MIS 431-001: Data Mining with Business Applications
Spring 2019 Course Syllabus
Tuesday 07:20 PM - 10:00 PM, Robinson Hall B228

This syllabus is tentative. I reserve the right to add, remove or alter this syllabus as needed.

Instructor: David Svancer
Office: Enterprise Hall 181
E-mail: dsvancer@gmu.edu
Office Hours: By Appointment

Prerequisites: Grade of C or higher in (MIS 301 or MIS 303) and in (BUS 310 or OM 210/211).

Course Textbooks and Software:

Textbook: An Introduction to Statistical Learning with Applications in R
Available for free at the following website: http://www-bcf.usc.edu/~gareth/ISL/

Textbook: R for Data Science by Hadley Wickham
Available for free at the following website: http://r4ds.had.co.nz/index.html

Textbook: R Tutorials by David Svancer
PDF tutorials that cover R programming, data manipulation, data visualization and implementation of data mining algorithms in R.
These will be uploaded to Blackboard as we cover topics. Most classes this semester will end with an “R workshop” where I demonstrate data mining techniques in R.

Software: R
Open source data science and statistical computing software.
Download at the following link: https://cloud.r-project.org/

Software: RStudio Desktop
An integrated development environment (IDE) for R.
Download at the following link: https://www.rstudio.com/products/rstudio/download/
Course Description
Data mining—the art of extracting useful information from large amounts of data—is of growing importance in today’s world. The amount of data flowing from, to, and through enterprises is enormous, and growing rapidly—more rapidly than the capabilities of organizations to use it. Enterprises are trying to make effective use of the abundance of data to which they have access: to make better predictions, better decisions, and better strategies. Therefore, managers now need to know about the possibilities and limitations of data mining. This course introduces data mining problems and tools to enhance managerial decision making. Students will learn how to ask the right questions and how to draw inferences from data by using the appropriate data mining tools. Students will acquire hands-on experience by applying data mining methods using R in the following broad categories:

1. Data understanding, exploration, and visualization
2. Applying statistical modeling and machine learning algorithms using R to solve complex business problems
3. Business interpretation and communication of model results

Overall, the course will enable students to approach business problems data-analytically, envision data mining opportunities in organizations, and follow up on ideas or opportunities that present themselves.

Course Objectives
- Introduce the steps involved in data mining, from goal definition to model deployment.
- Discuss data preparation, transformation, exploration, and visualization techniques.
- Examine supervised learning methods for regression and classification
  - Linear and Logistic Regression
  - Linear and Quadratic Discriminant Analysis, KNN regression and classification
  - Decision Trees and Ensemble Methods using Bootstrap Aggregation (Bagging)
- Examine unsupervised learning methods, such as Principal Components Analysis (PCA)

Undergraduate program learning goals (those in bold will be addressed in this class):

1. Our students will be competent in their discipline.
2. Our students will be aware of the uses of technology in business.
3. Our students will be effective communicators.
4. Our students will have an interdisciplinary perspective.
5. Our students will be knowledgeable about global business and trade.
6. Our students will recognize the importance of ethical decisions.
7. Our students will be knowledgeable about the legal environment of business.
8. Our students will be knowledgeable about team dynamics and the characteristics of effective teams.
9. Our students will understand the value of diversity and the importance of managing diversity in the context of business.
10. Our students will be critical thinkers.
Learning Goals of the Information Systems and Operations Management Program

a. Apply knowledge of information technology and business functions to understand its application in assessing, designing and improving business processes.

b. Develop data organization, storage and processing solutions to support organizational needs for information management. They will also have the option of developing skills in the area of supporting decision making through business intelligence solutions.

c. Use knowledge of computer networks as part of the IT solutions for improving business processes. They will also have option of developing more advanced skills in the areas of network and security.

d. Effectively manage information technology projects.

e. Understand the overall systems development life cycle and be able to recommend IT system solutions accordingly. They will also have option of learning appropriate development tools to develop prototype of IT solutions for business management.

Course Policies

- Access to the internet and a computer is required. Many of the course material will be online in our Blackboard course. The in-class exams will be administered through Blackboard. Hence, you must bring your laptop to class every day.
- It is necessary that you are able to install software applications – R and RStudio. These software applications are available for both Windows and Mac.
- Communication will be via Blackboard and/or your GMU e-mail only. Please make sure that you mention your class and section number in the subject. I will only respond to gmu.edu e-mails and will not respond to e-mails written on a private account.
- You must finish the assigned reading for the week before classes every week. Week 1 is an exception. You should be able to discuss the reading material in class.
- Attendance is mandatory. Every class will include applied examples using R with my added insights and clarifying comments. Any class material missed by the student is the student’s responsibility to acquire.
- There will be a midterm and final exam. The format of the tests will be discussed in class. No makeup examination will be given.
- All homework assignments will be due on Wednesday evenings, giving students a chance to ask questions on Tuesday during class. No assignment will be accepted more than 1 day late without my approval.
- All academic accommodations due to disability must be arranged through the Office of Disability Services (ODS). If you are a student with a disability and you require academic accommodations, please contact ODS at 703-993-2474. I will cooperate fully with the ODS to accommodate a student’s special needs.
- Students with differing abilities should arrange to meet with me by the end of the first week of classes to arrange for reasonable accommodations for their learning needs.
- Athletes with travel schedules should meet with me by the end of the first week of classes to discuss any necessary arrangements that need to be made.
- Arrangements for any religious observances or GMU sanctioned activity must be arranged with the instructor at least one week prior to the event.
- Students requiring accommodations due to medical conditions must provide medical documentation.
- By remaining registered in the course through drop/add period, you agree to all terms and policies set forth in the syllabus.
**Academic Integrity:** It is expected that students adhere to the George Mason University Honor Code as it relates to integrity regarding coursework and grades. The Honor Code reads as follows: “To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University community and with the desire for greater academic and personal achievement, we, the student members of the University Community have set forth this: Student members of the George Mason University community pledge not to cheat, plagiarize, steal and/or lie in matters related to academic work.” More information about the Honor Code, including definitions of cheating, lying, and plagiarism, can be found at the Office of Academic Integrity website at oai.gmu.edu

Mason takes instances of academic dishonesty very seriously. While the faculty have the authority to recommend the academic and educational sanctions for Honor Code violations listed below, there can be additional consequences based on the College your program is housed in. At the very least a **disciplinary record is created** whenever a student is found responsible for violating the honor code.

Typical academic sanctions include but are not limited to:

<table>
<thead>
<tr>
<th>Type of Violation</th>
<th>First Offense</th>
<th>Second Offense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plagiarism</td>
<td>An F in the class; referral to Writing Center; and Academic Integrity Seminar completion</td>
<td>An F in the class; referral to the Writing Center; Academic Integrity Seminar completion; termination from the School of Business; and at least one semester suspension or expulsion</td>
</tr>
<tr>
<td>Cheating</td>
<td>An F in the class; Academic Integrity Seminar completion</td>
<td>An F in the class, Academic Integrity Seminar completion; termination from the School of Business; and at least one semester suspension or expulsion</td>
</tr>
<tr>
<td>Lying</td>
<td>An F in the class; and Academic Integrity Seminar completion, and at least one semester suspension</td>
<td>An F in the class; Academic Integrity Seminar completion; termination from the School of Business; and at least one semester suspension or expulsion</td>
</tr>
</tbody>
</table>
Egregious Violation (e.g., stealing an exam; submitting coursework from another class as original work across multiple courses; lying to an employer about academic performance, false identification or posing as another, in person or online) | An F in the class; Academic Integrity Seminar completion; termination from the School of Business; and at least one year suspension
---|---
An F in the class; Academic Integrity Seminar completion; termination from the School of Business; and expulsion

### Methods of Student Evaluation

Students will be evaluated based on assignments, projects, and in-class exams.

<table>
<thead>
<tr>
<th>Item</th>
<th>Homework Assignments</th>
<th>Final Project</th>
<th>Midterm Exam</th>
<th>Final Exam</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>500</td>
</tr>
</tbody>
</table>

#### Course Grade

1. Students must be officially registered in this section to receive a grade. It is the sole responsibility of the student to verify their own registration status. (I will not verify your registration.) Specifically, you will not receive a grade if your name does not appear on the official class list. (Don’t wait until the end of the semester to be surprised.) Registration problems should be directed to either the SOM Office of Student Services or the Registrar’s Office.

2. The final letter grade is assigned objectively and strictly according to the numerical scores of all exams, quizzes, and assignments as a proportion of the total possible points in the course.

<table>
<thead>
<tr>
<th>92.5% - 100%</th>
<th>A</th>
<th>81.5% - 86.49%</th>
<th>B</th>
<th>69.5% - 77.49%</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.5% - 92.49%</td>
<td>A-</td>
<td>79.5% - 81.49%</td>
<td>B-</td>
<td>59.5% - 69.49%</td>
<td>D</td>
</tr>
<tr>
<td>86.5% - 89.49%</td>
<td>B+</td>
<td>77.5% - 79.49%</td>
<td>C+</td>
<td>below 59.49%</td>
<td>F</td>
</tr>
</tbody>
</table>

A+ may be awarded for exemplary performance
# TENTATIVE COURSE SCHEDULE

**Reading:**  
*ISLR* – Introduction to Statistical Learning, *RDS* – R for Data Science  
- Appropriate R Tutorials will be uploaded to Blackboard for each class/topic  
- Data mining methods in R will be demonstrated at the end of most classes this semester

<table>
<thead>
<tr>
<th>Class</th>
<th>Date</th>
<th>Topics</th>
<th>Reading</th>
<th>Assigned</th>
<th>Due</th>
</tr>
</thead>
</table>
| 1     | 01-22  | Introduction to Data Mining  
Introduction to R Programming                                               | ISLR - Chapters 1 and 2 R tutorial                                    | HW 1     | 02-06  |
| 2     | 01-29  | Advanced R programming  
Data Manipulation and Preprocessing                                        | RDS - Chapters 5 R tutorial                                           |          |        |
| 3     | 02-05  | Data Manipulation  
Joining and Restructuring Data                                              | RDS - Chapters 3, 12, 21 R tutorial                                   | HW 2     | 02-15  |
| 4     | 02-12  | Data Visualization                                                      | RDS - Chapters 3 R tutorial                                           |          |        |
| 5     | 02-19  | Applied Statistics – Distributions, Expected Value, Variance, Transformations  
Exploring and Summarizing Data with R                                        | R tutorial                                                           | HW 3     | 03-01  |
| 6     | 02-26  | Applied Statistics – Confidence Intervals, Hypothesis Testing, Correlation, Sampling Distributions, Central Limit Theorem | R tutorial                                                           |          |        |
|       | 03-05  | **Midterm Exam**                                                       |                                                                        | Final Project | 05-11  |
|       | 03-12  | **Spring Break – No Class**                                            |                                                                        |          |        |
| 7     | 03-19  | Regression – Simple Linear                                              | ISLR - Chapter 3 R tutorial                                           |          |        |
| 8     | 03-26  | Regression – Multiple Linear and Interactions  
Variable Selection Methods in R                                                | ISLR - Chapter 3 and 6 R tutorial                                    | HW 4     | 04-05  |
| 9     | 04-02  | Logistic Regression                                                     | ISLR - Chapter 4 R tutorial                                           |          |        |
| 10    | 04-09  | Discriminant Analysis and Naïve Bayes                                   | ISLR – Chapter 4 R tutorial                                           | HW 5     | 04-19  |
| 11    | 04-16  | KNN Regression and Classification  
Cross Validation                                                               | ISLR – Chapter 2 and 5 R tutorial                                    |          |        |
| 12    | 04-23  | Tree-Based Methods  
Decision Trees, Random Forests  
Ensemble Methods (Bagging)                                                     | ISLR – Chapter 8 R tutorial                                           | HW 6     | 05-11  |
| 13    | 04-30  | Principal Components Analysis (PCA)  
Dimension Reduction with LDA  
*Course Wrap-Up, Work on Final Projects*                                      | ISLR - Chapters 6 and 10 R tutorial                                  |          |        |
|       | 05/14  | **Final Exam 07:30 PM – 10:15 PM**                                       |                                                                        |          |        |

MIS 431, David Svancer