MSOM 622: Data Mining & Business Technology

Prerequisite(s): MSOM 606

Credits: 3

I. Course Description

This course covers the fundamental concepts and techniques in managing and presenting data for effective data-driven decision making through the use of standard exploratory data analysis techniques using quantitative statistical and graphical methods. Topics include association rules, k-NN classifier, cluster analysis, discriminant analysis, classification and regression trees, logistic regression, neural networks, data storage and indexing strategies, data warehousing, metadata management, and data visualization. The course combines theory and application to provide an up-to-date, practical learning opportunity that focuses on the needs of modern managers and professionals in an age of global competition and rapid technological change. Character and operational excellence principles related to management activities are also explored and applied.

II. Course Objectives

1. Identify and articulate the purpose of data mining in business decision making and strategy Apply different contexts of data mining, notably association, classification and prediction, identifying strengths, weaknesses, and distinguishing characteristics of data mining techniques
2. Determine necessary inputs, outcomes, and interpret data mining model output
3. Conduct multiple data mining modeling exercises, focusing on the seven classes of data mining
4. Utilize multiple software platforms to perform various analysis and modeling procedures

III. Course Textbooks and Reading Materials

NOTE: Students must purchase in advance texts and/or materials for this course.

> ***Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner, 3rd Edition***
> Authors: Shmueli, G., Bruce, P.C. & Patel, N.R.
> Publisher: Wiley
> ISBN 13: 9781118729274
>
> ***Management by Proverbs***
> Author: Michael Zigarelli
> Publisher: Xulon Press
> ISBN 13: 9781607910176
> **XLMiner** (available via download)

To download and install XLMiner:

1. Go to www.AnalyticSolver.com/Student
2. Fill out the form on this page. All of the information is required.
3. ***Download and Install****

- Go to www.analyticsolver.com and Log In, if you're not logged in already, with the account you just created. Once logged in, you will see a Download Analytic Solver button available along the top right. Click this button and download the installer.
- Make sure that Excel is closed (not running), then run the SolverSetup program.
- The SolverSetup program will prompt you to choose between Analytic Solver Platform, Risk Solver Platform and XLMiner. Choosing Analytic Solver Platform gives you all the features of Risk Solver Platform and XLMiner, so this is usually the best choice. You can change this choice later in Excel, by choosing a menu option Help – Change Product on the Ribbon.
- When the SolverSetup program finishes, start Excel (the last Setup dialog prompts you to do this). You should see new tabs on the Ribbon for Analytic Solver Platform or Risk Solver Platform, andXLMiner. Click the Solver Platform tab – you should see a "Welcome" dialog with various links. Use the Help dropdown menu to open Help text, the User Guide and Reference Guide, and load example workbooks.
- If you have problems, the best avenues to get help are to email support@solver.com (this creates a support ticket in Frontline's Help Desk) or start a Live Chat from any page on www.solver.com, or from within Excel (Help – Support Live Chat).

IV. Weekly Information

Module One – Week 1

Exploring Data Mining & Big Data; Visualization

Objectives

At the end of this Module, you should be able to:

1.1 Discuss the merits of the seven pillars for successful analytics implementation
1.2 Know the purpose of Data Mining for business
1.3 Articulate the difference between Prediction, Classification & Association
1.4 Distinguish the types of variables, both input and target, and the role they play in model selection
1.5 Match business problems to data mining techniques

Reading

Data Mining for Business Analytics - Chapters 1 – 3
7 Pillars for Successful Analytics Implementation (PDF provided)

Management by Proverbs Principle #1 – Maintain a Biblical Worldview

Lecture - PowerPoints

- Introduction to Data Mining & the seven categories of relevance
- On Visualization

Deliverables

Discussion:
- Introduction – Answer the questions so that we can begin to get to know each other.

Discussion:
- Which of the seven pillars for successful analytics implementation would likely be the most difficult to achieve? What can be done as a team lead to ensure a successful implementation?

Discussion:
- *Management by Proverbs* Principle #1 – Maintain a Biblical Worldview – discuss the differences between Pragmatism and Empiricism, and how they apply to data mining.

Assignment:
- Deliverable #1 - Knowledge checks – matching the business problem to the correct data mining category
  - Note – suggested response is located on Blackboard Module #1

Module Two – Week 2

Anomaly Detection; Data Reduction and Principal Components Analysis; Model Performance Objectives

At the end of this Module, you should be able to:
   2.1 Assess the value of data reduction and identifying outlier cases
   2.2 Evaluate model performance (which will carry throughout the course)
   2.3 Conduct a Principal Components Analysis for variable grouping
   2.4 Conduct a Principal Components Analysis using components as predictor variables

Reading

Data Mining for Business Analytics - Chapter 4
Management by Proverbs Principle #25 – Terminate the Troublemakers

Lecture - PowerPoints

- On Data Reduction, PCA and Outlier Detection

Demos

- PCA: Predicting the value of a diamond – data reduction and outlier detection

Deliverables

Discussion:
- *Naked Statistics* author Charlie Wheelan offers the example of Bill Gates walking into a bar. If you were taking a poll about income in the neighborhood, why would Bill Gates walking in the bar at the point be problematic?

Discussion:
- *Management by Proverbs* Principle #25 – Terminate the Troublemakers – Troublemaking in the workplace can be focal, or subtle. These employees are often outliers, and can be detrimental to team success. How can you apply Principle #25 to identify such outliers?

Assignment:
- Deliverable #2 – Problem 4.4 from *Data Mining for Business Analytics* (pp. 103) – *Sales of Toyota Corolla Cars*

Module Three – Week 3

Association & Rule Learning; Computational Mapping

Objectives

At the end of this Module, you should be able to:

3.1 Know the purpose of association rules in data mining

3.2 Interpret the role and variations of Computational Maps

3.3 Interpret output and perform analysis with association rules

3.4 Conduct an analysis using Computational Maps

Reading

Data Mining for Business Analytics - Chapter 14

Management by Proverbs Principle #17 – Value Employee Input

Lecture - PowerPoints

- On Computational Maps
- Association & Rule Learning

Demos

- Association Rules: Identifying rules for cosmetic purchases
- Computational Mapping: Determining the reasons for running an international road race

Deliverables

Discussion:
- Discuss some ways in which computational maps can help solve business problems.

Discussion:
- *Management by Proverbs* Principle #17 – Value Employee Input – Valuing the input of employees can be a challenge. Often times, effectively soliciting input takes a set of rules, whether these rules are formal or informal, to follow. How can you use Principle #17 to develop an effective rules set for inviting employee input?

Assignment:
- Deliverable #3 – Problem 14.2 from *Data Mining for Business Analytics* (pp. 332) – *Identifying Course Combinations*

Module Four – Week 4

Regression – Linear, Logit, and Regression Trees (CHAID)

Objectives

At the end of this Module, you should be able to:

4.1 Understand the role Linear Regression, Logistic Regression and Regression Trees in Data Mining

4.2 Interpret output from the different regression model techniques

4.3 Perform regression analysis and draw conclusions about business outcomes or strategy

Reading

Data Mining for Business Analytics - Chapters 6 – 9

Management by Proverbs Principle #4 – Assemble and Accountability Group

Lecture - PowerPoints

- Regression, Logit, and the Regression Tree

Demos

- Regression Trees: Predicting persistence at a public university – the role of switching majors

Deliverables

Discussion:
- What are the largest differences between a linear regression, a logistic regression, and a CHAID analysis?

Discussion:
- *Management by Proverbs* Principle #4 – Assemble an Accountability Group – One of the greatest challenges in statistics and data mining is proper interpretation. Paralleling this challenge, people in companies have been known to skew results to favor a certain outcome. How can you use Principle #4 to ensure there is accountability for your statistical results?

Assignment:
- Deliverable #4 – Problem 9.1 from *Data Mining for Business Analytics* (pp. 214) – *Competitive Auctions on eBay.com*

Module Five – Week 5

Classification (Part I) – Cluster Analysis, Classification and Regression Trees (CART Analysis)
Objectives

At the end of this Module, you should be able to:
- 5.1 Understand the role of classification and class modeling in data mining
- 5.2 Interpret output for a cluster analysis and CART analysis
- 5.3 Perform a cluster analysis and evaluate cluster model performance

Reading

Data Mining for Business Analytics - Chapter 15
Management by Proverbs Principle #8 – Design a Decision Tree

Lecture - PowerPoints

- On Classification - Clustering

Demos

- Cluster Analysis: Customer segmentation and Small Business Supplier Preferences

Deliverables

Discussion:
- Offer a scenario in which cluster analysis would be effective. What data would be needed to execute the analysis effectively?

Discussion:
- *Management by Proverbs* Principle #8 – Design a Decision Tree – How is statistical classification similar to the decision tree concept from Principle #8 in *Management by Proverbs*? How are the two linked?

Assignment:
- Deliverable #5 – Problem 15.1 from *Data Mining for Business Analytics* (pp. 360) – *University Rankings*

Module Six – Week 6

Classification (Part II) – k-Nearest Neighbor, Naïve Bayes, Neural Networks
Objectives

At the end of this Module, you should be able to:
- 6.1 Understand the purpose of advanced classification methods in data mining
- 6.2 Interpret the inputs and outcomes of advanced classification methods

6.3 Perform a Naïve Byes Classifier analysis to inform critical business decisions

Reading

Data Mining for Business Analytics - Chapters 7, 8 and 11
Management by Proverbs Principle #7 – Plan Persistently

Lecture - PowerPoints

- Advanced Classification Methods

Demos

- Naïve Bayes: Predicting flight delays
- K-Nearest Neighbor: How can we develop a true list of peer colleges and universities? An analysis of k-NN
- Neural Networks: How banks predict the probability of loan default

Deliverables

Discussion:

Using the demo of Pandora and the examples in the text (pp. 163-165) and other sources you might uncover, draw up an explanation for a five-year-old about how Pandora works.

Discussion:

- *Management by Proverbs* Principle #7 – Plan Persistently –As we saw in the discussion about Neural Networks, one of the key weaknesses is its inability to conduct proper variable selection. This implies an extensive amount of planning. How can you apply Principle #7 to making effective decisions about the variables in data mining, and in your life?

Assignment:
- Deliverable #6 - Problem 8.1 from *Data Mining for Business Analytics* (pp. 184) – *Personal Loan Acceptance*

<u>**Module Seven – Week 7**</u>

Time Series & Forecasting

Objectives

At the end of this Module, you should be able to:
7.1 Know the role and importance of time series analysis
7.2 Interpret multiple model outputs in time series and forecasting
7.3 Perform time series regression-based forecasting analysis

Reading

Data Mining for Business Analytics - Chapters 16 – 18

Management by Proverbs Principle #15 – Build Organizational Trust

Lecture - PowerPoints

- On Time Series, Regression Based Forecasting & Smoothing Methods

Demos

- Time Series & Regression Based Forecasting: Predicting Clothing Sales Over Time

Deliverables

Discussion:
- What are some explanations for why a time series trend might break?

Discussion:
- *Management by Proverbs* Principle #15 – Building Organizational Trust – As we should expect, building trust takes time. Even more so, building trust around the presentation of analytics results for a company that is used to anecdotal decision-making can be a challenge. Borrowing from Principle #15, how can you continue to build trust among the company over time?

Assignment:
- Deliverable #7 - Problem 17.3 from *Data Mining for Business Analytics* (pp. 398) – *Toys "R" Us Revenues*

V. Course Information and Grading

Course Requirements (Weekly)
Each Module will be based on the inclusion and grading of the following segments: Discussion Boards/ Assignments/ Quizzes.

Class Discussions

Successful online learning requires active and meaningful participation in the Discussion Board Forum. Critical thinking, appropriate use of judgment and professional comments are required for all responses. Simple agreement or disagreement with a fellow student does not constitute a quality response. Student opinions are important but must be supported by relevant data obtained in the literature or text.

1. One initial discussion board post is due by the fourth day (Thursday @ 11:59 pm) of each weekly module and two response posts are due by the last day (Sunday @ 11:59 pm) of each weekly module.
   a. Undergraduate Word Count Requirements:
      i. Initial post must be >200 words
      ii. Responses must be >150 words
   b. Graduate Word Count Requirements:
      i. Initial post must be >250 words
      ii. Responses must be >200 words

2. Participation – <u>Students must meet initial and response post deadlines to be eligible for full points</u>. Late initial posts may be accepted with a 33% reduction in the overall score assessed for each day late. Two response posts are required. If only one is received, then the overall score will be reduced by 25%. If no response posts are received, then the overall score will be reduced by 50%.

3. Students should be respectful and write in a proper and professional manner.

Please refer to the Discussion Board Rubric located within course syllabus for specific requirements and guidance on delivering a substantial discussion.

Late and Make-up Assignments

All deliverables are due by 11:59 PM on the due dates indicated. When substantive emergencies occur, acceptance of late submissions is left to the discretion of the course instructor.

Assignments that are submitted after the due date <u>with prior written approval</u> (text or email) from the instructor may receive the following deductions:

1. Late assignments are subject to a deduction of 10% of the available points for each day late.

2. Assignments submitted later than one week after due date will not be accepted and will receive a zero (0).

3. No work may be submitted after the last day of the course.

4. See Class Discussions for late discussion post policy.

Written Work

- All graded assignments must be typewritten, as designated by the professor of record for the course.
- All referenced materials must be presented according to the Publication Manual of the American Psychological Association (APA), based on the most current published edition. (current, 7th edition) (Except for English courses that specify use of MLA format.)
- All required written assignments are due on the last night of each lesson week at 11:59pm EST. This information is included in the course scheduler and calendar of every course.
- All written and graded assignments are considered the property of the College of Adult and Professional Studies and should be returned to the appropriate professor of record for the related course.
- SafeAssign may be used to check for plagiarism.

Grading Scale and Assigned Letter Grades
Charleston Southern Online Graduate classes follow the following numerical and letter grading scale. Grades will be posted by the Sunday night following the closing of the Module. No grades will be given to students over the phone.
A: 930 - 1,000 points

B+: 900 - 929 points
B: 830 - 899 points
C: 750 - 829 points
F: 0 - 749 points

The following table shows the assignment types contained within this course and the assigned weighting used to determine the final course grade.

Activity Types	Weights (%)
Discussion Forums	30%
Weekly Deliverables	70%
Total	100

Discussion Board Rubric – Introduction (First week only)

Criteria	Minimal	Effective
Percent Possible: 0 or 50 Initial Post Introduction.	Percent Possible: 0 No post or did not follow instructions at all.	Percent Possible: 50 Posted following instructions.
Percent Possible: 0 or 50 **Response Posts** Welcome.	Percent Possible: 0 No post or did not follow instructions at all.	Percent Possible: 50 Posted following instructions.

Discussion Board Rubric - Initial Post

Initial Post Rubric	Minimal	Competent	Effective	Mastery
Grade	**0-74%**	**75-82%**	**83-92%**	**93-100%**
Content	• Provides weak contribution to the discussion board • Does not convey concept understanding • Shows evidence of critical thought but is not concise or complete	• Makes mostly complete response to the questions • Mostly restatement of materials • Conveys thoughts adequately; however, need to further develop critical aspects	• Advances the discussion and body of knowledge providing relevant, original thoughts to the question • Conveys thoughts in a well-rounded manner that show understanding, critical thought, and application of concepts	• Relevant and thought-provoking viewpoints • Innovative and indicates a high level of thought • High level of critical thought, analysis, and application of concepts and ideas
Support & References	• Does not cite any peer-reviewed sources • Addresses questions in a manner that is unclear • Not appropriate APA format	• Student cites 1 peer-reviewed source • Not in appropriate APA format	• Student cites 2 peer-reviewed sources in the body of the post • Not in appropriate APA format	• Student cites 2 peer-reviewed sources in the body of the posts • Proper APA format
Mechanics/ Length	• Post is less than 196 • Poorly written • 0 points for posts submitted after 11:59 PM EST on Thursday	• Post is 197-221 words • A few writing errors	• Post 222-250 words • Well written	• Post exceeds 250 words • Very well written

Discussion Board Rubric - Peer Response

Peer Response Rubric	Minimal	Competent	Effective	Mastery
Grade	**0-74%**	**75-82%**	**83-92%**	**93-100%**
Content	• Shows minor comprehension of concepts and viewpoints from peers • Responses are vague and do not show understanding or comprehension	• Shows adequate understanding of the knowledge and concepts posted by peers • Responses are mostly, "I agree…" • Source is somewhat relevant and published in last 5 years	• Advances discussion providing additional commentary and critique • Responses are correct but mostly restate the peers' viewpoints without development of critical thought • Source is relevant and published in last 3 years	• Advances discussion providing additional commentary that is relevant or offers an alternative • Critical thought development is evidenced by thoughtful and meaningful responses • Source is relevant and published in last 3 years
Support & References	• Does not cite any peer-reviewed sources • Not appropriate APA format	• Student cites 1 peer-reviewed source • Not in appropriate APA format	• Student cites 2 peer-reviewed sources • Not in appropriate APA format	• Student cites 2 peer-reviewed sources in the body of the posts • Proper APA format
Mechanics/ Length	• Post is less than 158 words • Poorly written • 0 points for posts submitted after 11:59 PM EST on Sunday	• Post is 158-178 words • A few writing errors	• Post 178-200 words • Well written	• Post exceeds 200 words. • Very well written

Assignment/Case Study/Course Project Rubric

Assignment Rubric	Minimal	Competent	Effective	Mastery
Grade	**0-74%**	**75-82%**	**83-92%**	**93-100%**
Critical Thinking	• Consistently below expectations • Lacks significant details of the assignment • Lacks critical thought, analysis, and application • Partially addresses the topic	• Inconsistent comprehension • Marginal responses • Content, concepts, ideas/responses are not fully developed	• Demonstrates satisfactory understanding • Lacks answer to all questions/full scenario • Demonstrates some knowledge of content and professional vocabulary • Content needs additional focus for innovative responses other than restatement of text	• Relevant use of vocabulary throughout • Demonstrates superior knowledge of concepts and theories • Addresses problem/issue directly • Demonstrates active critical thinking relevant to problem identification and effective solutions
References & Organization	• Does not cite any peer-reviewed sources • Not appropriate APA format • 0-799 words • Poorly written • 0 points for late assignments	• Sources are outdated or ineffective • Lacks focus on content • Unorganized • Some writing errors • Not in appropriate APA format • 800-899 words	• Student cites some relevant, current peer-reviewed sources • Follows directions • APA formatting demonstrated but not extensive • 900-1000 words	• Professional writing structure • Well-organized response • Student cites appropriate, current peer-reviewed sources • Proper APA format • Exceeds 1000 words

Short Answer Rubric

Short Answer Rubric	Minimal	Competent	Effective	Mastery
Grade	**0-74%**	**75-82%**	**83-92%**	**93-100%**
Clarity	• Consistently below expectations • Lacks significant details of the assignment • Lacks critical thought, analysis, and application • Partially addresses the topic	• Inconsistent comprehension • Marginal responses • Content, concepts, ideas/responses are not fully developed	• Demonstrates satisfactory understanding • Lacks answer to all questions/full scenario • Demonstrates some knowledge of content and professional vocabulary • Content needs additional focus for innovative responses other than restatement of text	• Relevant use of vocabulary throughout • Demonstrates superior knowledge of concepts and theories • Addresses problem/issue directly • Demonstrates active critical thinking relevant to problem identification and effective solutions
Depth	• Lacked detail that demonstrates active thought • One-sentence answer	• Provided enough detail for marginal requirement • "Textbook" in nature (restatement of what was in textbook)	• Demonstrated some form of outside material used • Shows some critical thinking and active thought	• Articulates critical thoughts & analysis throughout the answer • Obvious use of outside resources to supplement resources
Quality	• Unorganized • Poorly written • 0 points for late assignments	• Lacks focus on content • Some writing errors	• Follows assignment instructions • Well-written • Few writing errors	• Professional writing structure • Well-organized response • Free of grammatical, spelling, & formatting errors