DSCI 560: Data Science Professional Practicum

Instructor: Young Cho, Ph.D. E-mail: youngcho@isi.edu

Course Description

This course allows students to work as teams to design and implement working systems that solve real-world problems using data science.

One of the fundamental principles of data science is the ability to live in the data. It is necessary to gain a level of immersion in the information environment to apply the diverse skill sets required to become an effective analyst and provide customers with solutions to hard problems. While students will get a sense of this paradigm in many of the courses focused on knowledge and skill enhancement throughout their degree matriculation, the goal of this course is to combine previously learned capabilities and apply them against actual data sets in a real data environment and toward solving difficult challenges for an external stakeholder.

The Data Science Professional Practicum is a capstone experience designed to expose students to the world of data analytics from the organization's perspective. Students will gain experience with tools and skills on realistic projects toward proposing, designing, implementing, and presenting their final project that solves real-world problems.

Required Textbook: None

Other Requirements/Recommendations: Experience in Linux, C programming, Virtual Environment, Web programming, Python and relevant Data Science related Python packages

Grading

- 25% Attendance and Pop Quizzes
- 10% Reading assignments (based on the scores for the summary presentation slides)
- 25% Laboratory assignments (based on the scores for all of the assignments)
- 40% Final Project

Preparation for Classes

- Students will be using a Linux-based system through the course. It is strongly recommended that the students become familiar with its navigation and use.
- Majority of the assignments will require the use of C/C++ and Python under various environments. It is recommended that students become familiar with the languages and relevant development environment.
- It is recommended that the students become familiar with cloud computing and virtual environments.

Grading Policies

- Late Policy: The score for the assignments turned in late will be deducted by 50%.
- Grade Adjustment: If you dispute any scoring of a problem on an exam or homework set, you have one week from the date that the graded paper is returned to request a change in the grade. After this time, no further alterations will be considered. All requests for a change in grade must be submitted in writing to me.
- Changes/Information: The student is responsible for all assignments, changes of tasks, announcements, lecture notes, etc. All such changes should be posted on the course website.
- Other: Per university guidelines published in SCampus, the academic integrity policy will be upheld.

Statement for Students with Disabilities:

Students requesting academic accommodations based on a disability must register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Statement on Academic Integrity
USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. The Student Guidebook contains the Student Conduct Code in Section 11.00, while the recommended sanctions are in Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/.

Lecture Contents

Lectures	Topics Covered
Lecture 1	Course Introduction
Lectures 2, 3	Data Science in Practice
Lecture 4	Ideals of Data Science
Lecture 5	Development Environment and Tools
Lecture 6	Data Sources and Destinations
Lecture 7, 8	Data Normalization, Filtering, and Extractions
Lectures 9	Data Processing and Visualization
Lecture 10, 11	Clustering and Classification
Lecture 12	Patterns, Partitions, and Discovery
Lecture 13	Knowledge Discovery
Lecture 14	Analysis with Physical Model
Lecture 15	Analysis with Machine Learning/Artificial Intelligence
Lecture 16	Collecting Data using Internet-of-Things
Lecture 17, 18	Embeddings and Their Applications
Lectures 19, 20	Large Language Models
Lecture 21, 22	Scalable Storage and Processors
Lecture 23	Hardware Accelerators for Data Analysis
Lecture 24	Final Lecture: Career Advice
Lectures 25, 26	Final Project Proposal/Practice
Lectures 27, 28	Final Project Presentations

Attendance and Quizzes
Lecture attendance is mandatory. There will be quizzes throughout the year on the topics discussed during the lecture that may not be found in the lecture notes.

Reading Assignments
There are weekly reading assignments for which summary slide presentations must be submitted. The presentation should be 3-4 pages, including the title and summary.

Laboratory Assignments
The Github commit records will be used to determine a portion of the lab grades.

1 Virtual Environment Configuration and Basic Data Science Tasks
2 Advanced Data Scraping
3 Stock Tick Data Collection and Trading Algorithms
4 Internet Data Clustering based on Semantic Analysis
5 Oil Well Data Analysis and Visualization

6 Internet-of-Things and the Cloud
7 Chatbot Design and Implementation

Final Project
The final project consists of a presentation and demonstrations. The grades will be based on the scores from all of the components.

Presentation: There will be approximately a 5-minute final slide presentation and a 2-minute Q&A session for each project group. All students will be required to participate and attend.

Project Demonstration Video: There will be weekly demonstrations of the project's progress until the project's final due date. Weekly milestones defined at the beginning of the project must be demonstrated for full credit.