**M461 - Data Science Analytics**

**M561 - Advanced Data Science Analytics**

**(Fall 2024)**

**Instructor information:**

Instructor: John Bardsley

Office: M 208

Email: bardsleyj@mso.umt.edu

Office hours: Monday and noon; Wednesday and Friday at 1 pm; or by appointment.

**Course Format:**

Meetings: Monday, Wednesday, Friday 9:00-9:50 a.m.

**Course content:**

Algorithms are the machinery behind the data analytics (the subject matter and focal point of the course). To be good at data analytics, one must be a competent programmer and have experience with the data and the algorithms of data science. To gain an understanding of algorithm design and good programming practices, students will work with a set of prototypical algorithms that are representative of data analytics. To learn how to function as a data scientist in a relatively short time, the student will be actively engaged in turning the algorithms into code and using them with real data

After successfully completing the course, you will have a good understanding of the following topics and their applications:

1. Data Visualization, Data Mapping and Data Reduction
2. Classification and Regression Algorithms
3. Ensemble Methods, Random Forests and Gradient Boosting
4. Naive Bayes and Text Classification
5. Clustering and Segmentation

**Learning outcomes:**

1. Understand the purpose of data reduction and information extraction (e.g., associative statistics and data mapping).
2. Develop understanding and practical experience regarding reduction of massive data sets and data streams.
3. Understand the mechanics of distributed computing.
4. Ability to implement algorithms for processing massively large data sets. Ability to compute histograms, correlation matrices, and linear regression estimators using massively large data sets.
5. Understand the objectives of multiple regression and examining model assumptions. Ability to carry out and interpret hypothesis tests.
6. Competency using Python.

**Textbook:** None

**Course GitHub repository:** <https://github.com/um-perez-alvaro/Data-Science-Practice>

**Getting Python**: You can download Python from Python.org. If you do not already have Python, I recommend instead installing the Anaconda distribution (www.anaconda.com), which already includes most of the tools that you need to do Machine Learning/Data Analysis.

A common issue during class is students having different versions of Python libraries. Before class, I suggest you update all packages if you’ve installed the Anaconda distribution. You just need to open the Anaconda Prompt and type:

>conda update --all

**Homework and Practice Problems:** Homework exercises emphasizing applications of the algorithms and/or theory will be assigned bi-weekly.

While the assignments will be the same, I expect more depth and rigor in the work submitted by graduate students (M561), who will be held to higher standards.

Practice problems will be assigned daily. Practice problems will be graded based on completion.

**Grading policy:** Your course grade will be based on homework, practice problems and a final project

|  |  |
| --- | --- |
| Item | Percentage |
| Homework | 60% |
| Practice problems | 10% |
| Final project | 30% |

**Student Conduct:** All students need to be familiar with the Student Conduct Code. You can find in the "A to Z Index" on the UM home page. All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University.

**Accommodation:** The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors, and the Office for Disability Equity (ODE). If you anticipate or experience barriers based on disability, please contact the ODE at (406) 243-2243/Aber Hall - Room 116, ode@umontana.edu, or visit [www.umt.edu/disability](http://www.umt.edu/disability) for more information. Retroactive accommodation requests will not be honored, so please, do not delay. As your instructor, I will work with you and the ODE to implement an effective accommodation, and you are welcome to contact me privately if you wish. If you would like to request reasonable accommodations, you are advised to provide your ODE verification letter to your instructor in the first week of class so appropriate arrangements can be made. If you decide after the semester begins to disclose your disability and request accommodations, you should provide documentation, if possible, at least 10 days before the upcoming assessment so I may prepare appropriately. It is the responsibility of students to make sure they understand the types of modifications available to them before assessments.

**Digital Access:** Digital devices (like laptops and cell phones) are becoming increasingly important to success in college. I recognize that some students are unable to afford the cost of purchasing digital devices and that other students rely on older, more problem-prone devices that frequently break down or become unusable. I also recognize that those technology problems can be a significant source of stress for students. Given those challenges, I encourage students to contact me if you experience a technology-related problem that interferes with your work in this course. Finally, students can check out laptops and other technology for free from the Mansfield Library and TRIO Student Support Services (if eligible).

**Important note:** Announcements made in class are considered addenda to this syllabus. Make sure you stay informed as the progress of the class.