Part 1: Data Mining, Machine Learning and Text Analytics

Exercise 1:

All exercise 1 screenshots support answers are in Appendix.

1. Data Mining and Machine Learning

   d.

   1) Status Category 96NK L level has the highest count.

   2) Yes, the variable Age contain 26,477 missing values.

   3) The average of Target Gift Amount is 15.62.

   e.

   1) Responders and non-responders each account for half of the data.

   2) 28,699 females responded to the campaign.

2. Training a Neural Network Model in SAS Visual Data Mining and Machine Learning

   h.

   1) 64,988 observations are used by algorithm.

   2) Because many modelling algorithms in SAS Visual Data Mining and Machine Learning operate under complete case analysis, which algorithms use only cases with no missing data. All observations here which are not used have missing data.

   3) The misclassification rate for the model created with default settings is 0.4751.

   i. The updated model produces a validation misclassification rate of 0.4379, which indicates a slight improvement over the previous result.

   j.

   2) The updated model produces a validation misclassification rate of 0.4393.

   3) The updated model produces a validation misclassification rate of 0.3755. This represents a significant improvement over previous result.

   4) The top 10% (percentile) of the data has about 1.58 times as many responders as a random ordering of the data, and 0.32 times fewer responders than a perfect ordering of the training data.

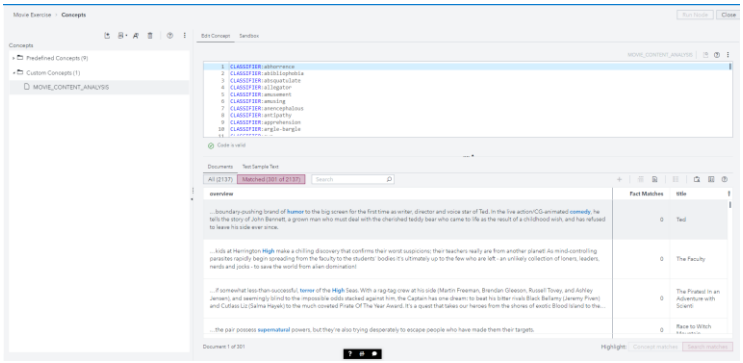
Exercise 2:

Comedies:

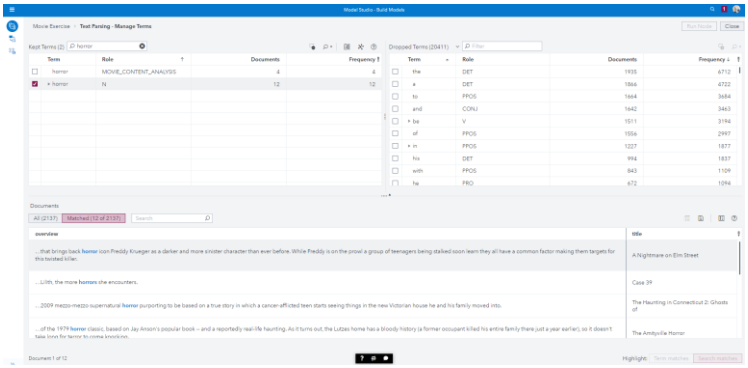
- The Meaning of Life
- Chasing Amy
- American Reunion

Horrors:

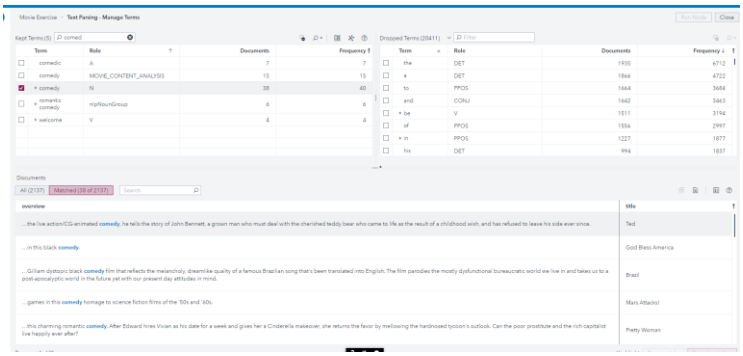
- Monster House
- Case 39
- Scary Movie 2



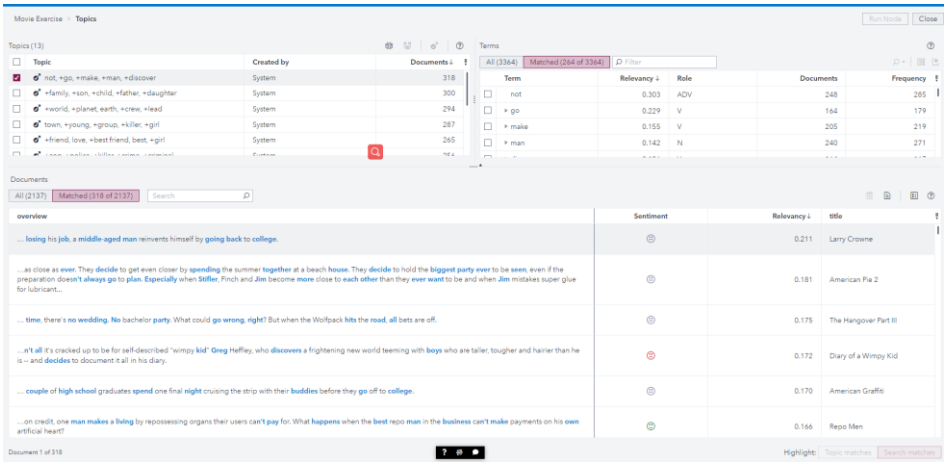
Concepts node



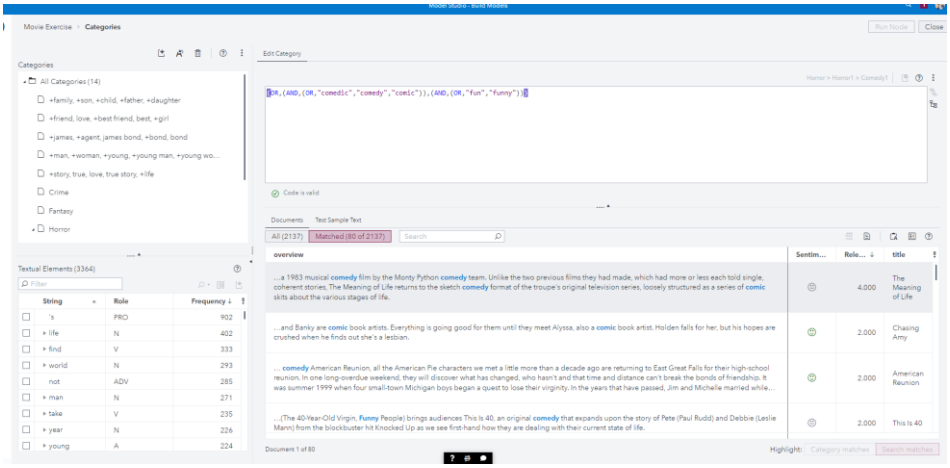
Text parsing node (Horrors)



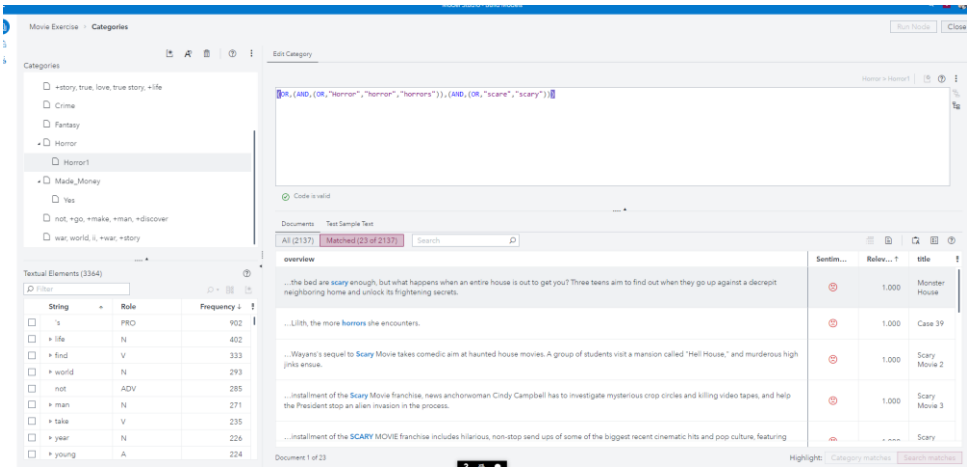
Text parsing node (Comedies)



Topic node



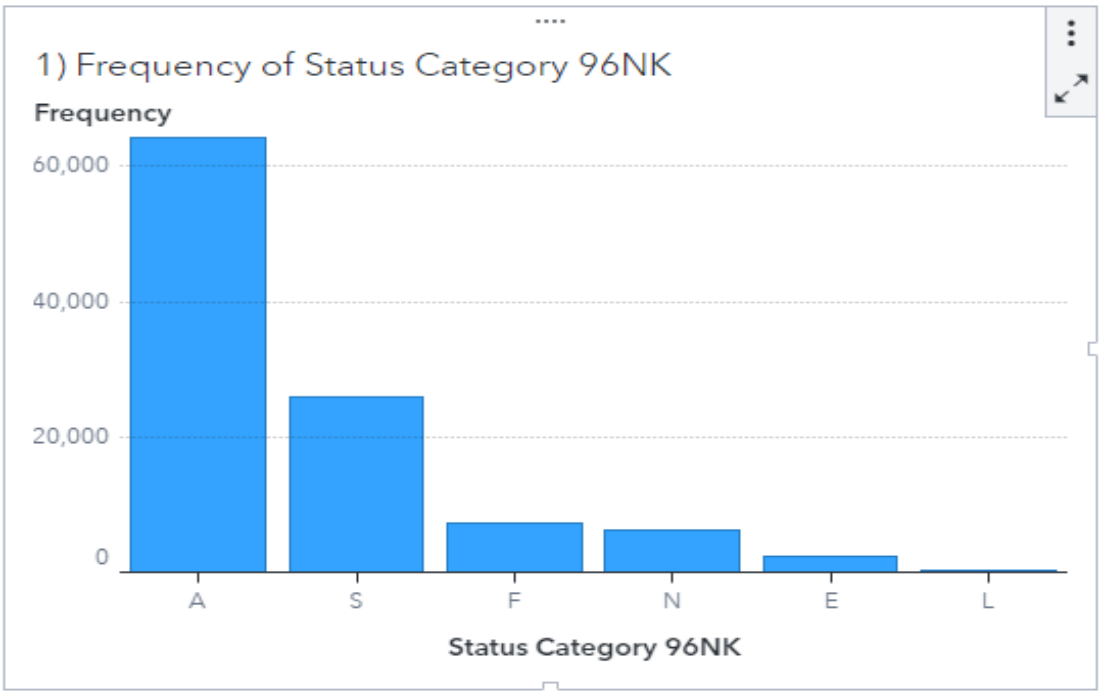
Category node (Comedies)



Category node (Horrors)

Appendix:

Part 1: Exercise 1

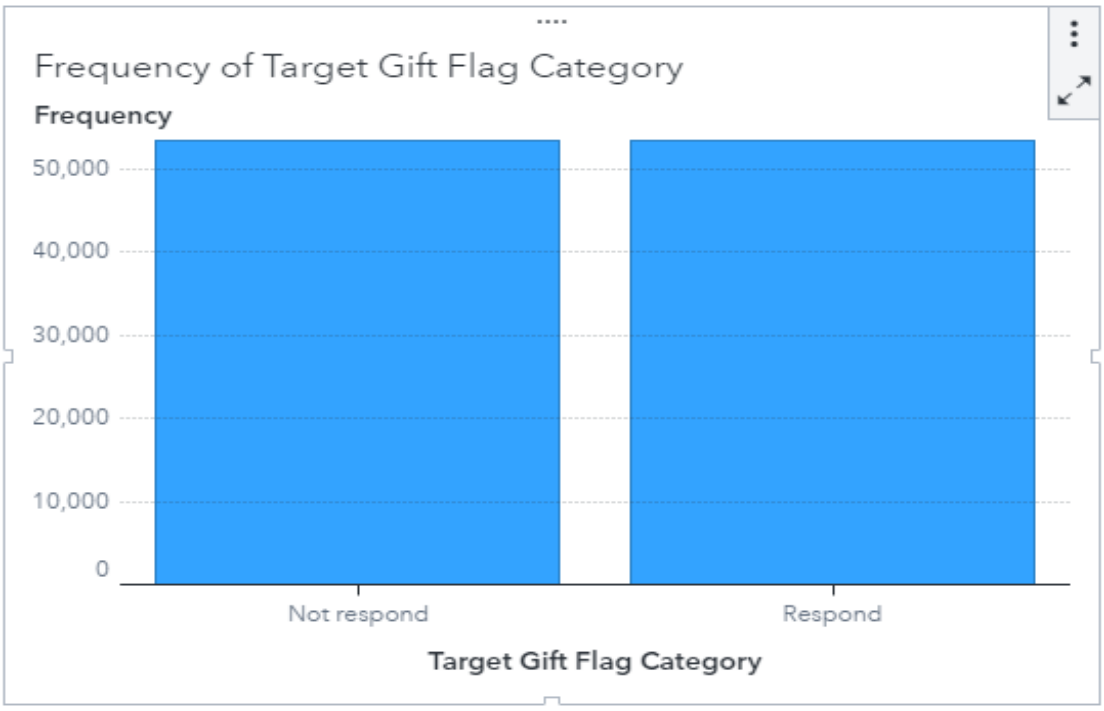


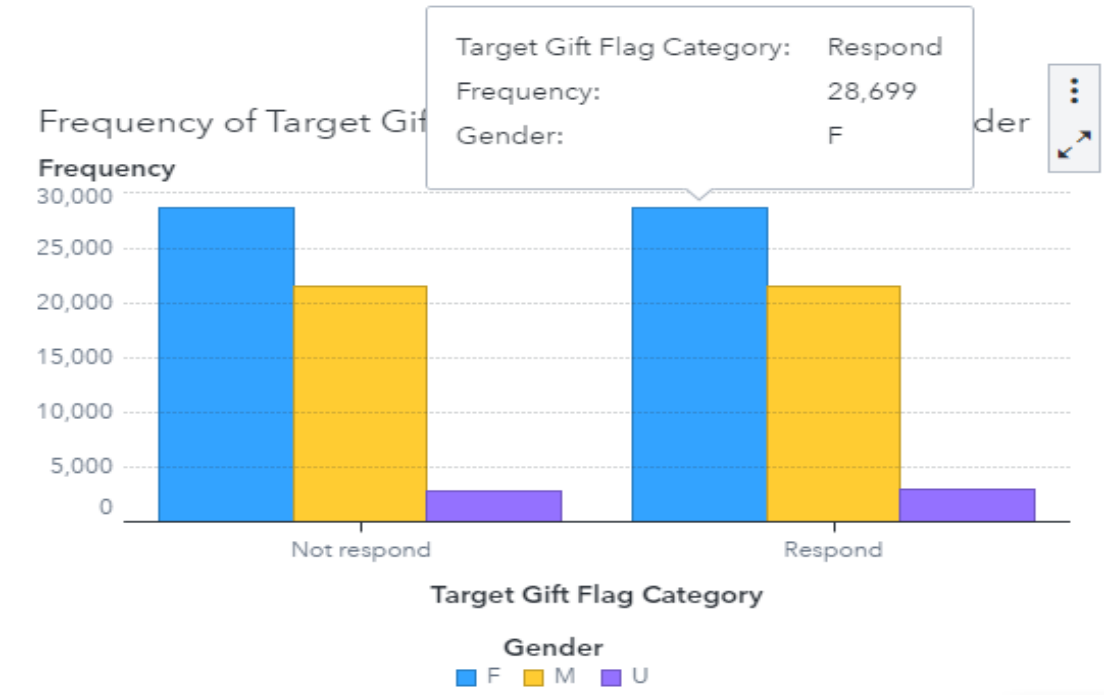
Name	Minimum
Age	0.00
Gift Amount Average 36 Months	0.00
Gift Amount Average All Months	1.50
Gift Amount Average Card 36 Months	1.33
Gift Amount Last	0.00

More information

Standard Deviation:	16.52
Standard Error:	0.06
Variance:	272.76
Distinct Count:	79
Number Missing:	26,477

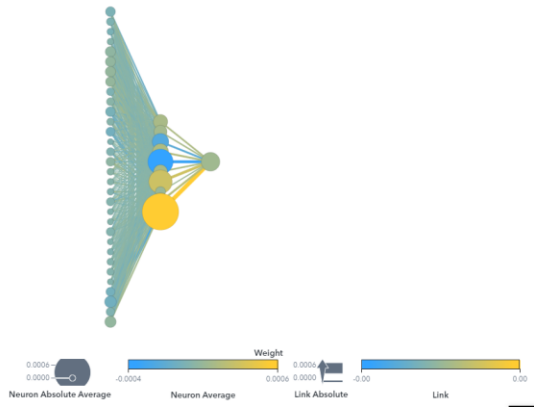
Name	Minimum	Maximum	Average	Sum
Target Gift Amount	1.00	200.00	15.62	832,355.70


Frequency of Target Gift Flag Category
Frequency
50,000
40,000
30,000
20,000
10,000
0
Not respond
Respond
Target Gift Flag Category

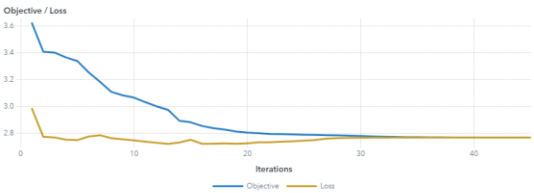

Target Gift Flag Category: Respond
Frequency: 28,699
Gender: F
Frequency
30,000
25,000
20,000
15,000
10,000
5,000
0
Not respond
Respond
Target Gift Flag Category
Gender
F
M
U

Neural Network **Target Gift Flag Category** (event=Respond) Misclassification Rate **0.4751** Observations Used **64,988** Unused **41,558** Create Pipeline

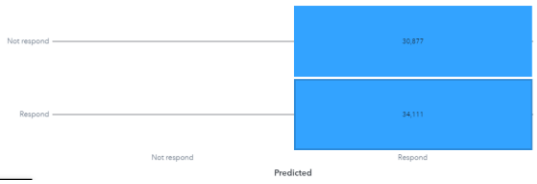
Network



Iteration Plot

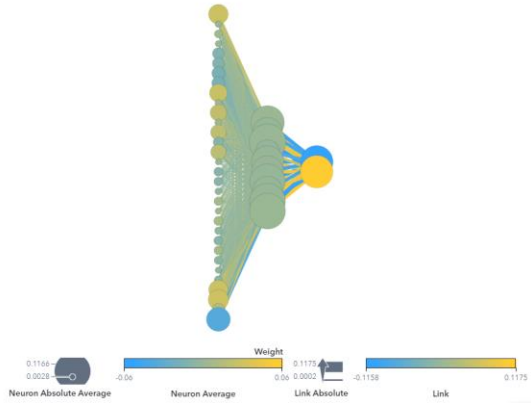


Confusion Matrix

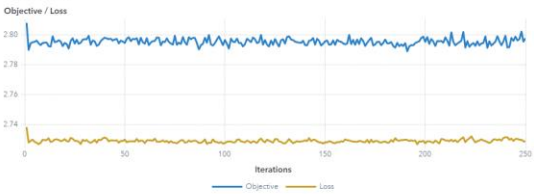


Neural Network **Target Gift Flag Category** (event=Respond) Misclassification Rate **0.4379** Observations Used **64,988** Unused **41,558** Create Pipeline

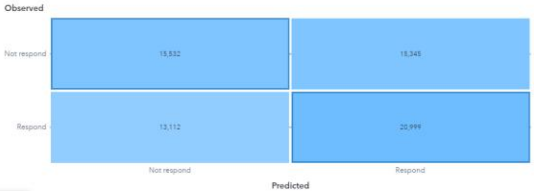
Network



Iteration Plot

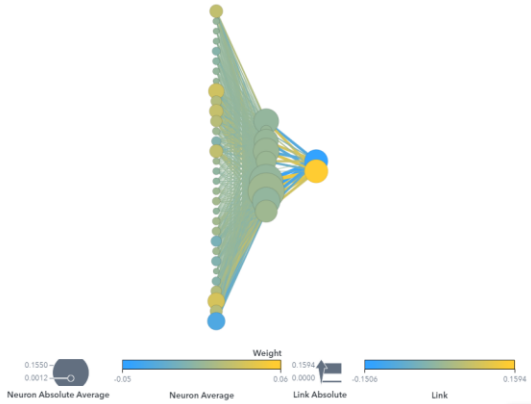


Confusion Matrix

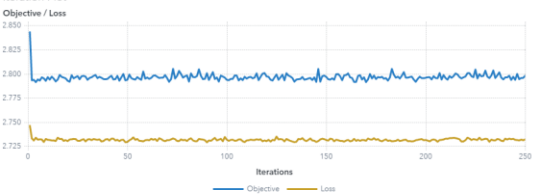


Neural Network **Target Gift Flag Category** (event=Respond) Validation Misclassification Rate **0.4393** Observations Used **64,988** Unused **41,558** Create Pipeline

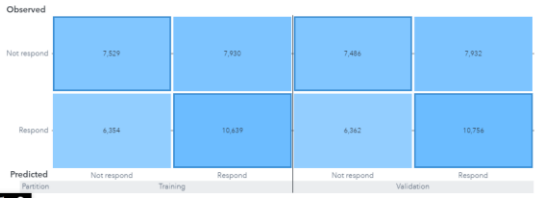
Network

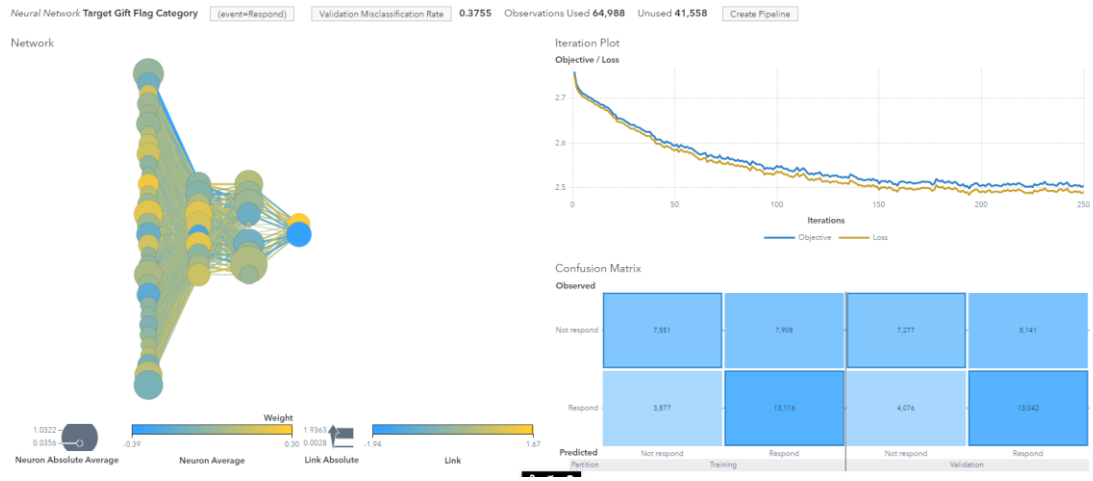


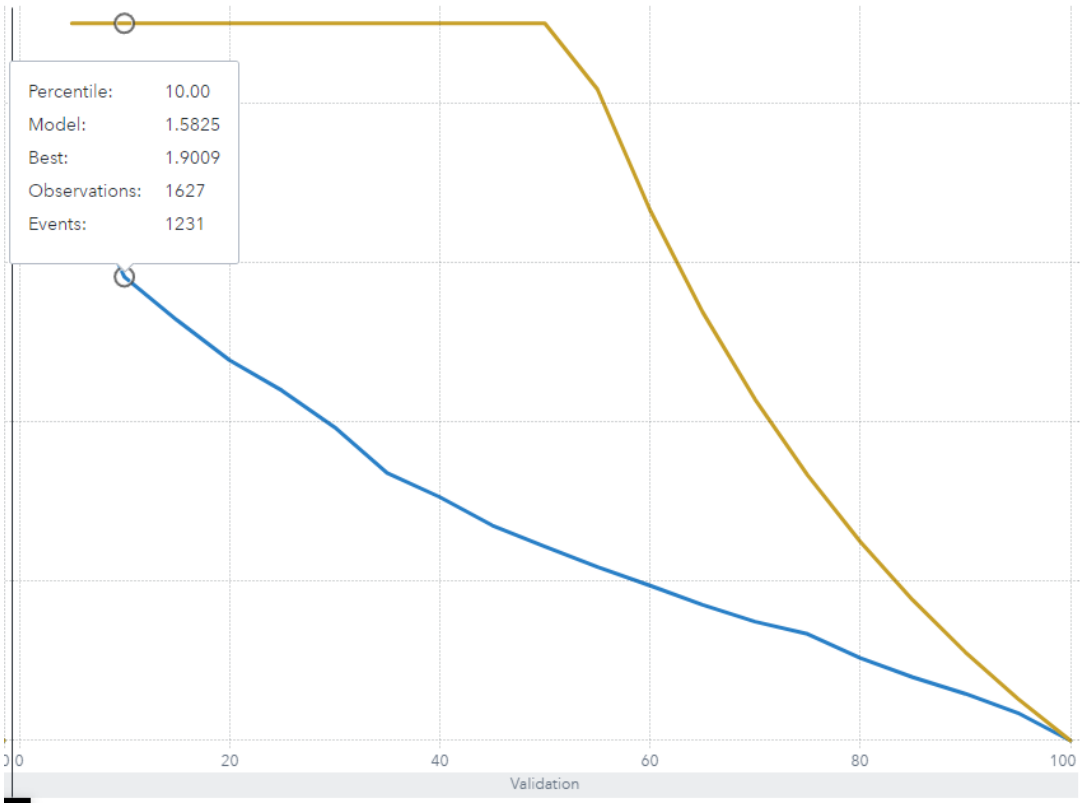
Iteration Plot



Confusion Matrix







Part 2: Exercise 2