

Classification in WEKA

2010/11/25

Petra Kralj Novak
Petra.Kralj@jys.si

Practice with Weka

1. Build a decision tree with the ID3 algorithm on the lenses dataset, evaluate on a separate test set
2. Classification on the CAR dataset
 - Preparing the data
 - Building decision trees
 - Naive Bayes classifier
 - Understanding the Weka output

Weka

Weka is open source software for machine learning and data mining.
<http://www.cs.waikato.ac.nz/ml/weka/>

Download version 3.6

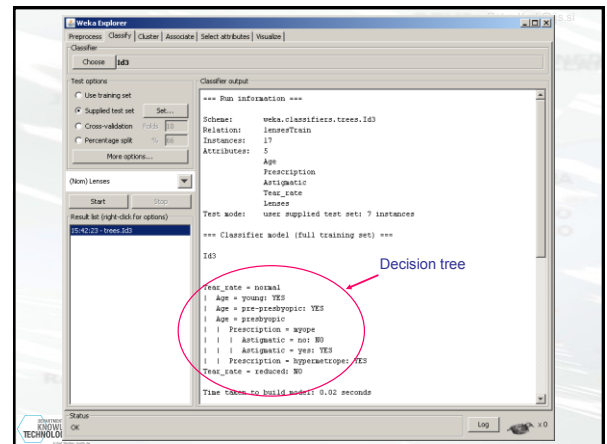
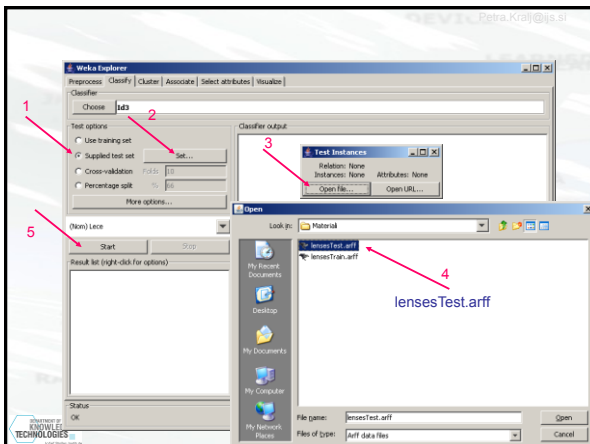
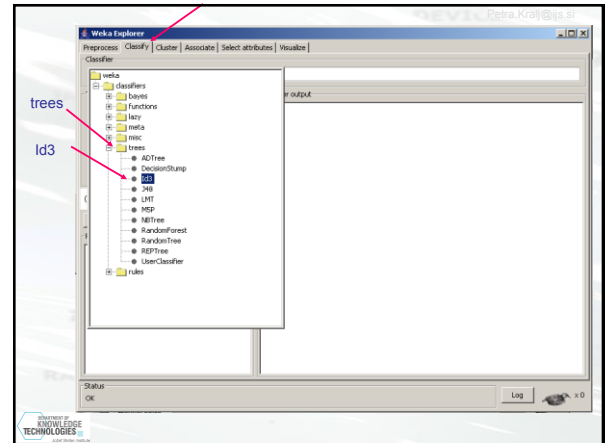
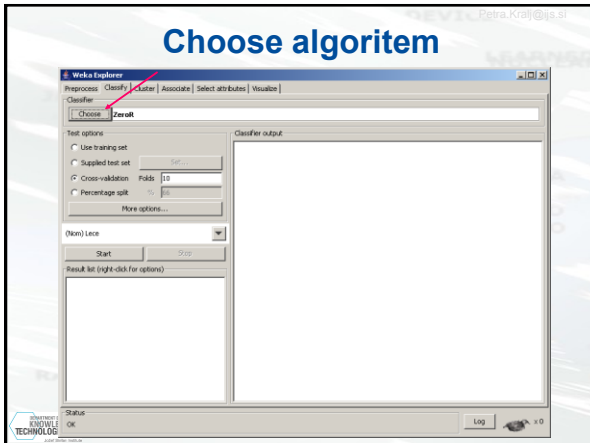
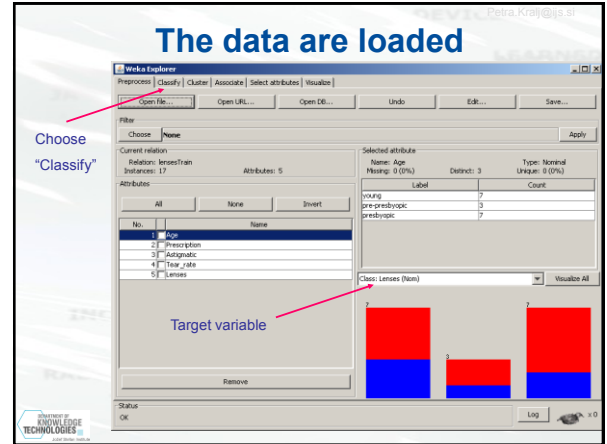
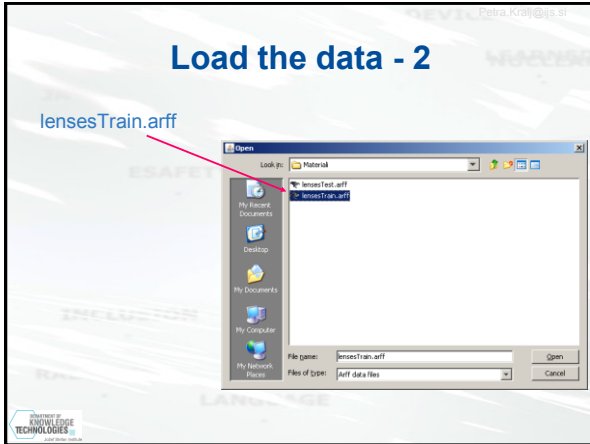
Run Weka

Choose Explorer

Exercise 1: 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



Classification accuracy

Time taken to build model:	0.02 seconds
--- Evaluation on test set ---	
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	107.2332 %
Total Number of Instances	7

Confusion matrix

```

*** Confusion Matrix ***
a b <- classified as
3 0 | a = "acc"
2 2 | b = "unacc"
    
```

Exercise 2: CAR dataset

- 1728 examples
- 6 attributes
 - 6 nominal
 - 0 numeric
- Nominal target variable
 - 4 classes: unacc, acc, good, v-good
 - Distribution of classes
 - unacc (70%), acc (22%), good (4%), v-good (4%)
- No missing values

Preparing the data for WEKA - 1

Data in a spreadsheet (e.g. MS Excel)

- Rows are examples
- Columns are attributes
- The last column is the target variable

Preparing the data for WEKA - 2

Save as ".csv"

- Careful with dots ".", commas ",", and semicolons ";"!

Load the data

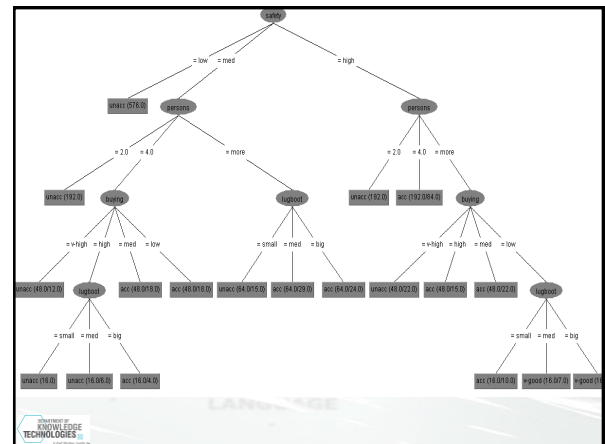
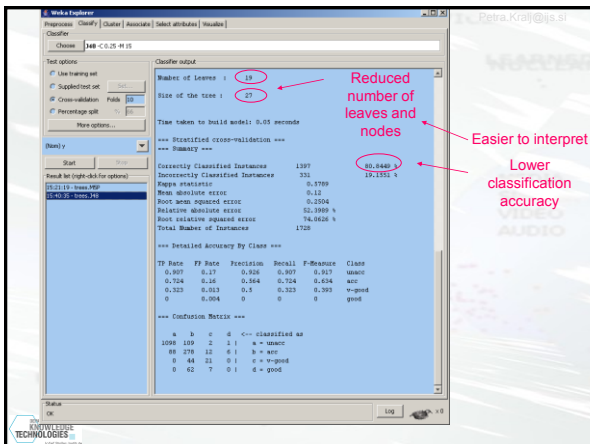
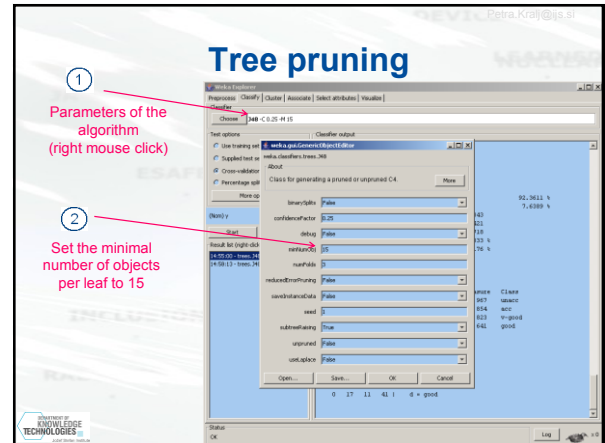
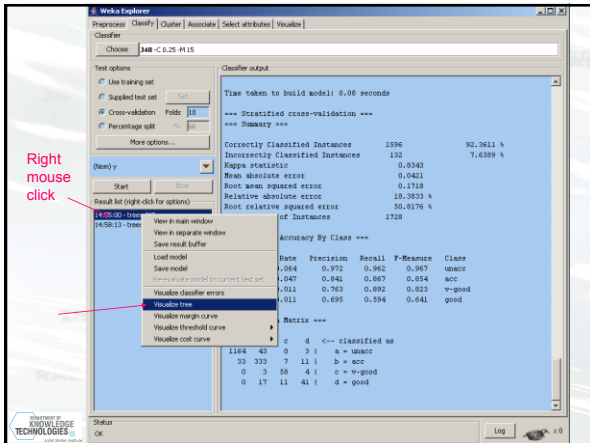
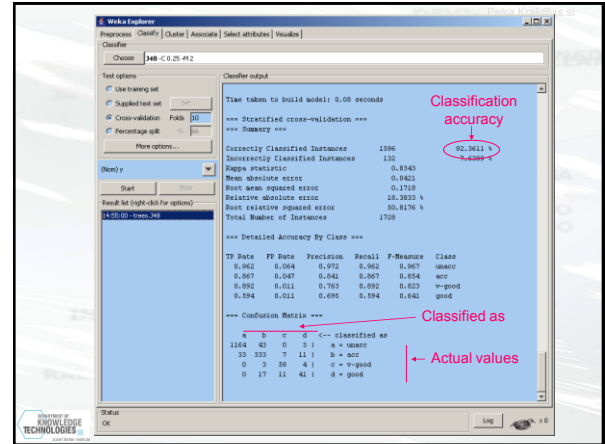
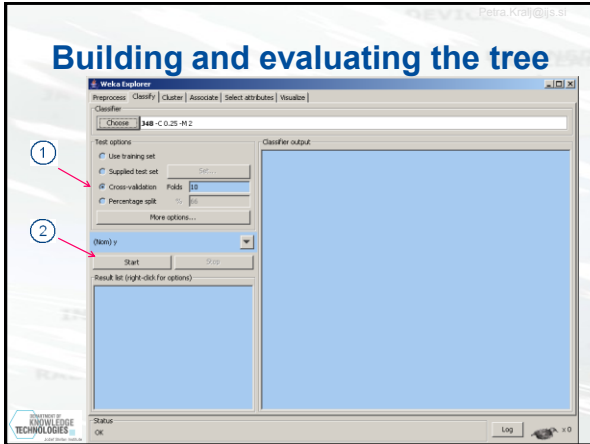
Car.csv

Target variable

Class	Count
unacc	1210
acc	384
v-good	95
good	89

Choose algorithm J48

1. Select 'Classifiers' in the tree view.
2. Select 'J48' in the list of classifiers.
3. Confirm the selection.



Naïve Bayes classifier

