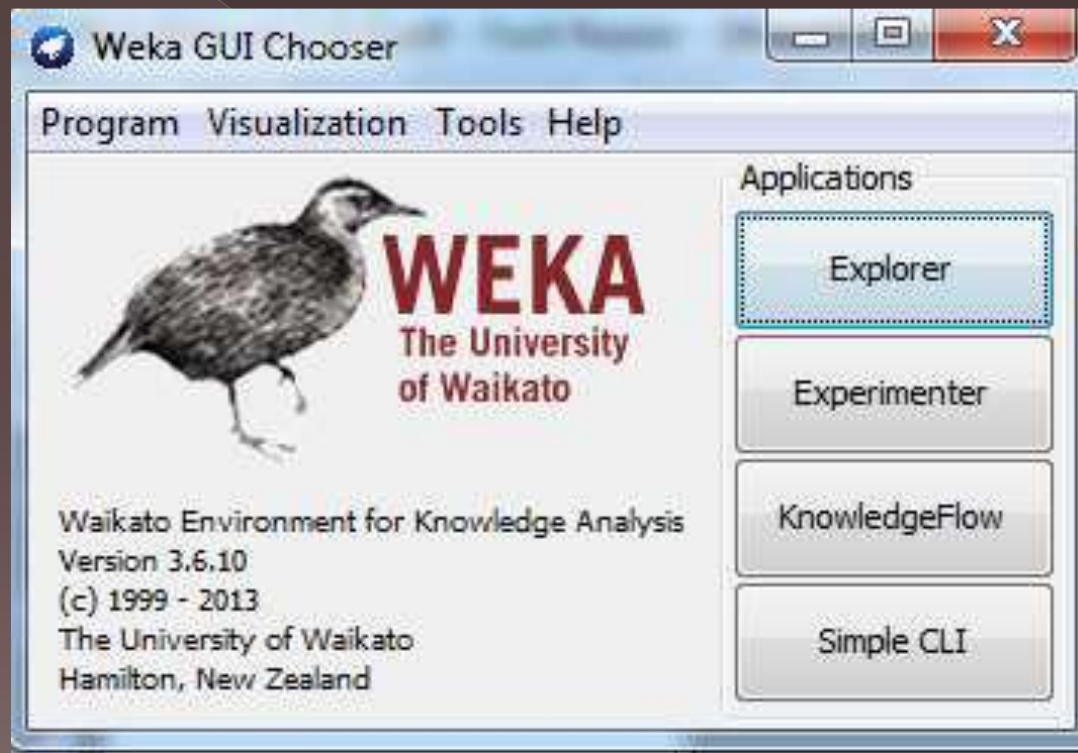
Pengenalan Weka

- Sekumpulan algoritme mesin *learning* untuk melakukan tugas-tugas data mining.
- Dikembangkan oleh University Of Waikato New Zealand.
- Bahasa pemrograman java dibawah lisensi GNU General Public License

Pengenalan Weka(Lanj.)

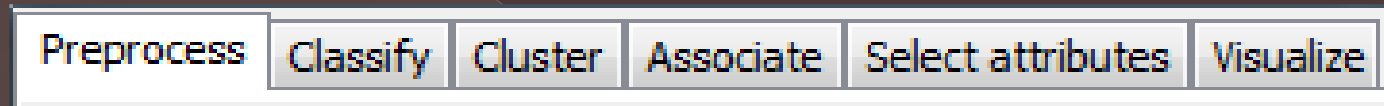
- ◉ Weka dapat digunakan untuk:
 - ◉ Praproses data.
 - ◉ Klasifikasi.
 - ◉ *Clustering*.
 - ◉ Regresi.
 - ◉ Aturan assosiasi.
 - ◉ Visualisasi.

Menu dan Fungsi Dasar Weka

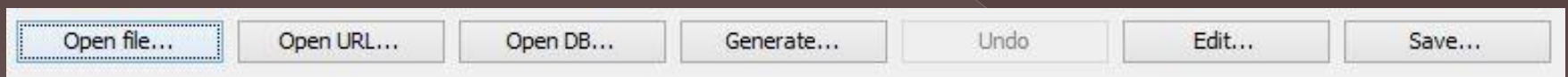


Menu dan Fungsi Dasar Weka(Lanj.)

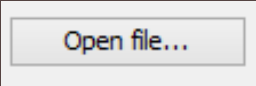
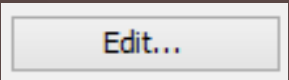
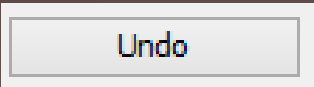
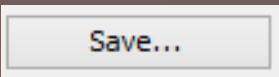
- Menu explorer



- Tab preprocess

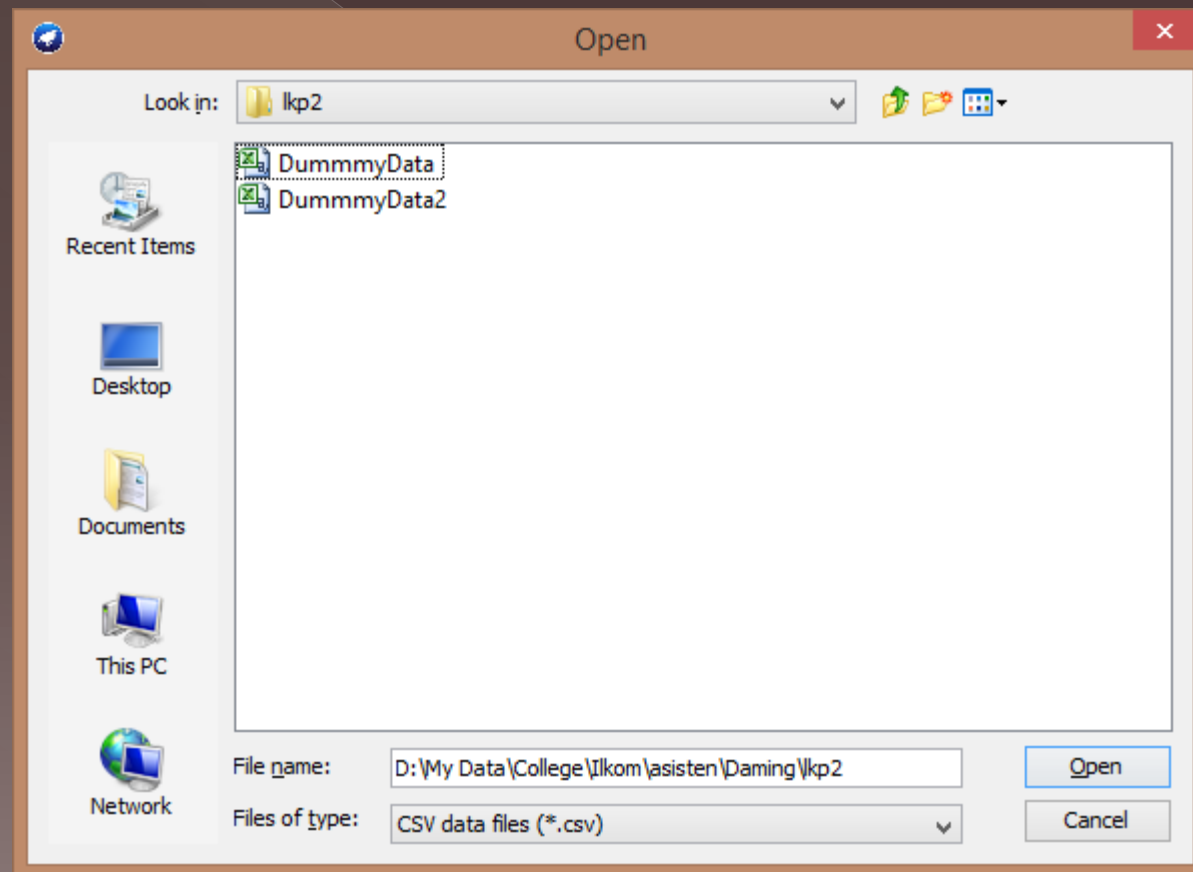


Menu dan Fungsi Dasar Weka(Lanj.)

-  Membuka file.
-  Edit file.
-  Membatalkan perubahan.
-  Menyimpan data.

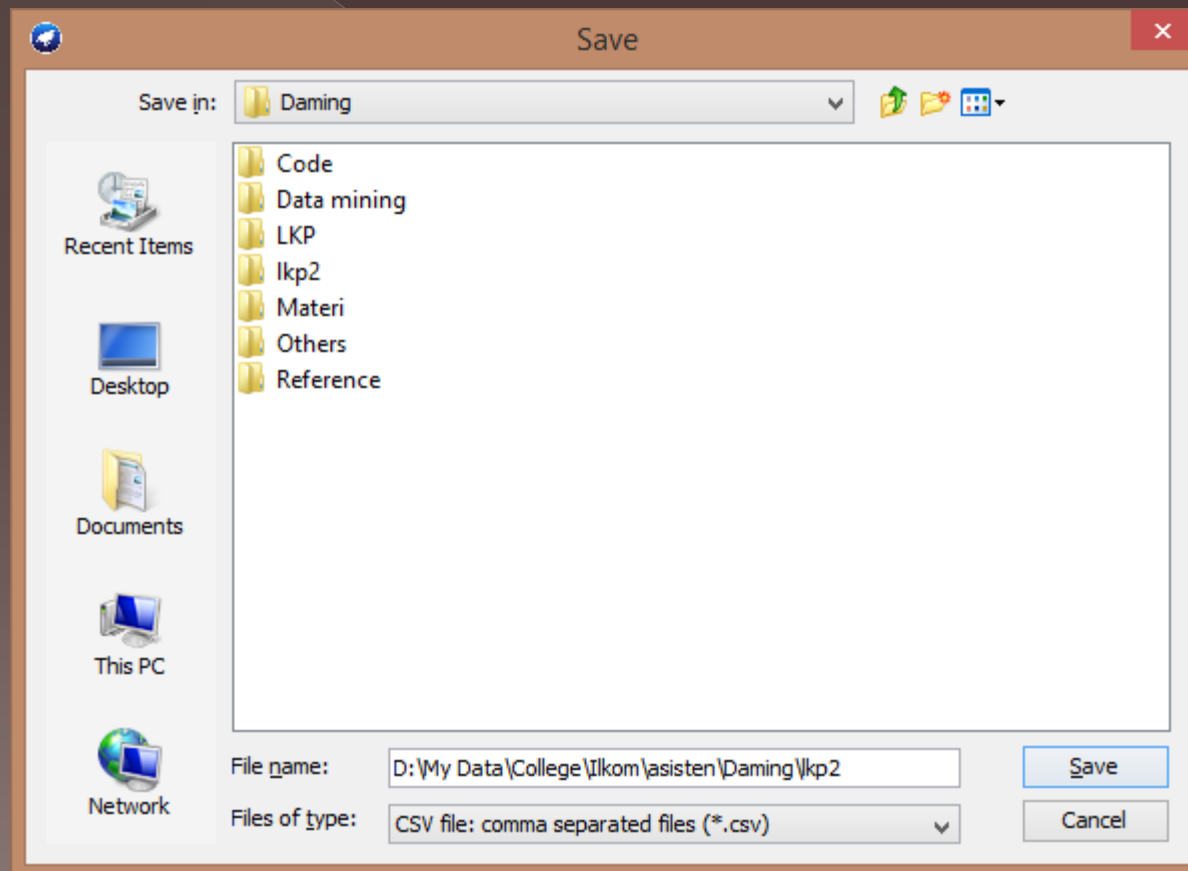
Import & Export .csv file

- ◉ Import: Klik tombol “Open file”



Import & Export .csv file

- Export: Klik tombol “Save”



Membuat .arff dari .csv

- ARFF (Attribute-Relation File Format) terdiri dari 2 bagian, yaitu:
 - Header: menjelaskan tipe atribut.
 - Bagian Data: meliputi data yang dipisah dengan tanda koma.

Membuat .arff dari .csv(Lanj.)

- Ubah tabel berikut ke file .csv

x	y	kelas
2	4	kuadrat
2	3	tambah
3	9	kuadrat
3	4	tambah
7	8	tambah
8	64	kuadrat

Membuat .arff dari .csv(Lanj.)

- Buka file .csv sebelumnya dengan text editor dan tambahkan:

```
@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
```

- Save file dengan ekstensi .arff.

Struktur File .arff

- ◉ Komentar pada awal dataset. Sebaiknya, berisi asal dan konten dataset.

```
% This is a toy example, the UCI weather dataset.  
% Any relation to real weather is purely coincidental.
```

- ◉ Nama internal dataset. Sebaiknya, nama tersebut dimengerti dan mendefinisikan banyak hal.

```
@relation golfWeatherMichigan_1988/02/10_14days
```

Struktur File .arff(Lanj.)

- Di sini, 2 atribut nominal didefinisikan, outlook dan windy. Single quotes (' ') digunakan untuk nilai nominal dengan karakter spesial, koma, atau spasi.

```
@attribute outlook {sunny, overcast, rainy}  
@attribute windy {TRUE, FALSE}
```

- Baris berikut mendefinisikan 2 atribut numerik. Selain real, integer dan numeric bisa digunakan.

```
@attribute temperature real  
@attribute humidity real
```

Struktur File .arff(Lanj.)

- Atribut terakhir adalah default target atau variabel kelas yang digunakan untuk prediksi. Pada contoh ini, atribut nominal digunakan.

```
@attribute play {yes, no}
```

Struktur File .arff(Lanj.)

- Sisa dataset terdiri atas token @data diikuti nilai untuk setiap atribut dipisahkan koma, satu baris per sampel.

```
@data
sunny,FALSE,85,85,no
sunny,TRUE,80,90,no
overcast,FALSE,83,86,yes
rainy,FALSE,70,96,yes
rainy,FALSE,68,80,yes
```


Ringkasan dan Visualisasi Data

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply

Current relation
Relation: iris
Instances: 150
Attributes: 5

Attributes
All None Invert Pattern

3

No.	Name
1	<input checked="" type="checkbox"/> sepal.length
2	<input type="checkbox"/> sepal.width
3	<input type="checkbox"/> petal.length
4	<input type="checkbox"/> petal.width
5	<input type="checkbox"/> class

Remove

Selected attribute
Name: sepal.length
Missing: 0 (0%)
Distinct: 35
Type: Numeric
Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

4

Class: class (Nom) Visualize All

5

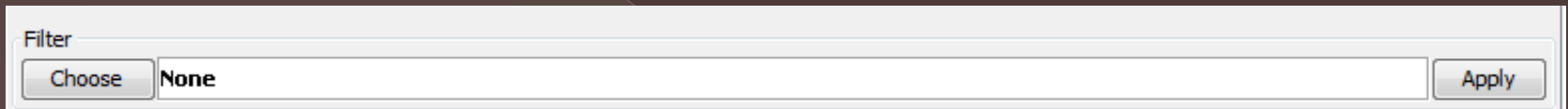
16 30 34 28 25 10 7

4.3 6.1 7.9

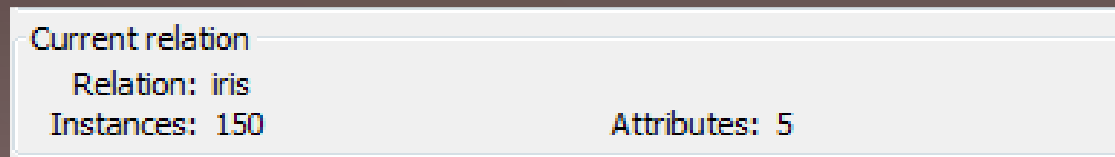
Status OK Log x 0

Ringkasan dan Visualisasi Data(Lanj.)

- ◉ (1) digunakan untuk mengolah data seperti diskretisasi.

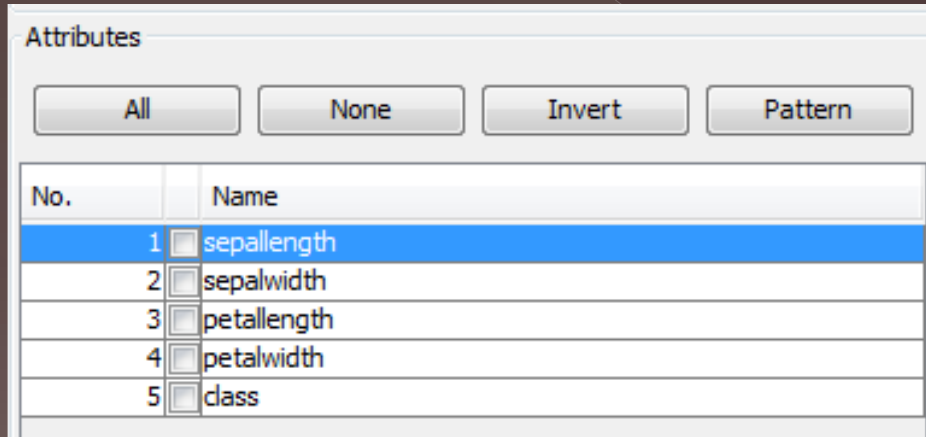


- ◉ (2) digunakan untuk melihat informasi relasi, jumlah data, dan jumlah atribut pada dataset.



Ringkasan dan Visualisasi Data(Lanj.)

- ◉ (3) digunakan untuk melihat isi setiap atribut.



The image shows a software dialog box titled "Attributes". At the top, there are four buttons: "All", "None", "Invert", and "Pattern". Below these buttons is a table with two columns: "No." and "Name". The table contains five rows of attributes, each with a checkbox in the "No." column. The first row, "1" and "sepallength", is highlighted in blue. The other rows are "2" and "sepalwidth", "3" and "petallength", "4" and "petalwidth", and "5" and "class".

No.	Name
1 <input checked="" type="checkbox"/>	sepallength
2 <input type="checkbox"/>	sepalwidth
3 <input type="checkbox"/>	petallength
4 <input type="checkbox"/>	petalwidth
5 <input type="checkbox"/>	class

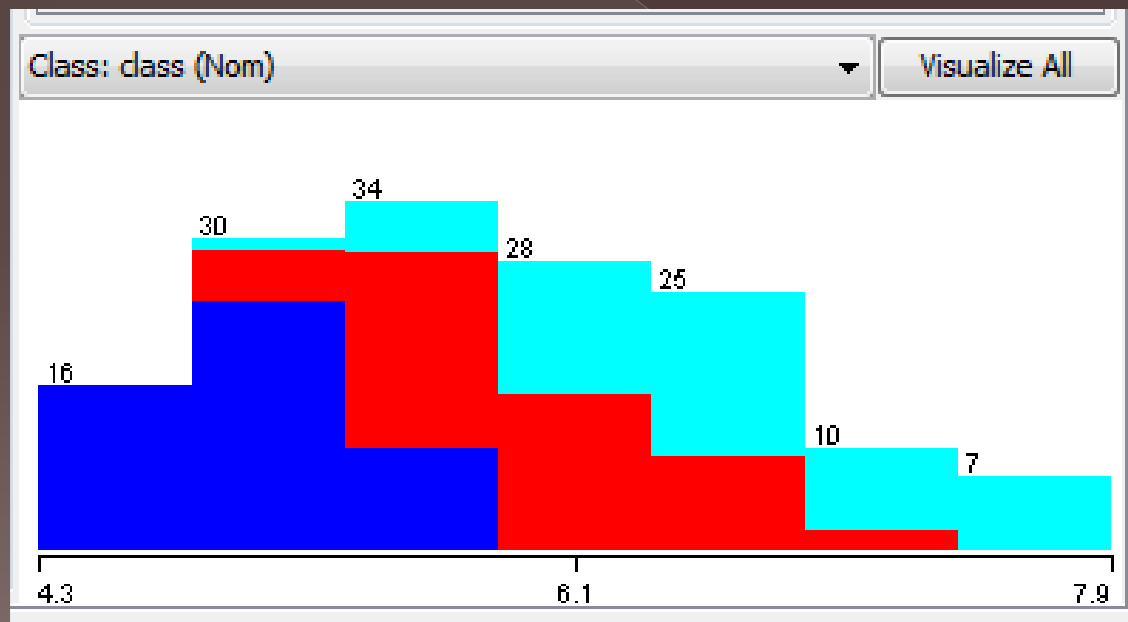
Ringkasan dan Visualisasi Data(Lanj.)

- ◉ (4) digunakan untuk melihat properties atribut: Nama, jumlah nilai yang hilang, jumlah nilai yang berbeda, jumlah dan persentase nilai unik, dan informasi statistik.

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

Ringkasan dan Visualisasi Data(Lanj.)

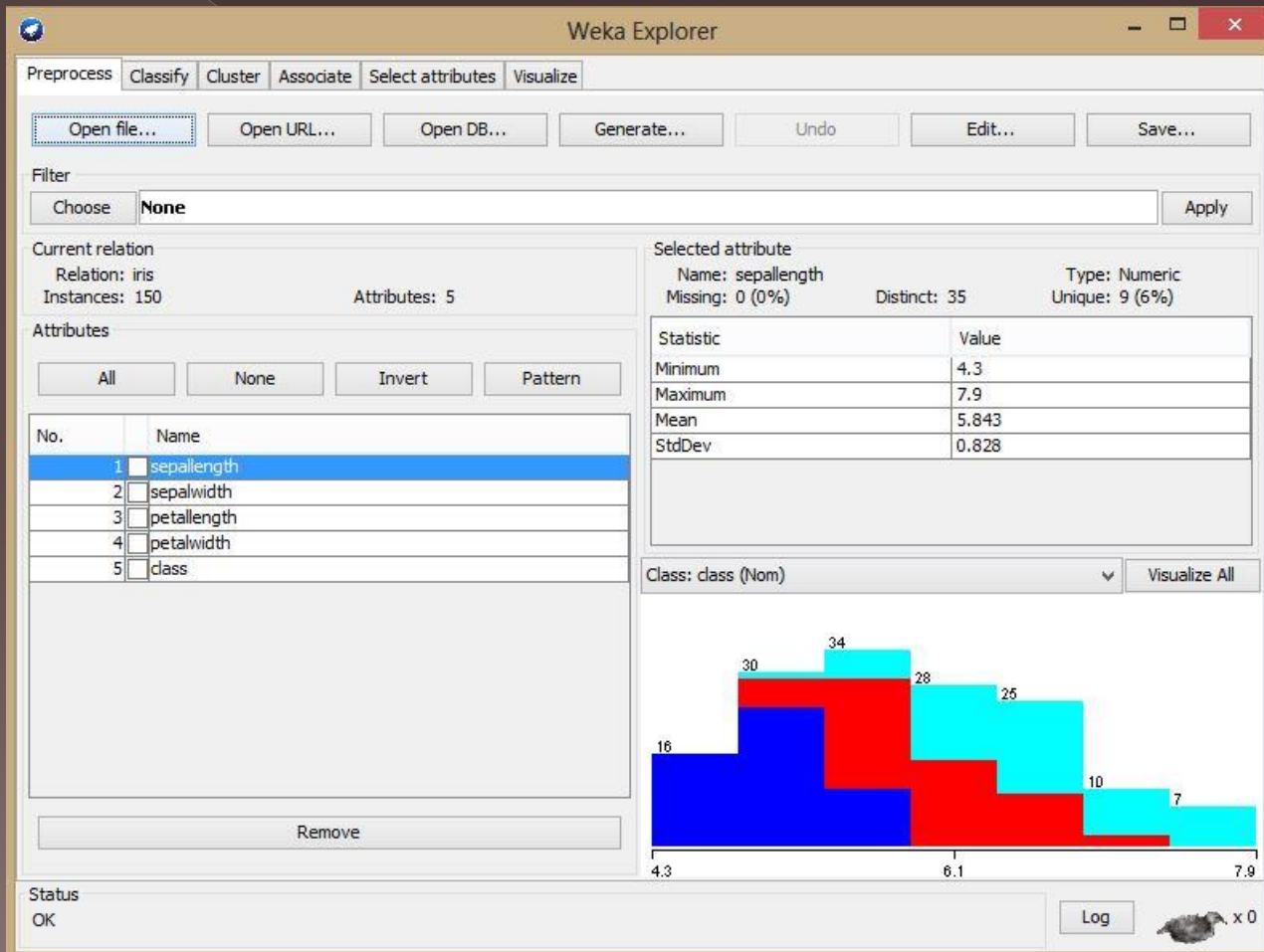
- ◉ (5) adalah visualisasi data. Warna menunjukkan label / kelas dataset.



Contoh Praproses Data

- Contoh yang diberikan adalah praproses sederhana untuk memilih atribut.
- Langkah-langkahnya:
 - Buka iris.arff
 - Misalkan, atribut sepallength dan sepalwidth ingin dihapus.
 - Pilih kedua atribut tersebut kemudian klik remove.
 - Atribut yang digunakan hanya tinggal petallength dan petalwidth saja.

Contoh Praproses Data(Lanj.)



Contoh Praproses Data(Lanj.)

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **None** Apply

Current relation
Relation: iris
Instances: 150
Attributes: 5

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> sepallength
2	<input checked="" type="checkbox"/> sepalwidth
3	<input type="checkbox"/> petallength
4	<input type="checkbox"/> petalwidth
5	<input type="checkbox"/> class

Remove

Selected attribute
Name: sepalwidth
Missing: 0 (0%)
Distinct: 23
Type: Numeric
Unique: 5 (3%)

Statistic	Value
Minimum	2
Maximum	4.4
Mean	3.054
StdDev	0.434

Class: class (Nom) Visualize All

Bin Range (sepalwidth)	Frequency
2.0 - 2.2	8
2.2 - 2.4	16
2.4 - 2.6	33
2.6 - 2.8	51
2.8 - 3.0	24
3.0 - 3.2	12
3.2 - 3.4	4
3.4 - 3.6	2

Status
OK

Log x 0

Contoh Praproses Data(Lanj.)

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **None** Apply

Current relation
Relation: iris-weka.filters.unsupervised.attribute.Remove-R1-2
Instances: 150 Attributes: 3

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> petallength
2	<input type="checkbox"/> petalwidth
3	<input type="checkbox"/> class

Remove

Selected attribute
Name: petallength Type: Numeric
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Class: class (Nom) Visualize All

Petal Length Range	setosa (blue)	versicolour (red)	virginica (cyan)
1.0 - 1.9	50	3	0
2.0 - 2.9	0	34	0
3.0 - 3.9	0	34	0
4.0 - 4.9	0	0	47
5.0 - 5.9	0	0	16
6.0 - 6.9	0	0	47

Status
OK

Log x 0

Latihan

- ◉ Silakan kerjakan latihan di LKP 2.
- ◉ Dikumpulkan perkelompok ke syeiva.nd@gmail.com.
- ◉ Subjek & nama file: [\[NamaKelompok\]_LKP2](#)
- ◉ Waktu pengerjaan: 1 jam.
- ◉ Toleransi pengumpulan sampai [pukul 20.00 WIB](#).