**Exercise 1**

Credit dataset (page 15-16).

Perform:  
- handling missing value (page 17)  
- discretization (page 18-20)  
- changing the labels of value (page 21-22)  
- choose 5 attributes (page 23 - 27).

Graphical user interface

Description automatically generated

Firstly, we need to clean any missing values.

Graphical user interface, application

Description automatically generated

Next, we will discretize the values in the dataset.

Graphical user interface, application

Description automatically generated

Then, we will conduct InfoGainAttribute to choose best 5 attributes.

3. Click to tab 'Classify' and choose J48. You will create a tree model for the dataset (page 39 & 44).

Graphical user interface

Description automatically generated

J48 Classifier using Cross Validation – 10 folds

Graphical user interface

Description automatically generated

J48 Classifier using Percentage Split 66%

Graphical user interface

Description automatically generated

J48 Classifier using Percentage Split 80%

4. Evaluate your model using cross validation and percentage split. Visualize classifiers error for both results. Discuss with your friend on the difference of the evaluation method (page 45)

It is shown that the accuracy is better when using Percentage Split 80% with 76.5 accuracy, followed by Percentage Split 66% with 73.5% and Cross Validation 10-folds 72.5%.

A picture containing text, antenna

Description automatically generated

Visualize Tree of Cross Validation – 10 folds

A picture containing text, antenna

Description automatically generated

Visualize Tree of Percentage Split 66%

A picture containing text, antenna

Description automatically generated

Visualize Tree of Percentage Split 80%

5. Check more results for detailed accuracy in cross validation (page 48 - 50). Find the TP, TN, FP and FN

|  |  |
| --- | --- |
| **Type** | **Values** |
| True Positive | 610 |
| True Negative | 185 |
| False Positive | 90 |
| False Negative | 115 |