Petra.Kralj.Novak@ijs.si

Classification rules and descriptive induction

12/11/2008

Voting dataset

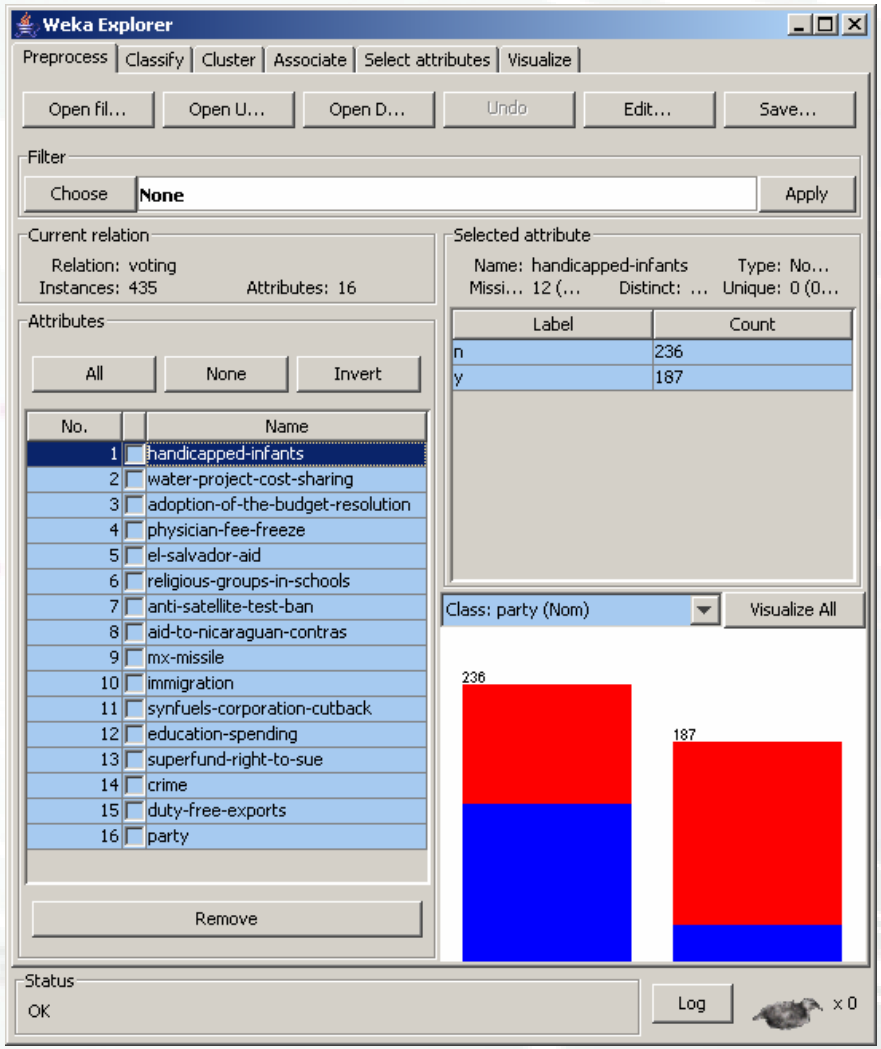
Iris dataset

DEPARTMENT OF KNOWLEDGE TECHNOLOGIES
Jožef Stefan Institute

Petra.Kralj.Novak@ijs.si

Voting dataset

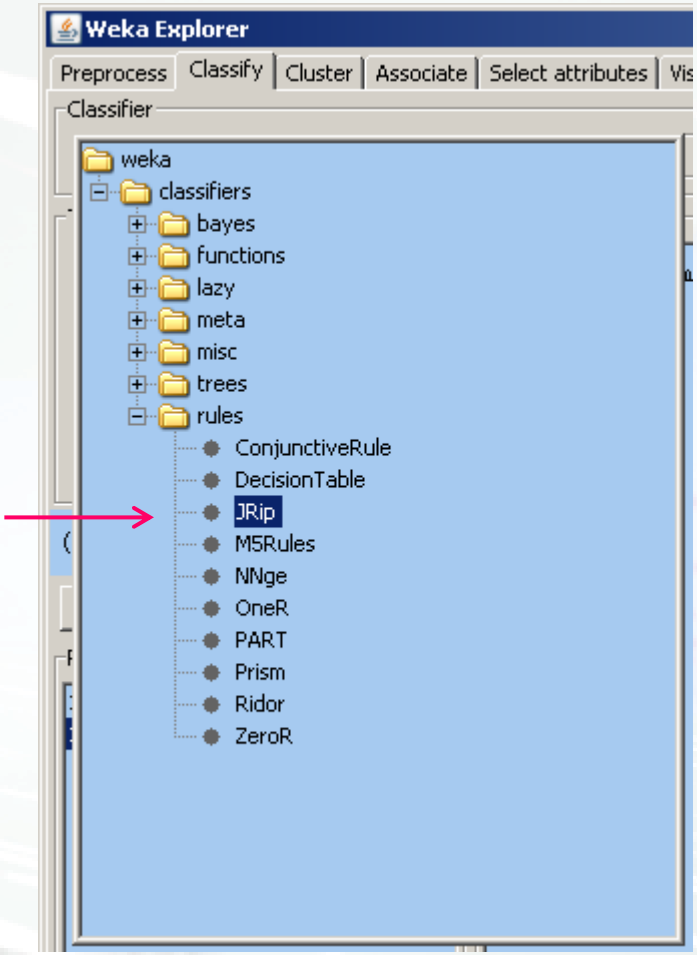
- 435 instances
- 16 attributes
  - 16 nominal attributes
  - 0 numeric attributes
- No target variable
- No missing values



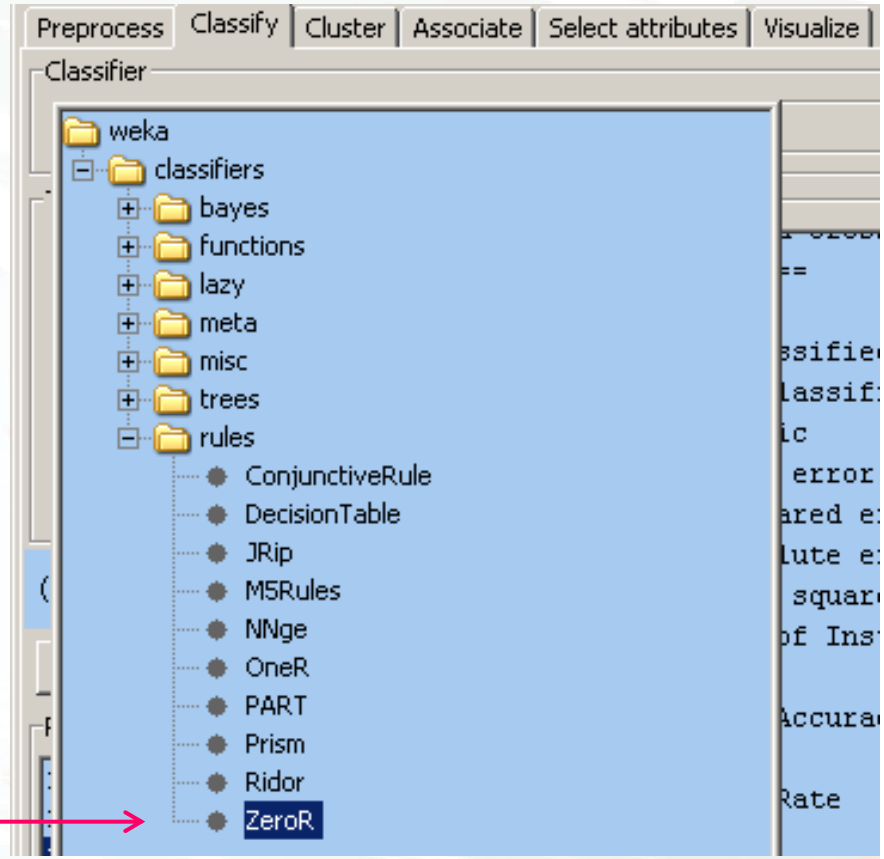
DEPARTMENT OF KNOWLEDGE TECHNOLOGIES
Jožef Stefan Institute

Petra.Kralj.Novak@ijs.si

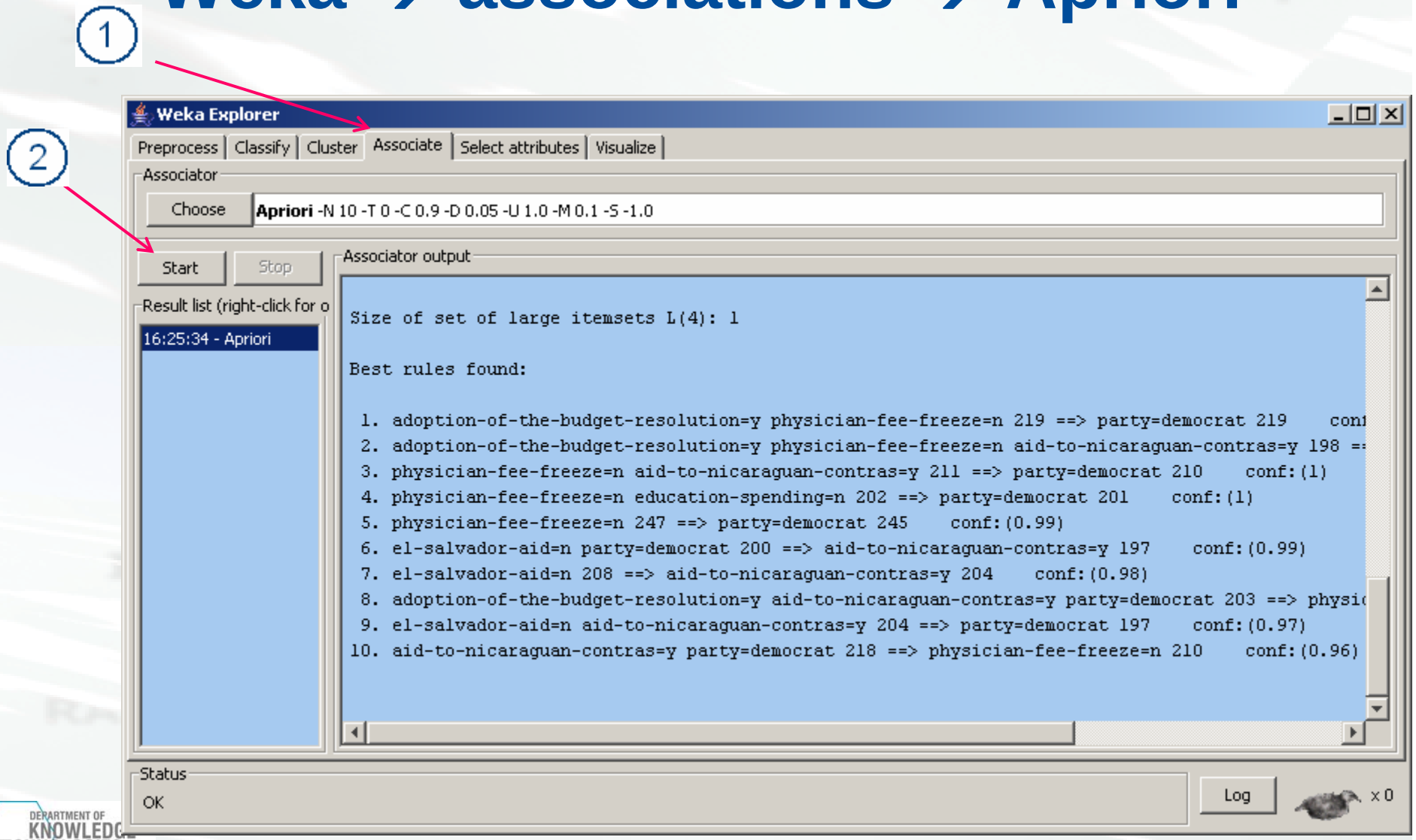
Classification rules: Weka → classifiers → rules → JRip



Baseline classifier:
Weka → classifiers → rules → ZeroR

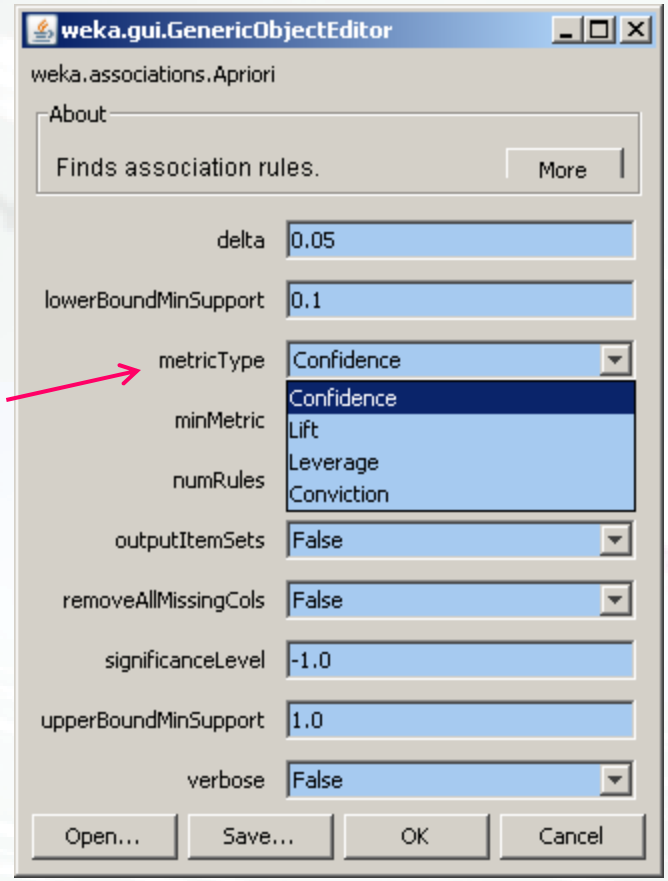


Association rules
Weka → associations → Apriori



Petra.Kralj.Novak@ijs.si

Quality of association rules



Petra.Kralj.Novak@ijs.si

Compare classification and association rules

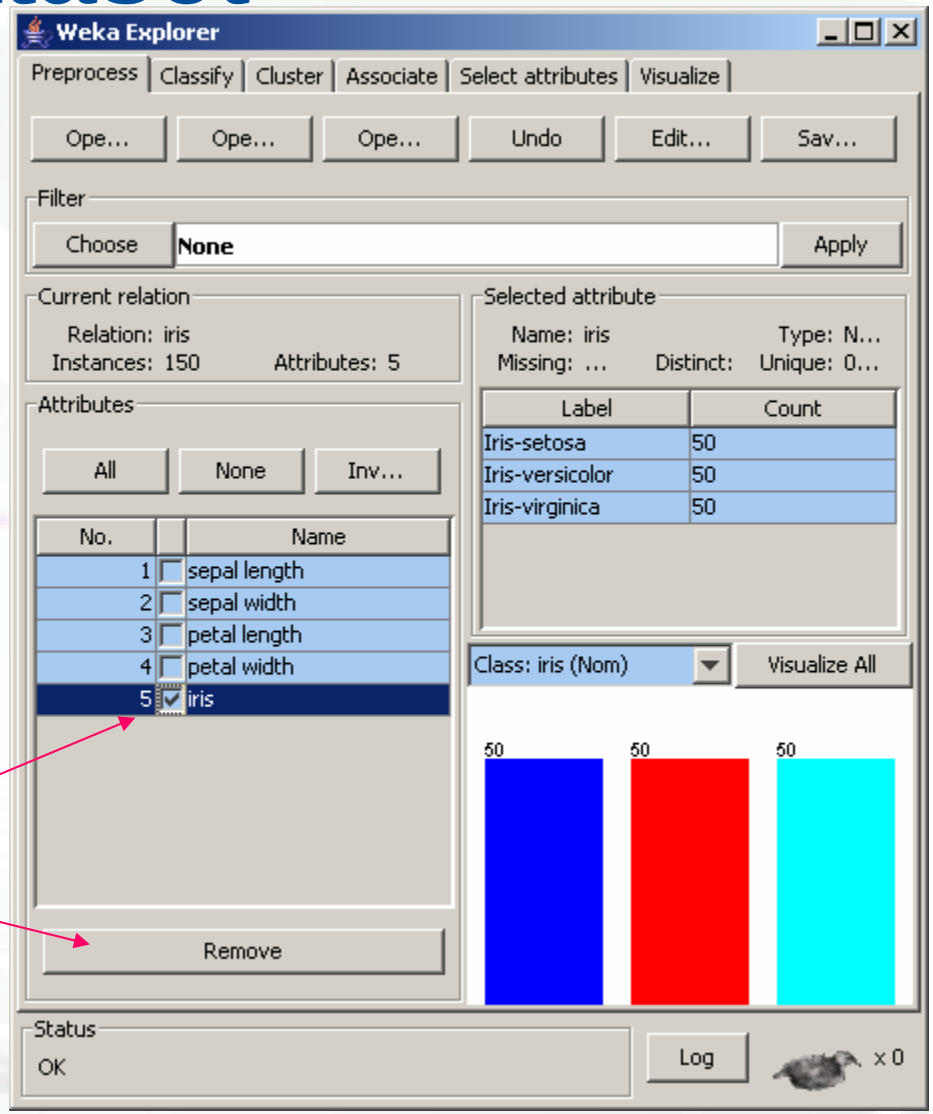
- Purpose
- Format
- Quality measure
- Ruleset / List of rules
- Exhaustiveness of algorithms

Compare classification and association rules

- Purpose
  - Classification rules: classification
  - Association rules: exploratory data analysis, descriptive induction
- Format
  - Both in the format X → Y
  - Classification rules: Y is a pair “target variable=class”
  - Association rules: both X and Y are itemsets ≅ conjunctions of attribute-value pairs
- Quality measure
  - Classification rules: classification accuracy of the ruleset, precision, rule accuracy, weighted relative accuracy
  - Association rules: support, confidence, lift, leverage, conviction
- Ruleset / List of rules
  - Classification rules: can be both: unordered sets of rules or ordered list of rules
  - Association rules: unordered set of rules
- Exhaustiveness of algorithms
  - Classification rules: heuristic algorithms
  - Association rules: exhaustive algorithms, guarantee the optimal results

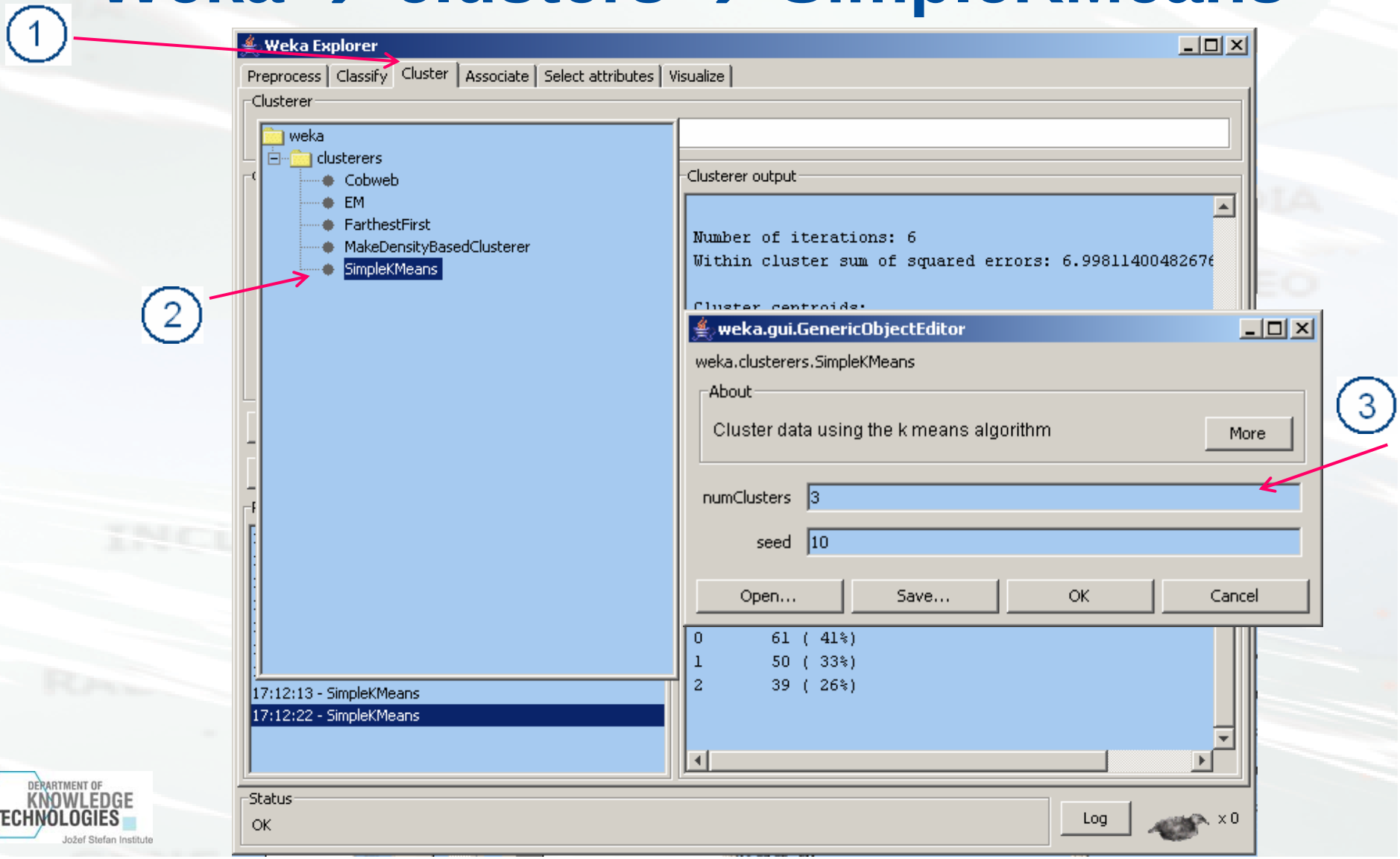
Iris dataset

- 150 instances
- 4 attributes
  - 0 nominal attributes
  - 4 numeric attributes
- Nominal target variable
  - 3 values:
    - Iris-setosa (30%)
    - Iris-versicolor (30%)
    - Iris-virginica (30%)
- No missing values



Petra.Kralj.Novak@ijs.si

Clustering
Weka → clusters → SimpleKMeans



Clustering visualization