Jacob Perrone

Assignment 4 Report

Part 1

Part C-D)

Graphical user interface, text, application

Description automatically generated

From the original dataset I went and sorted each column by largest value to see if there was any incorrect or missing data. I deleted the records that had incorrect data as there was only a few of them, the screen shot above shows the missing data handling from XLMiner which deleted records with missing data in any of the columns. Looking at the scatterplot matrix is not necessary as this is categorical data and dummies were not needed as this is a classification task.

Part E-I)

Table

Description automatically generated

Model 1 was chosen as the best model due to having the highest validation accuracy and high ROC value (validation) as well as other metrics such as low error and high precision and sensitivity.

Training Lift charts

Graphical user interface, chart, line chart

Description automatically generated

Validation Lift charts

Graphical user interface, chart, line chart

Description automatically generated

Scoring on New data

Table

Description automatically generated

Part 2

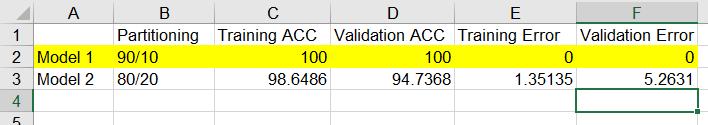
Part C-D)

Table

Description automatically generated

Going through the original dataset, I sorted each predictor column to see if there was any incorrect of missing values. There were none and data handling in XLMiner showed that no records were deleted for missing data. Scatterplot matrix and dummy creation were not necessary as all predictors are categorical and this is a classification problem. From the predictors, the Living predictor can be omitted moving forward as all of the records for that column are the same value which could throw off the model and lead to bad performance.

Part E-I)



Model 1 was chosen as the best model due to having the highest validation accuracy.

Scoring on new data

Table

Description automatically generated

The deployed model was not able to classify starfish, I suppose that this is due to the list of predictors that are used and thus changing which ones the model uses could yield to a classification for starfish.