

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

Université de WAIKATO

<http://www.cs.waikato.ac.nz/~ml/weka/>

Culture « machine learning »

Code source libre en JAVA (Licence GNU – 3.5.1)

Référence pour les publications

Site WEB très riche (données, doc.)

Exécution stand-alone sous Windows (JRE)

WEKA

Origine, architecture

Interface (1/2)

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comp

Expérimentations

Performances

Exploration graphique

```
WEKA Step1@CLI
Welcome to the WEKA Step1@CLI
Enter commands in the field at the bottom of
the window. The prompt will return to you
through previous commands.

> help
Command must be one of:
  java -classpath <steps>
  break
  kill
  exit
  help <command>

> help
```

WEKA Experiment Environment

Experiment Configuration Mode: Simple Advanced

Results Destination: DBSC database

Experiment Type: Cross-validation Data sets first Algorithm first

Algorithm: NeuralNetwork -1.0.3-M0.02-N 500-Y 0-S 0-E 20 -H a

Weka GUI Chooser

Waikato Environment for Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand

GUI

Simple CLI Explorer
Experimenter KnowledgeFlow

Weka Knowledge Explorer

Selected attribute: Name: spelllength, Type: Numeric, Minimum: 0.000, Maximum: 7.9, StdDev: 0.868

Weka KnowledgeFlow Environment

KnowledgeFlow Layout

WEKA

Origine, architecture

Interface (2/2)

Accès aux données

Définir les traitements

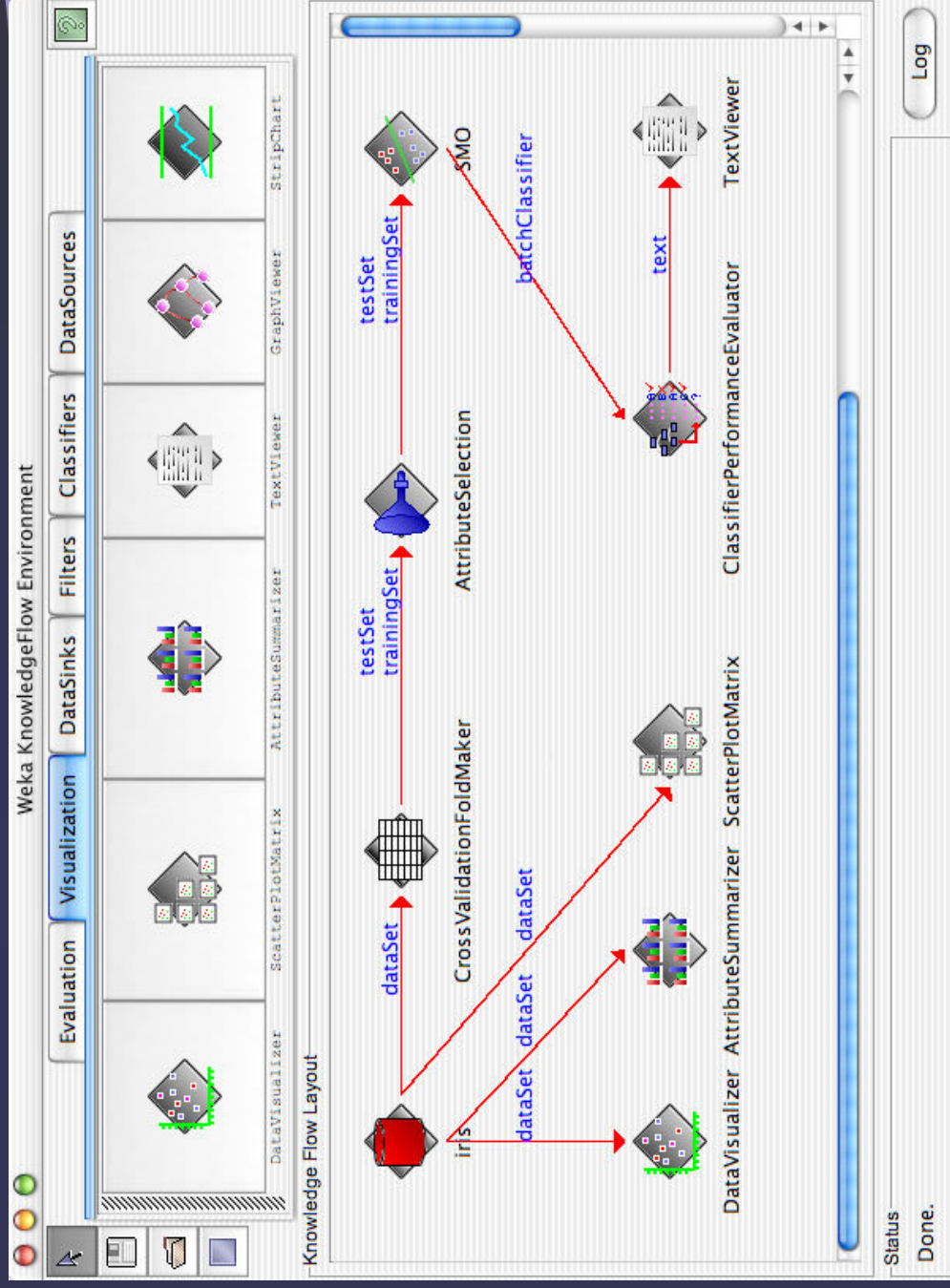
Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique



Très complet mais nécessite un apprentissage (ou une lecture attentive de la doc.)

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

Fichier texte, avec des balises spécifiques (format .arff)

```
@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
```

```
...
```



D'autres accès sont disponibles (CSV,...), y compris les accès à des SGBD via JDBC

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements (1/2)

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

The screenshot displays the Weka KnowledgeFlow Environment interface. At the top, a menu bar includes 'DataSources', 'DataSinks', 'Filters', 'Classifiers', 'Clusterers', 'Associations', and 'Evaluation'. Below the menu is a toolbar with icons for 'Data Visualizer', 'Scatter PlotMatrix', 'Attribute Summarizer', 'Model PerformanceChart', and 'Text Viewer'. The main workspace shows a 'Knowledge Flow Layout' with a workflow: 'iris' (Data) -> 'J48' (Classifier) -> 'Text Viewer' (Text) and 'Graph' (Graph). The 'J48' node is selected, and its 'Text Viewer' window is open, displaying the following decision tree structure:

```
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
  petalwidth <= 1.7
  | petalwidth <= 4.9: Iris-versicolor (48.0/1.0)
  | petalwidth > 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
Size of the tree : 9
```

Simultaneously, the 'Weka Classifier Tree Visualizer: 11:11:39 - Set 1 (iris) J48' window shows a graphical representation of the same decision tree. The root node is 'petalwidth <= 0.6', which splits into 'Iris-setosa (50.0)' and 'petalwidth > 0.6'. The 'petalwidth > 0.6' node further splits into 'petalwidth <= 1.7' and 'petalwidth > 1.7'. The 'petalwidth <= 1.7' node splits into 'petalwidth <= 4.9' (leading to 'Iris-versicolor (48.0/1.0)') and 'petalwidth > 4.9'. The 'petalwidth > 4.9' node splits into 'petalwidth <= 1.5' (leading to 'Iris-virginica (3.0)') and 'petalwidth > 1.5' (leading to 'Iris-versicolor (3.0/1.0)'). The 'petalwidth > 1.7' node leads to 'Iris-virginica (46.0/1.0)'.

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements (2/2)

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

Apprentissage des règles (JRIP) sur les descripteurs discrétisés

The screenshot shows the Weka KnowledgeFlow Environment interface. The top menu bar includes: DataSources, DataSinks, Filters, Classifiers, Clusters, Associations, Evaluation, Visualization. The main workspace displays a Knowledge Flow Layout with the following nodes and connections:

- Data Visualizer** (connected to **Scatter PlotMatrix**)
- Scatter PlotMatrix** (connected to **Attribute Summarizer**)
- Attribute Summarizer** (connected to **DiscretizeAttribute**)
- DiscretizeAttribute** (connected to **Class Assigner**)
- Class Assigner** (connected to **Training SetMaker**)
- Training SetMaker** (connected to **JRIP**)
- JRIP** (connected to **Text Viewer**)

The **Text Viewer** window displays the following output:

```
Result list
11:42:53 - Model: JRip
11:44:14 - Model: JRip
11:45:00 - Model: JRip
11:45:15 - Model: JRip
11:46:20 - Model: JRip

=== Classifier model ===
Scheme: JRip
Relation: iris-weka.filters.unsupervised.attribute.Discretize-E10-M-1.0-R
JRIP rules:
=====
(petalwidth = '[3.95-4.54]') => class=Iris-versicolor (26.0/1.0)
(petalwidth = '(3.36-3.95]') => class=Iris-versicolor (8.0/0.0)
(petalwidth = '(4.54-5.13]') and (petalwidth = '[1.3-1.54]') => class=Iris-versicolor (2.0/0.0)
(petalwidth = '(1.06-1.3]') => class=Iris-versicolor (3.0/0.0)
(petalwidth = '(1.54-1.78]') => class=Iris-versicolor (4.0/1.0)
(petalwidth = '(-inf-0.34]') => class=Iris-setosa (41.0/0.0)
(petalwidth = '(0.34-0.58]') => class=Iris-setosa (8.0/0.0)
=> class=Iris-virginica (48.0/2.0)

Number of Rules : 9
```

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

- Description rapide des données et stat. descriptives ☹️

- Traitement des données manquantes 😊

- Tests statistiques ☹️

- Sélection des variables pour le supervisé 😊😊😊

- Construction de variables 😊

- Apprentissage supervisé (et régression) 😊😊

- Apprentissage non-supervisé 😊

- Règles d'association 😊

- Méthodes factorielles ☹️

- Séries temporelles ☹️



Phénoménale pour le supervisé

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

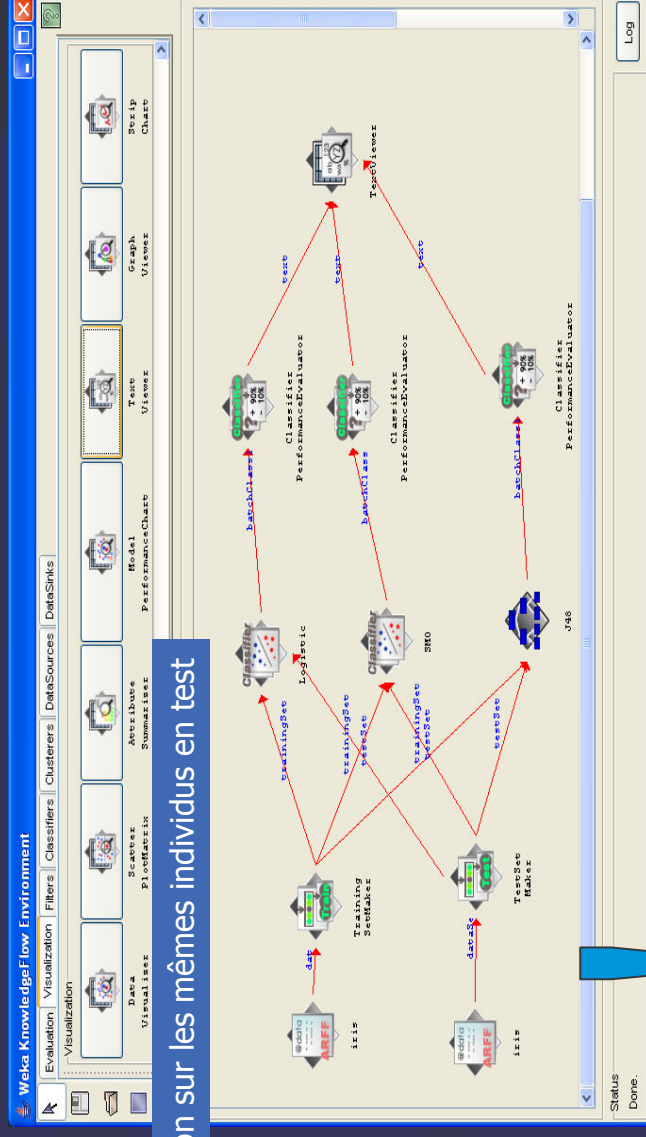
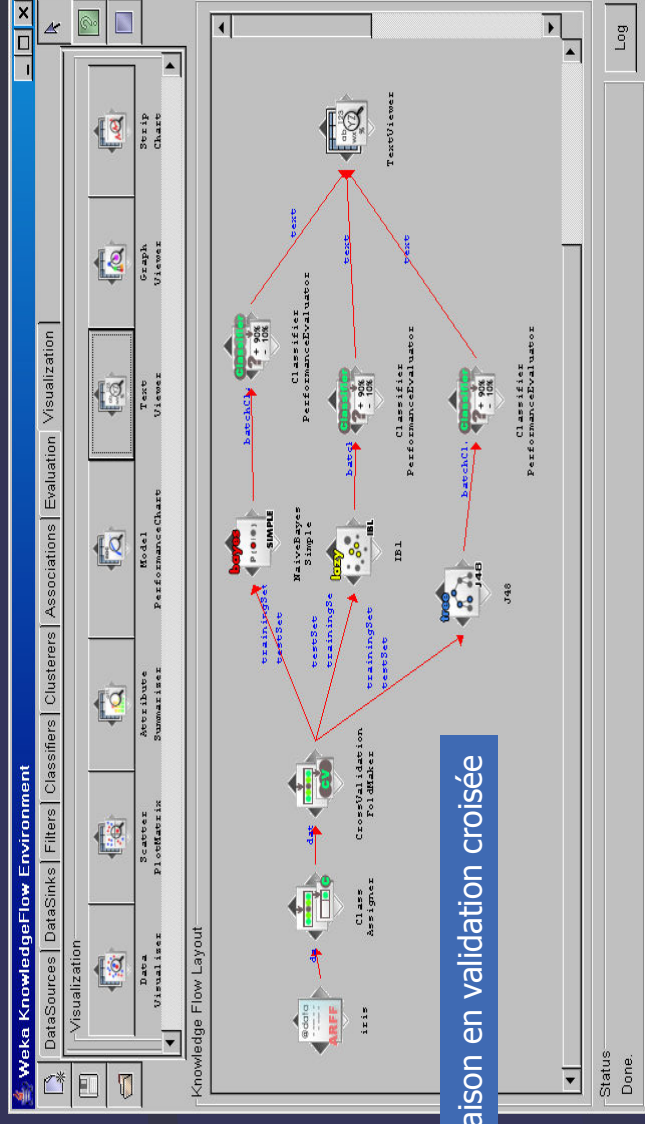
Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique



Complet pour les comparaisons

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

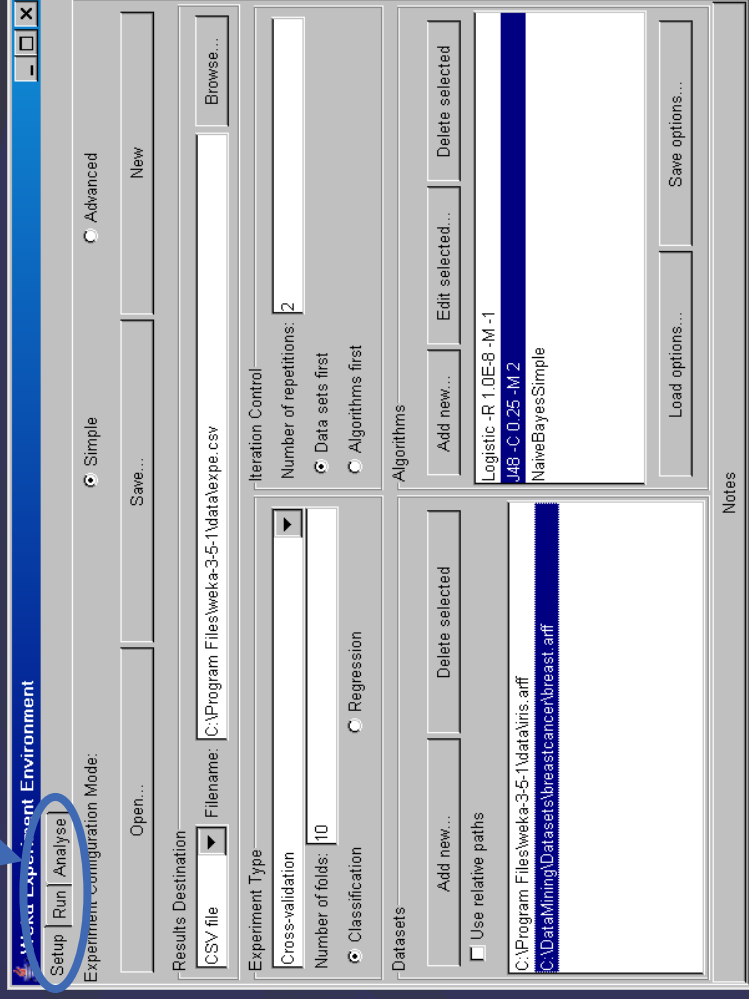
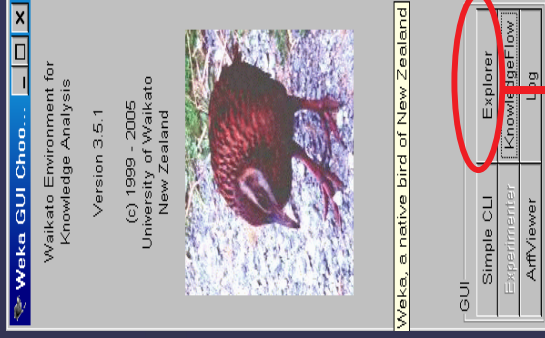
Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

- Définir les données et les traitements
- Lancer l'expérimentation
- Charger les résultats et réaliser des comparaisons



Royal pour les publications !

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

- Pas de limitation théorique (espace adressable) 😊
- Chargement de toutes les données en mémoire 😞
- Code bien écrit et organisé 😊
- **La vraie limitation est technologique (JAVA)**
 - Sur la taille de la base (taille du tas) 😞
 - Sur la rapidité 😞

WEKA

Origine, architecture

Interface

Accès aux données

Définir les traitements

Bibliothèque

Évaluation et comparaisons

Expérimentations

Performances

Exploration graphique

Weka Explorer

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

Open fil... | Open U... | Open D... | Generat... | Undo | Edit... | Save...

Filter: None

Current relation: iris
Instances: 150
Attributes: 5

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

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Class: class (Nom)

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

Weka Explorer

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

Plot Matrix

	sepal.length	sepal.width	petal.length	petal.width
sepal.length	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]
sepal.width	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]
petal.length	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]
petal.width	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]	[Scatter Plot]

PlotSize: [121]
PointSize: [5]
Jitter:
Colour: class (Nom)
SubSample %: [100]

Status: OK

Embryonnaire et inadapté pour les grandes bases