**Due**: 3/27  
**Note: Show all your work.**

**Additional file attached for calculations: Valencia\_Anthony\_HW7.xlsx**

**Assignment 7**

**Problem 1 (10 points)** This question is about a learning classifier system XCS which we discussed in the class. Consider the following population, which has the current set of rules:

1001 01

10#1 10

1#0# 01

#0#1 10

010# 10

1#0# 10

1001 10

Suppose that a sample 1001 10 is extracted from the training dataset.

(1). Generate the match set.

**1001 01**

**10#1 10**

**1#0# 01**

**#0#1 10**

**1#0# 10**

**1001 10**

(2). Determine the action from the match set.

**Action 10 is chosen**

(3). Generate the action set.  
**10#1 10**

**#0#1 10**

**1#0# 10**

**1001 10**

(4). Which rules are rewarded?

**10#1 10**

**#0#1 10**

**010# 10**

**1#0# 10**

**1001 10**

Which rules are not rewarded?

**1001 01**

**1#0# 01**

**Problem 2 (10 points).** This problem has two parts. Problem 2-1 uses Oracle and Problem 2-2 uses JMP Pro. You need to choose one of the two.

**Problem 2-1**. Follow the instructions in *oracle-classification-assignment.pdf* file. The submission requirements are indicated with “Required.”

**Problem 2-2.** For this part, you will use JMP Pro to build and test five classifier models – Naïve Bayes, KNN, Partition (decision tree), Boosted Tree, and Neural Network. Follow the instruction given below. You will use *german-bank.jmp* dataset for all five models. For more details about how to use each classifier model, refer to *Predictive and Specialized Modeling.pdf* documentation.

(1). Naïve Bayes

* Start JMP Pro
* Open *german-bank.jmp*
* Select Analyze > Predictive Modeling > Naïve Bayes
* Select *checking* through *foreign* and click *X,Factor*
* Select *class* and click *Y,Response*
* Select *Validation* (on the left) and click *Validation* (on the right)
* Click OK.

You will see Naïve Bayes Report (or output) that is similar to Figure 9.1 in *Predictive and Specialized Modeling.pdf* documentation.

* Capture the report screen and paste it in your submission. 
* The dataset’s class attribute has two possible values – 1 and 2.
* Using the confusion matrix of *Validation* in the report (There are two confusion matrices.

Make sure that you use the *Validation* confusion matrix), calculate the following measures for both classes (similar to those in Weka’s output window):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | Class |
| 0.857 | 0.143 | 0.810 | 0.857 | 0.833 | 1 |
| 0.525 | 0.475 |  |  |  | 2 |

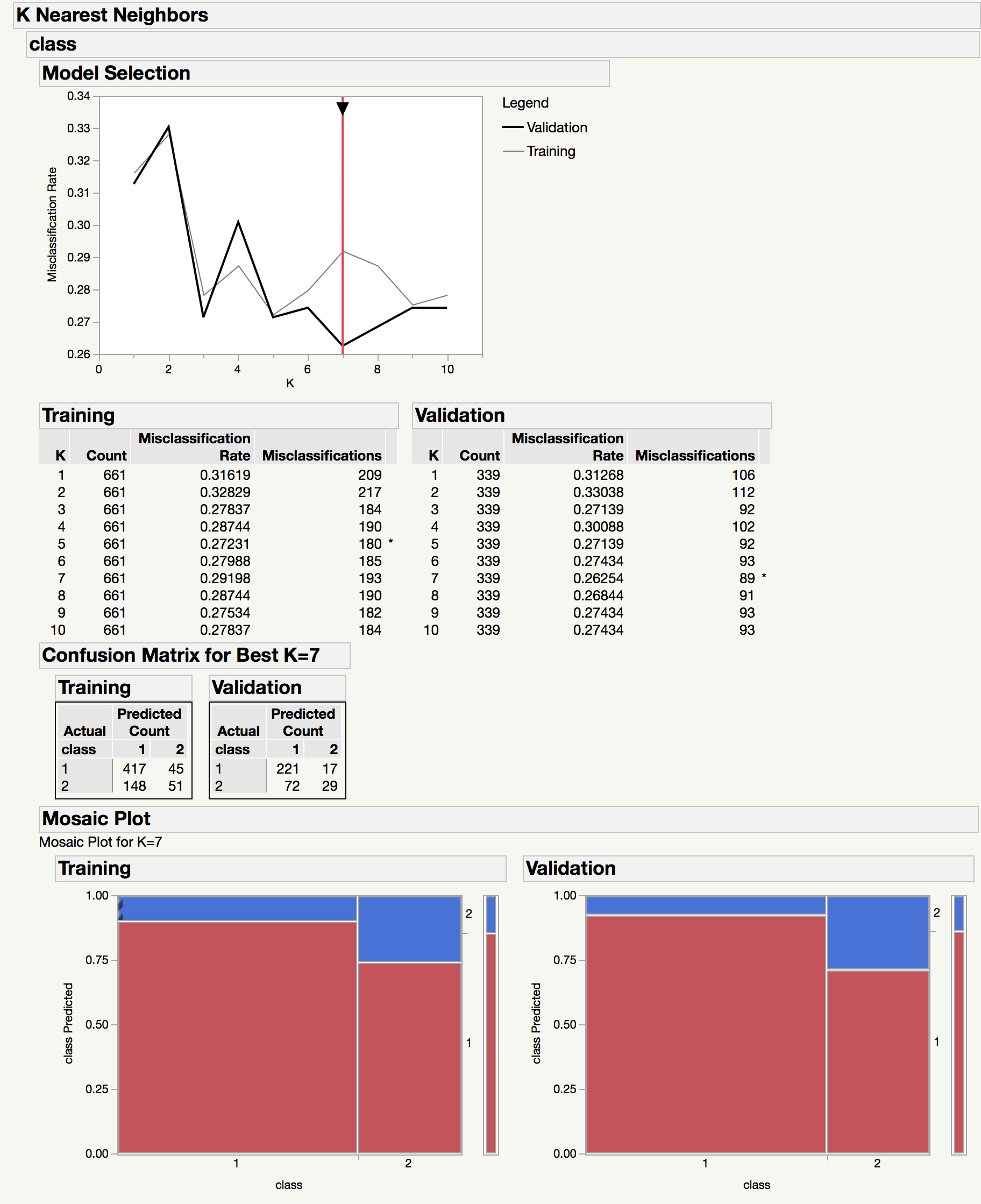
**Note**: the TP Rate for class 2 is actually the TN Rate and the FP Rate is the FN Rate, but labelling headers with different rows is difficult with the above table format.

(2). KNN

* Start JMP Pro
* Open *german-bank.jmp*
* Select Analyze > Predictive Modeling > K Nearest Neighbors
* Select *checking* through *foreign* and click *X,Factor*
* Select *class* and click *Y,Response*
* Select *Validation* (on the left) and click *Validation* (on the right)
* Click OK.

You will see K Nearest Neighbors Report (or output) that includes, among others, Model Selection, Misclassification Rate for both Training and Validation, and confusion matrices for best *K* value.

* Capture the report screen and paste it in your submission.



* What is the best K value? **7**
* The dataset’s class attribute has two possible values – 1 and 2.
* Using the confusion matrix of *Validation* in the report (There are two confusion matrices.

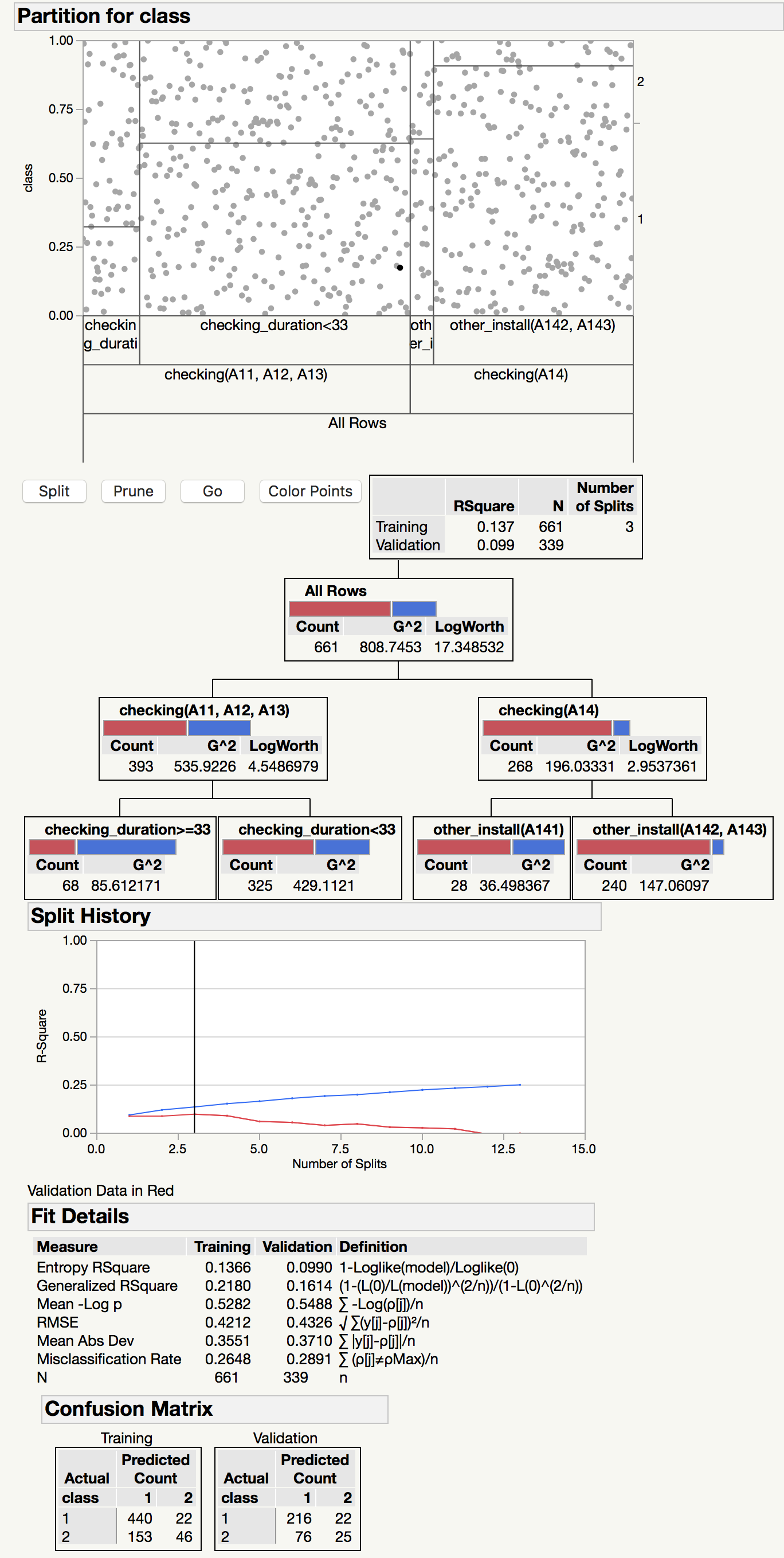
Make sure that you use the *Validation* confusion matrix), calculate the following measures for both classes (similar to those in Weka’s output window):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | page2image3749888Class |
| 0.929 | 0.071 | 0.754 | 0.929 | 0.832 | 1 |
| 0.287 | 0.713 |  |  |  | 2 |

**Note**: the TP Rate for class 2 is actually the TN Rate and the FP Rate is the FN Rate, but labelling headers with different rows is difficult with the above table format.

(3). Partition Model (decision tree)

* Start JMP Pro
* Open *german-bank.jmp*
* Select Analyze > Predictive Modeling > Partition
* Select *checking* through *foreign* and click *X,Factor*
* Select *class* and click *Y,Response*
* Select *Validation* (on the left) and click *Validation* (on the right)
* Click OK.
* In the output (this output is called platform report window), click *Go*. You will see a decision tree in the output.
* Click a red triangle next to *Partition for Class* and select *Show Fit Details*. Confusion matrices will be added to the output.
* Capture the output screen, which includes a decision tree and confusion matrices, and paste it in your submission.



* The dataset’s class attribute has two possible values – 1 and 2.
* Using the confusion matrix of *Validation* in the report (There are two confusion matrices.

Make sure that you use the *Validation* confusion matrix), calculate the following measures for both classes (similar to those in Weka’s output window):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | Class |
| 0.908 | 0.092 | 0.740 | 0.908 | 0.815 | 1 |
| 0.248 | 0.752 |  |  |  | 2 |

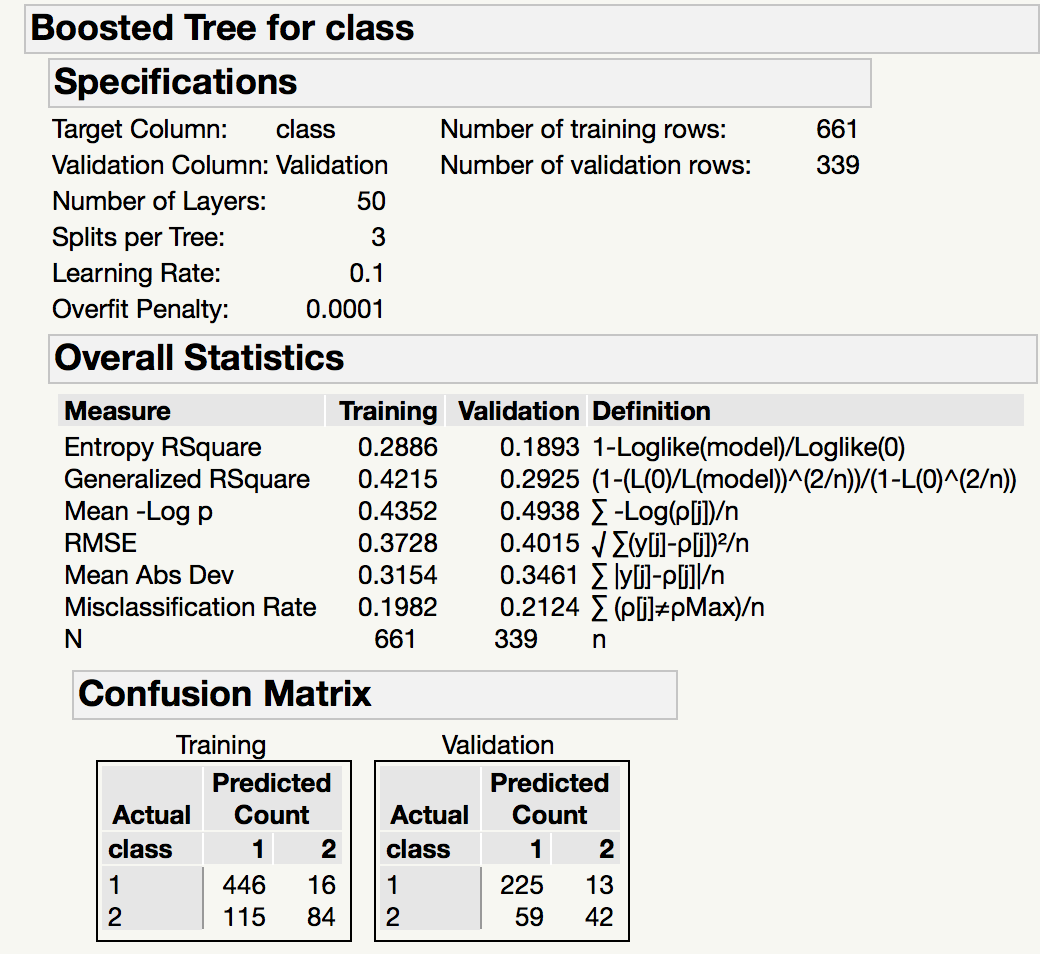
**Note**: the TP Rate for class 2 is actually the TN Rate and the FP Rate is the FN Rate, but labelling headers with different rows is difficult with the above table format.

(4). Boosted Tree

* Start JMP Pro
* Open *german-bank.jmp*
* Select Analyze > Predictive Modeling > Boosted Tree
* Select *checking* through *foreign* and click *X,Factor*
* Select *class* and click *Y,Response*
* Select *Validation* (on the left) and click *Validation* (on the right)
* Click OK.

Gradient-Boosted Tree Specification window appears.

* In the Reproducibility panel, select *Suppress Multithreading* and enter 123 for Random Seed.
* Click OK.
* Capture the output screen and paste it in your submission.



* The dataset’s class attribute has two possible values – 1 and 2.
* Using the confusion matrix of *Validation* in the report (There are two confusion matrices.

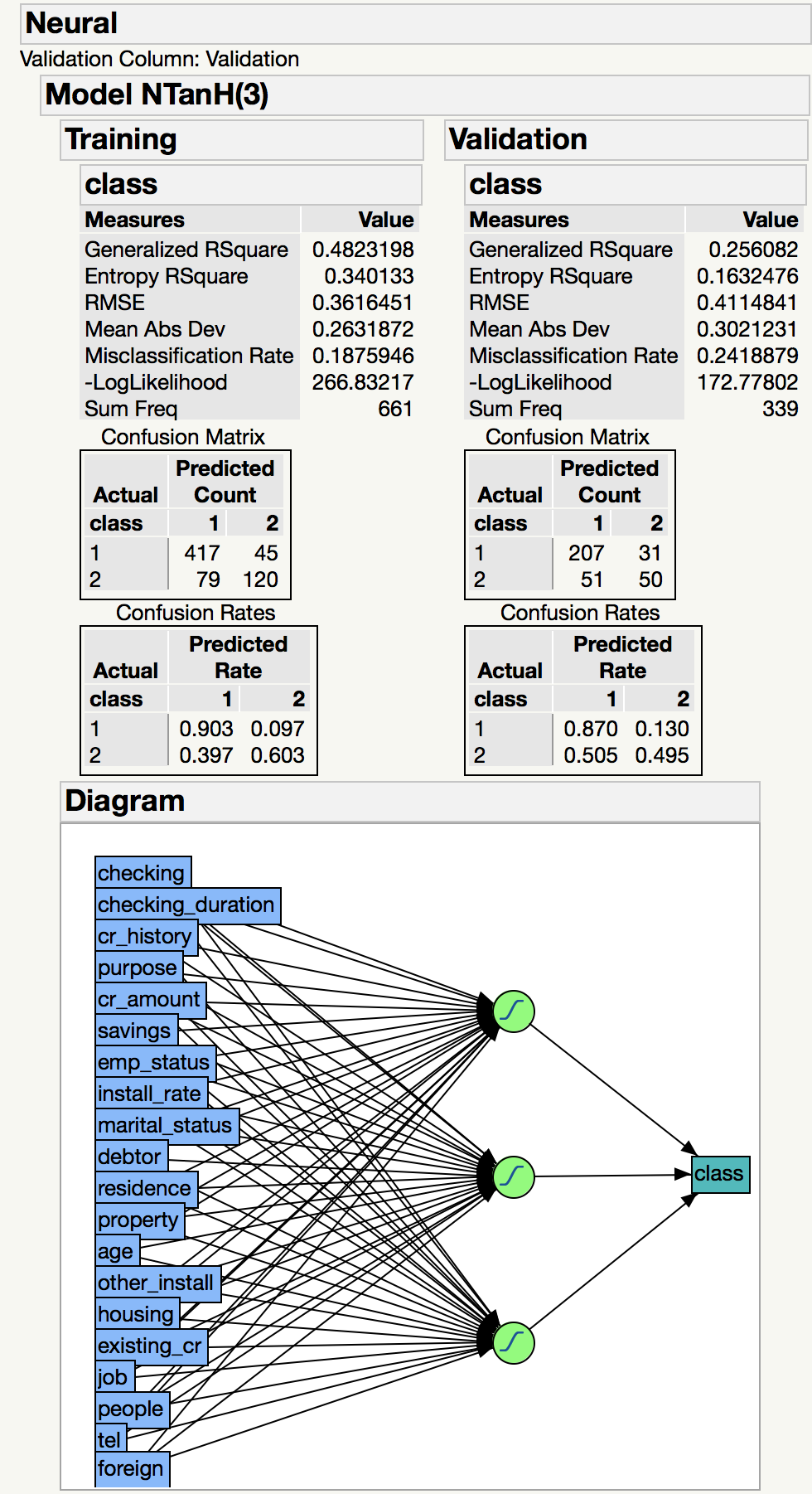
Make sure that you use the *Validation* confusion matrix), calculate the following measures for both classes (similar to those in Weka’s output window):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | page3image3763616Class |
| 0.945 | 0.055 | 0.792 | 0.945 | 0.862 | 1 |
| 0.416 | 0.584 |  |  |  | 2 |

**Note**: the TP Rate for class 2 is actually the TN Rate and the FP Rate is the FN Rate, but labelling headers with different rows is difficult with the above table format.

(5). Neural Network

* Start JMP Pro
* Open *german-bank.jmp*
* Select Analyze > Predictive Modeling > Neural
* Select *checking* through *foreign* and click *X,Factor*
* Select *class* and click *Y,Response*
* Select *Validation* (on the left) and click *Validation* (on the right)
* Click OK.
* In the next window (it is called Model Launch Control Panel), verify that 3 is in the box corresponding to *TanH* and *First*.
* Click GO.
* In the next screen, click red triangle next to *Model NTanH(3)* and select *Diagram*. You will see confusion matrices and a neural network diagram in the output window.
* Capture the output screen and paste it in your submission.



* How many hidden layers does the model have? **1**
* The dataset’s class attribute has two possible values – 1 and 2.
* Using the confusion matrix of *Validation* in the report (There are two confusion matrices.

Make sure that you use the *Validation* confusion matrix), calculate the following measures for both classes (similar to those in Weka’s output window):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| page4image3748016  TP Rate | FP Rate | Precision | Recall | F-Measure | Class |
| 0.870 | 0.130 | 0.802 | 0.870 | 0.835 | 1 |
| 0.495 | 0.505 |  |  |  | 2 |

**Note**: the TP Rate for class 2 is actually the TN Rate and the FP Rate is the FN Rate, but labelling headers with different rows is difficult with the above table format.