HLA 311: Data Science Applications in Healthcare

Instructor: Chris Malone
Office: Gildemeister 137 | Zoom: http://minnstate.zoom.us/my/chrismalone
Phone: Office: 507-457-2989
Email: cmalone@winona.edu
Websites: D2L: https://winona.learn.minnstate.edu/
               StatsClass.org: www.statsclass.org

Text: There is no text required for this course. The following references are provided.
- Few, Steven (2012); Show Me the Numbers, Analytics Press [Link: http://www.stephen-few.com/smtn.php]
- Nussbaumer Knaflic, Cole (2015); Storytelling with data, Wiley [Link: https://www.storytellingwithdata.com/]
- HBR Guide to Data Analytics Basics for Managers (2018); Harvard Business Review [Link: https://store.hbr.org/product/hbr-guide-to-data-analytics-basics-for-managers-ebook-tools/10232]
- Sharda, R, Delen, D. and Turban, E. (2018); Business Intelligence, Analytics, and Data Science: A Managerial Perspective, Pearson [Link: https://www.pearson.com/store/p/business-intelligence-analytics-and-data-science-a-managerial-perspective/P100000253187]

Course Description: The focus of this course will be on the methods and technologies used in the applications of data science to a variety of disciplines. The student will be exposed to the methods/processes for the curation and management of data, the summarization and visualization of data, machine learning algorithms, and the creation of data products for the communication of outcomes. The data science technologies used in this course will be accessible to students with a limited background in computing. Students will complete a data science project. Course not intended for DSCI or STAT majors.

Prerequisites: An introductory statistics course, e.g. at Winona State examples include STAT 301, STAT 110, STAT 305, PSY 231, ECON 222

Learning Outcomes: A student who has successfully completes this course will be able to:
1. The student will demonstrate the application of the data science cycle, e.g. for a project.
2. The student will demonstrate knowledge of common project development issues.
3. The student will perform, analyze, and critique various methods for the curation and management of data.
4. The student will calculate, organize, and evaluate data summaries and visualizations.
5. The student will be able to discuss and interpret outcomes from various machine learning algorithms.
6. The student will assemble and evaluate data products for the communication of outcomes.

Assessments:

Interactive Course Notes (Approximately 33% of grade)
A set of interactive course notes will be available for this course. The course notes are divided into modules and parts. These interactive notes may require that you complete various tasks and/or answer questions that are embedded within the notes. After you submit your notes, the answers to questions and/or tasks are provided so that you can check your understanding of the content being covered.

Homework Assignments (Approximately 67% of grade)
Most modules will include either a guided homework assignment and/or regular homework assignment. Guided homework assignments are setup where a particular data/analysis task is divided into small straight forward parts. My intention is that I "guide" you through a data/analysis the task to be completed. Guided homework assignments should take you about 30-60 minutes to complete. Some modules will include regular homework assignments. These are more traditional in nature and you will be asked to complete more substantial data/analysis task. You can expect regular homework assignments to take about 60-90 minutes to complete. All homework assignments will be distributed via Google Docs and completed assignments must be saved to your Google Drive.

Grades:

Your grade is determined by the completion and performance of the required work for this course.

You must complete the interactive course notes for each module / part. The course notes will be scored as follows:

- 0 pts: Did not complete tasks and/or answer questions posed
- 3 pts: Attempted to successfully complete tasks and/or answer questions posed

After completing the set of course notes for a module / part, you will be required to complete either 1) Guided Homework Assignment and/or 2) Regular Homework Assignment.

For the Guided Homework Assignment, scoring will be done as follows

- 0 pts: Did not make a reasonable attempt to complete the assigment by the due date
- 4 pts: Made a resonable attempt to sufficiently complete the assignment by the due date
- 7 pts: Sufficiently completed the assignment by the due date

Solutions to guided homework assignments will be provided upon submission. You should carefully review these solutions before moving onto the next content to be covered.

The Regular Homework Assignments will take more time to complete and are more typical in nature. These assignments will be worth between 20-25 points.

Note: You cannot submit homework assignments after solutions have been posted; thus, it is important that you submit your work before the specified deadline for the assignment.

Final grades will be determined using the following scale

- F: Less than 60%
- D: 60% - 70%
- C: 70% - 80%
- B: 80% - 90%
- A: 90% and above

Topic Outline:

1. Introduction to Data Science Cycle
    a. Formulate questions
    b. Curate/manage data
    c. Obtain appropriate summaries and visualizations
    d. Application of machine learning algorithms
    e. Creation of data products for communication of outcomes
2. Project Development
    a. Statement of problem
    b. Identification of consumers
    c. Data requirements
    d. Measuring success
    e. Ethical concerns

3. Data Curation and Management
    a. Common data structures, e.g. tables, databases
    b. Data documentation, e.g. dictionaries, data models
    c. Data verbs
    d. Aggregation
    e. Data cleaning and reshaping
4. Summaries and Visualization of Data
    a. Identification of appropriate data summaries
    b. Implementation of data visualization principles, e.g. EPTs, Gestalt
5. Introduction to Machine Learning Algorithms
    a. Types of algorithms, e.g. unsupervised and supervised
    b. Methods to evaluate algorithms
    c. Deployment Concerns, e.g. Architecture, Scalability, Automation,
6. Communication of Outcomes
    a. Creation and evaluation of data products
    b. Storytelling with data, e.g. Context, Design, Visuals, Delivery
    c. Effective feedback loop