

Preview: MBA6135 : Data Mining and Reporting

Syllabus

Course Overview

Business data mining involves the exploration and analysis of large quantities of data typically found in an organization's internal databases, using automated means to discover actionable patterns and rules. This course covers the methods and techniques for implementing this process to support tactical and strategic business decisions. It addresses skills needed to use industry-standard analytical tools for data access and management, data preparation, analysis, and reporting—the components of business data mining.

Business Intelligence Certification

This course also prepares you for elements of the Certified Business Intelligence Professional (CBIP) certification program. While the certification itself is optional, and not an endorsement by Capella, your pursuing the CBIP credential in the future will enhance your marketability as it tells the world—including current and prospective employers—that you are an ardent business intelligence professional. The link below provides details of the CBIP program and certification exams:

- [Certified Business Intelligence Professional \(CBIP\)](#).

VitalSource Bookshelf

This course offers e-books through the VitalSource Bookshelf. A link to your Bookshelf is provided in the left **Course Tools** menu. More information about your e-books can be found in the Unit 1 studies.

Course Competencies

(Read Only)

To successfully complete this course, you will be expected to:

- 1 Identify internal data sources appropriate to address global management decision support needs.

- 2 Use standardized data mining application software to retrieve decision support data.
- 3 Present results of analyses in a concise and compelling manner.
- 4 Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Course Prerequisites

Prerequisite(s): MBA6018, MBA6131, or MBA6140.

Syllabus >> Course Materials

Required

The materials listed below are required to complete the learning activities in this course.

Integrated Materials

Many of your required books are available via the VitalSource Bookshelf link in the courseroom, located in your Course Tools. Registered learners in a Resource Kit program can access these materials using the courseroom link on the Friday before the course start date. Some materials are available only in hard-copy format or by using an access code. For these materials, you will receive an email with further instructions for access. Visit the [Course Materials](#) page on Campus for more information.

Book

Please follow the instructions provided to you by the bookstore to download your e-books. Once procured, you may access your e-books via the VitalSource Bookshelf. A link to your Bookshelf is provided in the left Course Tools menu. The [Capella University Bookstore](#) shows any materials that may involve shipping.

Shmueli, G., Bruce, P. C., & Patel, N. R. (2016). *Data mining for business analytics: Concepts, techniques, and applications with XLMiner* (3rd ed.). Hoboken, NJ: Wiley. ISBN:

9781118729274.

Library

The following required readings are provided in the Capella University Library or linked directly in this course. To find specific readings by journal or book title, use [Journal and Book Locator](#). Refer to the [Journal and Book Locator library guide](#) to learn how to use this tool.

- Balijepally, V., Mangalaraj, G., & Iyengar, K. (2011). [Are we wielding this hammer correctly? A reflective review of the application of cluster analysis in information systems research](#). *Journal of the Association for Information Systems*, 12(5), 375–413.
- Nishi, D. (2005). [Market-basket mystery](#). *Chain Store Age*, May 2005 supplement, 12A–14A.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Cary, C., Wen, H. J., & Mahatanankoon, P. (2003). [Data mining: Consumer privacy, ethical policy, and systems development practices](#). *Human Systems Management*, 22(4), 157–168.
- Chakraborty, G., Pagolu, M., & Garla, S. (2013). [Text mining and analysis: Practical methods, examples, and case studies using SAS \(excerpt\).\[PDF\]](#). Cary, NC: SAS Institute.
- Kaggle: Your home for data science. (2017). Retrieved from <https://www.kaggle.com/>
- SAS Software. (2015). [Getting started with SAS Enterprise Miner: Transforming data and building parametric models \[Video\]](#). | [Transcript](#) Available from <https://www.youtube.com/watch?v=TnWRJQb5z4c>

Suggested

The following materials are recommended to provide you with a better understanding of the topics in this course. These materials are not required to complete the course, but they are aligned to course activities and assessments and are highly recommended for your use.

Optional

The following optional materials are offered to provide you with a better understanding of the topics in this course. These materials are not required to complete the course.

Integrated Materials

Software

The following software is not required to complete the learning activities in this course, but it may be helpful to you:

IBM SPSS Statistics Standard Grad Pack.

Microsoft Excel Analysis ToolPak add-on.

SAS Enterprise Miner.

XLMiner.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- SAS Institute Inc. (2014). [SAS OnDemand for academics: Create an account](https://odamid.oda.sas.com/SASODARegistration/). Retrieved from <https://odamid.oda.sas.com/SASODARegistration/>
- SAS Institute Inc. (n.d.). [SAS customer support](http://support.sas.com). Retrieved from <http://support.sas.com>
- SAS Software. (2011). [Manage all unstructured data with SAS text analytics \[Video\]](https://www.youtube.com/watch?v=NHAq8jG4FX4). | [Transcript](https://www.youtube.com/watch?v=NHAq8jG4FX4) Available from <https://www.youtube.com/watch?v=NHAq8jG4FX4>
- SAS Software. (2012). [Active learning with SAS Text Miner \[Video\]](https://www.youtube.com/watch?v=afJpLYjJhMA). | [Transcript](https://www.youtube.com/watch?v=afJpLYjJhMA) Available from <https://www.youtube.com/watch?v=afJpLYjJhMA>
- SAS Software. (2015). [Getting started with SAS enterprise miner: Exploring input data and replacing missing values \[Video\]](https://www.youtube.com/watch?v=EBFtrPKQ0zg). | [Transcript](https://www.youtube.com/watch?v=EBFtrPKQ0zg). Available from <https://www.youtube.com/watch?v=EBFtrPKQ0zg>
- SAS Software. (2015). [Getting started with SAS Enterprise Miner: Setting up an Enterprise Miner project \[Video\]](https://www.youtube.com/watch?v=489wJm2X0TY). | [Transcript](https://www.youtube.com/watch?v=489wJm2X0TY) Available from <https://www.youtube.com/watch?v=489wJm2X0TY>

Unit 1 >> Overview of Data Mining

Introduction

This unit introduces the topic of data mining and provides an overview of the process and tools used to support it. It covers the use of databases, database management systems, and data extraction and manipulation tools as they are used in obtaining data for subsequent analyses.

Learning Activities

u01s1 - Studies

VitalSource Bookshelf

After procuring the e-books for this course, you may access them via the VitalSource Bookshelf. A link to your Bookshelf is provided in the left **Course Tools** menu.

The interactive features built into the VitalSource Bookshelf help save time and enhance learning by allowing you to search, notate and highlight, organize, and share information about your texts. VitalSource Bookshelf is also available as a downloadable application. Once this application is on your computer or mobile device, your materials will be available even when not connected to the Internet. Visit [VitalSource: Support](#) to download the application.

Readings

Use your *Data Mining for Business Analytics* textbook to read the following:

- Chapter 1, "Introduction."
- Chapter 2, "Overview of the Data Mining Process: 2.7 Using Excel for Data Mining." This section in your text provides information about data mining software and an examination of the software market, including applications beyond Excel. This section will be useful in this unit's discussion on data mining analysis tools.
- Chapter 3, "Data Visualization."

These chapters introduce data mining and the processes and tools used to support it, such as databases, database management systems, data extraction and manipulation tools, and data visualization and summarization. This information will provide the foundation for your work throughout the course. You will delve deeper into these concepts in your unit discussions.

Assignment Preparation

Use the Internet to complete the following:

View the site, [Kaggle: Your Home for Data Science](#), and create a free account. You will use this Web site throughout the course. To prepare for this unit's assignment, familiarize yourself with the "Titanic: Machine Learning From Disaster" data set.

u01s2 - Course Preparation



[Bloom's Taxonomy](#)

 [Transcript](#)

Capella University Library

In today's marketplace, identifying, analyzing, and synthesizing information to support operational and strategic decisions for organizations is critical. While there are many resources readily available online, it is important to use appropriate and high quality information to support academic and professional activities. This process includes not only locating information but also ensuring that the information is sound, appropriate, and worthy of academic use.

- The [Research Guide – MBA Library Guide](#) provides guidance for accessing and using the rich resources available in the [Capella University Library](#) and beyond.
- The [Data Mining and Reporting: MBA6135](#) library guide is designed to help you learn search concepts for information in this course.

In particular, be sure to review the tutorials for finding scholarly articles. You will use the library several times throughout the course to locate and evaluate articles related to data mining.

Multimedia

Bloom's Taxonomy – Enhance Your Critical Thinking Skills

Critical thinking is an important skill to cultivate for both your coursework and your professional development. Many learners do not initially realize that there are different ways of thinking and levels

of depth in understanding. Bloom's taxonomy provides a structure to help conceptualize these different levels. Awareness of different ways to approach information helps you move beyond basic understanding to more effectively analyze, evaluate, and synthesize important concepts. It also helps you to clarify expectations and provide an appropriate level of response for your MBA coursework.

- Click **Bloom's Taxonomy** to see how the levels are defined and to explore how this can help in your academic and professional work.

Capella Writing Center

Professional communication is essential for a business leader. When you start the Capella MBA program, you complete a writing assessment. The purpose of this assessment is to provide feedback on your writing performance, and resources to enhance your writing skills. Visit the [Capella Writing Center](#) for a variety of tools to help you improve your written communication and presentation skills. You may also send papers to the [Smarthinking](#) tutoring service to receive feedback and revision suggestions prior to submitting assignments.

Note: Additional resources and discussions are also available in the Campus resource, [MBA Central](#) to help support your progress in the MBA program.

u01s3 - Optional – Software Installation and Tutorials

Throughout the course, you will use SAS Enterprise Miner or other relevant software such as SPSS or XLMiner with functionality to perform and complete several activities. Below are the instructions for downloading SAS Enterprise Miner.

XLMiner is an Excel add-on specifically for data mining. Instructions on how to download a trial license of this Microsoft Excel software add-on are included on the copyright page of the e-book provided at the Capella bookstore. **Note:** Both a textbook code and one-time course code are required to register and install XLMiner.

SAS Software

SAS Enterprise Miner is available for you to use during this course for your statistical calculations. There is no fee for you to use SAS Enterprise Miner. SAS is one of the most commonly used

statistical analysis tools in business, and as a Capella learner you have access to this valuable resource. **Note:** Mac users need a Windows virtual machine.

To have free access to this software you must first register with SAS using the following steps:

1. Register with your Capella e-mail address by going to the [SAS On Demand for Academics: Create an Account](#) registration page.
2. After registering and receiving your login ID and password, click on the course enrollment link provided to you by the instructor to access the SAS MBA6135 courseroom.
3. Login using your ID and password. Click the **Enroll** button to enter the course for access and use of SAS Enterprise Miner.

Note: You only need to access the SAS Web site to download the SAS Enterprise Miner software. If you have problems with the download or procedure, first consult with your instructor to determine if your problem is a hardware compatibility problem. To access this software in a Mac OSX environment you will need a Windows virtual machine. If you are having problems navigating the SAS Web site you can contact [SAS Customer Support](#).

SAS Enterprise Miner Tutorials

The following is a tutorial for importing Excel files into SAS Enterprise Miner:

- [Importing Excel Files Using SAS Enterprise Miner](#).

The following tutorials are some of the resources available to you from SAS:

- [Getting Started With SAS Enterprise Miner: Setting Up an Enterprise Miner Project | Transcript](#)
In this SAS Enterprise Miner video you will create a project, define a library and data source, create a process flow diagram, and place input data nodes on your process flow diagram.
- [Getting Started With SAS Enterprise Miner: Exploring Input Data and Replacing Missing Values | Transcript](#)
In this SAS Enterprise Miner video you will explore variable properties, create partition data sets, and work with missing data values.
- [Active Learning With SAS Text Miner | Transcript](#)
In this SAS video you will learn about active learning in Text Miner, using user feedback with modeling.
- [Manage All Unstructured Data With SAS Text Analytics | Transcript](#)
In this SAS Enterprise Miner video you will learn about using text analytics in decision making.
- [Getting Started With SAS Enterprise Miner: Transforming Data and Building Parametric Models | Transcript](#)
In this SAS Enterprise Miner video you will learn about imputing and transforming data, building regression models, and neural network models.

IBM SPSS Statistics Standard GradPack (Version 22 or Higher for Windows or Mac)

SPSS statistical analysis software is commonly used in business applications. As a Capella learner, you have access to the more robust IBM SPSS Statistics **Premium** GradPack arranged at an academic discount through a contracted vendor. Refer to the [Statistical Software](#) page on Campus for general information on SPSS software, including instructions for procuring the most recent version made available to Capella learners.

Analysis ToolPak Activation

Microsoft Excel Analysis ToolPak is a Windows PC add-in program that provides data analysis tools that allow you to complete more complex statistical functions. If you do not have the add-in available in Excel (Data tab), enter Analysis ToolPak into the Excel Help search box for installation instructions.

[StatPlus:mac LE](#) is the free equivalent for Apple users, but must be downloaded from AnalystSoft.

u01a1 - Data Exploration, Visualization, and Summarization

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Use standardized data mining application software to retrieve decision support data.
- Present results of analyses in a concise and compelling manner.
- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

In Chapters 2 and 3 of your readings for this unit, you were introduced to data exploration and visualization techniques. For this assignment, become familiar with a specific data mining software and tools for data exploration, including visualization and summarization.

After registering at Kaggle (linked in Resources), enroll in the getting started competition, "Titanic: Machine Learning Through Disaster." Read through the entire competition rules. Download the train.csv and test.csv data sets. All of the information you need is on the Web site. **Note:** These .csv files open in Excel.

Using only the train.csv data set, complete the following:

- Examine, clean, and reprocess a data file to prepare for analysis. You should impute/exclude missing data as appropriate.
- Create a table of summary statistics for all the variables in the data. Discuss and interpret.
- Create a scatter plot of two continuous variables in the data set. Discuss why this scatterplot might be interesting.
- Create a matrix of correlations for all continuous variables. Discuss which correlations are interesting.
- Create a management report summarizing your output to support the White Star Line (owner of the Titanic) or a law firm representing deceased passengers (your choice.)
- Submit your results in a single page Word document, pasting relevant outputs created to demonstrate your work.

Course Resources

[APA Style and Format](#)

[Data Mining and Reporting: MBA6135](#)

[Kaggle: Your Home for Data Science](#)

u01d1 - Data Mining Analysis Tools

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Use standardized data mining application software to retrieve decision support data.

Activity Instruction

This course introduces you to the commonly-used data mining software, SAS Enterprise Miner. For this discussion, search the Internet and find another tool that is a competitor to SAS Enterprise Miner. For your initial discussion post:

- Describe the tool, its vendor, its capabilities, its cost, and other requirements. Compare the tool features and functionality to SAS Enterprise Miner. Include an APA reference citation for any resources you used to analyze your selected tool.
- Review the [Data Mining and Reporting: MBA6135 library guide](#) for additional information to assist you with your tool search.
- Refer to the [Discussion Participation Scoring Guide](#) prior to completing your post, to ensure that you meet all evaluation criteria. Unless otherwise indicated, these criteria will apply for all of your discussion activities throughout the course. Post your initial discussion, then proceed to the [Response Guidelines](#) for instructions on completing your responses to your peers' initial discussion posts.

Response Guidelines

Read the posts of your peers and respond to at least two. If possible, respond to two learners whose tool selections are different from one other's and from yours. In each response, compare the features and functionality of the tools you and your peer selected. Also, discuss when you might want to use SAS Enterprise Miner rather than the other tools.

Course Resources

[Data Mining and Reporting: MBA6135](#)

Unit 2 >> Data Mining Techniques: Predictive Models

Introduction

This unit continues the overview of the various tools and techniques involved in data mining, introducing the nature of the problems for which each tool or technique is most suitable. The unit covers analysis approaches based on statistics, as well as approaches based on artificial intelligence.

Learning Activities

u02s1 - Studies

Readings

Use your *Data Mining for Business Analytics* textbook to read the following:

- Review Chapter 2, "Overview of the Data Mining Process." This chapter introduces you to a variety of data mining tools and techniques and their uses. The reading provides a context for applying these techniques in this unit and throughout the course.
- Chapter 6, "Multiple Linear Regression." This chapter introduces you to prediction methods based on linear regression analysis.

Remember, your e-books are available for reading via VitalSource Bookshelf. You can access Bookshelf from the left **Course Tools** menu or via the optional downloaded application. Remember, your e-books are available for reading via VitalSource Bookshelf. You can access Bookshelf from the left **Course Tools** menu or via the optional downloaded application.

Predictive Models Based on Regression Analysis

For your unit assignment, you will develop a predictive model based on linear multiple regression to predict clothing expenditures. You might wish to reference the tutorial, [Getting Started With SAS Enterprise Miner: Transforming Data and Building Parametric Models](#) | Transcript.

Assignment Preparation

Use the Internet to complete the following:

- Log in to [Kaggle: Your Home for Data Science](#) and familiarize yourself with the playground competition, "House Prices: Advanced Regression Techniques." For this unit's assignment, you will complete the analysis and prepare a report.

u02a1 - Building a Predictive Model

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Use standardized data mining application software to retrieve decision support data.
- Present results of analyses in a concise and compelling manner.
- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

Enroll and compete in the Playground competition, "House Prices: Advanced Regression Techniques," on Kaggle: Your Home for Data Science, linked in Resources, and successfully submit at least one model for consideration.

For this assignment, complete this analysis and prepare a 3–5 report of the results. Include the following:

- Describe which variables you used to forecast SalePrice (Kaggle.com scores) and what data cleaning you had to perform.
- Fit at least one predictive model based on multiple linear regression that could predict SalePrice.
- Use the models built on the training data set (train.csv) to forecast the test data set (test.csv). Submit the models to Kaggle.com. **Note:** Provide your Kaggle.com username and your scores with your report.
- Describe the results of your predictive models, and how well it predicts SalePrice.
- Based on the predictive model, analyze the implications of the results from this model for forecasting SalePrice.

Present the statistical output from your selected software, and the results of your analysis in a concise and compelling manner, using effective communication strategies and appropriate business terms. Adhere to current edition APA standards for citations and references.

Refer to the Building a Predictive Model Scoring Guide prior to submission to ensure that you meet all evaluation criteria, then submit your assignment.

Course Resources

[APA Style and Format](#)

[Kaggle: Your Home for Data Science](#)

u02d1 - Core Ideas of Data Mining

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Identify internal data sources appropriate to address global management decision support needs.

Activity Instruction

In Chapter 2 of your reading, you examined the core ideas of data mining, the two distinctions of data mining techniques (supervised and unsupervised methods), and the STEPS in data mining, otherwise referred to as SEMMA. For this discussion activity, select one of the following areas below and thoroughly discuss:

- The core ideas of data mining.
- Steps in the data mining process.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, compare your analysis of the steps and identify any gaps in your peer's analysis. If possible, relate any experience you have had in working with data mining.

u02d2 - Data Mining Application

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Use standardized data mining application software to retrieve decision support data.

Activity Instruction

Using the Capella University Library and the Internet, identify 2–3 software packages or systems that are among the leaders in each of the following categories:

- Data warehouses.
- Data marts.
- Online analytical processing (OLAP).

For each package or system, analyze the relative strengths and weaknesses of the following:

- Vendor.
- Cost (to the extent this is available).
- Environment in which it runs.
- Customer application examples (generally, vendors of these types of software packages will cite customer examples on their Web sites).

Explain when you would use each of these software packages or systems, and provide a clear rationale why. Present the results of your analyses in a concise and compelling manner, using effective communication and appropriate business terms.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, compare the effectiveness of the data mining initiative your peer analyzed with the initiative you analyzed. Discuss any similarities in the strengths and weaknesses of the application of data mining tools and processes.

Unit 3 >> Cluster Analysis: K-Means and Hierarchical Clustering

Introduction

This unit focuses on one widely used technique of data mining: cluster analysis. Cluster analysis helps identify meaningful, similar patterns in data through segmentation. This grouping of data provides a visual for identifying areas for business opportunities. The unit covers the concepts of, and approaches to, cluster analysis, the application of the technique, and examples of cluster analysis using representative software.

Learning Activities

u03s1 - Studies

Readings

Use your *Data Mining for Business Analytics* textbook to read the following:

- Chapter 5, "Evaluating Classification and Predictive Performance." This chapter describes how predictive and classification performance can be assessed, problems in overfitting to the training data, and performance metrics.
- Chapter 14, "Cluster Analysis." This chapter explains cluster analysis, a data mining technique that helps identify meaningful patterns in data. You will apply this technique in your unit assignment.

Use the Capella University Library to read the following:

- Balijepally, Mangalaraj, and Iyengar's 2011 article, "[Are We Wielding This Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research](#)," from the *Journal of the Association for Information Systems*, volume 12, issue 5, pages 375–413. You will use this article to analyze the implications of using cluster analysis to support decision making.

Assignment Preparation

Use the Internet to complete the following:

- Log in to [Kaggle: Your Home for Data Science](#) and review the "Titanic: Machine Learning From Disaster" data set. For this unit's assignment, you will apply cluster analysis to predict survival and prepare a report.

u03a1 - Performing Cluster Analysis

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Use standardized data mining application software to retrieve decision support data.
- Present results of analyses in a concise and compelling manner.
- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

Chapter 14 in your readings for this unit introduces the concept of cluster analysis. For this assignment, you will apply cluster analysis to predict survival on the Titanic and prepare a 3–5 page report of the results. To complete this assignment:

- Re-examine the "Titanic: Machine Learning From Disaster" train.csv data set, which you downloaded from Kaggle.com in Unit 1, to identify appropriate data needed for your analysis.
- Apply the k-means algorithm to the appropriate data file using standardized/normalized scores of input variables. Set k to 2, to reflect two clusters: died, survived. Note: Use SAS Enterprise Miner or other relevant data mining software, such as XLMiner with utility, to perform cluster analysis.
- Identify the groupings obtained from your analysis, and what they might reveal about survival from the perspective of the White Star Line and a law firm.
- Submit a single model to Kaggle.com and report the results along with your Kaggle.com user name.

Present your work, the statistical output from your selected software, and the results of your analysis. Your report should be written in a concise and compelling manner, using effective communication strategies and appropriate business terms (both the company and a law firm). Adhere to current edition APA standards for citations and references.

Refer to the Performing Cluster Analysis Scoring Guide prior to submission to ensure that you meet all evaluation criteria, then submit your assignment.

Course Resources

[APA Style and Format](#)

[Kaggle: Your Home for Data Science](#)

u03d1 - Cluster Analysis Concerns

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Use standardized data mining application software to retrieve decision support data.

Activity Instruction

In their 2011 article, "Are We Wielding This Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research," Balijepally, Mangalaraj, and Iyengar analyze the application of cluster analysis in information systems research. For this discussion, share how you feel about the use of cluster analysis in data mining for business decision support. Has the article changed your view on the application of cluster analysis?

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, identify additional implications for cluster analysis in data mining.

Course Resources

[Are We Wielding This Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research.](#)

Unit 4 >> Neural Networks Models for Data Mining

Introduction

This unit discusses various approaches to deriving and interpreting neural network models with application to business problems.

Learning Activities

u04s1 - Studies

Readings

Use your *Data Mining for Business Analytics* textbook to read the following:

- Chapter 11, "Neural Nets." This chapter provides an overview of neural nets, a method for classification and prediction. You will use neural net analysis for a real-world scenario in your unit assignment.
- Chapter 13, "Association Rules." This chapter describes the unsupervised learning methods of association rules where the goal is to identify item clustering in transaction-type databases.

Use the Capella University Library to read the following:

- Nishi's 2005 article, "[Market-Basket Mystery](#)," from *Chain Store Age*, pages 12A–14A. You will use this article to compare market-basket analysis to other data mining techniques.

Assignment Preparation

Use the Internet to complete the following:

- Log in to [Kaggle: Your Home for Data Science](#) and review the "Titanic: Machine Learning From Disaster" dataset. For this unit's assignment, you will use this information to perform neural net analyses and prepare report.

u04a1 - Neural Net Analysis

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Use standardized data mining application software to retrieve decision support data.
- Present results of analyses in a concise and compelling manner.

- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

For this assignment, you will use the information from this unit's readings to perform neural net analyses based on the Titanic data and prepare a 2–3 page report. Include the following:

- Using SAS Enterprise Miner or other relevant data mining software, run a neural net model for the "Titanic: Machine Learning From Disaster" data set to identify the types or characteristics of people who likely to have survived versus died.
- Use an option to normalize/transform quantitative data.
- Submit your work to Kaggle.com and report your scores.
- Compare the accuracy of your results with that of the cluster analysis you performed earlier.
- Interpret the implications of your results for the White Star Line or for a law firm representing some of the passengers.

Present your work, the statistical output from your selected software, and the results of your analysis. Your report should be written in a concise and compelling manner, using effective communication strategies and appropriate business terms. Adhere to current APA style and formatting standards for citations and references.

Refer to the Scoring Guide prior to submission to ensure that you meet all evaluation criteria, then submit your assignment.

Course Resources

[Kaggle: Your Home for Data Science](#)

[APA Style and Format](#)

u04d1 - Other Data Mining Techniques and Their Uses

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Use standardized data mining application software to retrieve decision support data.

Activity Instruction

For this discussion, compare assumptions, advantages, and disadvantages of the following data mining tools; provide an example of when using the tool would be most appropriate:

- K-nearest neighbor classifier.
- Discriminant analysis.
- Logistic regression.
- Neural nets.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, analyze the example scenarios and determine if they are appropriate to the specific tool.

u04d2 - Market-Basket Analysis

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Identify internal data sources appropriate to address global management decision support needs.

Activity Instruction

In his 2005 article, "Market-Basket Mystery," Nishi examines the use of market-basket analysis by grocery chains and other retailers. For this discussion, explain why market-basket analysis is less structured than other forms of data mining. Also, describe what grocery chains often do to gain more structured data.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, provide any additional insights you have gained about market-basket analysis from the Nishi article and Chapter 11 in your

readings for this unit.

Course Resources

[Market-Basket Mystery](#)

Unit 5 >> Modeling With Textual Data and Ethical Aspects of Data Mining

Introduction

A vast majority of business data is stored in text documents that are unstructured. This unit describes the process and analytical aspects of text mining. Moreover, some data mining practices raise issues of privacy violations, runaway marketing, and even financial harm. This unit also covers both the potential threats of data mining and approaches to protection against them.

Learning Activities

u05s1 - Studies

Readings

Use the Internet to complete the following reading:

- Cary, Wen, and Mahatanankoon's 2003 article, "[Data Mining: Consumer Privacy, Ethical Policy, and Systems Development Practices](#)," from Human Systems Management, volume 22, issue 4, pages 157–168. This article outlines ten data mining system development practices you will use to evaluate a data mining project in your unit assignment.
- The excerpt from Chakraborty, Pagolu, and Garla's 2013 text, "[Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS](#)."

Optional Video

- [Manage All Unstructured Data With SAS Text Analytics](#) | [Transcript](#) This video provides an overview of how analytics can make use of information that is embedded in your text

documents.

u05a1 - Ethics in Data Mining

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

Cary, Wen, and Mahatanankoon's article, "Data Mining: Consumer Privacy, Ethical Policy, and Systems Development Practices," introduces a variety of ethical issues associated with data mining, as well as strategies that can help mitigate these issues. For this assignment, examine these ethical implications and strategies in real-world practice. Complete the following:

- Locate a journal article from the Capella University Library (or another scholarly source) about a case in which a business data mining project was halted due to ethical issues. Be sure your topic is related to business!
- Describe the threat the data mining effort was posing and to whom.
- Assess the ethical implications of the data mining project and discuss the role of each of the following in the project cancellation:
 - Publicity.
 - Litigation.
 - Governmental regulation.
- Present the results of your analyses in a concise and compelling manner, using effective communication and appropriate business terms. Include an APA reference citation for each resource from which you draw information.

Review the Data Mining and Reporting: MBA6135 library guide for additional information to assist you with your article search.

Refer to the Ethics in Data Mining Scoring Guide prior to submission to ensure that you meet all evaluation criteria, then submit the assignment.

Course Resources

[APA Style and Format](#)

[Data Mining and Reporting: MBA6135](#)

[Data Mining: Consumer Privacy, Ethical Policy, and Systems Development Practices.](#)

u05d1 - Law Enforcement Data and Privacy

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Identify internal data sources appropriate to address global management decision support needs.

Activity Instruction

Improving coordination between various law enforcement agencies has long been a goal, especially since the terrorist attacks of September 11, 2001, which focused attention on the lack of such coordination. One crucial aspect of improving this coordination is the sharing of criminal data between the agencies. Shared data would allow law enforcement agencies to use data mining techniques to more effectively identify threats and investigate cases. However, many people are concerned that sharing data might pose a grave risk to individual rights to privacy.

For this discussion, complete a scan of the research on this issue as it is related to businesses and individuals, and then post your informed position on the question of data sharing and mining among law enforcement agencies. Be sure to include APA references for the resources you use. Review the Data Mining and Reporting: MBA6135 library guide for additional information to assist you with your research.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, evaluate your peer's position, and either further support or refute his or her arguments with sound reasoning and information from your research.

Course Resources

[Data Mining and Reporting: MBA6135](#)

u05d2 - Core Ideas of Text Mining

Activity Context

This discussion helps you develop the skills to master the following course competency:

- Identify internal data sources appropriate to address global management decision support needs.

Activity Instruction

Read the excerpt, "Text Mining and Analysis: Practical Methods, Examples, and Case studies Using SAS."

For this discussion activity, select either A or B of the following and thoroughly discuss:

A: Describe the Core

- Steps of the text mining process.

B: Discuss the Following Concepts as Related to Text Mining

- Information retrieval.
- Document classification.
- Ontology management.
- Information extraction.
- Sentiment analysis.

Response Guidelines

Read the posts of your peers and respond to at least two. In each response, compare your analysis of the steps and identify any gaps in your peer's analysis. If possible, relate any experience you have had in working with data mining.

Course Resources

[Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS \(excerpt\) \[PDF\]](#).

Unit 6 >> Integrative Case Project and Report

Introduction

In this final unit, you will integrate your knowledge gained from this course to make a recommendation for action in a case-based scenario. You will select an appropriate analysis method and run the necessary analyses to draw your conclusions in a formal memorandum report.

Learning Activities

u06s1 - Studies

Assignment Preparation

Use the Internet to complete the following:

- Log in to [Kaggle: Your Home for Data Science](#) and familiarize yourself with "House Prices: Advanced Regression Techniques." For this unit's assignment, you will be reviewing data about house sales in order to make a forecast.

u06a1 - Analyzing the Case

Activity Context

This assignment helps you develop the skills to master the following course competencies:

- Identify internal data sources appropriate to address global management decision support needs.
- Use standardized data mining application software to retrieve decision support data.
- Present results of analyses in a concise and compelling manner.
- Communicate in a manner that is professional and consistent with expectations for members of the business profession.

Activity Instruction

Throughout the course, you have examined and applied data mining strategies and tools. For this final assignment, you will integrate your learning into a report of the case study, "House Prices: Advanced Regression Techniques," about an Ames, Iowa real estate firm who is concerned with forecasting house prices.

Complete the following:

- Examine the data set to explore, and if necessary clean, the data. Summarize your findings from your data examination in your report's analysis.
- Develop tables and graphs to evaluate the effect of independent variables on the dependent variable ("SalePrice.") What do the tables and graphs indicate?
- Use appropriate data mining analysis technique to see whether any sales attributes can be used to predict customer sales price. What is the best model you can develop? What does it tell you, in regard to identifying sales price? Submit your model to Kaggle.com, and report on its predictive accuracy.
- What additional data would you want in order to complete your analysis of sales price before determining a model for sales prices? Explain why that data would be useful.
- What, if any, ethical concerns should be taken into consideration? There are always ethical concerns, so think about this question carefully.

Report Instructions

Present the results of your analyses in a concise and compelling manner, using effective communication and appropriate business terms and your report should:

- Be written in proper memorandum format. (Note: several appropriate templates are provided in Microsoft Word.)

- Begin with a short introductory paragraph stating the purpose of the memo, and end with a brief concluding paragraph.
- Contain (at least) the following subdivisions:
 - An analysis of sales price.
 - Your data mining analysis results and interpretation.
 - A discussion of additional data needed, and your rationale for using that data.
 - Your recommendation for action and your rationale for this recommendation.

Refer to the Analyzing the Case Scoring Guide prior to submission to ensure that you meet all evaluation criteria, then submit the assignment.

Note: Your instructor may also use the Writing Feedback Tool to provide feedback on your writing. In the tool, click the linked resources for helpful writing information.

Course Resources

[Kaggle: Your Home for Data Science](#)

[APA Style and Format](#)

[Writing Feedback Tool](#)

u06d1 - Course Reflection

Throughout the course, you have been building toward the achievement of the following competencies:

1. Identify internal data sources appropriate to address global management decision support needs.
2. Use standardized data mining application software to retrieve decision support data.
3. Present results of analyses in a concise and compelling manner.
4. Communicate in a manner that is professional and consistent with expectations for members of the business profession.

For this discussion, reflect on the knowledge and skills you have developed over the past six weeks. Address the following questions in your post:

- For each competency, how has your ability to perform these skills and apply this knowledge evolved?
- What concepts, skills, or insights were most relevant to you?
- How have you grown in your academic and professional goals? If you have made an action plan in a previous course, have you made progress on any items in the action plan?

Response Guidelines

When responding to others, provide encouraging feedback about how learners have grown over the past six weeks. If applicable, provide recommendations for how learners may continue to develop these skills and knowledge, or suggest additional strategies for leveraging the knowledge gained in the course.