CS 299-01 – INTRODUCTION TO DATA SCIENCE, SPRING 2018

COURSE INFORMATION

Instructor : Prof. Sampath Jayarathna, Web: http://www.cpp.edu/~ukjayarathna

Contact : Office: 8-46, Email: ukjayarathna@cpp.edu, Phone: (909) 869-3145

Office Hours : Monday and Tuesday, 11.00 AM – 1.00 PM, or email me for an appointment

Schedule : Monday and Wednesday, Room: 8-302, Time: 2.00 PM – 3.50 PM

Website : http://www.cpp.edu/~ukjayarathna/courses/s18/cs299

Piazza : www.piazza.com/csupomona/spring2018/cs299/home

Blackboard : https://blackboard.cpp.edu/

Prerequisites : CS 140 grade of C or better, or instructor's consent. Knowledge in linear algebra and statistics.

WHAT IS THIS COURSE ABOUT?

This course will introduce students to this rapidly growing field of Data Science and equip them with some of its basic principles and tools as well as its general mindset. Students will learn concepts, techniques and tools they need to deal with various facets of data science practices.

WHAT WILL YOU GET FROM THIS COURSE?

- Define and explain the key concepts and models relevant to data science.
- Understand the processes of data science: identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization.
- Develop an appreciation of the many techniques for data modeling
- Be comfortable using commercial and open source tools such as Python, the R language and associated libraries for data analytics and visualization.

REQUIRED/OPTIONAL MATERIALS:

- **Required textbook**. No textbook is required. All the key course content will be documented in slides, which will be available in the course website after each lecture.
- **List of optional but recommended materials.** You may find some of these optional textbooks helpful, though none are required:
    - Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, By William McKinney, O'Reilly; 2 edition (October 20, 2017)
    - R for Data Science, By Garrett Grolemund and Hadley Wickham, O'Reilly January 2017 1st Edition
    - Data Science from Scratch: First Principles with Python By Joel Grus, O'Reilly 1st edition, 2015
- **Bring Your Own Device (BYOD)**. You must have a computing device (Laptop, Tablet, or Phablet), we will do some activities in class and you should have a device in class to fully participate.

TENTATIVE COURSE SCHEDULE

Topics: The tentative topics are as follows. Topics and specific course activities may change as needed. PowerPoint slides will be available on the course web page after each lecture.

Week 1: Syllabus and Introductions
Week 2: Unstructured and Semi-Structured Data
Week 3: Data Cleaning
Week 4: Data Wrangling
Week 5: Text Data Analysis and Inference
Week 6: Evaluations
Week 7: Machine Learning on Data
Week 8: Delivering Results
Week 9: Recommender Systems
Week 10: Project Demo

WHAT YOU CAN EXPECT FROM ME:

I have an open-door policy i.e., office visits. My posted office hours are times when I will make concerted effort to be available. Occasionally administrative meetings or emergencies may interfere with these posted times. The open-door policy is: if my door is open, I am in and welcome walk-in visitations. I am committed to supporting students with disabilities. If you have challenges related to these issues or others I want to work with you to help you succeed. Please come and talk to me, since only you can properly communicate your situation to me.

WHAT YOU CAN GIVE TO THE CLASS:

It is extremely important for you to be engaged in the course. Otherwise, you will fall asleep and wonder what happened to your tuition dollars. So, I encourage you to ask questions during lecture and actively participate at the piazza forum. For the first few weeks, when asking a question at the class, state your name so that I know who you are.

Cell phones and Tardiness: You may have cell phones in class, but they must be on mute, or airplane mode and not answered until the end of class. You are expected to arrive on time so that you do not cause a disruption in the middle of class. I would like to start the class at the scheduled time. If you cannot make it on time or want to leave early for some reason, please let me know. Persistent tardiness will be noted.

COMMUNICATION

Piazza: All questions will be fielded through Piazza. The primary benefit is that for many questions everyone can see the answer and other students can answer as well. I will endorse good student responses. Additionally, I expect you to actively participate in online discussions at Piazza. You can post public or private messages that can only be seen by the instructor. You will be signed up with your cpp email, but you may switch to another email.

Blackboard: Blackboard will be used primarily for grade dissemination.

Email: If you send email to me, please be sure to include your name and the course number in the body of the e-mail. You should also use an appropriate subject line that looks like "CS299-HW1" etc. Failure to follow these guidelines may result in delayed response. *Again, email should only be used in rare instances*, I will probably point you back to Piazza if you have a question related to course materials and/or relevant to other students in the class.

The scores you receive on the various graded tasks in the class will be weighted as follows:

30%	Final Exam (Final is comprehensive): ***Monday, June 04, 1.50 pm – 3.50 pm***	
20%	Midterm Exam: ***Wednesday, April 25 during class time***	
15%	Homework Assignments (5)	
25 %	Term Project (1)	
10%	Quizzes (2) + in-class activities	
100%	Your Total Score for the class	

Project: The project is an opportunity to tackle a more challenging data science activity. Details, requirements and submission information will be on the course website. For the project, you will work in teams of either 2, or 3 students on a problem of your choosing (or instructor provided topics) that is interesting, significant, and relevant to Data Science. The ultimate goal of your course project is to develop a new tool to tackle some interesting real-world problem. At the end of the quarter, we will hold a competition during our regular class time for your project demonstration. All members of a group will receive the same grade on group work. Therefore, it is in your interest to choose other group members (ideally, first day of the class if possible) who have the same goal in the class as you do. It is also in your interest to work together and ensure that all tasks are completed effectively. Your scores on group work may be adjusted based on your contribution (peer-evaluation).

Final Exam: The final exam is comprehensive, closed books and will be held on **Monday, June 04 from 1.50 pm to 3.50 PM**. You may bring one standard 8.5" by 11" piece of paper with any notes you deem appropriate or significant (front and back).

Midterm Exam: The midterm exam will be held on **Wednesday, April 25 during class time**. For both exams, no iPads, iPhones, Blackberries, Android phones/tablets are allowed. Standard calculators are allowed.

Quizzes: You will have 2 quizzes spread across the quarter. I will use them to gauge what topics we need to devote more time to and to train you for the questions that can appear in mid-term and final exams.

In Class Activities: Attendance in class and participation in the discussion are both important to your success in the course. As one crude measure of your participation and course preparation, we will have in class activities related to lecture topics to supplement the learning. I will ask you to bring a computing device (laptop, tablet).

Homework: We will have 5 homework assignments, each worth 3% of your overall grade.

Final course grades are based on the overall average. You are guaranteed a grade based on a 10% window (e.g., 90-100% is an A). Overall class grade (not the individual grade) windows may be increased in size if the instructor finds it appropriate. Final score in % will be rounded to the nearest whole number. Assigning + or – grades may be made at instructor's discretion.)

A: 90-100, B: 80-89, C: 70-79, D: 60-69, Fail (Grade F): 0-59

Grading correction: Bring any assignment or exam grading correction requests to the instructor within 1 week of receiving the grade, or before the end of the quarter, whichever comes first. After that, your grade will not be adjusted. If you find a mistake in grading, please let the instructor know. Your grade will not be lowered.

ATTENDANCE, MAKE-UPS AND LATE POLICIES:

All project, homework assignments, are due at the beginning of class in all required forms (e.g., paper and/or submit on blackboard) on the due date. Changes to a submission's due dates will be avoided because they are unfair to those students who have organized their time to complete the assigned work. Individual accommodations will be discussed if you have a valid medical excuse.

Project due dates will be set to give ample time for completion of the project and will not be extended save for the unexpected and unlikely major, long-lived catastrophe. Start projects early--last minute computer malfunctions will not be accepted as a reason for delaying a project due date.

Unless otherwise specified by the instructor, only the final exam will be comprehensive, covering material from the entire course. There are no makeups or rescheduling of exams unless you have a plausible reason with appropriate document or verification for absence. Rescheduling of exams must be arranged at least one week in advance. An exam/quiz missed without an acceptable excuse will be recorded as a grade of zero (0). Please also be aware that no electronic devices are allowed during the exam.

For Homework assignments, each late submission will incur a 5 points penalty per day. A missed submission without an acceptable excuse will be recorded as a grade of zero (0). No submission will be accepted after 3rd day and will be recorded as a grade of zero (0). There will be no makeup for homework assignments or class activities.

ACADEMIC OFFENSES

Scholarly dishonesty, especially plagiarism, will not be tolerated. Plagiarism is defined as "Failing to credit sources used in a work product to pass off the work as one's own. Attempting to receive credit for work performed by another, including papers obtained in whole or in part from individuals or other sources." Students found to have engaged in plagiarism will be punished severely, typically earning an automatic F in the course and being reported to the Office of Student Conduct and Integrity.

Homework Assignments Collaboration Clarification: To clarify, your homework assignment is yours alone and you are expected to complete each independently. Your solution should be written by you without the direct aid or help of anyone else. However, I believe that collaboration and team work are important for facilitating learning, so I encourage you to discuss problems and general problem approaches (but not actual solutions) with your classmates. If you do have a chat with another student about a problem, you must inform me by writing a note on your submission (e.g., Bob pointed me to the relevant section for problem 3). The basic rule is that no student should explicitly share a solution with another student (and thereby circumvent the basic learning process), but it is okay to share general approaches, directions, and so on. If you feel like you have an issue that needs clarification, feel free to contact me.

DISABILITY RESOURCES

If you have a physical or a learning disability, please talk to me privately and/or contact the Disability Resource Center (DRC) at 909-869-3333. The location is at Bldg 9-103 to coordinate course accommodations. For further information, visit the DRC website at http://www.cpp.edu/~drc/index.shtml