Student number:B00060572

**Exercise 1**

* 1. What type of data is in the dataset? Is it mainly numeric attributes, polynominal attributes, or a mixture of both

a) Mixture of Both.

* 1. How many classes do you need to predict? And are they evenly balanced?

a) 3 classes, No they aren’t evenly balanced; Bad Profit =585, Bad Loss = 220, Good Risk = 195.

1.3 From your initial exploration of the dataset, how accurate do you expect a classification model to be? Justify your answer by looking at a scatter plot matrix.

a) The Income/Age plot shows some attributes do clearly indicate grouping. I expect the classification model to somewhat accurate.

**Exercise 2**: Explain the role of each of the operators in the X-Validation block.

a) **Decision Tree**: Generates a Decision Tree for classification of both nominal and numerical data.

**Apply Model**: This operator applies an already learnt or trained model on an Example Set.

**Performance Operator**: This operator is used for performance evaluation. It delivers a list of performance criteria values. These performance criteria are automatically determined in order to fit the learning task type.

**Exercise 3**

3.1. How accurate is the decision tree classifier?

a) accuracy: 74.50%.

3.2. Does it predict some classes more accurately than others?

a) it predicts bad profit more accurately than other classes.

3.3 What are the most significant attributes in determining the class label?

a) Loans and store car are at the top of the tree, these are good attributes for determining label.

**Exercise 4**

Exercise 4.1 How accurate is k-NN compared to DT?

a) k-NN = 56.90 DT = 74.5.

Exercise 4.2 Does normalising attributes improve model accuracy?

a) Yes normalizing the data for k-NN to improves accuracy by nearly 10% from 66% to 75%.

**Exercise 5**:

5.1 What is the optimal number for k?

a) 51

5.2 How does the accuracy of DT compare to k-NN accuracy when using the optimal value of K?

a) With k is set to 51 the DT accuracy is below the accuracy of k-NN. The accuracy percent of k-NN with k at 51% is 75.70% compared to the DT accuracy percent of 74.5%.