Experiment 8

Classification Using Naïve Bayes

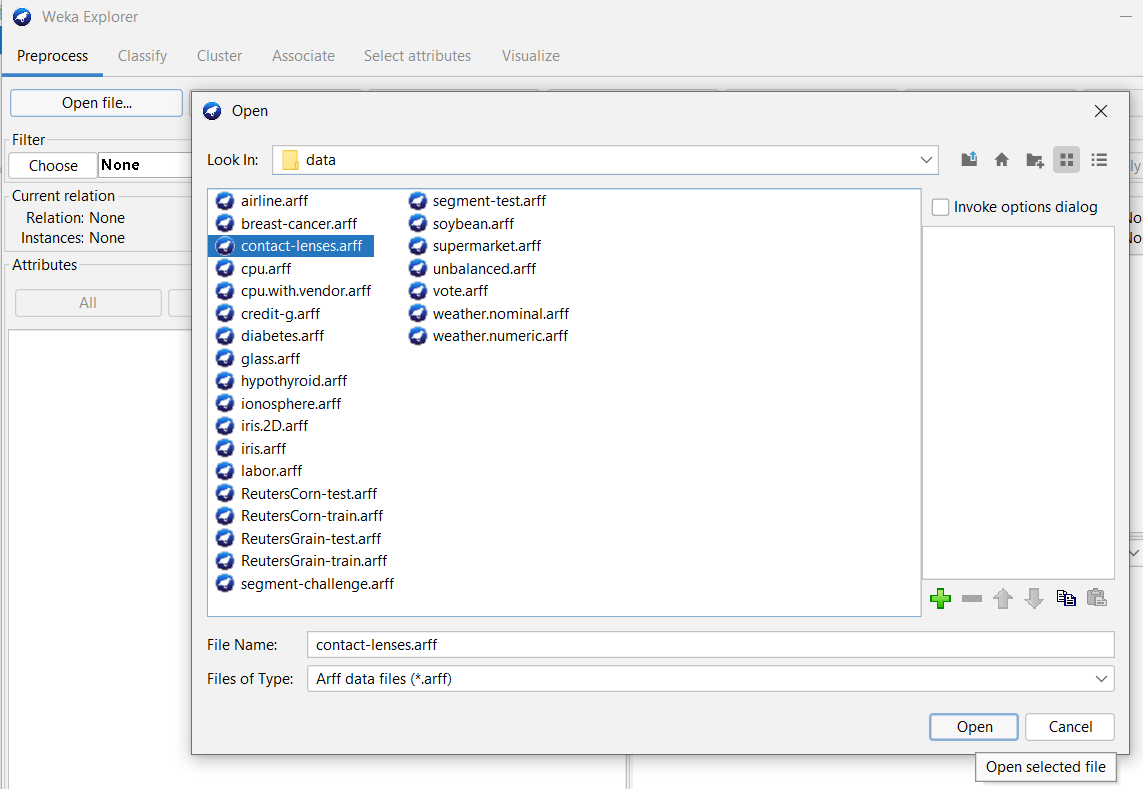
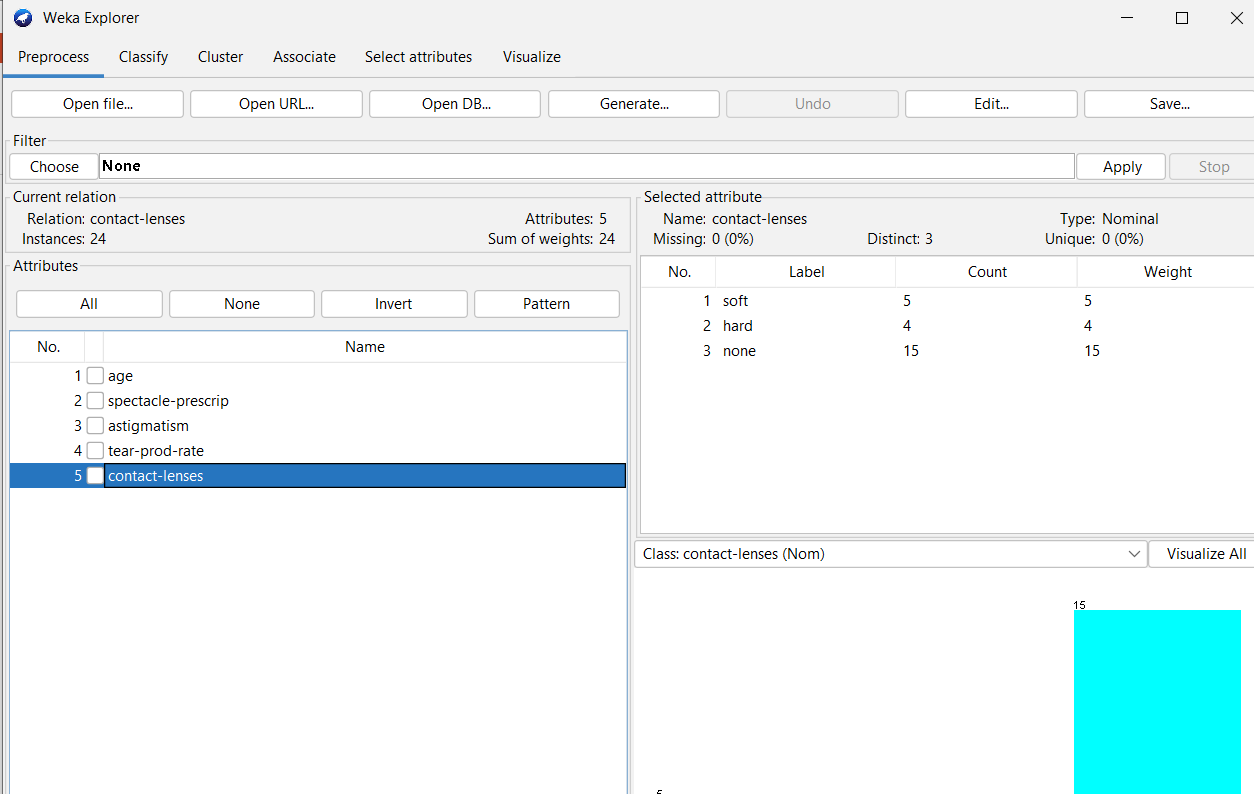
Aim: To demonstrate Classification process on contact-lenses.arff dataset using Naïve Bayes algorithm with cross-validation.

Tasks:

1. Load contact-lenses.arff dataset
2. Build a classification model using Naïve Bayes algorithm with k-fold cross-validation.
3. Make predictions on new data.

Task 1: Load contact-lenses.arff dataset

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

Task 2: Build a classification model using Naïve Bayes algorithm with k-fold cross validation.

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

Naive Bayes classifier is a popular machine learning algorithm based on Bayes' theorem. It is a probabilistic algorithm used for classification tasks. The "naive" in its name comes from the assumption that the features used to describe an instance are mutually independent, given the class label.

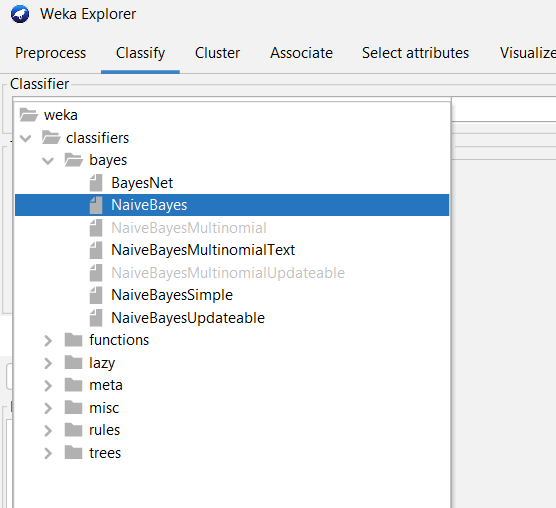
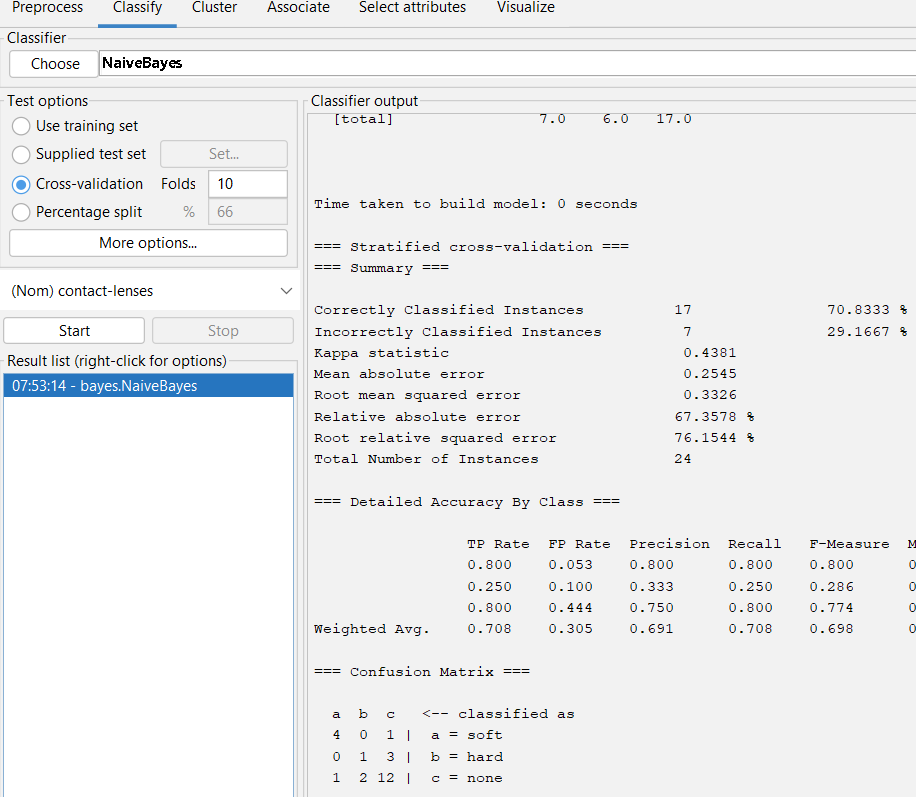
Assume the instances with n attributes and class label Then the prediction Where .

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 🡪 bayes 🡪 Naïve Bayes.
  3. Clock on Start.

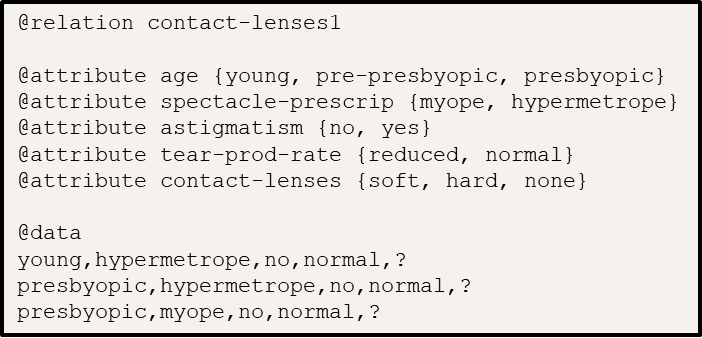
Observations:

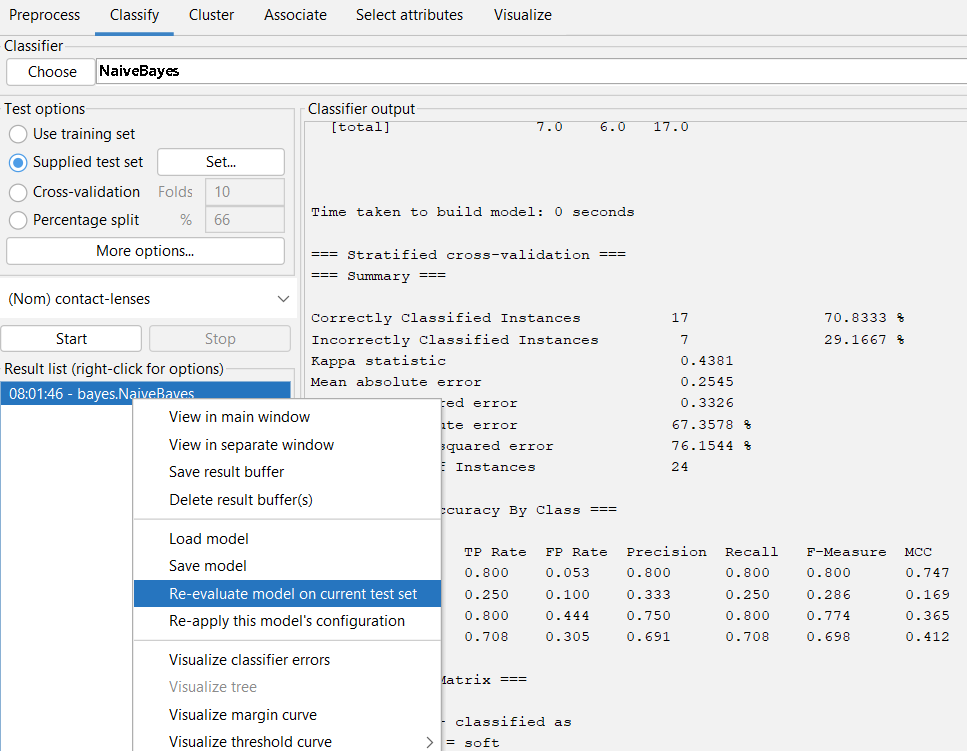
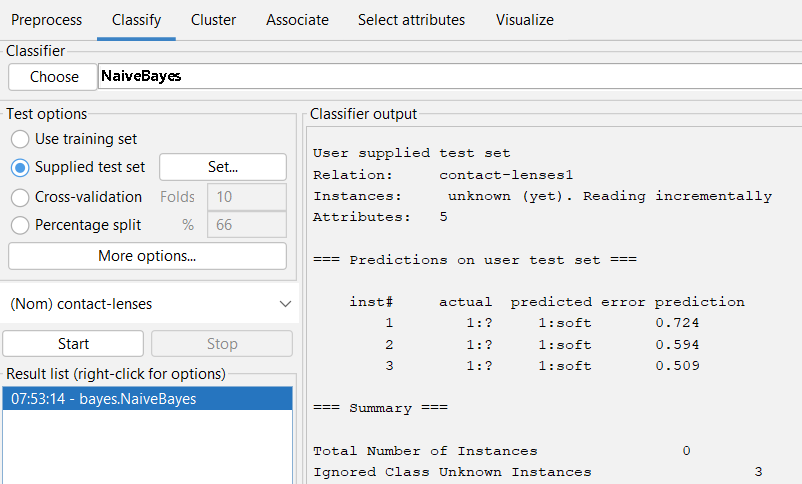
* + Total number of instances:
  + Correctly classified instances:
  + Incorrectly classified instances:
  + Accuracy:
  + Calculation of Accuracy from Confusion Matrix:

Task 3: 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: