Experiment 6

Classification Using ID3

Aim: To demonstrate Classification process on contact-lenses.arff dataset using ID3 algorithm with cross-validation.

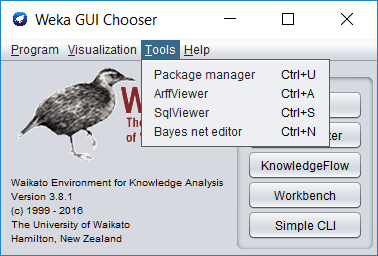
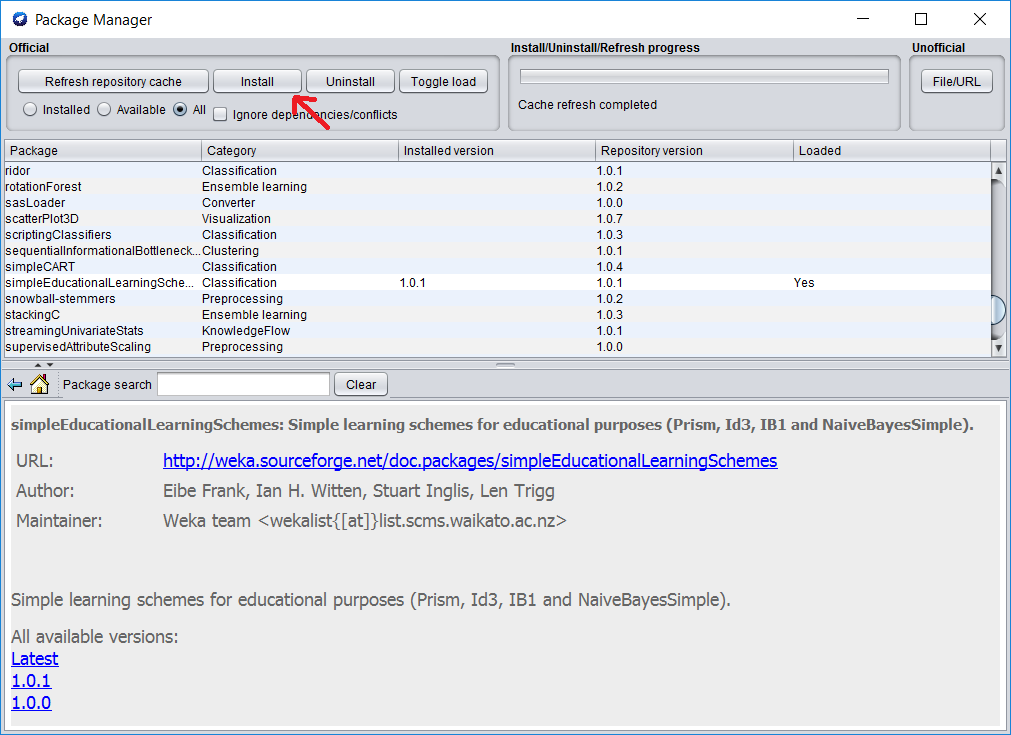
Tasks:

1. Install required package for ID3
2. Load contact-lenses.arff dataset
3. Build a classification model using ID3 algorithm with k-fold cross-validation.
4. Make predictions on new data.

Task 1: Install required package for ID3

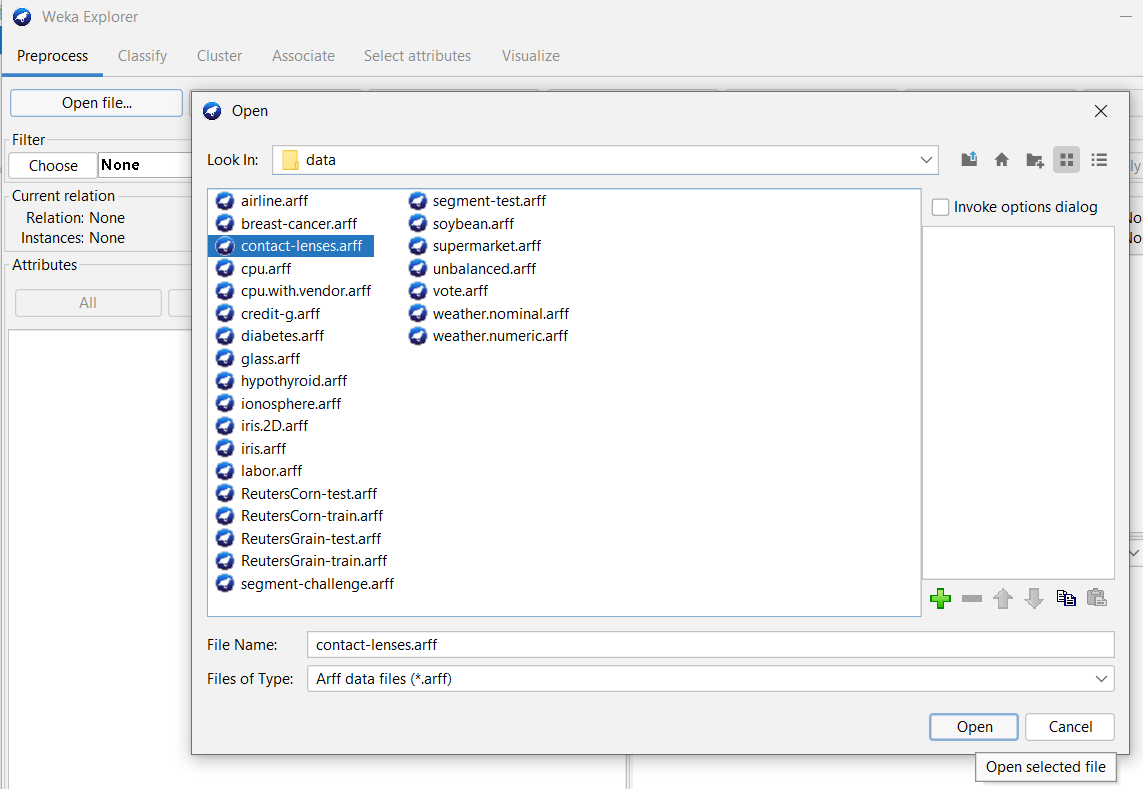
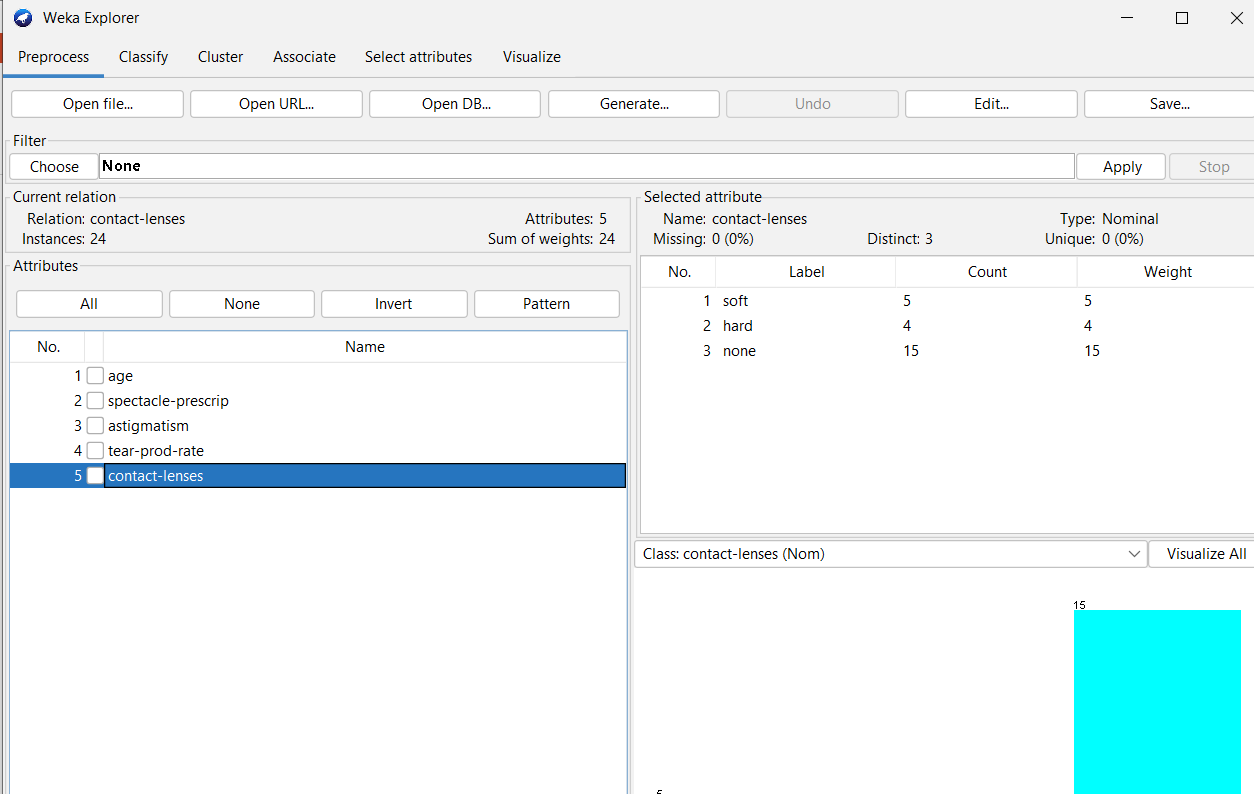
Select GUI Chooser 🡪 Tools 🡪 Package Manager

Search, select & install *simpleEducationalLearningSchemes* package to get ID3 classifier under trees group.

Task 2: Load contact-lenses.arff dataset

Load contact-lenses.arff from the Weka’s data folder.

Task 3: Build a classification model using ID3 algorithm with k-fold cross validation.

Classification is a process of determining the class (state) of the given instance.

ID3 stands for Iterative Dichotomiser 3 and is named such because the algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step. ID3 uses a top-down greedy approach to build a decision tree. ID3 is only used for classification problems with nominal features only.

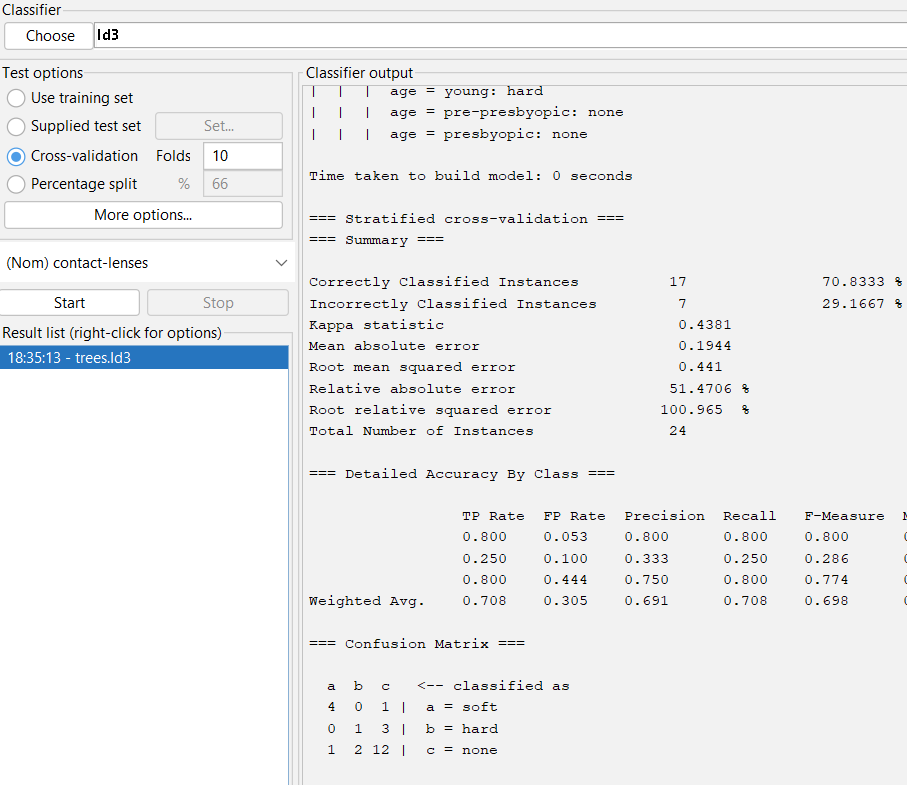
K-fold Cross Validation is a resampling procedure used to evaluate data mining models on a limited data set. It’s process is

* + 1. Split the input dataset into K groups
    2. For i from 1 to k
       - Take ith group as test dataset.
       - Use remaining K-1 groups as training dataset.
       - Fit the model using training set and evaluate its performance on test set.

Steps to build the model:

* 1. Click on Classify and select Cross-validation with some number folds under Test options group.
  2. Select Choose 🡪 classifiers 🡪 trees 🡪 ID3.
  3. Clock on Start.

Note that we can’t visualize ID3 tree.



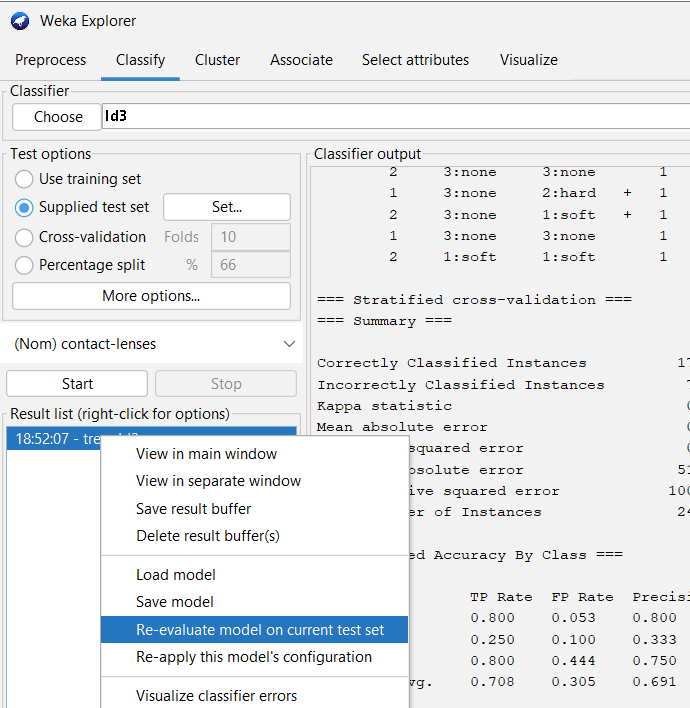
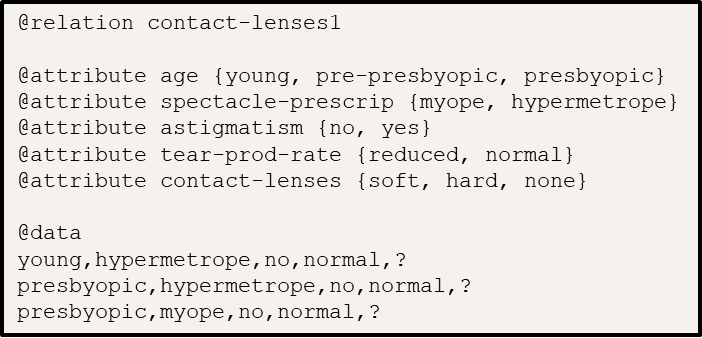
Observations:

* + Total number of instances:
  + Correctly classified instances:
  + Incorrectly classified instances:
  + Accuracy:
  + Calculation of Accuracy from Confusion Matrix:

Task 4: Make predictions on new data

Steps:

* 1. Create an ARFF file with unlabeled (use ? in the place of class label) instances.
  2. On the “Classify” tab, select the “Supplied test set” option in the “Test options” pane.
  3. Click the “Set” button, click the “Open file” button on the options window and select the new dataset.
  4. Click the “More options…” button and for the “Output predictions” option click the “Choose” button and select “PlainText”.
  5. Right click on the model in the “Results list” pane and Select “Re-evaluate model on current test set”.



Observations;

|  |  |
| --- | --- |
| Instance No. | Predicted class |
|  |  |

Conclusion: