



COLORADO STATE UNIVERSITY  
— GLOBAL —

## MIS510: DATA MINING AND VISUALIZATION

### **Credit Hours:** 3

**Contact Hours:** This is a 3-credit course, offered in accelerated format. This means that 16 weeks of material is covered in 8 weeks. The exact number of hours per week that you can expect to spend on each course will vary based upon the weekly coursework, as well as your study style and preferences. You should plan to spend 14-20 hours per week in each course reading material, interacting on the discussion boards, writing papers, completing projects, and doing research.

**Faculty Information:** Faculty contact information and office hours can be found on the faculty profile page.

---

## **COURSE DESCRIPTION AND OUTCOMES**

---

### **Course Description:**

This course will provide the basic framework for conducting various data and text mining methodologies, including logistic regression analyses, classical discriminant analyses, association rule, decision tree, support vector machine, neural networks, variable reduction, cluster analysis, text analytics, and web mining. In addition, this course teaches the essential and practical skills in visualization, including computer graphics, visual data representation, physical and human vision models, numerical representation of knowledge and concepts, pattern analysis, and computational methods.

### **Course Overview:**

In this course you will explore data mining and data visualization and examine the foundation for data mining in the era of technology changes impacting data analytics. This course will cover various data mining methodologies, and different analysis techniques. Topics include association rule, decision tree, support vector machine (SVM), and neural networks. Tools used for data visualization will be utilized.

### **Course Learning Outcomes:**

1. Evaluate various data mining methodologies.
2. Assess different analysis techniques such as logistic regression analyses, classical discriminant analyses, association rule, decision tree, support vector machine, neural networks, variable reduction, cluster analysis, text analytics, and web mining.
3. Apply various data mining tools and techniques to work with large data sets including extraction, transformation, integration, loading, and accessing to gain insight from the data.
4. Apply various tools and techniques for data visualization.

- Investigate how computer graphics, visual data representation, physical and human vision models, numerical representation of knowledge and concepts, pattern analysis, and computational methods work.

---

## PARTICIPATION & ATTENDANCE

---

Prompt and consistent attendance in your online courses is essential for your success at CSU-Global Campus. Failure to verify your attendance within the first 7 days of this course may result in your withdrawal. If for some reason you would like to drop a course, please contact your advisor.

Online classes have deadlines, assignments, and participation requirements just like on-campus classes. Budget your time carefully and keep an open line of communication with your instructor. If you are having technical problems, problems with your assignments, or other problems that are impeding your progress, let your instructor know as soon as possible.

---

## COURSE MATERIALS

---

### Required:

Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R., and Lichtendahl, K.C. (2018). *Data mining for business analytics: Concepts, techniques, and application sin R*. Wiley Publishing. ISBN: 9781118879337.

### Required Technology:

Please note the section at the bottom of this syllabus on “SAS Resources and Information on Certification.”

### Suggested:

Wickham, H., & Groelmund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. O’Reilly Media, Inc.

**NOTE:** All non-textbook required readings and materials necessary to complete assignments, discussions, and/or supplemental or required exercises are provided within the course itself. Please read through each course module carefully.

---

## COURSE SCHEDULE

---

### Due Dates

The Academic Week at CSU-Global begins on Monday and ends the following Sunday.

- Discussion Boards:** The original post must be completed by Thursday at 11:59 p.m. MT and peer responses posted by Sunday at 11:59 p.m. MT. Late posts may not be awarded points.
- Critical Thinking:** Assignments are due Sunday at 11:59 p.m. MT.

---

## WEEKLY READING AND ASSIGNMENT DETAILS

---

### Module 1

#### Readings

- Chapters 1 & 2 in *Data Mining for Business Analytics*
- Borysov, S., Geilhufe, R., & Balatsky, A. (2017). *Organic materials database: An open-access online database for data mining. PloS One, 12(2).*
- Prati, G., Pietrantonio, L., & Fraboni, F. (2017). *Using data mining techniques to predict the severity of bicycle crashes. Accident Analysis and Prevention, 101, 44-54*

#### Discussion (25 points)

#### Critical Thinking (65 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### Option #1: Categorical Attributes

In this Critical Thinking Assignment, you will install R and R Studio as well as explore a dataset and convert two categorical variables into binary dummies. Install R and R Studio by following the instructions here:

- R Studio Installation Guide
- R Installation Guide

Explore the Toyota Corolla used car data in the *ToyotaCorolla.csv* file by performing the following steps.

1. Follow the process under *Loading and Looking at the Data in R* in section 2.4 of our text.
2. Use the R code example shown in Table 2.3.
3. For your assignment submission, take screen shots showing the successful execution of the following functions. Your screen shots must include the system date and time.
  - a. `dim()`
  - b. `head()`
  - c. `View()`
  - d. Three other exploratory functions similar to those demonstrated in Table 2.3.
4. Convert the attributes *Fuel Type* and *Color* to binary dummies by following the process under *Handling Categorical Variables* in section 2.4 of our text.
5. Use the R code example shown in Table 2.6.
6. For your assignment submission, take a screen shot of your entire R Studio window showing your code that is similar to Table 2.6. Your screen shots must include the system date and time.

Your assignment submission must be one Word document that meets the following requirements:

- Conforms to CSU-Global Guide to Writing and APA
- Includes a cover page with the number and name of the Critical Thinking option you chose
- Does not include an abstract
- Includes a one-page description of what you did and what you learned
- Includes screen shots of your R code showing the entire R Studio window as well as the system clock date and time
- If the R code in your screen shots is too small to read, copy the code into the paper as a figure

- All figures, including screen shots and code, must meet APA labeling requirements

### **Option #2: Partitioning Data**

In this Critical Thinking Assignment, you will install R and R Studio as well as explore a dataset and partition data into training, validation, and test sets. Install R and R Studio by following the instructions here:

- R Studio Installation Guide
- R Installation Guide

Explore the Toyota Corolla used car data in the *ToyotaCorolla.csv* file by performing the following steps.

1. Follow the process under *Loading and Looking at the Data in R* in section 2.4 of our text.
2. Use the R code example shown in Table 2.3.
3. For your assignment submission, take screen shots showing the successful execution of the following functions. Your screen shots must include the system date and time.
  - a. `dim()`
  - b. `head()`
  - c. `View()`
4. Partition the data into training (50%), validation (30%), and test (20%) sets by following the process under *Creation and Use of Data Partitions* in section 2.5 of our text.
5. Use the R code example shown in Table 2.9.
6. For your assignment submission, take screen shots of your entire R Studio window showing the results of the `dim()` function for each partition set. Your screen shots must include the system date and time.

Your assignment submission must be one Word document that meets the following requirements:

- Conforms to CSU-Global Guide to Writing and APA
- Includes a cover page with the number and name of the Critical Thinking option you chose
- Does not include an abstract
- Includes a one-page description of what you did and what you learned
- Includes screen shots of your R code showing the entire R Studio window as well as the system clock date and time
- If the R code in your screen shots is too small to read, copy the code into the paper as a figure
- All figures, including screen shots and code, must meet APA labeling requirements

### **PORTFOLIO PROJECT REMINDER**

Be sure to review the Portfolio Project Description and Portfolio Project Rubric in the Module 8 folder for details.

## **Module 2**

### **Readings**

- Chapters 3 & 4 in *Data Mining for Business Analytics*
- Asadi, A., Verma, A., Yang, K., & Mejabi, B. (2017). Wastewater treatment aeration process optimization: A data mining approach. *Journal of Environmental Management*, 203(Pt 2), 630-639.
- Rubio-Escudero, C., Valverde-Fernandez, J., Nepomuceno-Chamorro, I., Pontes-Balanza, B., Hernandez-Mendoza, Y., & Rodriguez-Herrera, A. (2017). Data Mining Techniques Applied to Hydrogen Lactose Breath Test. *PLoS ONE*, 12(1).

### **Discussion (25 points)**

### Critical Thinking (65 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### Option #1: Hot vs Cold Cereals

In this Critical Thinking Assignment, you will explore and summarize a dataset as well as create several visualizations.

Explore breakfast cereals in the *Cereals.csv* file by performing the following steps.

1. Follow the process under *Summary Statistics* in section 4.4 of our text.
2. Use the R code example shown in Table 4.3 and Table 4.4.
3. For your assignment submission, take screen shots of your entire R Studio window showing the successful execution of the `data.frame()` function to compute the following values. Your screen shots must include the system date and time.
  - a. mean
  - b. standard deviation
  - c. min
  - d. max
  - e. median
  - f. length
  - g. sum of missing values for each of the quantitative attributes

Hint: You can select a set of data frame columns in the `sapply()` function. First, determine the quantitative attribute columns. These are the columns you want to keep. You can either positively include these or negatively exclude the other columns. For example, the following code excludes columns 6, 7, and 8 from the mean calculation.

```
sapply(cereals.df[-c[6:8], mean, na.rm=TRUE)
```

The `na.rm=TRUE` eliminates missing values, which cause errors in mathematical calculations.

4. Plot a histogram for each of the quantitative variables by following the process under *Distribution Plots: Boxplots and Histograms* in section 3.3.
5. Use the R code example shown in Figure 3.2.
6. For your assignment submission, take a screen shot of your entire R Studio window showing the following items. Your screen shots must include the system date and time.
  - a. Your code for creating all the histograms
  - b. At least one histogram in the Plot window
7. For your assignment submission, write a description of three things you can learn about this data from the histograms.
8. Plot a side-by-side boxplot comparing calories in hot versus cold cereals by following the process under *Distribution Plots: Boxplots and Histograms* in section 3.3.
9. Use the R code example shown in Figure 3.2 and Figure 3.3.
10. Add labels to your boxplot.
11. For your assignment submission, take screen shots of your entire R Studio window showing your code and the boxplot. Your screen shots must include the system date and time.
12. For your assignment submission, write a description of three things you can learn about this data from the boxplot.

Your assignment submission must be one Word document that meets the following requirements:

- Conforms to CSU-Global Guide to Writing and APA
- Includes a cover page with the number and name of the Critical Thinking option you chose
- Does not include an abstract

- Includes a one-page description of what you did and what you learned
- Includes screen shots of your R code showing the entire R Studio window as well as the system clock date and time
- If the R code in your screen shots is too small to read, copy the code into the paper as a figure
- All figures, including screen shots and code, must meet APA labeling requirements

### Option #2: Consumer Rating as Function of Shelf Height

In this Critical Thinking Assignment, you will explore and summarize a dataset as well as create several visualizations.

Explore breakfast cereals in the *Cereals.csv* file by performing the following steps.

1. Follow the process under *Summary Statistics* in section 4.4 of our text.
2. Use the R code example shown in Table 4.3 and Table 4.4.
3. For your assignment submission, take screen shots of your entire R Studio window showing the successful execution of the `data.frame()` function to compute the following values. Your screen shots must include the system date and time.
  - a. mean
  - b. standard deviation
  - c. min
  - d. max
  - e. median
  - f. length
  - g. sum of missing values for each of the quantitative attributes

Hint: You can select a set of data frame columns in the `sapply()` function. First, determine the quantitative attribute columns. These are the columns you want to keep. You can either positively include these or negatively exclude the other columns. For example, the following code excludes columns 6, 7, and 8 from the mean calculation.

```
sapply(cereals.df, -c[6:8], mean, na.rm=TRUE)
```

The `na.rm=TRUE` eliminates missing values, which cause errors in mathematical calculations.

4. Plot a histogram for each of the quantitative variables by following the process under *Distribution Plots: Boxplots and Histograms* in section 3.3.
5. Use the R code example shown in Figure 3.2.
6. For your assignment submission, take a screen shot of your entire R Studio window showing the following items. Your screen shots must include the system date and time.
  - a. Your code for creating all the histograms
  - b. At least one histogram in the Plot window
7. For your assignment submission, write a description of three things you can learn about this data from the histograms.
8. Plot a side-by-side boxplot of consumer rating as a function of the shelf height by following the process under *Distribution Plots: Boxplots and Histograms* in section 3.3.
9. Use the R code example shown in Figure 3.2 and Figure 3.3.
10. Add labels to your boxplot.
11. For your assignment submission, take screen shots of your entire R Studio window showing your code and the boxplot. Your screen shots must include the system date and time.
12. For your assignment submission, write a description of three things you can learn about this data from the boxplot.

Your assignment submission must be one Word document that meets the following requirements:

- Conforms to CSU-Global Guide to Writing and APA

- Includes a cover page with the number and name of the Critical Thinking option you chose
- Does not include an abstract
- Includes a one-page description of what you did and what you learned
- Includes screen shots of your R code showing the entire R Studio window as well as the system clock date and time
- If the R code in your screen shots is too small to read, copy the code into the paper as a figure
- All figures, including screen shots and code, must meet APA labeling requirements

## **Module 3**

### **Readings**

- Chapters 6, 9 (sections 9.1, 9.7-9.9), & 11 in *Data Mining for Business Analytics*
- Introduction through Output Formats in the *R Markdown Tutorial*. Retrieved from <https://rmarkdown.rstudio.com/lesson-1.html>
- Gruginskie, L., & Vaccaro, G. (2018). Lawsuit lead time prediction: Comparison of data mining techniques based on categorical response variable. *PloS One*, 13(6).
- Pourghasemi, H., Yousefi, S., Kornejady, A., & Cerdà, A. (2017). Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Science of the Total Environment*, 609, 764-775.

### **Discussion (25 points)**

### **Critical Thinking (65 points)**

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### **Option #1: Linear Regression Model**

In this Critical Thinking Assignment, you will install R Markdown, explore and summarize a dataset as well as create a linear regression model. Your assignment submission will be an R Markdown generated Word document.

Install R Markdown.

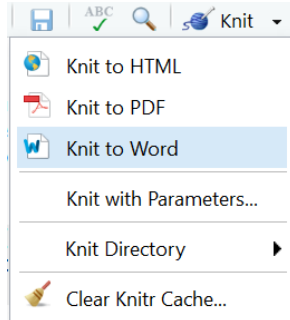
Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 3 CT Option 1* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

Explore Boston housing in the *BostonHousing.csv* file by performing the following steps.

1. Apply what you learned in Modules 1 and 2 about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.

3. Fit a multiple linear regression model to the median house price (MEDV) as a function of CRIM, CHAS, and RM by following the process under *Example: Predicting the Price of Used Toyota Corolla Cars* in section 6.3.
4. Use the R code example shown in Figure 6.3.  
Hint: You will need to remove the categorical variable CAT...MEDV prior to fitting the multiple linear regression model.
5. Create a scatter plot with the plot() function with the following attributes.
  - a. Use MEDV as the y-axis
  - b. Use the most significant attribute as the x-axis.
  - c. Use the abline() function to add a linear regression line to the scatter plot. Use the y-intercept as the y value and the factor value of the most significant attribute as the slope value.
6. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
7. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

### Option #2: Regression Tree

In this Critical Thinking Assignment, you will install R Markdown, explore and summarize a dataset as well as create a linear regression model. Your assignment submission will be an R Markdown generated Word document.

Install R Markdown.

Create a new R Markdown file by performing the following steps.

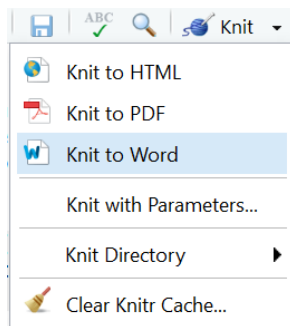
1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 3 CT Option 2* as the Title
4. Use your name as the Author



5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file

Explore Toyota Corolla used cars in the *ToyotaCorolla.csv* file by performing the following steps.

1. Apply what you learned in modules 1 and 2 about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
3. Run a regression tree with the following characteristics.
  - a. Outcome variable Price
  - b. Predictors Age\_08\_04, KM, Fuel\_Type, HP, Automatic, Doors, Quarterly\_Tax, Mfr\_Guarantee, Guarantee\_Period, Airco, Automatic\_Airco, CD\_Player, Powered\_Windows, Sport\_Model, and Tow\_Bar.
  - c. Keep the minimum number of records in a terminal node to 1
  - d. Keep the maximum number of tree levels to 30
  - e. Keep  $cp = 0.001$ , to make the run least restrictive.
4. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
5. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

## **Module 4**

### **Readings**

- Chapters 8, 9 (9.2 – 9.6), 10, & 12 in *Data Mining for Business Analytics*

- Chen, W., Zhang, S., Li, R., & Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Science of the Total Environment*, 644, 1006-1018.
- Luo, Y., Li, Z., Guo, H., Cao, H., Song, C., Guo, X., & Zhang, Y. (2017). Predicting congenital heart defects: A comparison of three data mining methods. *PloS One*, 12(5).

### Discussion (25 points)

#### Critical Thinking (65 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### Option #1: Logistic Regression on Banks

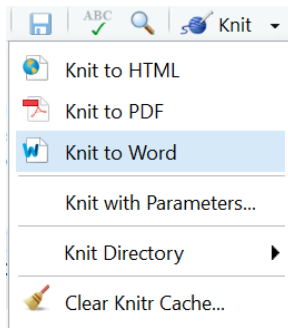
In this Critical Thinking Assignment, you will summarize a dataset as well as create a logistic regression model. Your assignment submission will be an R Markdown generated Word document.

Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 4 CT Option 1* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

Explore 20 banks in the *Banks.csv* file by performing the following steps.

1. Apply what you learned in Modules 1 and 2 about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
3. Run a logistic regression model by following the process described in section 10.3.
4. Use the R code example in table 10.2.
5. Do not partition the data. Instead, use the entire dataset for your model.
6. Model the *Financial Condition* attribute as a function of the other attributes. The financial condition of a bank is either strong (1) or weak (0).
7. Write the estimated logistic equation in the form of the example in equation 10.9.
8. Explain the estimated logistic equation in your one-page description.
9. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
10. Create a confusion matrix and lift chart as described in section 10.4.
11. Use the R code example in figure 10.6.
12. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
13. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

### **Option #2: Logistic Regression on Riding Mower Ownership**

In this Critical Thinking Assignment, you will summarize a dataset as well as create a linear regression model. Your assignment submission will be an R Markdown generated Word document.

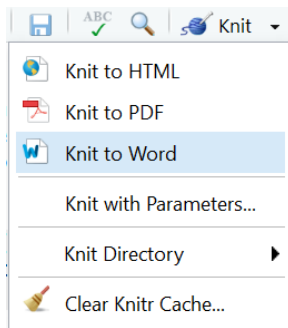
Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 4 CT Option 2* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

Explore riding mower ownership in the *RidingMowers.csv* file by performing the following steps.

1. Apply what you learned in modules 1 and 2 about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
3. Run a logistic regression model by following the process described in section 10.3.
4. Use the R code example in table 10.2.
5. Do not partition the data. Instead, use the entire dataset for your model.
6. Model the *Ownership* attribute as a function of the other attributes.
7. *Income* is in \$1000s and *Lot\_Size* is in 1000 ft<sup>2</sup>.
8. Write the estimated logistic equation in the form of the example in equation 10.9.
9. Explain the estimated logistic equation in your one-page description.
10. For your assignment submission, copy your commands into your R Markdown file.

- a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
11. Create a confusion matrix and lift chart for as described in section 10.4.
  12. Use the R code example in figure 10.6.
  13. For your assignment submission, copy your commands into your R Markdown file.
    - a. Include R comments on all your code.
    - b. Separate sections of R code by using appropriate R Markdown headings.
  14. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

### Portfolio Milestone (50 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### Option #1: Credit Data Mining

Submit a one-page written proposal on why you have chosen Portfolio Project Option #1. Include the following in your paper:

1. A description in your own words of what you will be doing in this option.
2. A listing of chapters and sections from our text that will help you complete this option.
3. Questions you have about this option that require clarification from the instructor.

Submit your assignment with a title page formatted according to the *CSU-Global Guide to Writing and APA*.

#### Option #2: Cross-Selling Data Mining

Submit a one-page written proposal on why you have chosen Portfolio Project Option #2. Include the following in your paper:

1. A description in your own words of what you will be doing in this option.

2. A listing of chapters and sections from our text that will help you complete this option.
3. Questions you have about this option that require clarification from the instructor.

Submit your assignment with a title page formatted according to the *CSU-Global Guide to Writing and APA*.

## **Module 5**

### **Readings**

- Chapter 5 in *Data Mining for Business Analytics*
- Sayago, A., González-Domínguez, R., Beltrán, R., & Fernández-Recamales, A. (2018). Combination of complementary data mining methods for geographical characterization of extra virgin olive oils based on mineral composition. *Food Chemistry*, 261, 42-50.
- Yang, H., Huang, M., Lai, C., & Jin, J. (2018). An approach combining data mining and control charts-based model for fault detection in wind turbines. *Renewable Energy*, 115, 808-816.

### **Discussion (25 points)**

#### **Critical Thinking (65 points)**

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### **Option #1: ROC vs. Lift Chart Paper**

Write a paper comparing and contrasting Receiver Operating Characteristic (ROC) curves and Lift Charts. Use example visualizations to illustrate and enhance your argument. Be sure to treat example visualizations as APA figures and cite their source.

Your paper should be 2-3 pages in length and conform to CSU-Global Guide to Writing and APA. Include at least two scholarly references in addition to the course textbook. The CSU-Global Library is a good place to find these references.

#### **Option #2: ROC vs. Lift Chart Presentation**

Create a PowerPoint presentation with extensive speaker notes comparing and contrasting Receiver Operating Characteristic (ROC) curves and Lift Charts. Use example visualizations to illustrate and enhance your argument. Be sure to treat example visualizations as APA figures and cite their source. Your presentation should be 6-8 slides and conform to CSU-Global Guide to Writing and APA. Include at least two scholarly references in addition to the course textbook. The CSU-Global Library is a good place to find these references.

## **Module 6**

### **Readings**

- Chapters 14 & 15 in *Data Mining for Business Analytics*
- Costa, E., Fonseca, B., Santana, M., De Araújo, F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247-256.
- Naghibi, S., Moghaddam, D., Kalantar, B., Pradhan, B., & Kisi, O. (2017). A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *Journal of Hydrology*, 548, 471-483.

- Rose, D. (2018). *Artificial intelligence foundations: Machine learning*. [video]. Available at <https://www.lynda.com/Data-Science-tutorials/Artificial-Intelligence-Foundations-Machine-Learning/601797-2.html>

### Discussion (25 points)

### Critical Thinking (65 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

#### Option #1: Course Selection Association Rules

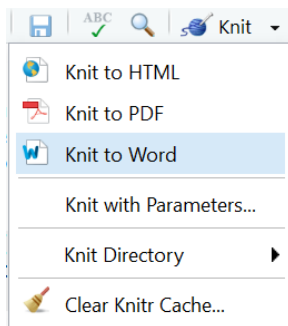
In this Critical Thinking Assignment, you will analyze association rules. Your assignment submission will be an R Markdown generated Word document.

Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 6 CT Option 1* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

The *CourseTopics.csv* file contains eight attributes which are courses that a student could take. Each data row provides the courses taken (1 in the attribute) by a single student. Analyze this data by performing the following steps.

1. Create and interpret 3-5 association rules.
2. Follow the process described in section 14.1.
3. Use the R code example in Table 14.4.
4. Use 0.01 for minimum support.
5. Run minimum confidence of 0.1 and 0.5 in the `apriori()` function.
6. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
7. Use the R Markdown *Knit* drop down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

### Option #2: Cosmetics Association Rules

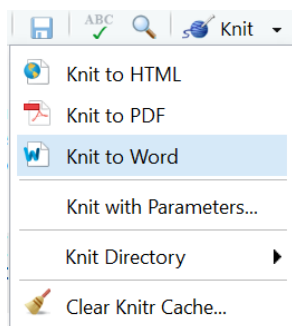
In this Critical Thinking Assignment, you will analyze association rules. Your assignment submission will be an R Markdown generated Word document.

Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *Module 6 CT Option 2* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

The *Cosmetics.csv* file contains 14 attributes which are cosmetic products. Each data row provides the products purchased (1 in the attribute) by a single customer. Analyze this data by performing the following steps.

1. Create and interpret 3-5 association rules.
2. Follow the process described in section 14.1.
3. Use the R code example in Table 14.4.
4. Use 0.01 for minimum support.
5. Run minimum confidence of 0.1 and 0.5 in the `apriori()` function.
6. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
7. Use the R Markdown *Knit* drop-down to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

## **Module 7**

### **Readings**

- Chapters 16, 17, & 18 in *Data Mining for Business Analytics*
- McCormick, K. (2018). Machine learning and AI foundations: Linear regression [video]. Available at <https://www.lynda.com/SPSS-tutorials/Time-series-forecasting/645049/745936-4.html>
- Ryoo, J. (2018). Data science & analytics career paths & certifications: First steps [video]. Available at <https://www.lynda.com/Data-Science-tutorials/Data-Science-Analytics-Career-Paths-Certifications-First-Steps/642486-2.html>

### **Discussion (25 points)**

### **Critical Thinking (60 points)**

### **Career Focus Paper**

Write a paper. Find three data mining job listings. Analyze the job listings and report on the requirements and qualifications of a data mining position. What types of preparation, such as education, training, certification, and skills, would you need to plan for success in these types of jobs? Write a paper to share your findings.

Your paper should be 2-3 pages in length and conform to CSU-Global Guide to Writing and APA. Include at least two scholarly references in addition to the course textbook. The CSU-Global Library is a good place to find these references.

## **Module 8**

### **Readings**

- Chapters 19 & 20 in *Data Mining for Business Analytics*
- Kimmons, R., McGuire, K., Stauffer, M., Jones, J., Gregson, M., & Austin, M. (2017). Religious identity, expression, and civility in social media: Results of data mining latter-day saint twitter accounts. *Journal for the Scientific Study of Religion*, 56(3), 637-657.
- Monkman, G., Kaiser, M., & Hyder, K. (2018). Text and data mining of social media to map wildlife recreation activity. *Biological Conservation*, 228, 89-99.
- Poulson, B. (2016). *Data science foundations: Data mining* [video]. Available at <https://www.lynda.com/Business-Intelligence-tutorials/Data-Science-Foundations-Data-Mining/475936-2.html>

### **Discussion (25 points)**



## Portfolio Project (300 points)

Choose one of the following two assignments to complete this week. Do not do both assignments. Identify your assignment choice in the title of your submission.

### Option #1: Credit Data Mining

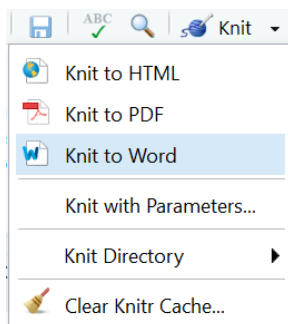
In this Portfolio Project, you will analyze credit worthiness. Your assignment submission will be an R Markdown generated Word document.

Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *MIS 510 Portfolio Project Option 1* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

Refer to section 21.2 for background on the *GermanCredit.csv* file. Explore *GermanCredit.csv* by performing the following steps.

1. Apply what you learned in this course about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
3. Divide the data into training and validation partitions.
4. Choose two of the following data mining techniques to explore classification models of this data.
  - a. Logistic regression
  - b. Classification trees
  - c. Neural networks.
5. Analyze your results.
6. Include appropriate visualizations in your analysis.
7. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
8. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

### Option #2: Cross-Selling Data Mining

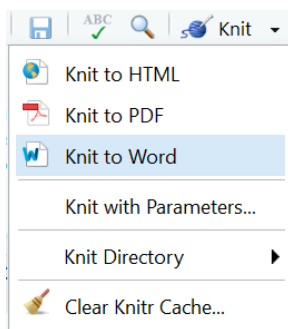
In this Portfolio Project, you will analyze catalog selling. Your assignment submission will be an R Markdown generated Word document.

Create a new R Markdown file by performing the following steps.

1. Open R Studio
2. Select File | New | R Markdown
3. Use *MIS 510 Portfolio Project Option 2* as the Title
4. Use your name as the Author
5. Select the Word output format
6. Delete all default content after the R Setup block of code, which is all content from line 12 through the end of the file.

Refer to section 21.8 for background on the *CatalogCrossSell.csv* file. Explore *CatalogCrossSell.csv* by performing the following steps.

1. Apply what you learned in this course about data exploration by selecting and running appropriate data exploration functions. Run at least five functions.
2. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
3. Analyze your data by creating and interpreting association rules.
4. Include appropriate visualizations in your analysis.
5. For your assignment submission, copy your commands into your R Markdown file.
  - a. Include R comments on all your code.
  - b. Separate sections of R code by using appropriate R Markdown headings.
6. Use the R Markdown *Knit* drop-down menu to select *Knit to Word* to create the Word document for your assignment submission.



Your assignment submission must be one Word document that meets the following requirements:

- Is an R Markdown generated Word document containing all R code used in this assignment, appropriate R comments on code, and appropriate R Markdown headings.
- Does not include a cover page.
- Does not include an abstract.
- Includes a one-page description of what you did and what you learned. Add this description to the end of the R Markdown document as a new page. This page must conform to CSU-Global Guide to Writing and APA.

---

## COURSE POLICIES

---

### Course Grading

20% Discussion Participation  
45% Critical Thinking Assignments  
35% Final Portfolio Project

Grading Scale	
A	95.0 – 100
A-	90.0 – 94.9
B+	86.7 – 89.9
B	83.3 – 86.6
B-	80.0 – 83.2
C+	75.0 – 79.9
C	70.0 – 74.9
D	60.0 – 69.9
F	59.9 or below

---

## IN-CLASSROOM POLICIES

---

For information on late work and incomplete grade policies, please refer to our [In-Classroom Student Policies and Guidelines](#) or the Academic Catalog for comprehensive documentation of CSU-Global institutional policies.

### **Academic Integrity**

Students must assume responsibility for maintaining honesty in all work submitted for credit and in any other work designated by the instructor of the course. Academic dishonesty includes cheating, fabrication, facilitating academic dishonesty, plagiarism, reusing/repurposing your own work (see CSU-Global Guide to Writing & APA for percentage of repurposed work that can be used in an assignment), unauthorized possession of academic materials, and unauthorized collaboration. The CSU-Global Library provides information on how students can avoid plagiarism by understanding what it is and how to use the library and Internet resources.

### **Citing Sources with APA Style**

All students are expected to follow the CSU-Global Guide to Writing & APA when citing in APA (based on the most recent APA style manual) for all assignments. A link to this guide should also be provided within most assignment descriptions in your course.

### **Disability Services Statement**

CSU-Global is committed to providing reasonable accommodations for all persons with disabilities. Any student with a documented disability requesting academic accommodations should contact the Disability Resource Coordinator at 720-279-0650 and/or email [ada@CSUGlobal.edu](mailto:ada@CSUGlobal.edu) for additional information to coordinate reasonable accommodations for students with documented disabilities.

### **Netiquette**

Respect the diversity of opinions among the instructor and classmates and engage with them in a courteous, respectful, and professional manner. All posts and classroom communication must be conducted in accordance with the student code of conduct. Think before you push the Send button. Did you say just what you meant? How will the person on the other end read the words?

Maintain an environment free of harassment, stalking, threats, abuse, insults, or humiliation toward the instructor and classmates. This includes, but is not limited to, demeaning written or oral comments of an ethnic, religious, age, disability, sexist (or sexual orientation), or racist nature; and the unwanted sexual advances or intimidations by email or on discussion boards and other postings within or connected to the online classroom.

If you have concerns about something that has been said, please let your instructor know.

### **Information about SAS Certification**

Taking SAS certification exams help you validate your skills and increase your value to an employer. You can choose SAS certifications across many subjects, including programming, data management, and analytics, to name a few. For more information on SAS certification go here: [https://www.sas.com/en\\_us/certification.html](https://www.sas.com/en_us/certification.html). All students, teachers, professors or staff associated with an academic institution qualify for 50% discount on all SAS certification exams. Please contact [certification@sas.com](mailto:certification@sas.com) to receive the discount code that will reduce the exam fee by 50% during the registration process.

### **Resources for Learning SAS**

SAS Certification Prep Guides: [https://www.sas.com/store/books/categories/certification-guide/cBooks-cbooks\\_categories-cbooks\\_categories\\_12-p1.html](https://www.sas.com/store/books/categories/certification-guide/cBooks-cbooks_categories-cbooks_categories_12-p1.html)

Visit SAS Communities Visit our online sites to share and connect with other SAS users and build your SAS skills. Don't miss key communities including: SAS Certification, SAS Training, SAS Academy for Data Science, SAS Programming, New SAS User, SAS Analytics U and SAS Viya for Learners. <https://communities.sas.com/t5/Learn-SAS/ct-p/learn>