Spring 2-1-2018

M 467.01: Big Data Analytic Projects

Brian M. Steele
University of Montana - Missoula, brian.steele@umontana.edu

Syllabus: Math 467 Data Analytics Projects

Instructor: Brian Steele. Office: Math 314. 243-5396. `brian.steele@umontana.edu`

Course Format: Three meetings per week MWF 11:00-11:50 a.m. Class time is split (25/75) between lecture and team solutions to data analytic problems.

Learning Outcomes: At the completion of this course, the student will

1. Understand the objectives and basic methods of predictive analytics.
2. Be able to program core algorithms of predictive analytics.
3. Be able to apply the core predictive analytic algorithms to new problems.
4. Have gained experience in planning a data analytic project.
5. Have gained experience describing and communicating solutions and problems related to team-oriented problem solving.
6. Have gained experience writing reports on data analytic projects.
7. Have gained experience and understanding related to team-based problem solving.

Course Content: The first half of the course focuses on several core algorithms of predictive analytics. Undergraduates will focus on implementing the algorithms in `Python` and applying them to practical data analytic problems. The second half of the course centers on solving practical data analytic problems. Teams are comprised of three undergraduate students for the purpose of solving problems originating from three sources: businesses, researchers, and competitions (private and public). A central focus of the course is communicating solutions to others in written and oral form.

M 567(G) co-convenes with M 467(U). Course content differs however. In particular, theoretical and foundational aspects of predictive algorithms is a significant component of M 567 but a negligible component of M 467. Problems assigned to graduate student teams are more difficult. Furthermore, M 567 requires a significantly more sophisticated level of writing compared to M 467.

Textbook: Algorithms for Data Science (ISBN-10: 3319457950)

Prerequisites: M 461 or consent of the instructor.

Homework: Exercises emphasizing application of algorithms will be assigned weekly during the first half of the semester.

Projects: Projects originate from three sources: businesses, researchers, and competitions (private and public) sponsored by companies and researchers. Students are organized in teams of usually three undergraduate students for developing solutions to a problem. Teams are necessary since the solutions are involved, sometimes requiring disparate methods from statistics, mathematics, and computer science. Solutions may require some innovation—some problems are not *textbook* problems with a methodological solution apparent from the start and no hidden difficulties. Students will work on two projects during the course of the semester.

Grading: Your course grade will be based on homework, written project reports and oral presentations. Homework assignments are worth 30% of the course grade. The first project, due near the midpoint of the semester is worth 20% and the remainder is attached to the second project due at the end of the semester.

Project grade is based on written reports and oral presentations (mostly short reports on progress). Written reports must discuss objectives, methods, results, and a conclusion. Oral presentations at the project end are expository and aimed at communicating methods and results.

Details

Our primary focus is on a project brought by Edulog, a company based in Missoula. The project will be introduced by a representative from Edulog in the near future. It's unclear exactly how complex the problem is, but I believe that everyone in the class will work on one or more aspects of the problem.

Students will work on another project as time permits in teams of ~ 3. The problems are Kaggle problems and brief descriptions can be found at

1. `https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs`
2. `https://www.kaggle.com/keplersmachines/kepler-labelled-time-series-data`