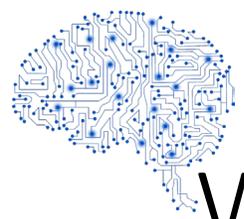


# *Робота з Века*

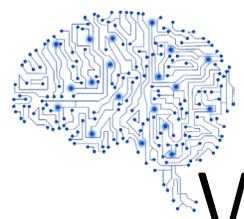
Професор, д.е.н. Ставицький А.В.

*Середовище Века*



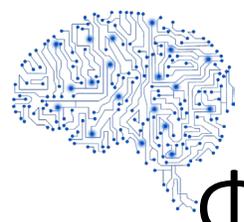
# Weka

- Програмне забезпечення для машинного навчання/інтелектуального аналізу даних, написане на Java (розповсюджується за публічною ліцензією GNU)
- Використовується для досліджень, освіти та застосування
- Містить повний набір інструментів попередньої обробки даних, алгоритмів навчання та методів оцінювання
- Має графічний інтерфейс користувача (включаючи візуалізацію даних)
- Має модуль для порівняння алгоритмів навчання



# WEKA

- Скачування: [https://waikato.github.io/weka-wiki/downloading\\_weka/](https://waikato.github.io/weka-wiki/downloading_weka/)
- Остання стабільна версія – 3.8.



# Файли

```
@relation heart-disease-simplified
```

```
@attribute age numeric
```

```
@attribute sex { female, male}
```

```
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
```

```
@attribute cholesterol numeric
```

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

```
@attribute class { present, not_present}
```

```
@data
```

```
63,male,typ_angina,233,no,not_present
```

```
67,male,asympt,286,yes,present
```

```
67,male,asympt,229,yes,present
```

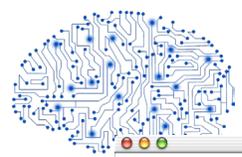
```
38,female,non_anginal,?,no,not_present
```

```
...
```

Числовий атрибут

Номінальний атрибут

Файл у  
ARFF форматі



```
Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.

> help

Command must be one of:
java <classname> <args>
break
kill
cls
exit
help <command>
```

**Weka GUI Chooser**

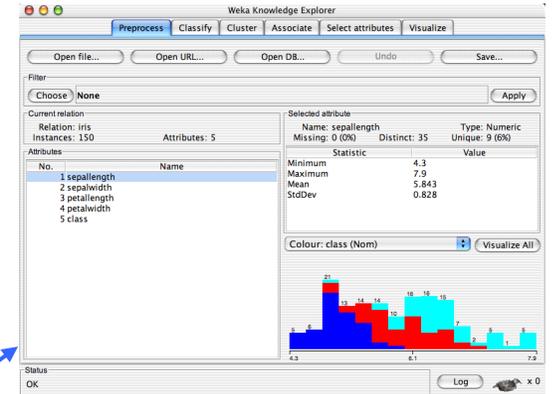
Program Visualization Tools Help

**Applications**

- Explorer
- Experimenter
- KnowledgeFlow
- Workbench
- Simple CLI

**WEKA**  
The University of Waikato

Waikato Environment for Knowledge Analysis  
Version 3.8.5  
(c) 1999 - 2020  
The University of Waikato  
Hamilton, New Zealand



**Weka Experiment Environment**

Setup Run Analyse

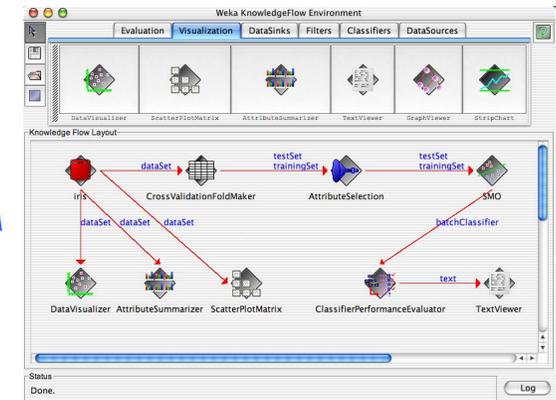
Experiment Configuration Mode: Simple Advanced

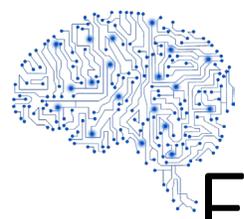
Results Destination: JDBC database URL: jdbc:tdb=experiments.prp

Experiment Type: Cross-validation Number of folds: 10

Iteration Control: Number of repetitions: 10

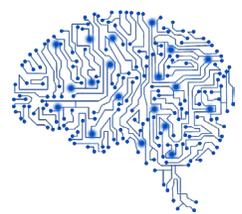
Algorithms: J48 -C 0.25 -M 2 NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a NaiveBayes





# Explorer: попередня обробка даних

- Дані можна імпортувати з файлу в різних форматах: ARFF, CSV, C4.5, бінарний
- Дані також можна читати з URL-адреси або з бази даних SQL (за допомогою JDBC)
- Інструменти попередньої обробки в WEKA називаються «фільтрами»
- WEKA містить фільтри для:
  - дискретизація,
  - нормалізація,
  - повторна дискретизація,
  - вибір атрибутів,
  - перетворення,
  - комбінування атрибутів



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter:  **None**

<b>Current relation</b> Relation: None Instances: None Attributes: None	<b>Selected attribute</b> Name: None Missing: None Distinct: None Type: None Unique: None
--	--

Attributes

---

Status: Welcome to the Weka Knowledge Explorer  x 0



### Weka Knowledge Explorer

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

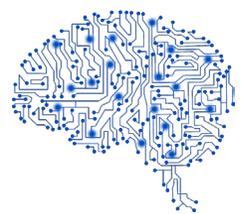
Open file... | Open URL... | Open DB... | Undo | Save...

Filter:  **None**

Current relation		Selected attribute	
Relation: None	Attributes: None	Name: None	Type: None
Instances: None		Missing: None	Distinct: None
			Unique: None

Attributes:

Status: Welcome to the Weka Knowledge Explorer  x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter:  **None**

**Current relation**

Relation: iris  
Instances: 150      Attributes: 5

**Attributes**

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

**Selected attribute**

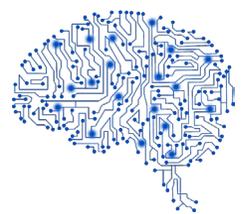
Name: sepalength      Type: Numeric  
Missing: 0 (0%)      Distinct: 35      Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom)

Bin Range	Count
4.3 - 4.6	5
4.6 - 4.9	6
4.9 - 5.2	21
5.2 - 5.5	13
5.5 - 5.8	14
5.8 - 6.1	14
6.1 - 6.4	10
6.4 - 6.7	16
6.7 - 7.0	16
7.0 - 7.3	15
7.3 - 7.6	7
7.6 - 7.9	2
7.9 - 8.2	5
8.2 - 8.5	1
8.5 - 8.8	5

Status: OK  x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter:  **None**

Current relation

Relation: iris  
Instances: 150      Attributes: 5

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

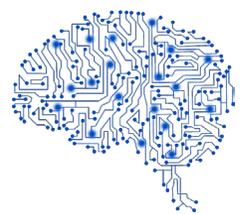
Name: sepalength      Type: Numeric  
Missing: 0 (0%)      Distinct: 35      Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom)

Bin Range	Count
4.3 - 4.6	5
4.6 - 4.9	6
4.9 - 5.2	21
5.2 - 5.5	13
5.5 - 5.8	14
5.8 - 6.1	14
6.1 - 6.4	10
6.4 - 6.7	16
6.7 - 7.0	16
7.0 - 7.3	15
7.3 - 7.6	7
7.6 - 7.9	2
7.9 - 8.2	5
8.2 - 8.5	1
8.5 - 8.8	5

Status: OK  x 0



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...   Undo   Save...

Filter:

**Current relation**

Relation: iris  
Instances: 150                      Attributes: 5

**Attributes**

No.	Name
1	sepalwidth
2	sepalwidth
3	petalwidth
4	petalwidth
5	class

**Selected attribute**

Name: class                      Type: Nominal  
Missing: 0 (0%)                  Distinct: 3                      Unique: 0 (0%)

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

Colour: class (Nom)

50

50

50

Status: OK  x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter:

**Current relation**

Relation: iris  
Instances: 150      Attributes: 5

**Attributes**

No.	Name
1	sepalwidth
2	sepalwidth
3	petalwidth
4	petalwidth
5	class

**Selected attribute**

Name: class      Type: Nominal  
Missing: 0 (0%)      Distinct: 3      Unique: 0 (0%)

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

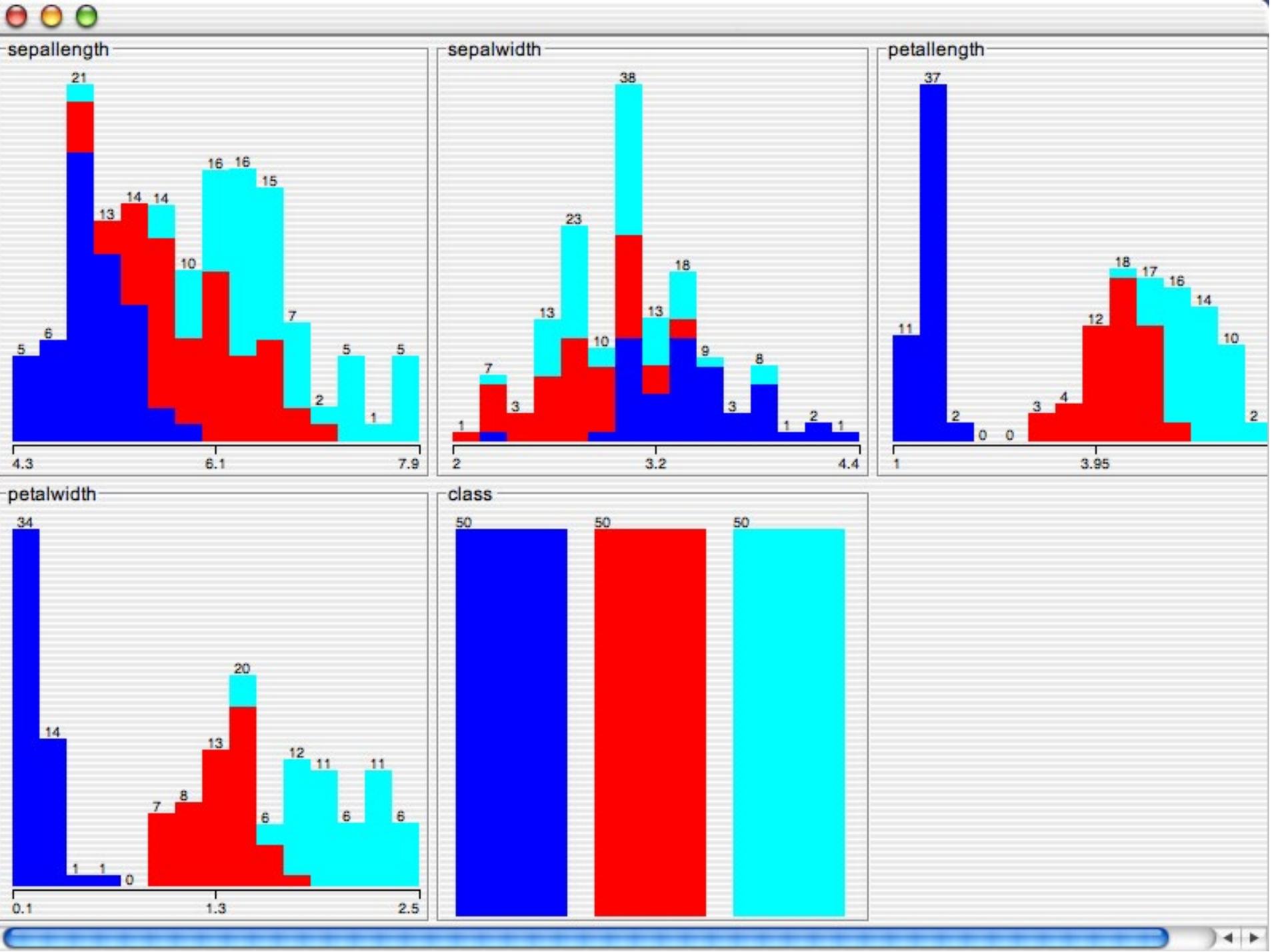
Colour: class (Nom)

50

50

50

Status: OK  x 0





### Weka Knowledge Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...UndoSave...

Filter

Choose None Apply

Current relation

Relation: iris  
Instances: 150      Attributes: 5

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

Attributes

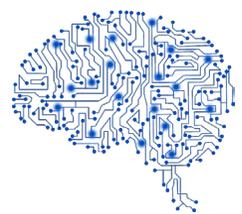
No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Colour: class (Nom) ▼Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status

OK Log x 0



### Weka Knowledge Explorer

**Preprocess**   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...   Undo   Save...

Filter:

Current relation: Relation: iris   Instances: 150   Attributes: 5

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

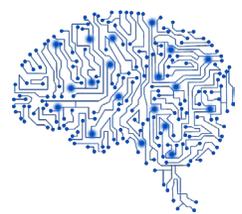
Attributes:

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Colour: class (Nom)

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status: OK  x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter

- weka
  - filters
    - unsupervised
      - attribute
      - instance

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

Colour: class (Nom) Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status

OK Log x 0



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...   Undo   Save...

Filter

- weka
  - filters
    - unsupervised
      - attribute
      - instance

Apply

Selected attribute

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

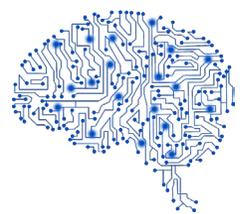
Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom)   Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status: OK

Log   x 0



### Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

**Filter**

- weka
  - filters
    - unsupervised
      - attribute
        - Add
        - AddCluster
        - AddExpression
        - AddNoise
        - Copy
        - Discretize**
        - FirstOrder
        - MakeIndicator
        - MergeTwoValues
        - NominalToBinary
        - Normalize
        - NumericToBinary
        - NumericTransform
        - Obfuscate
        - PKIDiscretize
        - Remove
        - RemoveType

**Selected attribute**

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

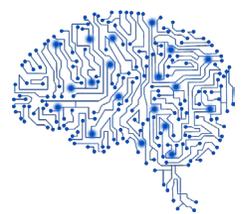
Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom) Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

**Status**

OK Log x 0



### Weka Knowledge Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...UndoSave...

Filter

Choose **Discretize -B 10 -R first-last** Apply

Current relation

Relation: iris  
Instances: 150      Attributes: 5

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

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

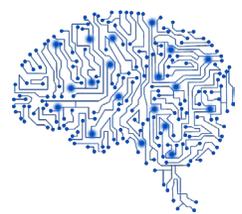
Colour: class (Nom) Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status

OK

Log x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter:  **Discretize -B 10 -R first-last**

Current relation: Relation: iris, Instances: 150, Attributes: 5

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

Attributes:

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Colour: class (Nom)

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	3
3.5 - 4.0	4
4.0 - 4.5	12
4.5 - 5.0	18
5.0 - 5.5	17
5.5 - 6.0	16
6.0 - 6.5	14
6.5 - 7.0	10
7.0 - 7.5	2
7.5 - 8.0	4

Status: OK  x 0



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...   Undo   Save...

Filter: **Discretize -B 10 -R first-last**

Current relation: Relation: iris, Instances: 150, Attributes:

No.	Name
1	sepal.length
2	sepal.width
3	petal.length
4	petal.width
5	class

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About: An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. [More](#)

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: False

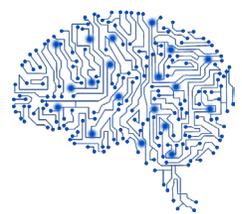
: Numeric  
: 10 (7%)

[Visualize All](#)

Open...   Save...   OK   Cancel

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	18
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status: OK [Log](#) x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter: Choose **Discretize -B 10 -R first-last** Apply

Current relation: Relation: iris Instances: 150 Attributes: !

No.	Name
1	sepal.length
2	sepal.width
3	petal.length
4	petal.width
5	class

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About: An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: False

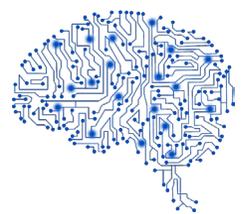
Visualize All

Open... | Save... | OK | Cancel

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	2
2.0 - 2.5	0
2.5 - 3.0	0
3.0 - 3.5	3
3.5 - 4.0	4
4.0 - 4.5	12
4.5 - 5.0	10
5.0 - 5.5	2
5.5 - 6.0	4

Status: OK

Log x 0



### Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter: Choose **Discretize -B 10 -R first-last** Apply

Current relation: Relation: iris Instances: 150 Attributes: !

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About: An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: True

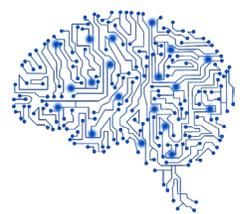
Open... Save... OK Cancel

: Numeric  
: 10 (7%)

Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	2
2.0 - 2.5	0
2.5 - 3.0	0
3.0 - 3.5	3
3.5 - 4.0	4
4.0 - 4.5	12
4.5 - 5.0	10
5.0 - 5.5	2
5.5 - 6.0	4

Status: OK Log x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter: Choose **Discretize -B 10 -R first-last** Apply

Current relation: Relation: iris, Instances: 150, Attributes:

No.	Name
1	sepalwidth
2	sepalwidth
3	petalwidth
4	petalwidth
5	class

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

About: An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: True

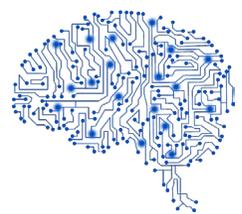
Visualize All

OK

Bin Range	Count
1.0 - 1.9	11
2.0 - 2.9	2
3.0 - 3.9	0
4.0 - 4.9	0
5.0 - 5.9	3
6.0 - 6.9	4
7.0 - 7.9	12
8.0 - 8.9	4
9.0 - 9.9	10
10.0 - 10.9	2
11.0 - 11.9	4

Status: OK

Log x 0



### Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter: Choose **Discretize -F -B 10 -R first-last** Apply

**Current relation**  
Relation: iris  
Instances: 150      Attributes: 5

**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

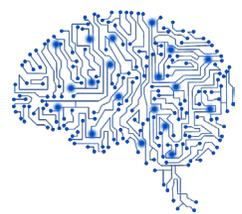
**Attributes**

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Colour: class (Nom) Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	0
3.5 - 4.0	3
4.0 - 4.5	4
4.5 - 5.0	12
5.0 - 5.5	18
5.5 - 6.0	17
6.0 - 6.5	16
6.5 - 7.0	14
7.0 - 7.5	10
7.5 - 8.0	2
8.0 - 8.5	4

Status: OK Log x 0



### Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter: **Discretize -F -B 10 -R first-last** Apply

**Current relation**  
Relation: iris  
Instances: 150      Attributes: 5

**Attributes**

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

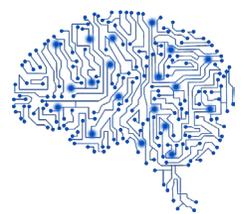
**Selected attribute**  
Name: petallength      Type: Numeric  
Missing: 0 (0%)      Distinct: 43      Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom) Visualize All

Bin Range	Count
1.0 - 1.5	11
1.5 - 2.0	37
2.0 - 2.5	2
2.5 - 3.0	0
3.0 - 3.5	3
3.5 - 4.0	4
4.0 - 4.5	12
4.5 - 5.0	18
5.0 - 5.5	17
5.5 - 6.0	16
6.0 - 6.5	14
6.5 - 7.0	10
7.0 - 7.5	2
7.5 - 8.0	4

Status: OK Log x 0



### Weka Knowledge Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...UndoSave...

Filter

ChooseDiscretize -F -B 10 -R first-lastApply

Current relation

Relation: iris-weka.filters.unsupervised.attribute.Disc...

Instances: 150      Attributes: 5

Selected attribute

Name: petallength      Type: Nominal

Missing: 0 (0%)      Distinct: 10      Unique: 0 (0%)

Label	Count
'(-inf-1.45]'	23
'(1.45-1.55]'	14
'(1.55-1.8]'	11
'(1.8-3.95]'	13
'(3.95-4.35]'	14
'(4.35-4.65]'	15
'(4.65-5.05]'	18

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

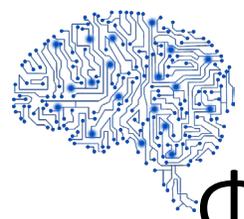
Colour: class (Nom) Visualize All

Label	Count
'(-inf-1.45]'	23
'(1.45-1.55]'	14
'(1.55-1.8]'	11
'(1.8-3.95]'	13
'(3.95-4.35]'	14
'(4.35-4.65]'	15
'(4.65-5.05]'	18

Status

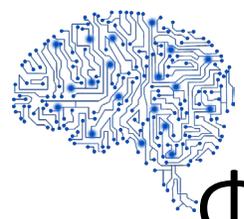
OK

Log x 0



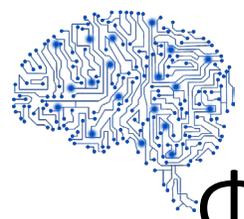
# Фільтри: Supervised-Attribute

- AddClassification - Додає до набору даних класифікаційні прогнози з іншого набору даних.
- AttributeSelection - Відбирає найважливіші атрибути для подальшого використання.
- Discretize - Змінює неперервні атрибути на дискретні.



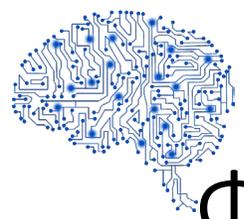
# Фільтри: Supervised-instance

- **ClassBalancer** - цей фільтр балансує розподіл класів у наборі даних, шляхом зменшення кількості записів в домінуючому класі і збільшення кількості записів у менш представлених класах.
- **Resample** - цей фільтр дозволяє збільшувати або зменшувати кількість записів в наборі даних, використовуючи підхід з випадковим відбором записів з дублюванням або видаленням.
- **SpreadSubsample** - цей фільтр зменшує кількість записів у наборі даних, зберігаючи при цьому більш точне представлення розподілу класів, порівняно з **Resample**. Він використовується для швидкої попередньої обробки великих наборів даних перед подальшою обробкою, наприклад, перед побудовою моделей.
- **StratifiedRemoveFolds** - цей фільтр видаляє частину записів з набору даних, розділяючи його на кілька фолдів (складів) за допомогою критерію стратифікації за класом. Він зазвичай використовується для оцінки моделей, коли потрібно виконати перехресну перевірку, але необхідно забезпечити, щоб кожен фолд мав приблизно однакову кількість записів з кожного класу.



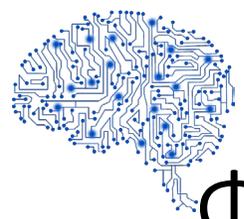
# Фільтри: Unsupervised-Attribute – 1

- Add - Додає новий атрибут до даних зі значенням за замовчуванням або обчисленим виразом.
- AddCluster - Додає атрибут, що вказує, до якої кластерної групи належать екземпляри.
- AddExpression - Додає атрибут на основі виразу, який може використовувати інші атрибути.
- AddID - Додає атрибут з унікальним ідентифікатором для кожного екземпляру.
- AddNoise - Додає шум до числових атрибутів.
- AddUserFields - Додає атрибути з введеними користувачем значеннями.
- AddValues - Додає нові значення до наявних атрибутів.
- CartesianProduct - Об'єднує два або більше набори даних в один, створюючи всі можливі комбінації екземплярів.
- Center - Центрує дані відносно середнього значення або медіани.
- ChangeDateFormat - Змінює формат дати з одного формату на інший.
- ClassAssigner - Додає клас до даних, використовуючи значення з іншого атрибуту.
- ClusterMembership - Додає атрибут, який вказує, до якої кластерної групи належить екземпляр.
- Copy - Створює копію атрибуту або набору даних.
- DateToNumeric - Перетворює дату в числовий формат.
- Discretize - Кількісний атрибут дискретизується у кілька дискретних атрибутів.



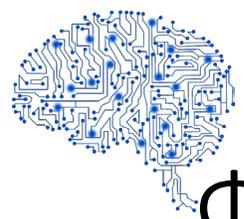
# Фільтри: Unsupervised-Attribute – 2

- FirstOrder - Перетворює часові ряди в перший порядок.
- FixedDictionaryStringToWordVector - Перетворює рядки у вектори термів з використанням заздалегідь заданого словника.
- InterquartileRange - Видаляє значення, що виходять за межі міжквартильного діапазону.
- KernelFilter - Фільтрує дані за допомогою ядрової функції.
- Makeindicator: створює нові бінарні атрибути на основі значень існуючих категоріальних атрибутів.
- MathExpression: обчислює новий атрибут на основі математичного виразу, що використовує існуючі атрибути.
- MergeInfrequentNominalValues: замінює рідкі значення категоріальних атрибутів на значення "рідкі", щоб зменшити розмір даних.
- MergeManyValues: замінює групу значень категоріального атрибуту на одне значення.
- MergeTwoValues: замінює два значення категоріального атрибуту на одне значення.
- NnominalToString: перетворює категоріальні значення на рядкові.
- NominalToBinary: перетворює категоріальні атрибути на бінарні, зберігаючи тільки два можливі значення.
- NominalToString: перетворює категоріальні значення на рядкові.
- Normalize: нормалізує числові атрибути до заданого діапазону.
- NumericCleaner: видаляє некоректні значення з числових атрибутів.
- NumericToBinary: перетворює числові атрибути на бінарні за допомогою порогового значення.



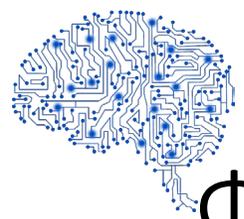
# Фільтри: Unsupervised-Attribute – 3

- NumericToDate: перетворює числові значення на дати за заданим форматом.
- NumericToNominal: перетворює числові значення на категоріальні за допомогою бінування.
- NumericTransform: обчислює нові числові атрибути на основі існуючих атрибутів та математичних формул.
- Obfuscate: замінює значення атрибутів на випадкові значення для забезпечення конфіденційності.
- OrdinalToNumeric: перетворює категоріальні атрибути на числові, зберігаючи порядок категорій.
- PartitionedMultiFilter: виконує послідовну обробку даних декількома фільтрами з використанням різних наборів параметрів.
- PKIDiscretize: Дискретизація числових атрибутів з використанням методу Pekalskiego і Jaroszewicza.
- Principalcomponents: Виконання аналізу головних компонент (PCA) для зменшення розмірності даних.
- Randomprojection: Застосування випадкових проєкцій для зменшення розмірності даних.
- Randomsubset: Вибір випадкового підмножини даних.
- Rename Attribute: Перейменування атрибутів даних.
- Remove: Вилучення з даних об'єктів або атрибутів за заданими критеріями.
- RemoveByName: Вилучення атрибутів даних за назвою.
- RemoveType: Вилучення атрибутів зазначеного типу (числові, категоріальні тощо).
- RemoveUseless: Вилучення атрибутів з малим значенням інформативності для класифікації або кластеризації.



# Фільтри: Unsupervised-Attribute – 4

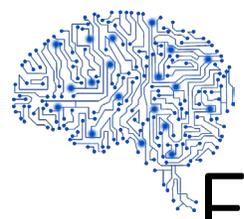
- `RenameNominalValues`: Перейменування категоріальних значень атрибутів даних.
- `Reorder`: Зміна порядку атрибутів даних.
- `ReplaceMissingValues`: Заміна пропущених значень у даних.
- `ReplaceMissingWithUserConstant`: Заміна пропущених значень у даних заданим користувачем константним значенням.
- `ReplaceWithMissingValue`: Заміна значень у даних на пропущені значення.
- `SortLabels`: Сортування категорійних значень у заданому порядку.
- `Standardize`: Стандартизація числових атрибутів даних до середнього значення 0 та стандартного відхилення 1.
- `StringToNominal`: Конвертація строкових атрибутів у категоріальні.
- `StringToWordVector`: Конвертація строкових атрибутів в числові вектори, що відображають частоту використання слів.
- `Swarvalues`: Заміна значень двох атрибутів між собою.
- `TimeSeriesDelta`: Обчислення змін у часових рядках за попередніми спостереженнями.
- `TimeSeriesTranslate`: Трансформація часових рядків шляхом зсуву значень в часі.



# Фільтри: Unsupervised-instance

- NonSparseToSparse - перетворення неповних матриць у матриці розрідженого формату.
- Randomize: цей фільтр випадковим чином перемішує інстанси у наборі даних.
- RemoveDuplicates - вилучення дублікатів екземплярів.
- RemoveFolds: цей фільтр видаляє задану кількість згорток (folds) у наборі даних. Залишається тільки одна згортка для тестування.
- RemoveFrequentValues: цей фільтр видаляє атрибути, які мають часті значення. Дозволяє зменшити розмірність даних.
- RemovePercentage: цей фільтр видаляє випадкові відсотки інстансів у наборі даних.
- RemoveRange: цей фільтр видаляє інстанси з заданих діапазонів значень для вказаних атрибутів.
- RemoveWithValues - вилучення екземплярів, які мають певні значення для заданого атрибуту.
- Resample: цей фільтр зменшує або збільшує розмір набору даних шляхом випадкового вибору інстансів для збереження або видалення.
- SubsetByExpression: цей фільтр дозволяє створювати підмножини даних на основі виразу, який використовує атрибути та їх значення. Наприклад, можна вибрати всі інстанси, у яких значення атрибута "age" більше 30 і менше 50.

# Класифікатори



# Explorer: створення «класифікаторів»

- Класифікатори у WEKA є моделями для прогнозування номінальних або числових величин
- Реалізовані схеми навчання включають:
  - Деревя рішень і списки, класифікатори на основі екземплярів, опорні векторні машини, багатошарові перцептрони, логістичну регресію, мережі Байєса, ...
- До «мета»-класифікаторів належать:
  - Пакетування, посилення, стекування, вихідні коди з виправленням помилок, локально зважене навчання...



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Classifier output

Status

OK

Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose ZeroR

Test options

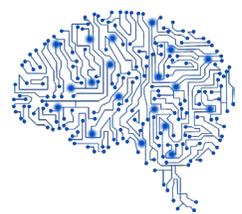
- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output

Status: OK   x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

- weka
  - classifiers
    - bayes
    - functions
    - lazy
    - meta
    - misc
    - trees
      - adtree
        - DecisionStump
        - Id3
      - j48
        - J48**
      - lmt
      - m5
        - RandomForest
        - RandomTree
        - REPTree
        - UserClassifier
    - rules

Classifier output

Status

OK

Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Classifier output

Status

OK

Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) class

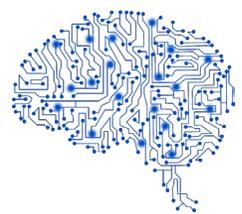
Start Stop

Result list (right-click for options)

Classifier output

Status

OK Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) class

Result list (right-click for options)

Status: OK

Log  x 0

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

binarySplits

confidenceFactor

minNumObj

numFolds

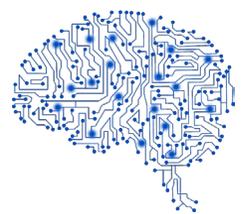
reducedErrorPruning

saveInstanceData

subtreeRaising

unpruned

useLaplace



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:

- Use training set
- Supplied test set (Set...)
- Cross-validation Folds: 10
- Percentage split %: 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Status: OK

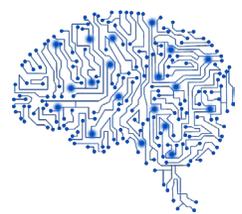
Log x 0

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

- binarySplits: False
- confidenceFactor: 0.25
- minNumObj: 2
- numFolds: 3
- reducedErrorPruning: False
- saveInstanceData: False
- subtreeRaising: True
- unpruned: False
- useLaplace: False

Open... Save... **OK** Cancel



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier  
Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Classifier output

Status  
OK

Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output

Status: OK

 x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

(Nom) class

Result list (right-click for options)

Classifier output

Status

OK   x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

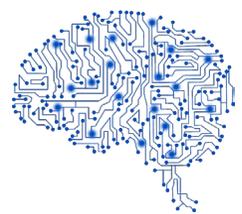
(Nom) class

Result list (right-click for options)

Classifier output

Status: OK

 x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output

Classifier evaluation opt

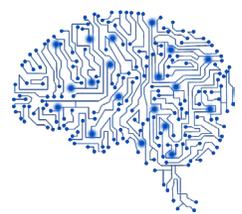
- Output model
- Output per-class stats
- Output entropy evaluation measures
- Output confusion matrix
- Store predictions for visualization
- Output text predictions on test set
- Cost-sensitive evaluation

Random seed for XVal / % Split

Status

OK

 x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output

Classifier evaluation opt

- Output model
- Output per-class stats
- Output entropy evaluation measures
- Output confusion matrix
- Store predictions for visualization
- Output text predictions on test set
- Cost-sensitive evaluation

Random seed for XVal / % Split

Status: OK

x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose J48 -C 0.25 -M 2

Test options:

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

(Nom) class

Result list (right-click for options)

Classifier output

Status: OK

 x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier  
Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) class

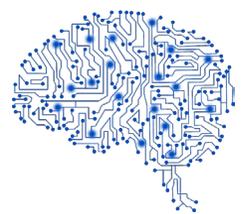
Start Stop

Result list (right-click for options)

Classifier output

Status  
OK

Log  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set **Set...**
- Cross-validation Folds **10**
- Percentage split % **66**

**More options...**

(Nom) class **▼**

**Start** **Stop**

Result list (right-click for options)

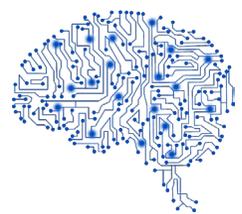
- 11:49:05 - trees.j48.J48

Classifier output

```
=== Run information ===  
  
Scheme:      weka.classifiers.trees.j48.J48 -C 0.25 -M 2  
Relation:    iris  
Instances:   150  
Attributes:  5  
             sepallength  
             sepalwidth  
             petallength  
             petalwidth  
             class  
Test mode:   split 66% train, remainder test  
  
=== Classifier model (full training set) ===  
  
J48 pruned tree  
-----  
  
petalwidth <= 0.6: Iris-setosa (50.0)  
petalwidth > 0.6  
|   petalwidth <= 1.7  
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)  
|   |   petallength > 4.9  
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)  
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)  
|   |   petalwidth > 1.7: Iris-virginica (46.0/1.0)  
|  
Number of Leaves :      5
```

Status: OK

**Log**  x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set **Set...**
- Cross-validation Folds **10**
- Percentage split % **66**

**More options...**

**(Nom) class** ▾

**Start** **Stop**

Result list (right-click for options)

**11:49:05 - trees.j48J48**

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.trees.j48.J48 -C 0.25 -M 2
Relation:    iris
Instances:   150
Attributes:  5
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
Test mode:   split 66% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
-----

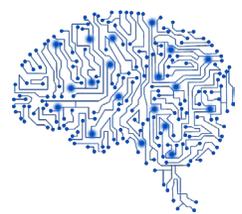
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)
|   |   petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves :      5
```

Status: OK

**Log**  x 0





Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set **Set...**
- Cross-validation Folds **10**
- Percentage split % **66**

**More options...**

**(Nom) class**

**Start** **Stop**

Result list (right-click for options)

- 11:49:05 - trees.j48.J48**

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

=== Detailed Accuracy By Class ===

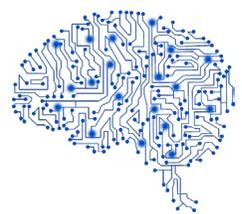
TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

=== Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status: OK

**Log** x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options

- Use training set
- Supplied test set **Set...**
- Cross-validation Folds **10**
- Percentage split % **66**

**More options...**

(Nom) class **▼**

**Start** **Stop**

Result list (right-click for options)

- 11:49:05 - trees.j48.J48**

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

=== Detailed Accuracy By Class ===

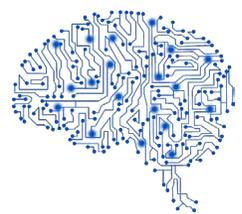
TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

=== Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status: OK

**Log** x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66

(Nom) class

Classifier output

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

=== Detailed Accuracy By Class ===

Recall	F-Measure	Class
1	1	Iris-setosa
1	0.95	Iris-versicolor
0.882	0.938	Iris-virginica

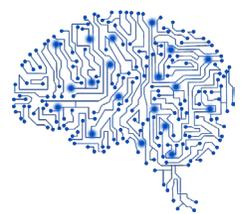
Result list (right-click for options)

- 11:49:05 - trees.j48.J48

- View in main window
- View in separate window
- Save result buffer
- Load model
- Save model
- Re-evaluate model on current test set
- Visualize classifier errors
- Visualize tree**
- Visualize margin curve
- Visualize threshold curve
- Visualize cost curve

Status: OK

x 0



Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **J48 -C 0.25 -M 2**

Test options:  
 Use training set  
 Supplied test set  
 Cross-validation  
 Percentage split  
More options

(Nom) class

Start

Result list (right-click for details):  
11:49:05 - trees.j48.J48

Weka Classifier Tree Visualizer: 11:49:05 - trees.j48.J48 (iris)

Tree View

```
graph TD
    A(petalwidth) -- "<= 0.6" --> B(Iris-setosa 50.0)
    A -- "> 0.6" --> C(petalwidth)
    C -- "<= 1.7" --> D(petalwidth)
    C -- "> 1.7" --> E(Iris-virginica 46.0/1.0)
    D -- "<= 4.9" --> F(Iris-versicolor 48.0/1.0)
    D -- "> 4.9" --> G(petalwidth)
    G -- "<= 1.5" --> H(Iris-virginica 3.0)
    G -- "> 1.5" --> I(Iris-versicolor 3.0/1.0)
```

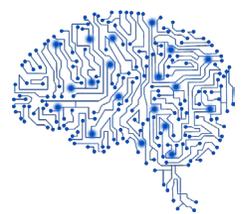
96.0784 %  
3.9216 %

ass  
is-setosa  
is-versicolor  
is-virginica

15	0	0		a = Iris-setosa
0	19	0		b = Iris-versicolor
0	2	15		c = Iris-virginica

Status: OK

Log x 0



### Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

---

**Classifier**  
Choose **J48 -C 0.25 -M 2**

---

**Test options**

Use training set  
 Supplied test set Set...  
 Cross-validation Folds   
 Percentage split %

More options...

**(Nom) class** ▼

Start Stop

---

**Classifier output**

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===  
=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

=== Detailed Accuracy By Class ===

Recall	F-Measure	Class
1	1	Iris-setosa
1	0.95	Iris-versicolor
0.882	0.938	Iris-virginica

---

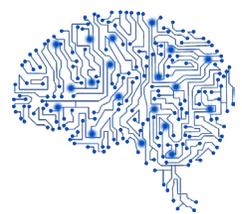
**Result list (right-click for options)**

11:49:05 - trees.j48J48	<ul style="list-style-type: none"><li>View in main window</li><li>View in separate window</li><li>Save result buffer</li><li>Load model</li><li>Save model</li><li>Re-evaluate model on current test set</li><li><b>Visualize classifier errors</b></li><li>Visualize tree</li><li>Visualize margin curve</li><li>Visualize threshold curve</li><li>Visualize cost curve</li></ul>
-------------------------	--

---

**Status**  
OK

Log  x 0



### Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

**Classifier**  
Choose **J48 -C 0.25 -M 2**

**Test options**  
 Use training set  
 Supplied test set  
 Cross-validation  
 Percentage split

X: petallength (Num) Y: petalwidth (Num)  
Colour: class (Nom) Select Instance

Reset Clear Save Jitter

More options (Nom) class Start

Result list (right-click for details)  
11:49:05 - trees.j48

Plot: iris\_predicted

lass  
ris-setosa  
ris-versicolor  
ris-virginica

Class colour  
Iris-setosa Iris-versicolor Iris-virginica

```
0 19 0 | b = Iris-versicolor  
0 2 15 | c = Iris-virginica
```

Status  
OK Log x 0



### Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

---

**Classifier**  
Choose **J48 -C 0.25 -M 2**

**Test options**  
 Use training set  
 Supplied test set Set...  
 Cross-validation Folds   
 Percentage split %   
More options...

(Nom) class ▼

Start Stop

**Result list (right-click for options)**  
**11:49:05 - trees.j48.J48**

**Classifier output**

```
Time taken to build model: 0.24 seconds

=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances      49      96.0784 %
Incorrectly Classified Instances     2       3.9216 %
Kappa statistic                     0.9408
Mean absolute error                  0.0396
Root mean squared error              0.1579
Relative absolute error              8.8979 %
Root relative squared error         33.4091 %
Total Number of Instances           51

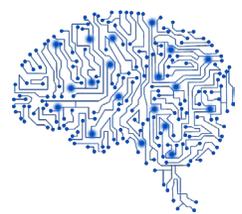
=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
1         0         1          1         1          Iris-setosa
1         0.063    0.905     1         0.95       Iris-versicolor
0.882    0         1          0.882    0.938     Iris-virginica

=== Confusion Matrix ===

 a  b  c  <-- classified as
15  0  0  | a = Iris-setosa
 0 19  0  | b = Iris-versicolor
 0  2 15  | c = Iris-virginica
```

**Status**  
OK Log  x 0



### Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

---

**Classifier**  
Choose **J48 -C 0.25 -M 2**

**Test options**

Use training set  
 Supplied test set Set...  
 Cross-validation Folds   
 Percentage split %

More options...

**(Nom) class** ▼

Start Stop

**Result list (right-click for options)**

11:49:05 - trees.j48.J48
--------------------------

---

**Classifier output**

Time taken to build model: 0.24 seconds

=== Evaluation on test split ===  
=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

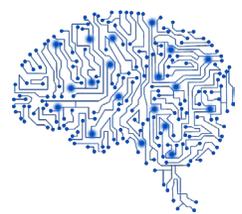
=== Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

---

**Status**  
OK

Log  x 0



### Weka Knowledge Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

---

#### Classifier

- weka
  - classifiers
    - bayes
      - functions
        - LeastMedSq
        - LinearRegression
        - Logistic
        - neural
          - NeuralNetwork**
        - pace
        - supportVector
          - SimpleLinearRegression
          - SimpleLogistic
          - VotedPerceptron
          - Winnow
      - lazy
      - meta
      - misc
      - trees
      - rules

Time taken to build model: 0.24 seconds

Evaluation on test split ===  
Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Absolute error	0.0396	
Mean squared error	0.1579	
Relative absolute error	8.8979 %	
Relative squared error	33.4091 %	
Number of Instances	51	

Confusion Matrix By Class ===

	Actual Class	FP Rate	Precision	Recall	F-Measure	Class
1	a	0	1	1	1	Iris-setosa
2	b	0.063	0.905	1	0.95	Iris-versicolor
3	c	0	1	0.882	0.938	Iris-virginica

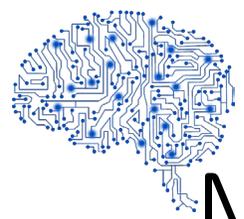
Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

---

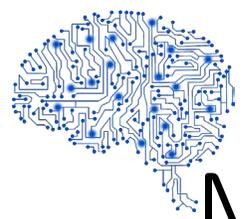
Status: OK

Log x 0



# Методи класифікації у Weka

- 1.Дерева рішень (Decision Trees): J48, RandomTree, REPTree.
- 2.Байєсовські класифікатори (Bayesian Classifiers): BayesNet, NaiveBayes, NaiveBayesMultinomial, NaiveBayesSimple.
- 3.Функції (functions): Logistic, MultilayerPerceptron, SMO, SimpleLinearRegression.
- 4.Ліниві алгоритми (lazy): IBk, LWL.
- 5.Мета-фільтри (meta): AdaBoostM1, Bagging, FilteredClassifier.
- 6.Класифікація на основі правил (Rule-based Classification): JRip, OneR, PART.
- 7.Інші методи класифікації: Holt-Winters, SerializedClassifier.



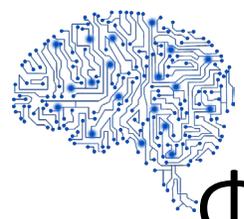
# Методи дерев рішень

1. J48: це реалізація алгоритму C4.5 для побудови дерев рішень. Дерева побудовані з рекурсивним діленням інформації, використовуючи алгоритм ID3, який заснований на ентропії.
2. LMT (Logistic Model Trees): це метод, який поєднує дерева рішень та логістичну регресію. Цей метод використовує високопродуктивну процедуру обчислення, яка дозволяє побудувати набагато більш точні моделі.
3. M5P: це метод, який побудовує дерево рішень на основі моделі регресії. Цей метод зазвичай використовується для задачі прогнозування числових значень.
4. RandomTree: це метод, який побудовує дерева рішень за допомогою вибірки випадкових атрибутів та порогових значень.
5. RERTree: це метод, який побудовує дерева рішень з мінімальним додатковим витратам і забезпечує хорошу ефективність для великих наборів даних.
6. HoeffdingTree: це метод, який побудовує дерева рішень із змінною глибиною. Цей метод підходить для потокових даних, де ми можемо побудувати дерево, не враховуючи всіх даних.



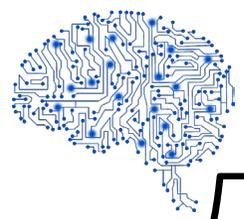
# Байєсовські класифікатори

1. Naive Bayes: Це один з найпростіших байєсовських класифікаторів. Він припускає, що значення атрибутів незалежні одне від одного. Для класифікації він використовує ймовірності віднесення об'єкту до кожного класу.
2. Naive Bayes Multinomial: Цей класифікатор підтримує мультиплікативну модель для вірогідності появи термінів. Він використовує вірогідності появи термінів в документі для прийняття рішення про класифікацію.
3. Naive Bayes Updateable: Цей класифікатор може навчатися на нових даних, додаючи їх до попередніх даних, які він вже мав. Він також використовує мультиплікативну модель для вірогідності появи термінів.
4. Bayes Net: Цей класифікатор використовує байєсовську мережу для моделювання залежностей між атрибутами і класами. Він використовує ймовірності появи атрибутів в мережі для прийняття рішення про клас



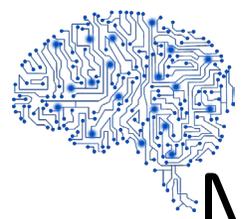
# Функції

1. Logistic: цей метод використовує логістичну регресію для знаходження ймовірності належності об'єкта до конкретного класу.
2. MultilayerPerceptron: цей метод використовує штучну нейронну мережу з кількома прихованими шарами для вирішення задач класифікації.
3. SMO: цей метод використовує опорні вектори для вирішення задач класифікації.
4. SimpleLogistic: цей метод є простішою версією логістичної регресії.
5. VotedPerceptron: цей метод використовує алгоритм "голосування перцептрона" для вирішення задач класифікації.
6. SGD: цей метод використовує стохастичний градієнтний спуск для вирішення задач класифікації.
7. SimpleLinearRegression: цей метод є простою версією лінійної регресії.



# Лінійні алгоритми

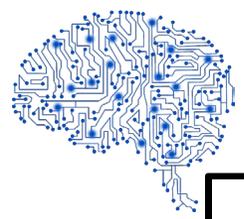
1. **1V $k$** : Використовує метод  $k$ -найближчих сусідів, де  $k$  - це кількість найближчих сусідів, які використовуються для класифікації нових прикладів.
2. **LazyKStar**: Використовує різноманітні метрики для знаходження  $k$  найближчих сусідів, а також застосовує фільтрацію даних, щоб зменшити вплив шуму.
3. **LWL**: Використовує зважену методику  $k$ -найближчих сусідів, де вага кожного сусіда залежить від його відстані від нового прикладу.



# Мета-фільтри

Секція "meta" у Weka включає в себе класифікатори, які поєднують кілька інших класифікаторів або виконують інші мета-операції для покращення результатів класифікації. Деякі методи у цій групі включають:

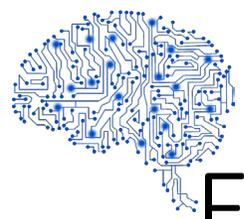
- AdaBoostM1: використовує адаптивне зважування для покращення точності класифікації.
- Bagging: використовує багато екземплярів одного й того ж класифікатора, навчених на випадкових підвбірках даних для покращення стійкості та зменшення дисперсії.
- Stacking: поєднує прогнози декількох базових класифікаторів з використанням іншого класифікатора, названого мета-класифікатором.
- MultiScheme: може використовувати різні алгоритми класифікації на різних підмножинах даних для покращення результатів.
- FilteredClassifier: дозволяє виконувати фільтрацію даних перед їх класифікацією. Фільтрація включає в себе видалення непотрібних атрибутів, а також виправлення та попередню обробку даних.
- ClassificationViaRegression: використовує регресійні моделі для класифікації даних.
- Decorate: використовує алгоритми штучного інтелекту та машинного навчання для класифікації даних.
- RandomSubSpace: використовує випадкові підмножини атрибутів для класифікації даних.



# Правила

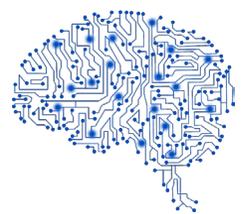
1. PART (Partial Decision Trees): алгоритм вибудовує дерево рішень для кожного класу, використовуючи частини дерева, що мають високу точність. Результатом є набір правил, що покривають всі класи.
2. JRip (Repeated Incremental Pruning): алгоритм JRip використовує дерева рішень та розраховує множину правил шляхом ітеративного зменшення дерева, що дозволяє уникнути перенавчання та зменшити кількість правил.
3. OneR (One Rule): алгоритм OneR використовує техніку, що дозволяє побудувати правило для кожної змінної, що дозволяє визначити, яка змінна найбільше пов'язана з цільовою змінною.
4. ZeroR (Zero Rule): цей алгоритм не використовує жодних змінних та завжди вибирає найбільш популярний клас.
5. DecisionTable: цей алгоритм перетворює дані у формат таблиці рішень та використовує її для побудови правил класифікації.
6. M5Rules: цей алгоритм використовує M5 дерева рішень та конвертує їх у набір правил для кожного класу.

# Кластеризація



# Explorer: кластеризація даних

- WEKA містить «кластеризатори» для пошуку груп подібних екземплярів у наборі даних
- Реалізовані схеми:
  - k-Means, EM, Cobweb, X-means, FarthestFirst
- Кластери можна візуалізувати та порівняти з «справжніми» кластерами (якщо вони є)
- Оцінка на основі функції максимальної правдоподібності, якщо схема кластеризації дає розподіл ймовірностей



Weka Knowledge Explorer

Preprocess   Classify   **Cluster**   Associate   Select attributes   Visualize

Clusterer

Choose `EM -I 100 -N -1 -S 100 -M 1.0E-6`

Cluster mode

- Use training set
- Supplied test set
- Percentage split    %
- Classes to clusters evaluation

Store clusters for visualization

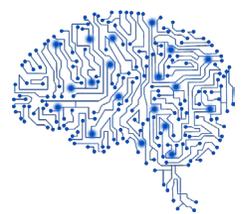
Result list (right-click for options)

Clusterer output

Status

OK

 x 0



Weka Knowledge Explorer

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

Clusterer

Choose `EM -I 100 -N -1 -S 100 -M 1.0E-6`

Cluster mode

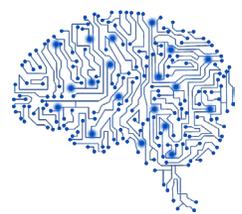
- Use training set
- Supplied test set
- Percentage split %
- Classes to clusters evaluation
- Store clusters for visualization

Result list (right-click for options)

Clusterer output

Status

OK   x 0



Weka Knowledge Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer

- weka
  - clusters
    - EM
    - SimpleKMeans
    - Cobweb**
    - FarthestFirst
    - XMeans

77387815

Clusterer output

Status  
OK

Log  x 0



Weka Knowledge Explorer

Preprocess   Classify   **Cluster**   Associate   Select attributes   Visualize

Clusterer

Choose **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode

- Use training set
- Supplied test set   Set...
- Percentage split   % 66
- Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

Start   Stop

Clusterer output

Result list (right-click for options)

Status

OK

Log  x 0



Weka Knowledge Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer  
Choose **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode

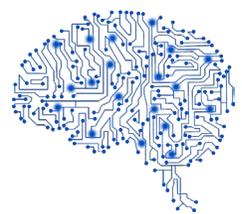
- Use training set
- Supplied test set
- Percentage split %
- Classes to clusters evaluation
- Store clusters for visualization

Result list (right-click for options)

Clusterer output

Status  
OK

 x 0



Weka Knowledge Explorer

Preprocess   Classify   **Cluster**   Associate   Select attributes   Visualize

Clusterer  
Choose **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode

- Use training set
- Supplied test set   Set...
- Percentage split   % 66
- Classes to clusters evaluation  
(Nom) class
- Store clusters for visualization

Ignore attributes

Start   Stop

Result list (right-click for options)

Clusterer output

Status  
OK

Log  x 0



Weka Knowledge Explorer

Preprocess   Classify   **Cluster**   Associate   Select attributes   Visualize

Clusterer  
Choose **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode

- Use training set
- Supplied test set   Set...
- Percentage split   % 66
- Classes to clusters evaluation  
(Nom) class
- Store clusters for visualization

Ignore attributes

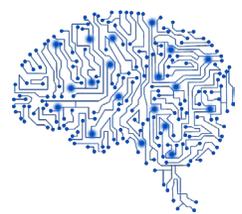
Start   Stop

Result list (right-click for options)

Clusterer output

Status  
OK

Log   x 0



Weka Knowledge Explorer

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

Clusterer:  Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set

Percentage split % 66

Classes to clusters evaluation

(Nom) class

Store clusters for visualization

Result list (right-click for options)

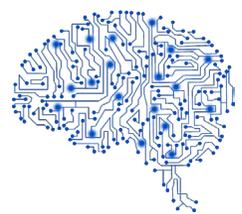
16:05:58 - Cobweb

Clusterer output

```
=== Run information ===  
Scheme:      weka.clusterers.Cobweb -A 1.0 -C 0.002820947917  
Relation:    iris  
Instances:   150  
Attributes:  5  
             sepallength  
             sepalwidth  
             petallength  
             petalwidth  
  
Ignored:     class  
Test mode:   Classes to clusters evaluation on training data  
  
=== Clustering model (full training set) ===  
Number of merges: 0  
Number of splits: 0  
Number of clusters: 3  
  
node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]  
  
=== Evaluation on training set ===
```

Status: OK

x 0



Weka Knowledge Explorer

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

Clusterer:  Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set

Percentage split % 66

Classes to clusters evaluation

(Nom) class

Store clusters for visualization

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

```
=== Run information ===
Scheme:      weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation:    iris
Instances:   150
Attributes:  5
              sepallength
              sepalwidth
              petallength
              petalwidth

Ignored:
              class

Test mode:   Classes to clusters evaluation on training data

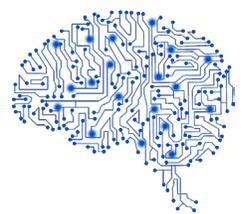
=== Clustering model (full training set) ===

Number of merges: 0
Number of splits: 0
Number of clusters: 3

node 0 [ 150]
|  leaf 1 [ 96]
node 0 [ 150]
|  leaf 2 [ 54]

=== Evaluation on training set ===
```

Status: OK  x 0



Weka Knowledge Explorer

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

Clusterer:  **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode

Use training set

Supplied test set

Percentage split %

Classes to clusters evaluation

Store clusters for visualization

Result list (right-click for options)

16:05:58 - Cobweb

Clusterer output

Number of clusters: 3

node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]

Clustered Instances

1	100	( 67%)
2	50	( 33%)

Class attribute: class

Classes to Clusters:

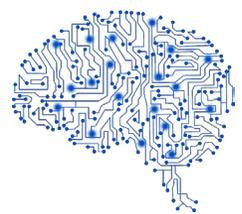
1	2	<-- assigned to cluster
0	50	Iris-setosa
50	0	Iris-versicolor
50	0	Iris-virginica

Cluster 1 <-- Iris-versicolor  
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances : 50.0 33.3333 %

Status: OK

 x 0



### Weka Knowledge Explorer

Preprocess   Classify   **Cluster**   Associate   Select attributes   Visualize

---

**Clusterer**  
Choose

**Cluster mode**

- Use training set
- Supplied test set
- Percentage split    %
- Classes to clusters evaluation
- Store clusters for visualization

**Result list (right-click for options)**

16:05:58 - Cobweb
-------------------

**Clusterer output**

```
Number of clusters: 3
node 0 [ 150]
| leaf 1 [ 96]
node 0 [ 150]
| leaf 2 [ 54]

Clustered Instances
1      100 ( 67%)
2       50 ( 33%)

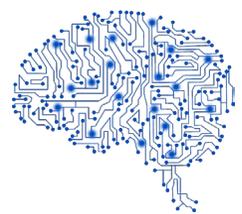
Class attribute: class
Classes to Clusters:
  1  2  <-- assigned to cluster
  0  50 | Iris-setosa
  50  0 | Iris-versicolor
  50  0 | Iris-virginica

Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances :      50.0      33.3333 %
```

**Status**  
OK

 x 0



Weka Knowledge Explorer

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

Clusterer:  Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set

Percentage split % 66

Classes to clusters evaluation

(Nom) class

Store clusters for visualization

Result list (right-click for options)

16:05:58 - Cobweb

- View in main window
- View in separate window
- Save result buffer
- Load model
- Save model
- Re-evaluate model on current test set
- Visualize cluster assignments
- Visualize tree**

Clusterer output

```
=== Run information ===
Scheme:      weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation:    iris
Instances:   150
Attributes:  5
              sepallength
              sepalwidth
              petallength
              petalwidth

Ignored:
              class

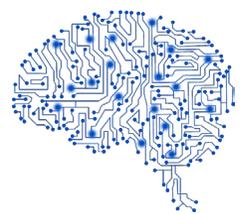
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3

training set ===
```

Status: OK

 x 0



Weka Knowledge Explorer

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

Clusterer: Choose **Cobweb -A 1.0 -C 0.0028209479177387815**

Cluster mode:  
 Use training set  
 Supplied test set  
 Percentage split  
 Classes to cluster  
(Nom) class  
 Store clusters for visualization

Clusterer output: Weka Classifier Tree Visualizer: 16:05:58 - Cobweb (iris)

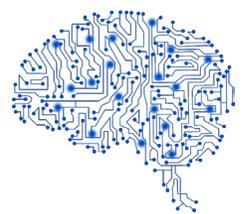
Tree View

```
graph TD; node0([node 0 (150)]) --- leaf1[leaf 1 (96)]; node0 --- leaf2[leaf 2 (54)];
```

Result list (right-click for details):  
16:05:58 - Cobweb

Status: OK

Log x 0



Weka Knowledge Explorer

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

Clusterer:  Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

- Use training set
- Supplied test set
- Percentage split % 66
- Classes to clusters evaluation
- Store clusters for visualization

Result list (right-click for options)

- 16:05:58 - Cobweb
  - View in main window
  - View in separate window
  - Save result buffer
  - Load model
  - Save model
  - Re-evaluate model on current test set
  - Visualize cluster assignments**
  - Visualize tree

Clusterer output

```
=== Run information ===
Scheme:      weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation:    iris
Instances:   150
Attributes:  5
              sepallength
              sepalwidth
              petallength
              petalwidth

Ignored:     class

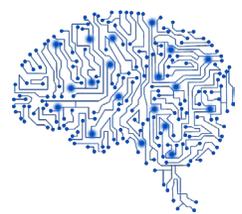
Test mode:   Classes to clusters evaluation on training data

=== Clustering model (full training set) ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3

... on training set ===
```

Status: OK

x 0



Weka Knowledge Explorer

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

Clusterer: **Cobweb -A 1.0 -C 0.0028209479177387815**

Weka Clusterer Visualize: 16:05:58 - Cobweb (iris)

Cluster mode:

- Use training set
- Supplied test set
- Percentage split
- Classes to cluster

X: petallength (Num) | Y: petalwidth (Num)

Colour: Cluster (Nom) | Select Instance

Reset | Clear | Save | Jitter:

(Nom) class

Store clusters for visualization

Log | Start

Result list (right-click for details): 16:05:58 - Cobweb

Plot: iris\_clustered

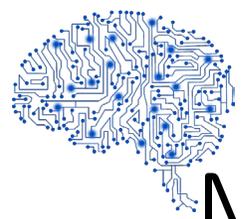
Class colour:

cluster0      cluster1      cluster2

=== Evaluation on training set ===

Status: OK

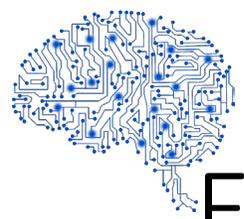
Log x 0



# Методи кластеризації у Weka

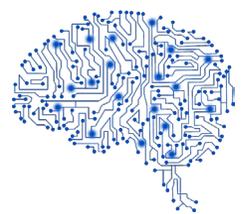
1. Алгоритми ієрархічної кластеризації: Simple Hierarchical Clustering, Complete Linkage Clustering.
2. Методи кластеризації на основі центрів мас: SimpleKMeans, XMeans.
3. Методи кластеризації на основі розподілу ймовірностей: Expectation-Maximization (EM).
4. Методи кластеризації на основі правил: Cobweb.
5. Методи кластеризації на основі розбиття на класи: FilteredClusterer, Canopy.

*Асоціації*



# Explorer: пошук асоціацій

- WEKA містить реалізацію алгоритму Apriori для вивчення правил асоціації
  - Працює тільки з дискретними даними
- Може визначити статистичні залежності між групами атрибутів:
  - молоко, масло, хліб, яйця (з надійністю 0,9 і підтримкою 2000)
- Apriori може обчислити всі правила, які мають задану мінімальну підтримку та перевищують задану надійність



Weka Knowledge Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

Start Stop

Associator output

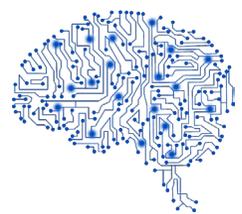
Result list (right-click for options)

Status

OK

Log  x 0

Detailed description: This is a screenshot of the Weka Knowledge Explorer software interface. The window title is 'Weka Knowledge Explorer'. At the top, there are several tabs: 'Preprocess', 'Classify', 'Cluster', 'Associate' (which is currently selected and highlighted in blue), 'Select attributes', and 'Visualize'. Below the tabs, the 'Associate' panel is active. It features a 'Choose' button followed by a text field containing the command: 'Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0'. Below this, there are 'Start' and 'Stop' buttons. The main area is divided into two panes: 'Result list (right-click for options)' on the left and 'Associator output' on the right. Both panes are currently empty. At the bottom of the window, there is a 'Status' bar showing 'OK' and a 'Log' button with a small bird icon and 'x 0' next to it.



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 10 -T 0 -C 0.9 -E 0.05 -U 1.0 -M 0.1 -S -1.0

Start Stop

Result list (right-click for options)

Associator output

Status

OK

Log  x 0

A red arrow points from the 'Preprocess' button to the 'Apriori' text in the command line.



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   Visualize

Open file...   Open URL...   Open DB...   Undo   Save...

Filter: Choose **None** Apply

Current relation: Relation: vote   Instances: 435   Attributes: 17

Selected attribute: Name: handicapped-infants   Type: Nominal  
Missing: 12 (3%)   Distinct: 2   Unique: 0 (0%)

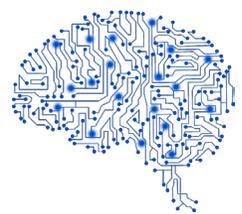
Label	Count
n	236
y	187

Attributes:

No.	Name
1	handicapped-infants
2	water-project-cost-sharing
3	adoption-of-the-budget-resolution
4	physician-fee-freeze
5	el-salvador-aid
6	religious-groups-in-schools
7	anti-satellite-test-ban
8	aid-to-nicaraguan-contras
9	mx-missile
10	immigration
11	synfuels-corporation-cutback
12	education-spending
13	superfund-right-to-sue
14	crime
15	duty-free-exports
16	export-administration-act-south-africa
17	Class

Colour: Class (Nom) Visualize All

Status: OK   Log   x 0



Weka Knowledge Explorer

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

Open file... | Open URL... | Open DB... | Undo | Save...

Filter: Choose None Apply

Current relation: Relation: vote, Instances: 435, Attributes: 17

Selected attribute: Name: handicapped-infants, Type: Nominal, Missing: 12 (3%), Distinct: 2, Unique: 0 (0%)

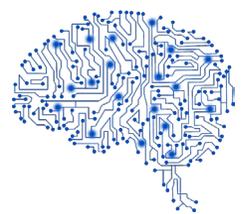
Label	Count
n	236
y	187

Attributes:

No.	Name
1	handicapped-infants
2	water-project-cost-sharing
3	adoption-of-the-budget-resolution
4	physician-fee-freeze
5	el-salvador-aid
6	religious-groups-in-schools
7	anti-satellite-test-ban
8	aid-to-nicaraguan-contras
9	mx-missile
10	immigration
11	synfuels-corporation-cutback
12	education-spending
13	superfund-right-to-sue
14	crime
15	duty-free-exports
16	export-administration-act-south-africa
17	Class

Colour: Class (Nom) Visualize All

Status: OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

Start Stop

Associator output

Result list (right-click for options)

Status

OK

Log  x 0

Detailed description: This is a screenshot of the Weka Knowledge Explorer software interface. The window title is 'Weka Knowledge Explorer'. At the top, there are several tabs: 'Preprocess', 'Classify', 'Cluster', 'Associate' (which is currently selected and highlighted in blue), 'Select attributes', and 'Visualize'. Below the tabs, the 'Associate' panel is active. It features a 'Choose' button followed by a text box containing the command 'Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0'. Below this, there are 'Start' and 'Stop' buttons. The main area is divided into two panes: 'Result list (right-click for options)' on the left and 'Associator output' on the right, both of which are currently empty. At the bottom of the window, there is a 'Status' bar showing 'OK' and a 'Log' button with a small bird icon and 'x 0' next to it.



Weka Knowledge Explorer

Preprocess   Classify   Cluster   **Associate**   Select attributes   Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

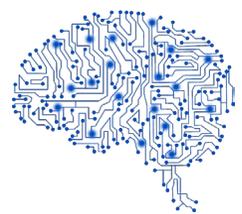
Start   Stop

Result list (right-click for options)

Associator output

Status  
OK

Log  x 0



Weka Knowledge Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

Start Stop

Result list (right-click for options)

- 16:29:37 - Apriori

Associator output

Minimum metric <confidence>: 0.9  
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20  
Size of set of large itemsets L(2): 17  
Size of set of large itemsets L(3): 6  
Size of set of large itemsets L(4): 1

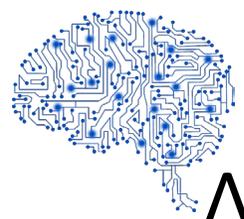
Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 conf:(0.99)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 conf:(0.99)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 conf:(0.99)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf:(0.99)
5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 conf:(0.98)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 200 conf:(0.98)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 conf:(0.98)
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210 conf:(0.98)

Status

OK

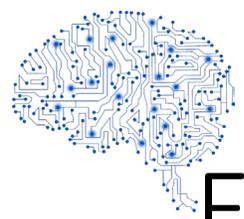
Log x 0



# Асоціативні методи у Weka

1. Apriori: який використовує техніку генерації кандидатів та підтримку для пошуку наборів асоціаційних правил.
2. FPGrowth: який використовує структуру дерева підтримки для пошуку наборів асоціаційних правил.
3. FilteredAssociator - це не метод пошуку асоціацій, а метод, який дозволяє застосовувати фільтри до даних перед тим, як вони будуть передані асоціаційному алгоритму. Після того, як фільтри будуть застосовані до даних, "FilteredAssociator" використовує один з методів пошуку асоціацій.

*Вибір атрибутів*



# Explorer: вибір атрибутів

- Панель, яку можна використовувати для дослідження того, які (підмножини) атрибутів є найбільш прогнозованими
- Методи вибору атрибутів складаються з двох частин:
  - Метод пошуку: найкращий, прямий відбір, випадковий, вичерпний, генетичний алгоритм, ранжування
  - Метод оцінки: на основі кореляції, обгортка, отримання інформації, хі-квадрат, ...
- Дуже гнучкий: WEKA дозволяє (майже) довільні комбінації цих ДВОХ





Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator  
Choose CfsSubsetEval

Search Method  
Choose BestFirst -D 1 -N 5

Attribute Selection Mode  
 Use full training set  
 Cross-validation Folds 10 Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

Attribute selection output

Status  
OK

Log  x 0



Weka Knowledge Explorer

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

Attribute Evaluator  
Choose **CfsSubsetEval**

Search Method  
Choose **BestFirst -D 1 -N 5**

Attribute Selection Mode  
 Use full training set  
 Cross-validation    Folds: 10    Seed: 1

(Nom) Class

Start    Stop

Result list (right-click for options)  
16:39:40 - BestFirst + CfsSubsetEval

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode:  evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found: 0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
                    physician-fee-freeze
```

Status  
OK

Log  x 0



Weka Knowledge Explorer

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

Attribute Evaluator: Choose **CfsSubsetEval**

Search Method: Choose **BestFirst -D 1 -N 5**

Attribute Selection Mode:  
 Use full training set  
 Cross-validation

Folds: 10  
Seed: 1

(Nom) Class

Start Stop

Result list (right-click for options):  
16:39:40 - BestFirst + CfsSubsetEval

Attribute selection output:

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found: 0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
  physician-fee-freeze
```

Status: OK

Log x 0



Weka Knowledge Explorer

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

### Attribute Evaluator

- weka
  - attributeSelection
    - CfsSubsetEval
    - ClassifierSubsetEval
    - WrapperSubsetEval
    - ConsistencySubsetEval
    - ReliefFAttributeEval
    - InfoGainAttributeEval**
    - GainRatioAttributeEval
    - SymmetricalUncertAttributeEval
    - OneRAttributeEval
    - ChiSquaredAttributeEval
    - PrincipalComponents
    - SVMAttributeEval

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode: evaluate on all training data

Attribute Selection on all input data ==
Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found: 0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
  physician-fee-freeze
```

Status: OK

Log  x 0



Weka Knowledge Explorer

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

Attribute Evaluator

Choose **InfoGainAttributeEval**

Search Method

- weka
  - attributeSelection
    - BestFirst
    - ForwardSelection
    - RaceSearch
    - GeneticSearch
    - RandomSearch
    - ExhaustiveSearch
    - Ranker**
    - RankSearch

308 -N -1

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode: evaluate on all training data
```

Attribute Selection on all input data ==

Search Method:

```
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 83
Merit of best subset found: 0.729
```

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):  
CFS Subset Evaluator

Selected attributes: 4 : 1  
physician-fee-freeze

Status

OK

Log  x 0





Weka Knowledge Explorer

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

Attribute Evaluator  
Choose **InfoGainAttributeEval**

Search Method  
Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode  
 Use full training set  
 Cross-validation    Folds: 10    Seed: 1

(Nom) Class

Start    Stop

Result list (right-click for options)  
16:39:40 - BestFirst + CfsSubsetEval  
**16:43:05 - Ranker + InfoGainAttributeEval**

Attribute selection output  
Information Gain Ranking Filter

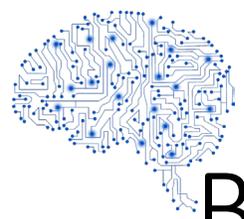
Ranked attributes:

0.7078541	4	physician-fee-freeze
0.4185726	3	adoption-of-the-budget-resolution
0.4028397	5	el-salvador-aid
0.34036	12	education-spending
0.3123121	14	crime
0.3095576	8	aid-to-nicaraguan-contras
0.2856444	9	mx-missile
0.2121705	13	superfund-right-to-sue
0.2013666	15	duty-free-exports
0.1902427	7	anti-satellite-test-ban
0.1404643	6	religious-groups-in-schools
0.1211834	1	handicapped-infants
0.1007458	11	synfuels-corporation-cutback
0.0529956	16	export-administration-act-south-africa
0.0049097	10	immigration
0.0000117	2	water-project-cost-sharing

Selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16

Status  
OK

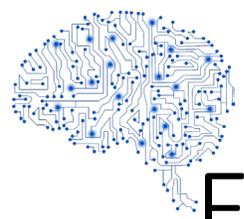
Log  x 0



# Вибір атрибутів у Weka

1. CfsSubsetEval - використовує кореляцію між атрибутами та класами для відбору підмножини атрибутів, яка є корисною для класифікації.
2. CorrelationAttributeEval - використовує кореляцію між атрибутами для відбору підмножини атрибутів, які мають високу кореляцію з класом.
3. InfoGainAttributeEval - використовує інформаційний приріст для відбору підмножини атрибутів, які мають високий приріст інформації про класи.
4. GainRatioAttributeEval - використовує співвідношення інформаційного приросту та ентропії для відбору підмножини атрибутів, які мають високий приріст інформації про класи.
5. ReliefAttributeEval - використовує набір випадкових об'єктів для визначення важливості атрибутів відносно класів.
6. OneRAttributeEval - використовує алгоритм OneR для відбору атрибутів, які мають найвищу точність при класифікації на основі одного атрибута.
7. WrapperSubsetEval - використовує зовнішній класифікатор
8. SymmetricalUncertAttributeEval - оцінює неоднозначність та ентропію між атрибутами та цільовим класом.

# *Візуалізація даних*



# Explorer: візуалізація даних

- Візуалізація дуже корисна на практиці: наприклад, допомагає визначити складність навчальної проблеми
- WEKA може візуалізувати окремі атрибути (1-d) і пари атрибутів (2-d)
- Кольорове кодування значень класу
- Опція «тремтіння» для роботи з номінальними атрибутами (і виявлення «прихованих» точок даних)
- Функція «Збільшення».



### Weka Knowledge Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...UndoSave...

Filter

Choose **None** Apply

Current relation

Relation: Glass  
Instances: 214      Attributes: 10

Selected attribute

Name: RI      Type: Numeric  
Missing: 0 (0%)      Distinct: 178      Unique: 145 (68%)

Statistic	Value
Minimum	1.511
Maximum	1.534
Mean	1.518
StdDev	0.003

Attributes

No.	Name
1	RI
2	Na
3	Mg
4	Al
5	Si
6	K
7	Ca
8	Ba
9	Fe
10	Type

Colour: Type (Nom) ▼Visualize All

Bin Range	Count
1.510 - 1.511	3
1.511 - 1.512	3
1.512 - 1.513	1
1.513 - 1.514	5
1.514 - 1.515	45
1.515 - 1.516	48
1.516 - 1.517	46
1.517 - 1.518	18
1.518 - 1.519	7
1.519 - 1.520	15
1.520 - 1.521	9
1.521 - 1.522	5
1.522 - 1.523	1
1.523 - 1.524	1
1.524 - 1.525	4
1.525 - 1.526	1
1.526 - 1.527	0
1.527 - 1.528	0
1.528 - 1.529	1
1.529 - 1.530	0
1.530 - 1.531	1

StatusLog x 0

OK



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   **Visualize**

**Plot Matrix**

	RI	Na	Mg	Al	Si	K
Type						
Fe						

PlotSize: [100]

PointSize: [1]

Jitter:

Colour: Type (Nom)

SubSample % :

Class Colour

```
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
```

Status

OK

x 0



### Weka Knowledge Explorer

Preprocess   Classify   Cluster   Associate   Select attributes   **Visualize**

**Plot Matrix**

	RI	Na	Mg	Al	Si	K
Type						
Fe						

PlotSize: [100]

PointSize: [1]

Jitter:

Colour: Type (Nom)

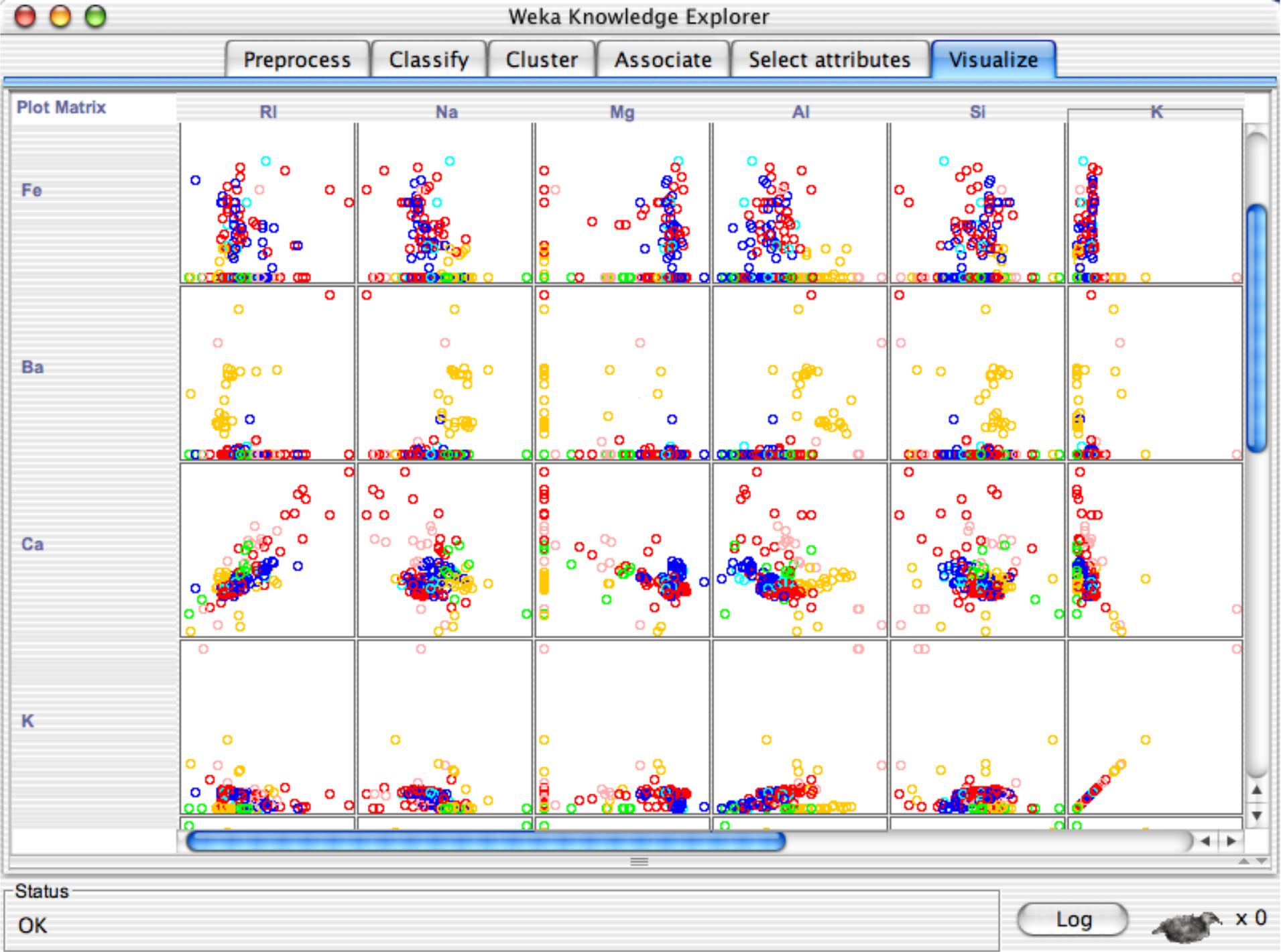
SubSample % :

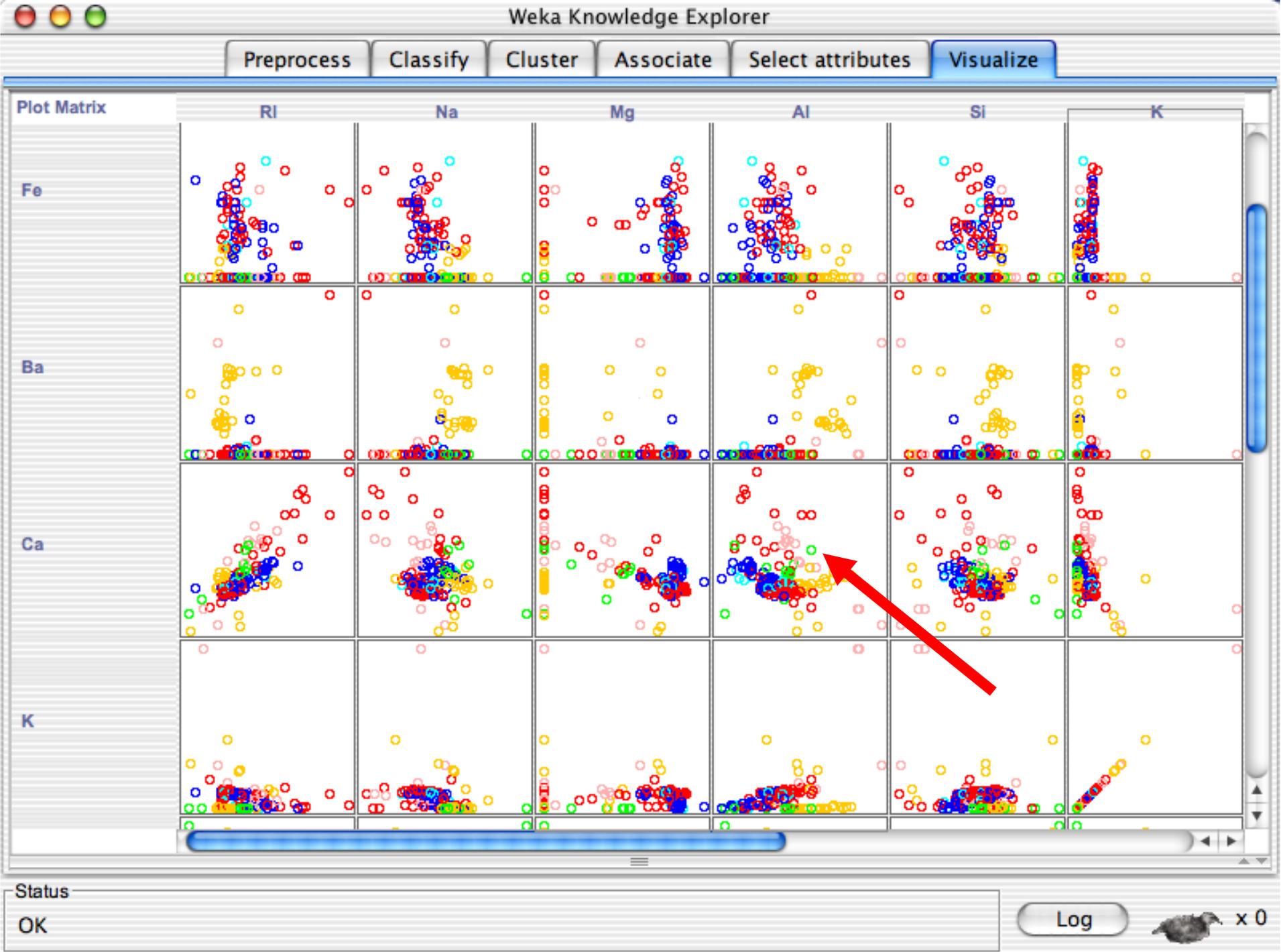
Class Colour

```
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
```

Status

OK  x 0





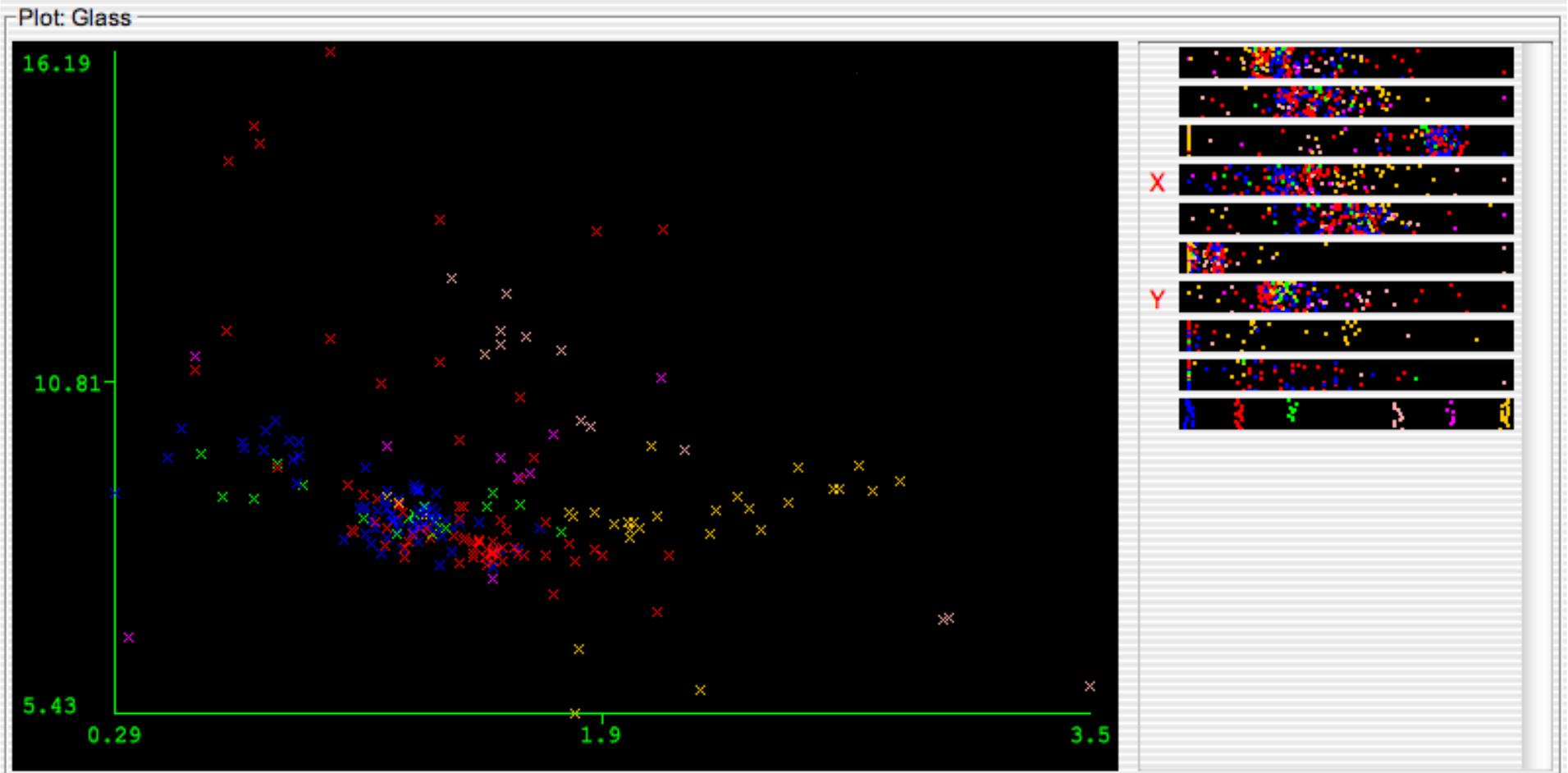


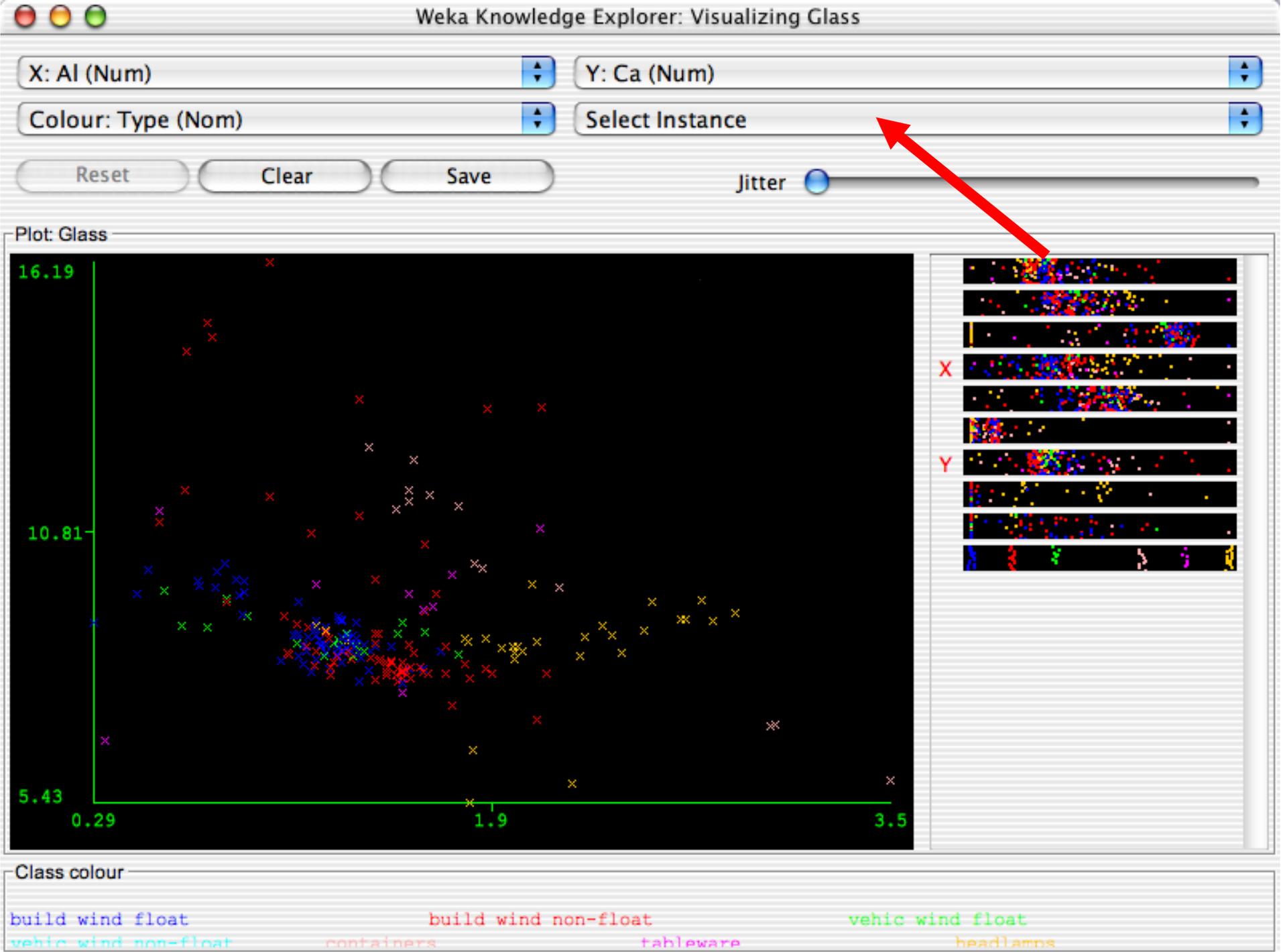
Weka Knowledge Explorer: Visualizing Glass

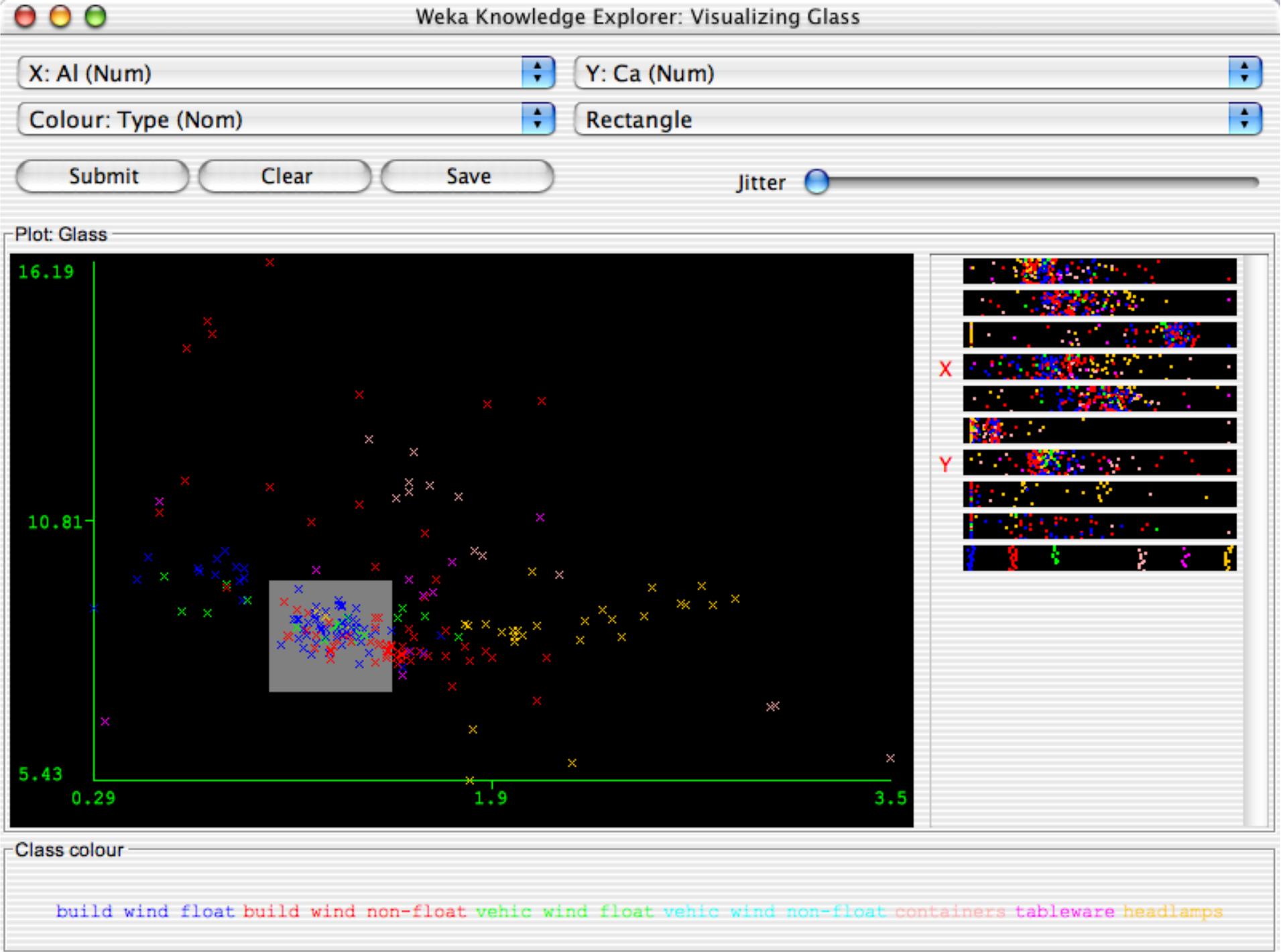
X: Al (Num)    Y: Ca (Num)

Colour: Type (Nom)    Select Instance

Reset    Clear    Save    Jitter









### Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)    Y: Ca (Num)

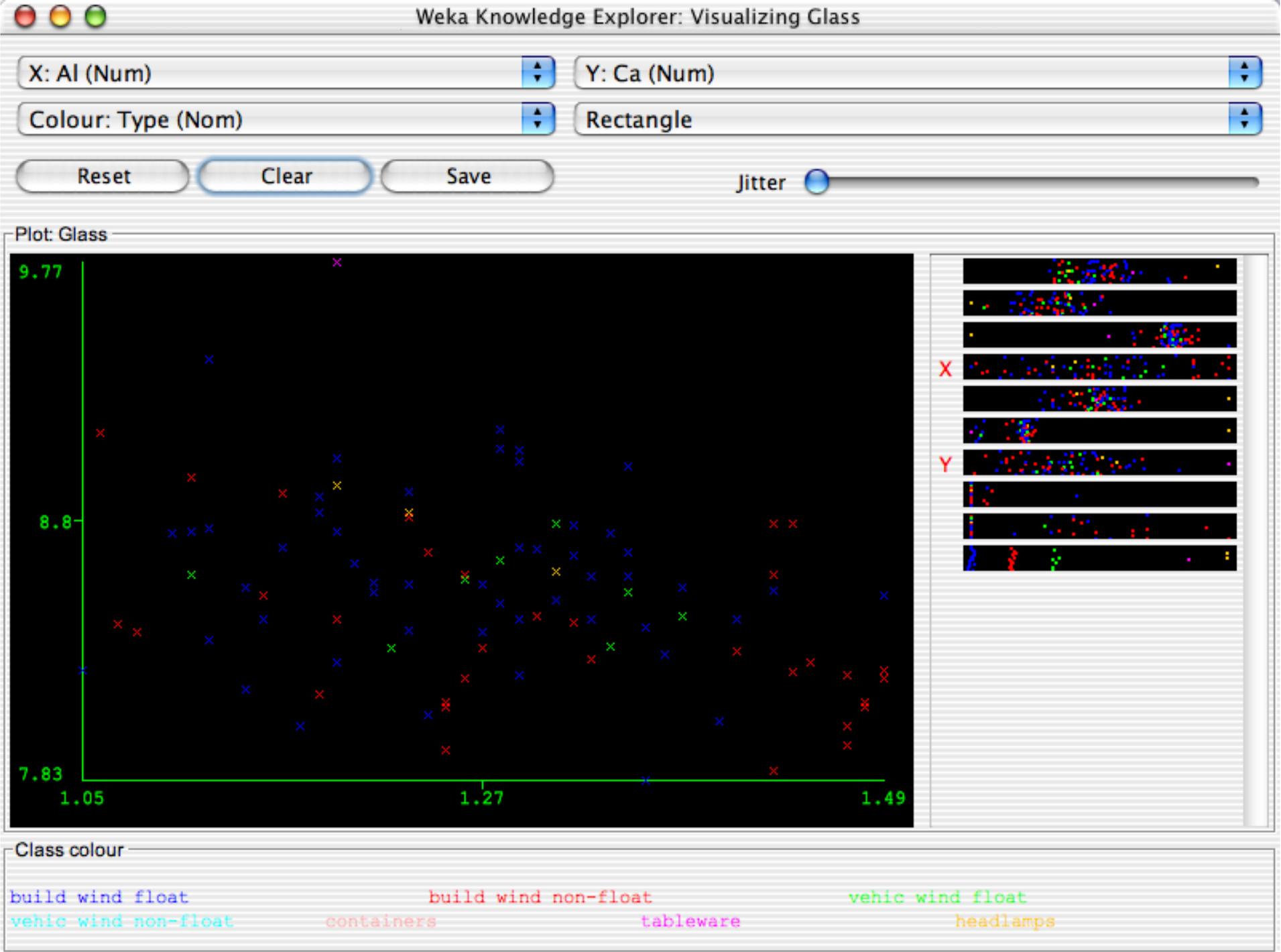
Colour: Type (Nom)    Rectangle

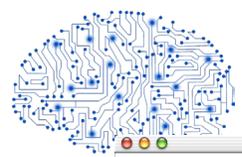
        Jitter

Plot: Glass

Class colour

```
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps
```





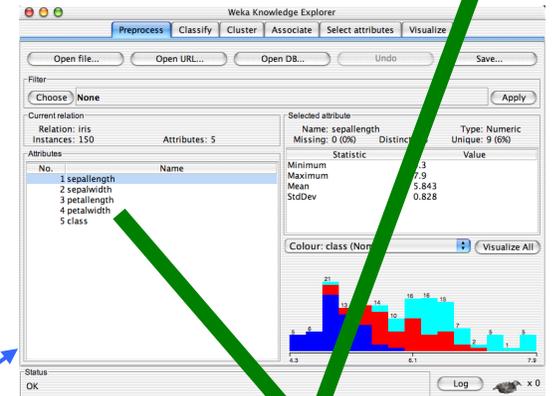
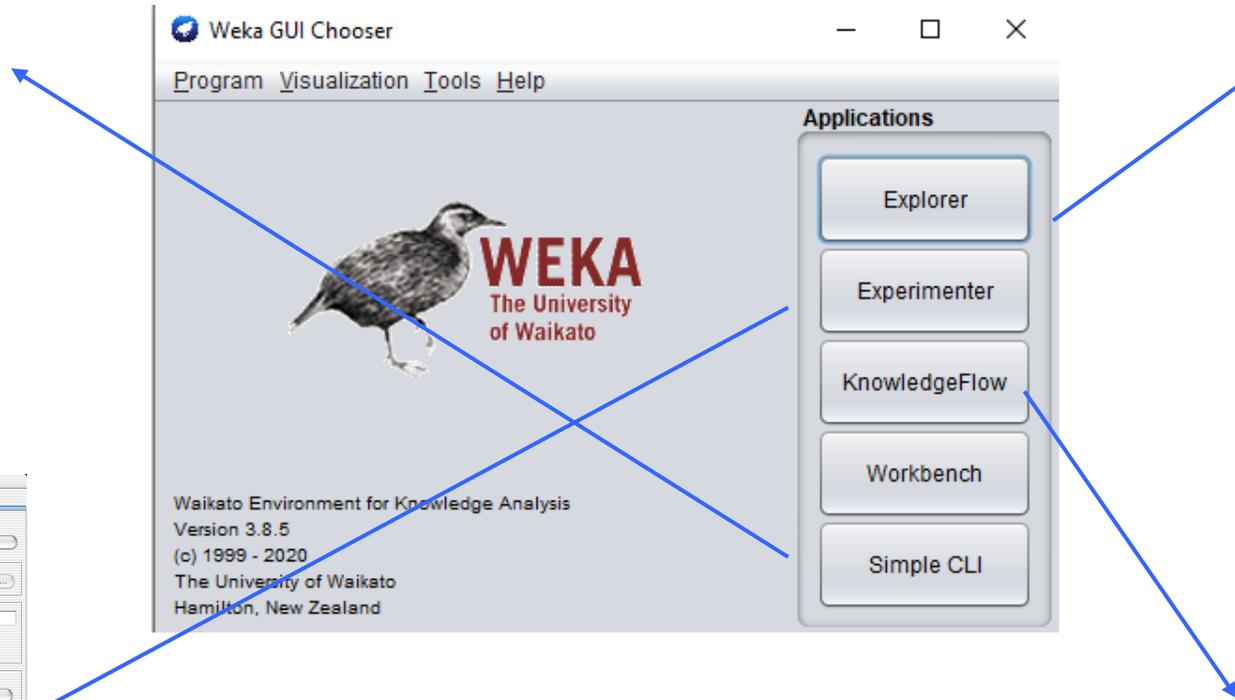
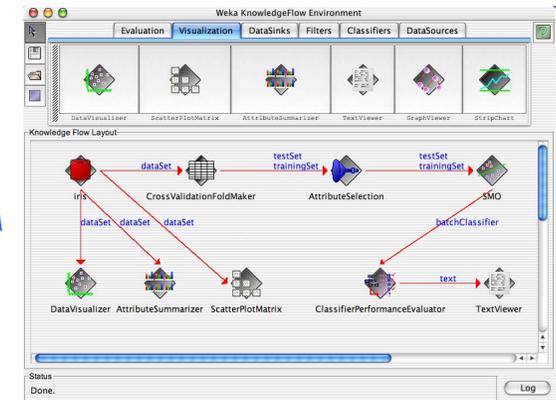
```

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.

> help
Command must be one of:
  java <classname> <args>
  break
  kill
  cla
  exit
  help <command>

```

*Питання?*