# **MAT 303 Module Six Problem Set Report**

Decision Trees

[Your Full Name]

[Your SNHU Email]

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## 1. Introduction

*Discuss the statement of the problem with regard to the statistical analyses that are being performed. Address the following questions in your analysis:*

* *What is the data set that you are exploring?*
* *How might your results be used?*
* *What types of analyses will you be running in this problem set?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## 2. Data Preparation

*There are some important variables that you have been asked to analyze in this problem set. Identify and explain these variables. Address the following questions in your analysis:*

* *What are the important variables in this data set?*
* *How many rows and columns are present in this data set?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## 3. Classification Decision Tree

### Reporting Results

* *Use set.seed(705526) and split the credit card default data set into training and validation sets using 70% and 30% split, respectively. How many rows are in the original data set, the training set, and the validation set?*
* *Use set.seed(705526) and create a classification decision tree for the default variable using missed payment, credit utilization, and assets as predictors. Include the cost-complexity (cp) table.*
* *Plot the validation error against the cost-complexity parameter (cp). What is an appropriate cp value to use in pruning the tree?*
* *Use set.seed(705526) and prune the tree using the appropriate cp value and include the plot of the resulting decision tree.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### Evaluating Utility of Model

*Evaluate the utility of the classification decision tree. Address the following questions in your analysis:*

* *Obtain the confusion matrix and report the counts for true positives, true negatives, false positives, and false negatives.*
* *Report the following:*
  + *Accuracy*
  + *Precision*
  + *Recall*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### Making Predictions Using Model

*Make predictions using the regression model. Address the following questions in your analysis:*

* *What is the prediction for defaulting on credit for an individual who has not missed payments, owns a car and a house, and has a 30% credit utilization?*
* *What is the prediction for defaulting on credit for an individual who has missed payments, does not have any assets, and has a 30% credit utilization?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## 4. Regression Decision Tree

### Reporting Results

* *Use set.seed(705526) and split the credit card default data set into training and validation sets using 80% and 20% split, respectively. How many rows are in the original data set, the training set, and the validation set?*
* *Use set.seed(705526) and create a regression decision tree for wage growth using economy, unemployment, and gdp as predictors. Include the cost-complexity (cp) table.*
* *Plot the validation error against the cost-complexity parameter (cp). What is an appropriate cp value to use in pruning the tree?*
* *Use set.seed(705526) and prune the tree using the appropriate cp value and include the plot of the resulting decision tree.*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### Evaluating Utility of Model

*Evaluate the utility of the classification decision tree. Address the following question in your analysis:*

* *What is the root mean squared error for the regression decision tree? Interpret this value.*

Caution sign icon Answer the question in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

### Making Predictions Using Model

*Make predictions using the regression model. Address the following questions in your analysis:*

* *What is the predicted wage growth if the economy is not in recession, unemployment is at 3.4%, and the GDP growth rate is 3.5%?*
* *What is the predicted wage growth if the economy is in recession, unemployment is at 7.4%, and the GDP growth rate is 1.5%?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## 5. Conclusion

*Describe the results of the statistical analyses and address the following questions:*

* *Fully describe what these results mean for your scenario using proper descriptions of statistical terms and concepts.*
* *What is the practical importance of the analyses that were performed?*

Caution sign icon Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.

## 6. Citations

*You are not required to use external resources for this report. If none were used, remove this entire section. However, if you used any resources to help you with your interpretation, you must cite them. Use proper APA format for citations.*

*Insert references here in the following format:*

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.