

Predictive Analytics and Data Mining 2024 Spring Semester

*The goal is to turn data into information and information into insight.
~Carly Fiorina, former Hewlett Packard CEO*

Instructor

Name: Andrea Hupman

Phone number: 314-516-4746

Email address: hupmana@umsl.edu

Campus office: 220 ESH

Office Hours: Wednesdays, 1-2:00 pm via Zoom

(or in-person with min 24 hr advance request via email)

Office Hours Zoom Link (also on Canvas):

<https://umssystem.zoom.us/j/93635258424?pwd=dElhNHRVM0tuVStrVHM2TEM5TDFkdz09>

Additional office hours by appointment

Class Sessions

In-Person: Mondays, 6-7:15 pm, ABH 003

Online: Asynchronous content, complete weekly (posted weekly on Tuesdays)

Course Website

All course material (except for the textbook) will be available through Canvas.

Prerequisites

A probability/statistics course and a minimum campus GPA of 2.0

Required Text

Data Science for Business by Provost & Fawcett (O'Reilly, 2013)

Optional Supplemental Text (for Python)

Python for Data Analysis by Wes McKinney (O'Reilly, 2018)

(A PDF of this text has been posted online at the following blog:

<https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf>)

Course Overview and Objectives

This course focuses on predictive analytics in business settings. Topics include applications of multivariate analyses to problems in marketing, finance, and supply chain management. The course covers use of decision trees, regression and logistic regression to explain phenomena and predict future outcomes. Students acquire experience in the use of modern tools for data mining including the use of Python.

Throughout the semester, students will

- Understand the data mining process
- Apply data mining and predictive analytic techniques to business problems
- Evaluate predictive models
- Describe the advantages and disadvantages of analytic techniques

- Use Python to conduct analyses
- Explain the business insights provided by classification and prediction algorithms

Technology Requirements

To be successful in this course, you are expected to have reliable internet access. If you have computing problems, it is your responsibility to address these or to use campus computing labs. Problems with your computer or other technology issues are not an excuse for delays in meeting expectations and missed deadlines for the course. If you have a problem, get help in solving it immediately. At a minimum, you will need the following software/hardware to participate in this course:

1. Computer with an updated operating system (e.g. Windows, Mac, Linux)
2. Updated Internet browser of your choice
3. Ability to navigate Canvas (Learning Management System)
4. Minimum Processor Speed of 1 GHz, 2 or more GHz recommended.
5. DSL or Cable Internet connection or a connection speed no less than 10 MB/s
6. Media player such as Windows Media Player to open course media. Flash player may be required by some aspects of the course and is available as a free download (<http://get.adobe.com/flashplayer/>)
7. Adobe Acrobat to open PDF files throughout the course – available as a free download (<http://get.adobe.com/reader/>)
8. A webcam and/or microphone is **recommended to enable remote work if needed**.
9. Access to a full copy Microsoft Excel. Students can get a version for free (<http://products.office.com/en-us/student/office-in-education>)
10. Ability to download and install Anaconda, a Python integrated development environment (ide), available for free at <https://www.anaconda.com/products/individual>

Grading and Assignments

Homework & In Class Assignments

Class sessions will include hands-on learning and assignments that apply the concepts being learned. Some assignments will require presenting your work to the class and may have points allocated to participation in class discussions. Poll Everywhere will be used and participation grades may be assigned. Homework assignments will provide an additional opportunity to apply the concepts. Assignments may also include written summaries and explanations that must be completed by students individually and in their own words.

Course Project

The course project is an opportunity to work on real world problems and will be completed in small groups. There will be 2 topics for class projects that will be presented by guest speakers from a local company. Students will indicate their preferred topic. Groups will be assigned by the instructor based on student interest.

You will receive a set of masked real-world data from a local company along with a statement of the business problem. Your task will be to use the methods and tools from class to analyze the data and create a recommended solution that includes the development of a model based on class concepts. Project updates will be assigned throughout the semester. The final deliverables will include a report and a presentation made during the scheduled final exam time for the class. A representative from the local company will be present at the presentations.

Data Mining in the World / Application Assignment

Data mining has been used in numerous applications over the past two decades. Your task in this assignment is to identify an article describing a real-world application of data mining. You may use more than one information source, as several applications are quite famous and have garnered a lot of attention. You will synthesize the information, providing a short presentation describing the scenario, including business

problem and data mining approach, in a short presentation (5 min max) for the class. Your presentation will take place in-class on April 15.

Importantly, only one student may present on any specific real-world project. A discussion board will be opened. Once you have settled on a topic/example, you must post a brief description of the topic with at least one information source that you plan to use. For example: "I plan to present on company ABC's use of DEF data to predict GHI. I plan to use the Harvard Business Review article, 'Article Title.'" If two students present on the same data mining example, the student who first posted their topic on the discussion board will receive credit for their submission; the other student(s) will receive a "0."

Exams

There will be two in-person exams administered during our regular class meeting time on the following days:

- Exam 1: Feb. 26
- Exam 2: April 8

Grading

Your final grade will be calculated as a weighted average with the following percentages:

- Homework, In Class Assignments, & Application Presentation: 35%
- Course project: 25%
- Exam 1: 20%
- Exam 2: 20%

Final letter grades will be assigned based on the following scale; however, the instructor reserves the right to modify the scale below based on class performance. Plus/minus grades will be assigned for scores within 2% points of the grade cutoff scores.

| | | | |
|---------|---|--------|---|
| 90-100% | A | 60-69% | D |
| 80-89% | B | 0-59% | F |
| 70-79% | C | | |

Illness, Emergent Situations & Class Attendance

This class includes in-person meetings. We will observe the following safety policies:

- **Do not attend class in-person if you feel ill, have a fever, or have recently tested positive for an illness that is communicable in a classroom setting (e.g. COVID or flu).**

If you are ill and unable to attend class, you may request a make-up assignment for any in-class assignments. Such requests must be made PRIOR to any missed class.

If you should experience any health event(s) that affect your ability to complete assignments virtually, the instructor will refer you to the UMSL CARES team which can collect medical documentation. Please do not submit medical documentation directly to the instructor. Leniency on due dates will be extended as needed. However, *it is very difficult to catch-up on class material if you fall behind. It is strongly advised that students do not request an illness-related extension unless it is actually necessary.*

Course Policies

We will follow all policies in the UMSL *Student Conduct Policy* (<http://www.umsl.edu/~studentconduct/Student%20Conduct%20Policy/index.html>)

Accommodations

Any student requiring special accommodations for any reason should contact the instructor as soon as possible. Students with disabilities who believe that they may need accommodations in this course are encouraged to contact Disability Access Services in 144 Millennium Student Center at 314-516-6554. Information about your disability is confidential.

Academic Integrity

- You are responsible for being attentive to and observant of University policies about academic honesty as stated in
 - the University's Campus Policies, <http://umsl.edu/services/academic/policy/academic-dishonesty.html>, and
 - the Code of Student Conduct, <https://bulletin.umsl.edu/studentconduct/>, found in the UMSL Bulletin
- Academic dishonesty is a serious offense that may lead to probation, suspension, or dismissal from the University. One form of academic dishonesty is plagiarism – the use of an author's ideas, statements, or approaches without crediting the source. Academic dishonesty also includes such acts as cheating by using any unauthorized sources of information and providing or receiving unauthorized assistance on any form of academic work or engaging in any behavior specifically prohibited by the faculty member (e.g., copying someone else's answers on tests and quizzes). Unauthorized possession or distribution of academic materials is another type of academic misconduct. It includes the unauthorized use, selling or purchasing of examinations or other academic work, using or stealing another student's work, unauthorized entry or use of material in a computer file, and using information from or possessing exams that an instructor did not authorize for release to students. Falsification is any untruth, either verbal or written, in one's academic work. Facilitation is knowingly assisting another to commit an act of academic misconduct. **Plagiarism, cheating, and falsification are not acceptable.**
- Accessing online discussion boards to discuss solutions to class assignments is not permitted.
- Copying responses from generative artificial intelligence for concept questions, or any items other than writing Python code, is not permitted.
- Copying from an unauthorized source and replacing words with synonyms is not permitted.
- All instances of academic dishonesty will be reported to the Office of Academic Affairs who will determine whether you will appear before the Student Conduct Committee for possible administrative sanctions such as dismissal from the university. The instructor will make an academic judgment about the student's grade on that work in this course. The campus process regarding academic dishonesty is [described in the "Policies" section of the Academic Affairs website](#)

Excused Absences

You may complete make up activities for an in-person class if you are unable to attend in-person. However, class sessions will not be recorded, and the student is responsible for learning all material presented in class.

Mandatory Reporting

Under Title IX, all UMSL faculty, staff, and administrators (with limited exception) are obligated to report any incidents of sexual harassment, sexual misconduct, sexual assault, or gender discrimination to the Title IX office and/or other University officials. This ensures that all parties are protected from further abuses and that victim(s) are supported by trained counselors and professionals. Note: There are several offices at UMSL (e.g. Counseling Services, Health Services, Community Psychological Service, Center for Trauma Recovery, and Student Social Services) whose staff are exempt from Title IX mandatory reporting when the information is learned in the course of a confidential communication.

Email Communications

All email from the instructor will be sent to each student's UMSL/MST/UMSYSTEM email address. Check this email address regularly. Treat all email as professional correspondence with an appropriate salutation and closing.

Late Assignments

No late assignments are accepted without advance written permission from the instructor. Extension requests must be submitted via email no less than 24 hours prior to the assignment due date. In true emergency situations (e.g. you have been rendered unconscious by an accident), requests should be submitted as soon as feasibly possible.

Schedule

The schedule for the course topics is tentative and presented by week. Revisions will be made as necessary.

DSB = *Data Science for Business* (course textbook)

| Week | Date (Slides) | Topics |
|------|----------------|--|
| 1 | Jan. 15 | Martin Luther King, Jr. Day – <i>No class!</i> |
| | Jan. 16-21 | Video: 1-Introduction & Data Mining Process Read Chp. 2 in DSB |
| 2 | Jan. 22 | Intro to Data Mining Activities / Customer Churn Example |
| | Jan. 23-28 | Video: 2-Correlation to Supervised Segmentation Read Chp. 3 in DSB |
| 3 | Jan. 29 | Introduction to Python Practice Problems HW1 Assigned |
| | Jan. 30-Feb. 4 | Video: 3-Fitting Parametric Models Read Chp. 4 in DSB |
| 4 | Feb. 5 | Fitting models in Python Variable Selection Algorithms for Multiple Regression HW 2 Assigned |
| | Feb. 6-11 | Video: 4-Overfitting |
| 5 | Feb. 12 | Discussion/practice with overfitting 5-Data Cleaning |
| | Feb. 13-18 | Video: 6-Data Visualization in Python |
| 6 | Feb. 19 | Introduce projects Review for Exam HW 3 Assigned |
| | Feb. 20-25 | Project work time/exam study time |

| | | |
|----|--|---|
| 7 | Feb. 26 | Exam 1 (in class) |
| | Feb. 27-March 3 | Video: 7-Similarity, Neighbors & Clusters Read Chp. 6 in DSB |
| 8 | March 4 | Practice Problems & Application in Python HW 4 Assigned |
| | March 5-10 | Video: 8- Decision Analytic Thinking 1: What is a good model? (confusion matrix, unbalanced data, costs/benefits) Read Chp. 7 in DSB |
| 9 | March 11 | 9-Visusalizing Model Performance Example problems & Python Project updates due HW 5 Assigned |
| | March 12-17 | Video: 10 – Evidence & Probabilities Read Chps. 8-9 in DSB |
| 10 | March 18 | <i>No in-person class</i> Video: 10-Evidence & Probabilities Video: 11-Ensemble Methods & Oversampling HW 6 Assigned Read Chp. 9 in DSB Read the Bias, Variance, and Ensemble Methods section of Chp. 12 |
| | March 19-22 | Project work time |
| | <i>UMSL Spring Break – March 23 – March 31</i> | |
| 11 | April 1 | Synthesis & Review |
| | April 2-7 | Exam study time |
| 12 | April 8 | Exam 2 (in class) (Chps. 6-9, 12 & additional topics covered in class) |
| | April 9-14 | Writing a data science report |
| 13 | April 15 | Student presentations – Data mining applications Project work time |
| | April 16-21 | Project work time |
| 14 | April 22 April 23-28 | Project work time |
| 15 | April 29 April 30-May5 | Project work time |
| 16 | May 6 | Final project presentations, May 6, 5:30 – 7:30 pm |