

USC Viterbi

School of Engineering

DATA SCIENCE PROGRAM

DSCI 560: Data Science Professional Practicum

(4 units)

Syllabus

Term-Day-Time

Spring 2023- Monday 4:00-7:20pm

Location: [GFS118](#)

Instructor: Dr. Anna Farzindar

Email: farzinda@usc.edu

Professor's Office Hours:

Office: Physical or virtual address

Office Hours: Monday Before class

Students are advised to make appointments with the professor ahead of time in any event and be specific with the subject matter to be discussed. Students should also be prepared for their appointment by bringing all applicable materials and information.

Catalogue Description:

Student teams working on external customer data analytic challenges; project/presentation based; real client data, and implementable solutions for delivery to actual stakeholders; capstone to degree.

Expanded Course Description:

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 truly apply the diverse skill sets necessary to both become an effective analyst, and provide customers 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 real data environment, and toward solving difficult challenges for an external stakeholder.

The Data Science Professional Practicum is a capstone experience designed to allow students exposure to the world of data analytics from the perspective of the organization. Students will work with external stakeholders on a project that answers an organizational problem. Each semester, the data science program will partner with an external organization, which will pose a broad business requirement to the class. Students will team and propose projects around this requirement and be responsible for delivering a data science-based solution to the stakeholder. This product will consist of a not less than 15-page report detailing the proposal, methods utilized in the analysis, the output from the analysis, and conclusions and recommendations based on the study. Students will also be responsible for a not less than 20-minute presentation on their project to the stakeholder.

A possible example of a capstone project would be: Google volunteers to be the external partner. They obviously are one of the main collectors and brokers of information in the world. One of their areas of philanthropic interest is disease monitoring. Google might pose the question “How can the data Google collects in general could be better utilized to inform on disease, or help mitigate disease, in various areas in the world?” It would be up to the student teams to proposal a project plan, design a study, implement the analysis, and report conclusions and recommendations.

The Professional Practicum (DSCI 560) provides an opportunity to face real-world challenges and to develop your personal innovative project in cooperation with the members of your team. The combination of industry and academic experiences will provide a peek at what to expect as a data science experts and data scientist, in addition to providing project management techniques such as lean six sigma aimed at efficiency and effectiveness.

Learning Objectives

- Expected learning outcomes for students are:
- To understand how to apply the various engineering and business principles studied in Data Science curriculum toward solving an organizational challenge
- To handle difficulties associated with defining and organizing a realistic problem Statement
- To manage impediments in obtaining information and approval
- To present and sell ideas to higher-level management
- To understand the importance of the need for a continuous exchange between engineers, management and employees in solving an existing problem, given a set of constraints
- To meet aggressive deadlines in a multidisciplinary team effort
- To improve project-based presentation skills, both in-class and in company settings
- To understand the requirements and objectives of customers, how these vary, and how one must tailor a solution to the expectations of a customer
- To understand how data science are viewed in diverse domains; and how solutions and approaches will differ in various fields
- To understand how to work with individuals from diverse domain (engineering, business, etc.) to accomplish a common goal

Students are expected to have foundational knowledge in data management, machine learning, data mining, and data visualization, as well as other topics associated with the Data Science degree program. The course can be taken by students in the ISE Analytics program and/or the Marshall Business Analytics program (being created) as well as the Data Science program. Course is usually taken after other courses in the program. This class will be primarily group work, with some assigned readings, and a major project and presentation that will count for the primary grade in the course.

Assignments/Reports/Project

The major deliverable from this course is a written project report that reflects a proposed solution, and subsequent results, for an external stakeholder. At the first class meeting, the

external partner will be introduced, and a problem/requirement posed. Students will then divide into teams. All projects are team projects. All topics will be unique, and there should be high levels of collaboration on the projects with team members and the professor. Students are required to meet with the professor not less than three times during the semester. Students are required to submit three progress reports throughout the course. There will be a midterm.

There will be three inputs to students' final grades: 1) instructor evaluation; 2) peer review; and 3) evaluation by external stakeholder.

NO PROJECTS WILL BE ACCEPTED LATE, AND STUDENTS MUST PRESENT THEIR FINDINGS ON THE DAY SELECTED DURING THE SEMESTER.

Class Communication:

USC D2L, Piazza and Blackboard at USC will be used for class communication.

Grading Schema:

Research paper presentation	5%
Assignments:	15%
Quizzes:	15%
Midterm:	15%
Monthly Reports and Instructor Meetings:	5%
Midterm Sponsor review and report	10%
Final Sponsor review for Course Project Report and Demo Presentation:	25%
Final Report and Findings:	10%
Total	100%

Grades will range from A through F. The following is the breakdown for grading:

94 - 100 = A 74 - 76 = C
90 - 93 = A- 70 - 73 = C-
87 - 89 = B+ 67 - 69 = D+
84 - 86 = B 64 - 66 = D
80 - 83 = B- 60 - 63 = D-
77 - 79 = C+ Below 60 is an F

Class Structure & Schedule:

Class sequence, dates, topics and guest speakers are subject to change as the semester proceeds. Any revisions will be noted and announced in class in advance.

Week	Topic	Readings/Notes	Homework
Week 1 01/09	Orientation, Requirement Outline, Developing Project Ideas, Fundamentals of Project Management	Overview of project management in Data Science Post-class, teams will be formed	Ref: Fundamentals of Project Management (Worksmart) : Joseph Heagney
Week 2 01/16	No Class Martin Luther King Jr. Day		Team Formation
Week 3 01/23	External Partner presentation	The external sponsor will present and challenge the class	
Week 4 01/30	Lean Six sigma, Define, Deliverables, DMAIC ¹ , Analytics in the “real world”	DMAIC (Define) Proposals briefs (2-3 pages outlining “big idea”) will be completed	Proposal, project Charter (assignment 1)
Week 5 02/06	Value stream mapping , Quality, Variability Root Cause Analysis, Teams, People, Lean Engineering	DMAIC (Measure and Analyze)	SIPOC and Process Mapping (assignment 2)
Week 6 02/13	Understanding the Analytics Product in Business, Improve	DMAIC (Improve and Control)	VOC and Root Cause Analysis (assignment 3)
Week 7 02/20	No Class Presidents’ Day		

¹ DMAIC refers to a data-driven quality strategy for improving processes, and is an integral part of the company's Six Sigma Quality Initiative. DMAIC is an acronym for five interconnected phases: Define, Measure, Analyze, Improve, and Control.

Week 8 02/27	Agile Methodology and Mid-Term Review	Agile Data Science Methodologies Half of class dedicated to mid-term review	Monthly report (assignment 4)
Week 9 03/06	Mid-Term Exam		
Week 10 03/13	Spring Break		
Week 11 03/20	Team Presentation on Monthly report	Review progress, answer questions, Students will provide the presentations.	Team Presentation on Monthly report
Week 12 03/27	Technical presentation		Second Monthly Report Due
Week 13 04/03	Invited speaker or discussion	Presentation on topics of Data analytics, Management, Leadership and entrepreneurship.	
Week 14 04/10	Technical presentation		
Week 15 04/17	Dress Rehearsal	For those who desire, they will be able to go over their reports and presentations with the instructor, and receive feedback before presenting to the customer	Team presentation
Week 16 04/24	Demo Day Final report	Last class- Student presentations, Demo Final project submission (May 01, 2023)	Final Project close out

Books and Readings:

All books, papers or reports will be available to students in one of three ways: 1) in the USC bookstore; 2) via a CD that the instructor will provide at the beginning of class; and/or 3) via the web.

J. Heagney “Fundamentals of Project Management (Worksmart)” Fourth Edition Edition, 2011.

S. Shaffie, S. Shahbazi, “The McGraw-Hill 36-Hour Course: Lean Six Sigma (McGraw-Hill 36-Hour Courses) Paperback – April 30, 2012.

Russell Journey, “Agile Data Science 2.0”, Publisher(s): O'Reilly Media, Inc. 2017.

C. E. Knoeppel, “Installing Efficiency Methods,” The Engineering Magazine Company, republished by Google Books, 1917.

Assigned Reading and Media List:

A. Farzindar and D. Inkpen "Natural language processing for social media 3rd. Edition" second edition, Morgan & Claypool Publishers, 2020.

M. George, J. Maxey, D. Rowlands, M. Price “The Lean Six Sigma Pocket Toolkit: A Quick Reference Guide to 100 Tools for Improving Quality and Speed and Complexity. January 1, 2004
J. Lean Six Sigma for Supply Chain Management, Second Edition: The 10-Step Solution Process 2nd Edition

McKinsey Global Institute, “The challenge – and opportunity – of ‘big data,’” in McKinsey Quarterly, May 2011.

Richard T. Kouzes, Gordon A. Anderson, Stephen T. Elbert, Ian Gorton, and Deborah K. Gracio, “The Changing Paradigm of Data-Intensive Computing,” in IEEE Computer, January 2009.

David Bollier, “The Promise and Peril of Big Data,” through The Aspen Institute Communications and Society Program, 2010.

James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, “Big data: The next frontier for innovation, competition, and productivity,” published through McKinsey Global Institute, May 2011.

Viktor Mayer-Schonberger, Kenneth Cukier, “Big Data: A revolution that will transform how we live, work and think,” Emaon Dolan/Houghton Mifflin Harcourt, March 2013.

Eric Siegel, “Predictive Analytics: The power to predict who will click, buy, lie, or die,” Wiley, February 2013.

Lev Manovich, "Trending: the Promise and the Challenges of Big Social Data," Debates in the Digital Humanities, July 2011.