

# Hand on Weka: Just a Taste

2012/11/20

Petra Kralj Novak

Petra.Kralj.Novak@ijs.si

# Data Mining Tools

- Weka <http://www.cs.waikato.ac.nz/ml/weka/>
- Orange <http://orange.biolab.si/>
- Knime <http://www.knime.org/>
- Taverna <http://www.taverna.org.uk/>
- Rapid Miner <http://rapid-i.com/content/view/181/196/>
- ClowdFlows <http://clowdflows.org/>

# Weka (Waikato Environment for Knowledge Analysis)

- Collection of machine learning algorithms for data mining tasks
- The algorithms
  - Can be applied directly to a dataset
  - Can be called from Java code (library)
- Weka contains tools for
  - Data pre-processing
  - Classification
  - Regression
  - Clustering
  - Association rules
  - Visualization
- Weka is open source software issued under the GNU General Public License

# Practice with Weka

1. Build a decision tree with the ID3 algorithm on the lenses dataset, evaluate on a separate test set


# Weka: Install

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

Weka 3 - Data Mining with Open Source Machine Learning Software in Java - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.cs.waikato.ac.nz/ml/weka/

 **WEKA**  
The University of Waikato

**Software**

[project](#) ▪ [software](#) ▪ [book](#) ▪ [publications](#) ▪ [people](#) ▪ [related](#)

Home

**Getting started**

[Requirements](#)

[Download](#)

[Documentation](#)

[FAQ](#)

[Citing Weka](#)

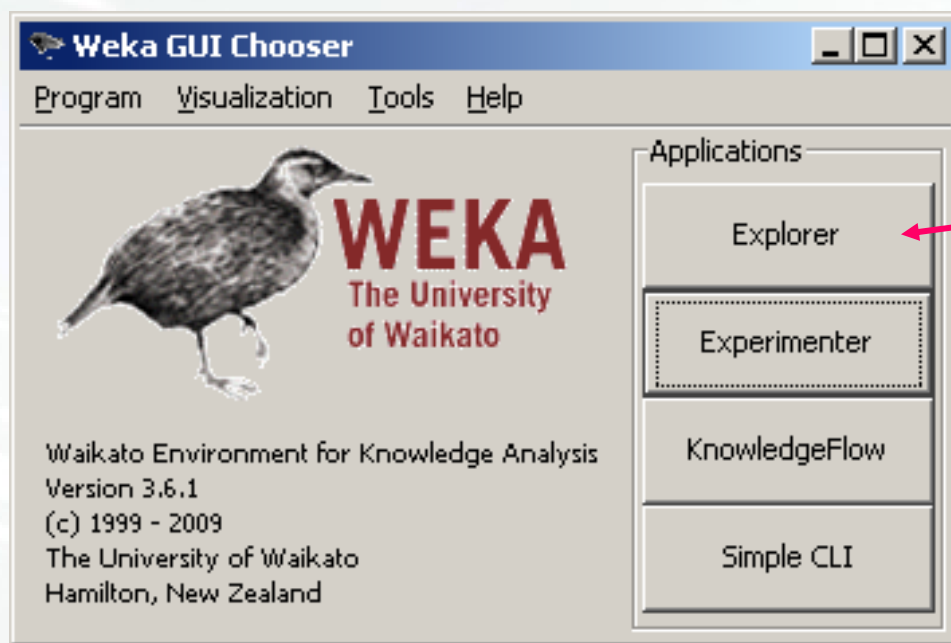
**Weka 3: Data Mining Software in Java**

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Weka is open source software issued under the [GNU General Public License](#).

Download  
version  
3.6

# Weka: Run Explorer



Choose Explorer

# Exercise: Lenses dataset

- In the Weka data mining tool, induce a decision tree for the lenses dataset with the ID3 algorithm.
- Data:
  - lensesTrain.arff
  - lensesTest.arff
- Compare the outcome with the manually obtained results.

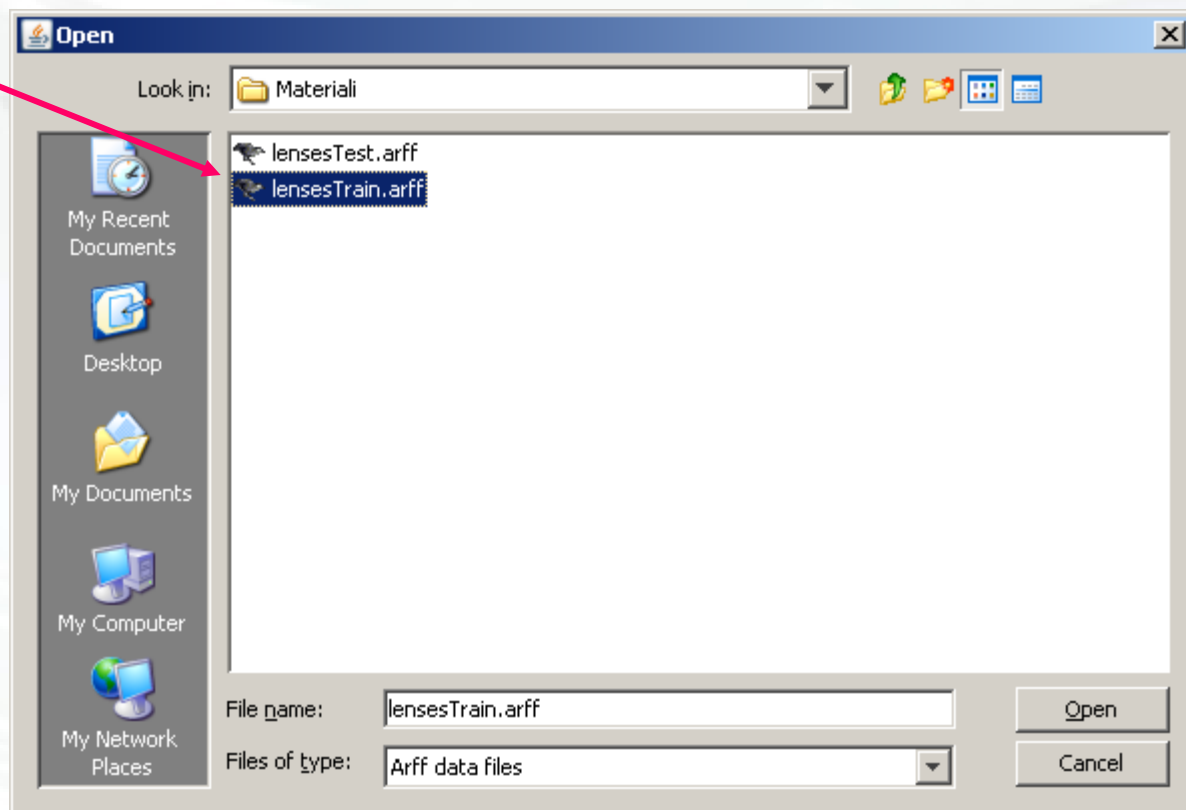
# Load the data

The screenshot shows the Weka Explorer application window. A red arrow points to the 'Open file...' button in the top toolbar. The interface includes several panels: 'Preprocess' (selected), 'Classify', 'Cluster', 'Associate', 'Select attributes', and 'Visualize'. Below the toolbar are buttons for 'Open file...', 'Open URL...', 'Open DB...', 'Undo', 'Edit...', and 'Save...'. The 'Filter' section has a 'Choose' button and a text field containing 'None', with an 'Apply' button. The 'Current relation' section shows 'Relation: None', 'Instances: None', and 'Attributes: None'. The 'Attributes' section has 'All', 'None', and 'Invert' buttons. The 'Selected attribute' section shows 'Name: None', 'Missing: None', 'Type: None', 'Distinct: None', and 'Unique: None'. A 'Visualize All' button is at the bottom right of the main area. The 'Status' bar at the bottom says 'Welcome to the Weka Explorer' and includes a 'Log' button and a small icon with 'x 0'.



# Load the data - 2

lensesTrain.arff



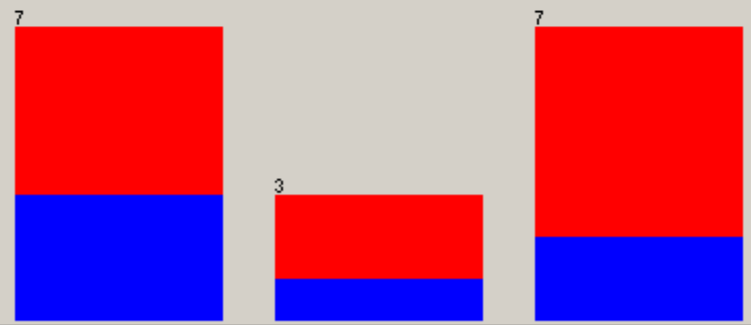
# The data are loaded

Choose  
"Classify"

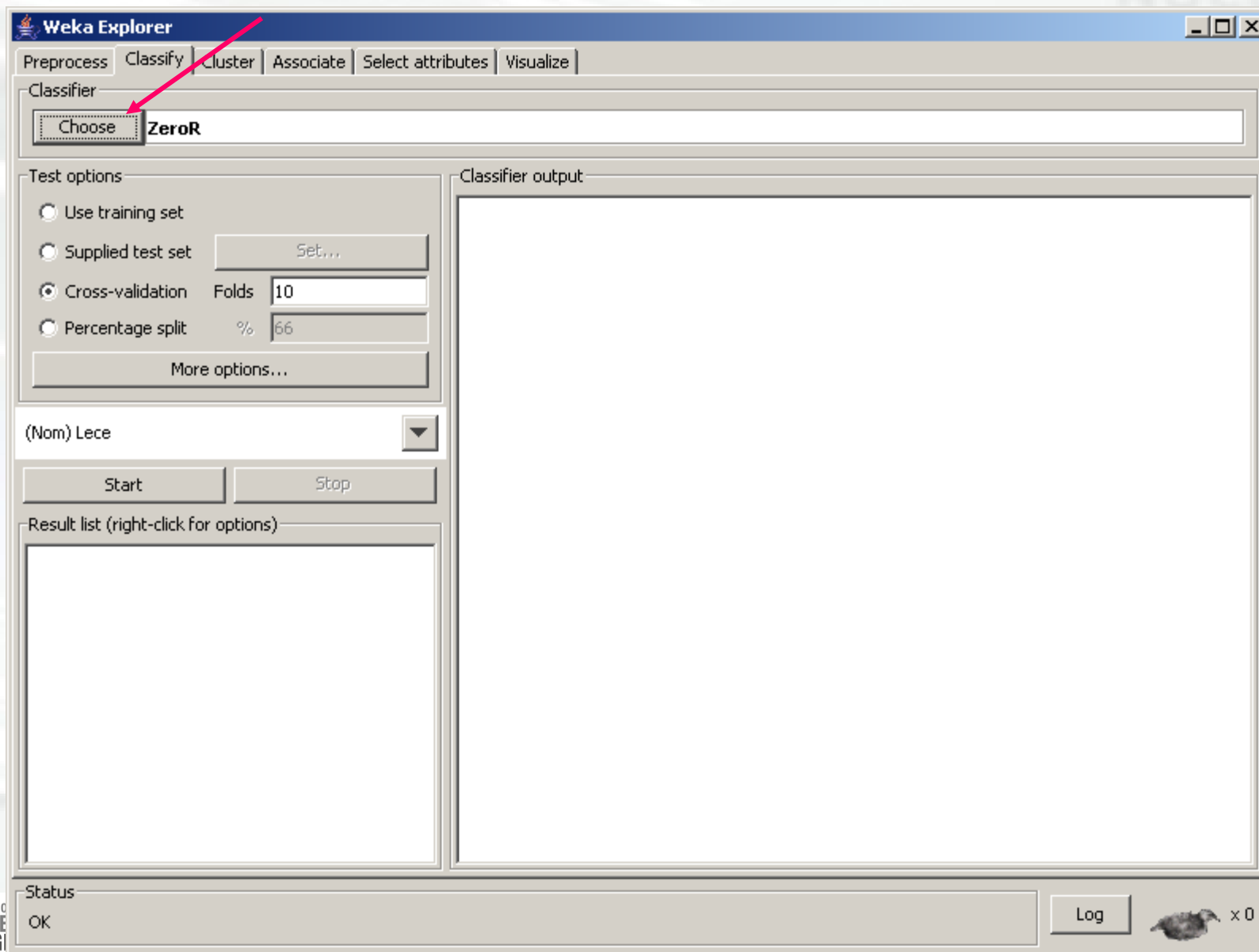
The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected in the top menu. A red arrow points to the 'Classify' tab. Below the menu, there are buttons for 'Open file...', 'Open URL...', 'Open DB...', 'Undo', 'Edit...', and 'Save...'. The 'Filter' section shows 'None' selected. The 'Current relation' section displays 'Relation: lensesTrain' and 'Instances: 17'. The 'Attributes' section lists five attributes: Age, Prescription, Astigmatic, Tear\_rate, and Lenses. A red arrow points to the 'Lenses' attribute, which is labeled as the 'Target variable'. The 'Selected attribute' section shows 'Age' with a table of counts for 'young', 'pre-presbyopic', and 'presbyopic'. The 'Class: Lenses (Nom)' dropdown is also visible. At the bottom, there is a 'Status' bar showing 'OK' and a 'Log' button.

Target variable

Label	Count
young	7
pre-presbyopic	3
presbyopic	7



# Choose algoritem



The screenshot shows the Weka Explorer application window. The 'Classifier' tab is active, and the 'ZeroR' classifier is selected. A red arrow points to the 'Choose' button next to 'ZeroR'. The 'Test options' section is visible, with 'Cross-validation' selected and 'Folds' set to 10. The 'Classifier output' area is empty. The status bar at the bottom shows 'Status OK' and a 'Log' button.

**Weka Explorer**

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

Classifier

Choose **ZeroR**


Test options

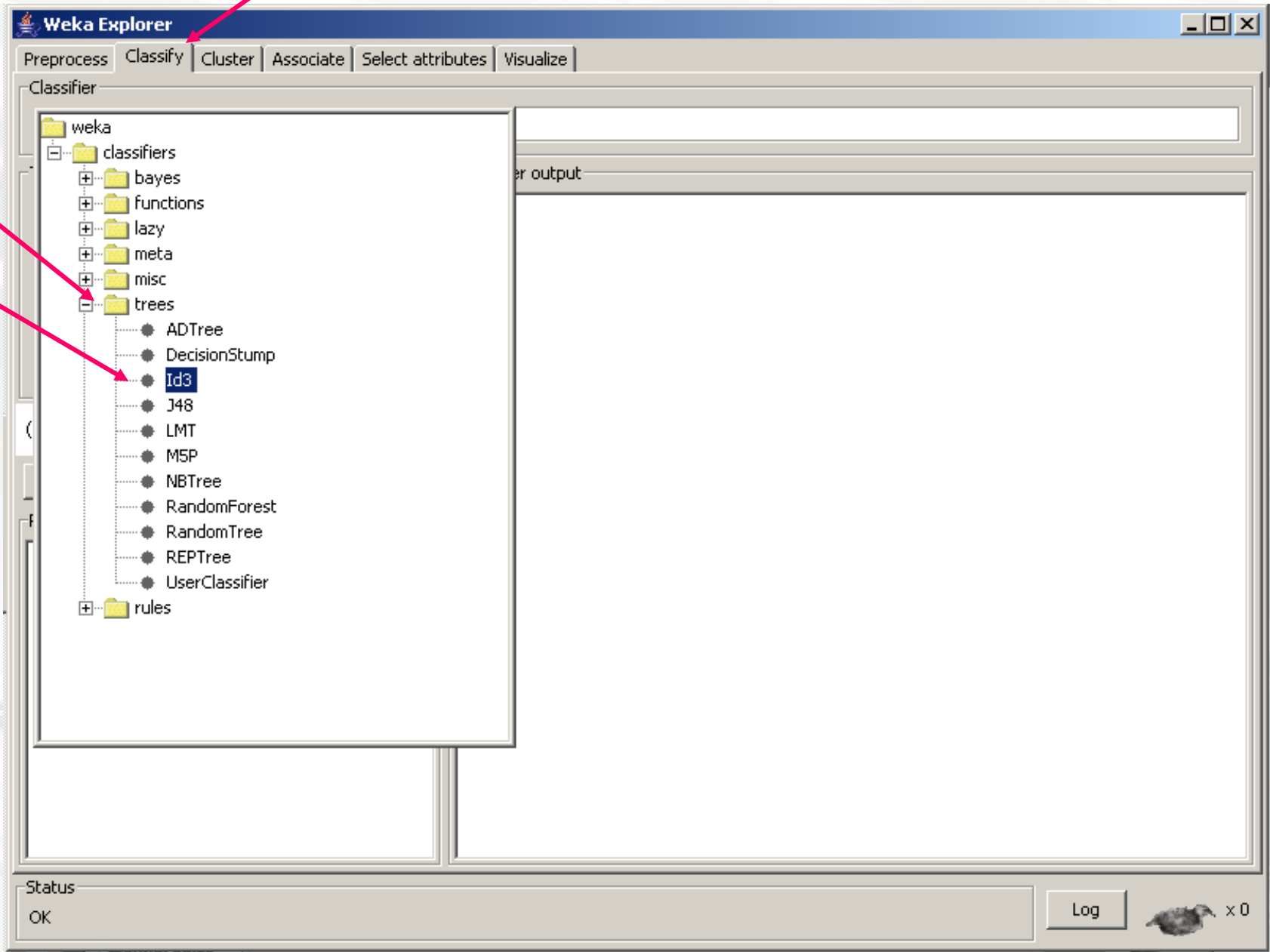
- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %

(Nom) Lece

Result list (right-click for options)

Status  
OK

 x 0



trees

Id3

The image shows the Weka Explorer software interface. The 'Classifier' tab is active, with 'Id3' selected. The 'Test options' section has 'Supplied test set' selected, and a 'Set...' button is visible. The 'Classifier output' area is empty. An 'Open' dialog box is overlaid on the main window, showing the 'Materiali' folder containing 'lensesTest.arff' and 'lensesTrain.arff'. The 'lensesTest.arff' file is selected. The 'File name' field contains 'lensesTest.arff' and the 'Files of type' dropdown is set to 'Arff data files'. Red arrows with numbers 1 through 5 point to specific elements: 1 points to the 'Choose' button, 2 to the 'Set...' button, 3 to the 'Open file...' button in the 'Test Instances' dialog, 4 to the 'lensesTest.arff' file in the 'Open' dialog, and 5 to the 'Start' button.

**Weka Explorer**

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

Classifier: Choose **Id3**

Test options

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

More options...

(Nom) Lenses ▼

Start Stop

Result list (right-click for options)

15:42:23 - trees.Id3

Classifier output

```

=== Run information ===

Scheme:      weka.classifiers.trees.Id3
Relation:    lensesTrain
Instances:   17
Attributes:  5
              Age
              Prescription
              Astigmatic
              Tear_rate
              Lenses

Test mode:   user supplied test set: 7 instances


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

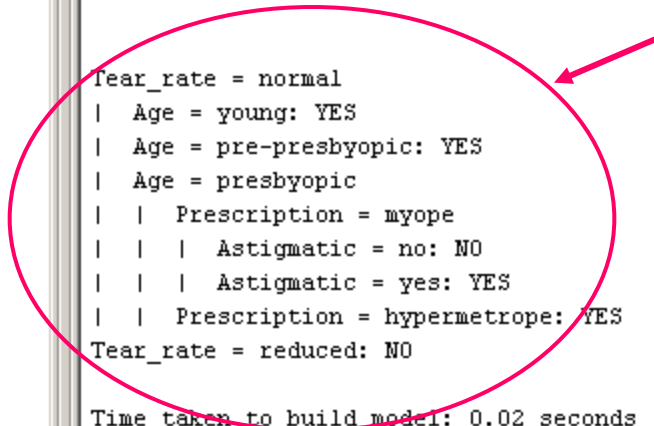
Id3

Tear_rate = normal
| Age = young: YES
| Age = pre-presbyopic: YES
| Age = presbyopic
| | Prescription = myope
| | | Astigmatic = no: NO
| | | Astigmatic = yes: YES
| | Prescription = hypermetrope: YES
Tear_rate = reduced: NO

Time taken to build model: 0.02 seconds
  
```

Decision tree

Status: OK Log  x 0



**Weka Explorer**

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

Classifier: Choose **Id3**

Test options:
 

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

 More options...

(Nom) Lece

Start Stop

Result list (right-click for options):
 

- 15:42:23 - trees.Id3
- 15:45:48 - trees.Id3

Classifier output:

Time taken to build model: 0.02 seconds

=== Evaluation on test set ===

=== Summary ===

Correctly Classified Instances	5	71.4286 %
Incorrectly Classified Instances	2	28.5714 %
Kappa statistic	0.4615	
Mean absolute error	0.2857	
Root mean squared error	0.5345	
Relative absolute error	59.375 %	
Root relative squared error	107.2232 %	
Total Number of Instances	7	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0.5	0.6	1	0.75	YES
0.5	0	1	0.5	0.667	NO

=== Confusion Matrix ===

```

a b <-- classified as
3 0 | a = YES
2 2 | b = NO
  
```

Classification accuracy

Confusion matrix

Status: OK

Log x 0