COSC 481 - Introduction to Data Science

Spring 2019
Syllabus

Would you like to predict the future?

Just the Facts

Course Number: COSC 481
Title: Introduction to Data Science
Semester: Spring 2019
Meeting Time: TR 2-3:50pm
Locale: Schaefer 165
Instructor: Alan Jamieson
Office: Schaefer 154
Office Hours: MWF 10:30am-noon, TR 1-2pm
Email: acjamieson@smcm.edu
Google Messenger: acjamieson@smcm.edu
Slack: via COSC481Group
Online Office Hours: Most evenings and weekends
Recommended Textbooks: *Grus, Data Science from Scratch, O'Reilly*
O'Neil and Schutt, Doing Data Science: Straight Talk from the Frontline, O'Reilly
O'Neil, Weapons of Math Destruction, Random House
Website: http://ripark.github.io/s19/cosc481

Catalog Description: This course serves as an introduction to the programming and data analysis techniques required for data science. Topics include: basic data science tools and techniques, data acquisition and analysis, basic predictive modelling, and ethics in data science. Not open to students who have received credit for COSC 480 Data Science. Prerequisites: COSC 251 or permission of the instructor.

Overview: Data science is one of the hottest subfields of computer science and mathematics over the past few years. The ability of a programmer to take large sets of data and glean some useful nugget for predictive analysis and decision-making has been sought after by the largest and smallest companies. In this course, we will do an introduction to data science, focusing on the algorithmic techniques required in Python. We will also work on examining data sets and formatting them for analysis.

Purpose: Data science and "Big Data" blend programming, statistics, and, sometimes, social sciences in order to find connections in chaos and make predictions for the future. It can be used for great things, or it can be used to tear societies apart. The goal of this course is to give you the tools to be able to solve these great problems, and to introduce the ethical questions behind many of the decisions we make as computer scientists.

Grade Distribution:

Labs/quizzes/homework - 20%
Miniprojects (2) - 15% each
Major Project - 40%
Major Project Presentation - 10%

You will be expected to participate in class by asking questions and answering questions posed by myself and those in class. Rather than a drab lecture, the class sessions will be run in a discussion style environment. Activity and debate are highly encouraged. In addition to this, there will be coding exercises and every student will be expected to submit at least one solution throughout the semester. Failure to do so will result in a grade penalty.

Learning Objectives: At the completetion of COSC 481, students will be able to:

construct a predictive model utilizing large data sets.
explain basic data science concepts.
design a complex predictive model.
design a presentation on an implementation of a predictive model.
construct a large data set from available sources.

Final Information: There will not be a final exam in this class, but teams will be presenting their final projects during the final period Saturday, May 4 at 2pm in Schaefer 165. Failure to attend your presentation will result in an F for the course.

Assignments: There will be three projects in this course, two smaller and one major. The major project will be a half-semester team project. Each team will take on some data science problem related to a larger societal problem, procure the data sets, and then present the results during the final period. Your group will also prepare a position paper arguing your results to be sent to appropriate officials in the hopes of instigating change. In addition, there will be a smattering of in-class and homework programming assignments.
Blackboard Use: I will be utilizing Blackboard primarily for your grades in this course.

GitHub Use: Major project files and some other sources will be maintained on GitHub. Any questions, concerns or objections should be noted during the first week of classes. You will be given a primer on how to use GitHub during the first part of the semester.

Policies

Cell Phones: Please, turn off or turn to silent any cell phones prior to getting to class. If they go off in class they are distraction not only to myself, but to everyone else in the class as well. Habitual offenders will be excused from the class with a 0 for any quizzes that day.

Computer Use: Computer use in this lab is for academic use only. If you bring a laptop with you to this class I expect you to be only using it for purposes related to this class. The same goes for the computers in this lab.

Attendance and Tardiness: Attendance is highly recommended. Missing a class not only causes you to miss the information disseminated in that lecture, but can cause you to miss important information in regards to exams and assignments and the potential of receiving a 0 for a quiz that day. I start class promptly on the hour and expect the students to be in class at that time. If you have circumstances that can prevent you from being in class on time, please let me know as soon as possible. Habitual offenders will be excused from the class with a 0 for any quizzes that day.

Exams and Quizzes: There will be no exams in this class. Every class has the potential of having a quiz to reinforce the ideas from the lecture the previous class. These will not be announced ahead of time. They will be 1-3 question quizzes that can be easily done in 15 minutes either at the start or the end of the class period.

Assignments: Assignments and other outside of class work should be done on an individual basis unless otherwise specified in the description of the assignment. Assignments and other outside of class work will be taken late only under the conditions listed in the Late Policy section.

Late Policy: You are allowed 2 "slip-days" throughout the semester. This means that you may turn in an assignment late, where each day it is late will reduce your number of slip-days by 1. So, you could turn in a project 2 days late, but then you wouldn't have any further slip-days left for the rest of the semester. Once you are out of slip-days, if you turn in the assignment late, you will earn a 0 for that assignment. As a further encouragement to turn in assignments on-time, each slip-day you have left at the end of the semester will add 0.5% to your final average. You may not use a slip day for the second project.

Extra Credit: I may or may not be offering any extra credit opportunities in this class.

Communication: The simplest way to get in touch with me is by coming by my office during my office hours or contacting me via email. The easiest way to get in touch with me "after hours" is to send me an email. I habitually check my St. Mary's email account all hours of the day. If you come by my office and the door is open, feel free to stop in to chat. The open door indicates that I'm not working on anything that has to keep my undivided attention at that time so do not feel that you are interrupting me or anything like that. I do make appointments if you have a certain time that you'd like to meet with me. If it fits in my schedule (meaning I'm not teaching class during that time) I will be happy to meet with you.

Academic Honesty: Academic misconduct policies are covered in the Student Code and Student Rights and Responsibilities, Article III. Pay close attention to the definitions of academic misconduct noted in Section 1. This can be found in the Student Handbook.

Disability: If you have any kind of disability that can affect your performance in this class, please let me know privately through email or stopping by my office.

Title IX and Sexual Misconduct: As stated in the St. Mary's Way, the College is a place where we strive to foster relationships based upon mutual respect, honesty, integrity, and trust. As such, we are committed to providing an educational, living, and working environment free from all forms of harassment and discrimination for all members of our community. The College prohibits all forms of sexual or gender-based harassment, discrimination or misconduct, including sexual assault, sexual harassment, relationship violence, and stalking.

If you or someone you know has experienced sexual misconduct, you may find information about resources and options on the Campus Rights webpage (www.smcm.edu/campus-rights) or by contacting the College's Title IX Coordinator, Michael Dunn (mkdunn@smcm.edu or 240-895-4105). Under College policy, faculty members are required to share any reports of sexual misconduct with Michael in order to make sure that the College is responding appropriately to address the health and safety needs of members of our community.

There are on-campus confidential resources available, including the counselors at the Wellness Center (240-895-4289) and the Sexual Misconduct Advocacy and Resource Team student-run 24/7 hotline (301-904-2015). More information about on- and off-campus confidential resources, as well as medical treatment, law enforcement, and other support services, may be found on the Campus Rights webpage.

Schedule: The schedule for the class is attached to the syllabus. The schedule is subject to change (multiple times).

Closing: The most important thing in any of my classes is that you are learning and expanding your horizons. If you are having any undue difficulty with your work as it pertains to this class, please contact me as soon as possible. Always remember that professors win when you don't need us any longer. I want you to be bouncing ideas off of each other throughout the class and it is my hope that by the end of the semester that you are driving the class session rather than me.

Provisional Schedule

Week 1: What is Data Science?
Week 2: Python libraries.
Week 3: Exploratory data analysis.
Week 4: Data acquisition and cleanup.
Week 5: Miniproject 1.
Week 6: Data scraping.
Week 7: Correlation, causation, inference.
Week 8: Gradient descent.
Week 9: Regression and miniproject 2.
Week 10: Major project introduction.
Week 11: Regression part 2.
Week 12: Classification.
Week 13: Major project checkin.
Week 14: Ethics in data science.