

Xavier University

Exhibit

Management Information Systems Syllabi

Management Information Systems

2017

INFO 680-81 Introduction to Data Mining

Joel Asay
Xavier University

Follow this and additional works at: https://www.exhibit.xavier.edu/management_information_systems_syllabi

Recommended Citation

Asay, Joel, "INFO 680-81 Introduction to Data Mining" (2017). *Management Information Systems Syllabi*. 348.

https://www.exhibit.xavier.edu/management_information_systems_syllabi/348

This Restricted-Access Syllabus is brought to you for free and open access by the Management Information Systems at Exhibit. It has been accepted for inclusion in Management Information Systems Syllabi by an authorized administrator of Exhibit. For more information, please contact exhibit@xavier.edu.

INFO 680 / MKTG 680: Introduction to Data Mining (Summer 2017)

CLASS LOCATION AND TIME: G28 Smith Hall, T/R, 6:00pm – 9:15pm

INSTRUCTOR

Name: Joel Asay
Email: asayj@xavier.edu
Office Location: 221 Smith Hall or 315 Smith Hall
Office Hours: **M, T, W, R: 4:00PM – 6:00PM**
Or by appointment

Telephone: 513-475-2938 (office)
Website: canvas.xavier.edu
Section: 81

Williams College of Business Mission:

“We educate students of business, enabling them to improve organizations and society, consistent with the Jesuit tradition.”

Required Text:

Fundamentals of Predictive Analytics with JMP 2nd edition by Ron Klimberg and B.D. McCullough
(ISBN: 978-1-629-59856-7)

Optional Texts:

Practical Data Analysis with JMP 2nd edition by Robert Carver
(ISBN: 978-1-612-90823-6)

Data Mining for Business Analytics by Shmueli, Bruce, Stephens and Patel
(ISBN: 978-1-118-87743-2)

Prerequisite:

STAT 500

Course Description:

This introductory course will familiarize students with popular data mining methods for generating new knowledge from existing data. While soft skills related to data mining will be touched upon, the focus of this class is on developing applied data mining prowess with a software package. The course will explore popular data mining methods and the managerial uses thereof. Topics discussed include: the process of data mining for knowledge discovery, data management and cleaning, data mining methods such as exploratory, descriptive, and predictive analysis, decision support systems and contemporary issues in data mining. Each method and concept is taught using real data from various organizations. Data mining output is directly tied to potential business strategies and outcomes.

Learning Objectives:

Upon completion of this course, you should be able to:

- Recognize popular data mining methods and know when to apply them
- Be familiar with and properly use a commercial data mining software
- Apply appropriate data mining techniques to generate new knowledge from data
- Employ data mining output in problem solving
- Communicate derived findings to management in a meaningful way
- Develop business strategy founded in evidence sourced from data mining output
- Correctly answer data mining related questions on the CAP exam

WCB Learning Goals and Objectives:

This course reinforces the following MBA program learning goals:

Strategic Thinking and Leadership

- Ability to demonstrate the appropriate knowledge of data mining in strategic thinking

Ethics and Social Responsibility

- Ability to foster an ethical climate in their roles and responsibilities in business and society

Critical Thinking

- Ability to clarify problems, generate and evaluate alternatives using appropriate analytical and quantitative techniques, and draw conclusions

Effective Written and Oral Communication

- Ability to communicate complex subjects and solutions in accessible ways to others



General Course Policies:

- I take attendance every class. This will only factor into your grade if attendance is exceptionally poor. If you miss a class, it will be extremely hard for you to catch-up because of the “learning-by-doing” emphasis. If you miss class, it is **your** responsibility to cover the material and get additional notes from your classmates.
- Assignments are to be submitted on the due date. Late assignments will **not** be accepted unless prior arrangements have been made with the instructor. A score of 0 will be recorded for any assignment received beyond the due date however I will still provide feedback where appropriate.
- Grade tracking is the responsibility of the student. Canvas will be kept up-to-date for your convenience. Please communicate questions or concerns to me as soon as possible.
- All communication from me will be through Canvas and email. I usually respond to email within an hour of receipt if I am not sleeping. I usually respond within 12 hours. If I do not respond within 12 hours, I have likely missed your email and you should bother me again. I expect you to respond with 24 hours.

Class Technology Policies:

This class uses Microsoft Excel and a statistical package by SAS called JMP. All software we use in class will be provided to student personal computers free of charge. It is **STRONGLY** recommended that students bring and use their own laptop computers in class for assignments and class work.

Windows PCs are preferred, but the software is available for Macs as well. All examples in class will be shown with the Windows versions of the software. If you do not have a personal laptop available to you, accommodations will be made so you can still succeed in the class.

If you use an Apple computer, I can provide a copy of Microsoft Windows that can be installed **in addition to** your OS X operating system. If you would like help with this process, or if you desire other Microsoft software for free including Access, Project, SQL Server, etc., please let me know and I can provide it to you for free via the Microsoft Imagine Developer Network.

I expect technology use to be **appropriate in nature**. If I observe another student becoming distracted with your technology use, I may ask you to adjust the behavior or leave the classroom.

Plagiarism, cheating and privacy with regards to data files:

Excel and JMP are amazing applications that keep track of any actions performed. This information is also written to their saved data files, making it incredibly easy to identify plagiarism by the digital fingerprint left behind on these files. **PLEASE** do not be tempted to submit a classmate's data file as your own! Direct and unattributed use of another's efforts is cheating! The penalty for violation of this policy will be a zero for the assignment if it is a first offense. **Subsequent violation will result in an F for the course and other college disciplinary action.**

Several of the data files used in this class are copyrighted and should not be re-distributed by you. Some of the files we use may also require a non-disclosure agreement to be signed before working with them. If you are unwilling or unable to sign such a non-disclosure agreement, please notify me as soon as possible.

The Xavier virtual PC is also available for use and has the required software installed in the "Online Lab Xtra." The virtual PC can be slow and frustrating...it works but consider it a last resort!

The class will consist of the following methods of assessment

Exams:

There will be one midterm exam and a comprehensive final exam. These exams cover material from the textbook, readings, software package, data mining methods and the practical application of data mining output. Both exams will be completed out-of-class (take-home) and they are open book, note, internet but NOT open buddy/classmate/partner. These tests are designed to challenge even the best students in the course!

Individual or Group Project:

There will be a course project that uses JMP and Excel. You will be required to find your own data for the project; however, I will present several examples throughout the course. You are welcome to work on the project alone or with **one** partner. You must inform me what the data are and where they came from by class on **Thursday, June 1**. See the project assignment sheet for more information.

Homework:

Each week there will be a homework assignment. These assignments will require you to use Excel and JMP, again reinforcing the earlier suggestion you install JMP on a personal computer. You will also read and respond to articles related to real-world data mining use and their ethical/managerial implications. These responses may be in the form of discussion board posts or essays to me. Homework will be assigned as we cover the material and will take **2-3 hours per week** for an average student.

Reflections:

In keeping with our Ignatian paradigm, reflections on the course and your learning will be submitted to me twice during the term. These short (2-3 paragraphs) are meant for you think back on your reaction to the topics discussed in class, the skills you are acquiring, the structure of the course and how you make these experiences **your own**. Reflections might include a thoughtful reconsideration of the material, your experience in the class, your reaction to the arguments for and against an idea or strategy proposed or additional material you seek for further learning. I especially encourage you to internalize the unintended consequences of these data mining tools and especially the ethical considerations thereof. Reflections will be submitted via canvas and I will be the only one to read them.

In-Class Work:

We will be completing several tasks in class that provide the foundation for your homework and project. Thus, it is imperative you **attend class regularly**, otherwise you will find yourself with **significantly more homework!** While I do not specifically grant points for class participation, I reserve the right to **raise** your final grade if you show exceptional class engagement.

If you miss a class, it is your responsibility to obtain notes from your classmates, master the material, and submit the relevant homework assignments. The tasks we complete in-class will **not** be made available for download in a completed form—you must follow along in-class with your own computer. Video tutorials may be available for some of the tasks, but do not count on these always being available! **Remember** the exams are open book and note so following along and completing the tasks as we perform them in class will be **very** helpful on the exams!

Assessment (Scored out of 1000 total points)

Grading Criteria

Midterm	25%
Final Exam	25%
Homework	25%
Project	20%
Reflections	5%

Grade Distribution

A	94-100%	C+	77-79%
A-	90-93%	C	74-76%
B+	87-89%	C-	70-73%
B	84-86%	D	60-69%
B-	80-83%	F	Below 60%

Qualified students with disabilities who will require disability accommodations in this class are encouraged to make their requests to me by sharing their Accommodation Letters with me at the beginning of the semester either during office hours or by appointment. Disability related information is confidential. If you have not previously contacted Disability Services, I encourage you to do so by phone at 513-745-3280, in person on the Fifth Floor of the Conaton Learning Commons, Room 514, or via e-mail to Cassandra Jones at jonesc20@xavier.edu, to coordinate reasonable accommodations as soon as possible as accommodations are not retroactive.

It is my goal that this class be an accessible and welcoming experience for all students. If you are a student with a disability who may have trouble participating or effectively demonstrating learning in this course, contact me to arrange an appointment to share your Accommodation Letters from Disability Services and to discuss your needs. Disability related information is confidential. If you have not contacted Disability Services (located in the Learning Assistance Center) to arrange accommodations, I encourage you to do so by contacting Cassandra Jones, by phone at 513-745-3280, in person on the Fifth Floor of the Conaton Learning Commons, Room 514, or via e-mail at jonesc20@xavier.edu as soon as possible as accommodations are not retroactive.

Class Schedule

(Subject to change)

Date:	Class Topics:	Textbook Chapters:
5/16	Introduction to Data Mining and the CAP Credential Introduction to JMP, Excel, and Data Types Statistics Review and Data Cleaning	Klimberg: 1 & 2
5/18	Statistics Review and Data Cleaning (cont.) <u>Begin Exploratory Analysis</u> Exploratory Analysis and Visualization (In Excel & JMP)	Klimberg: 3 & 4
5/23	Scales, Measurement and Latent Variables Exploratory Factor Analysis (PCA) Exploratory Cluster Analysis (Hierarchical & K-means)	Klimberg: 7 & 9
5/25	<u>Begin Descriptive Analysis</u> Multivariable Statistics The Problem of Causal Inference and Dependence	Klimberg: 5, 7 & 9
5/30	Assumptions of Regression and the Linear Model Multivariable Regression and Transformations	Klimberg: 5
6/1	<u>Begin Predictive Analysis</u> Cross Validation and Regression as a Tool for Prediction Predictive and Prescriptive Analytics Project Work Time	Klimberg: 5 Data explanation & summary due Midterm Exam Assigned
6/6	Odds Functions, Probability and Classification Logistic Regression	Klimberg: 6 Midterm Reflection Due Midterm Exam Due
6/8	Logistic Regression (continued) CART (Classification and Regression Trees)	Klimberg: 6 & 10
6/13	Neural Networks & Machine Learning Project Work Time (Prof available for questions)	Klimberg: 12
6/15	Model Comparison	Klimberg: 14 Final Exam Assigned
6/20	Contemporary Issues in Data Mining (Topics TBD) Project Work Time	
6/22	Final Project Presentations	Final Reflection Due Final Exam Due