Syllabus for DS 6015: Data Science Capstone Project

Last updated: May 16, 2024

Course Description

The purpose of the capstone experience is to provide students with an understanding of data science as it may be practiced in the context of real-world problems. It allows the student to collaborate as a member of a small team to pull together knowledge from the diverse areas of data science and integrate them in an effective and professional manner. The focus is on addressing an actual client's need by building a data product that can be shared with the client. The course covers the steps of the data science pipeline, which can be summarized in four phases: (1) framing the question, problem or value proposition, (2) acquiring, transforming, and exploring the data, (3) analyzing and modeling the data, and interpreting the results, and (4) packaging, communicating, and sharing the results. During the capstone course, students will be introduced to both technical and professional challenges and will be guided by basic project management protocols in meeting them. The course sequence culminates in the production of a publishable paper and a presentation of results to the client.

Purpose of this Document

This document provides general guidelines for each individual capstone section.

Learning Objectives

1. Gain an understanding of data science as it may be practiced in the context of real-world problems

2. Collaborate as a member of a team to pull together knowledge from the diverse areas of data science and integrate them in an effective and professional manner

3. Address an actual client's need by building a data product that can be shared with the client

4. Demonstrate oral and written communication skills through a formal paper and presentation of project outcomes

Course Format

Each capstone course will consist of a faculty mentor and their assigned projects. Each week, the mentor and students from these projects will meet at the time and place assigned by SIS. On paper, these classes will be two hours long; in practice, the mentor may decide to meet for a shorter period and delegate the rest of the time to individual project meetings. In addition, student teams may wish to meet outside of class.

Activities

The semester will be structured around the following goals:

- Re-establish communication plan with your sponsor if you have been out of communication
- Establish a meeting schedule for your team based on the semester's course schedule, etc.
- Complete the process of establishing your data and preparing it for the modeling and/or deep exploration phase. This will include various forms of data reduction, feature engineering, and data shaping to meet the computational requirements of your models and tools.
- Prepare intermediate progress reports **to be shared with sponsor and faculty mentor**
- Develop and test one or more models, or other data products, that meet your project requirements and expectations; revise these expectations if necessary

Deliverables

- Package and share all data products and research artifacts with clients and other stakeholders
- Produce a six-page publishable paper on your work summarizing your team's research and findings
- Produce a slide deck to be presented by all students at the Capstone Presentation event
- Submit the link to your GitHub repository containing all finalized, documented code. Check with the sponsor if this repo can be public; otherwise it can be kept private to stakeholders.

Information on all assignments and deliverables will be described in the course GitHub repo. Note that aside from these assignments (see Grading below for a list), this course is not organized like a standard academic course. In place of a strictly defined schedule of inputs (lectures, etc.) and outputs (homeworks, quizzes, etc.), you will complete your project work according to the pace dictated by the dynamics of your project.

Participation

You are expected to attend all class meetings. If you have a good reason to miss a meeting, notify your mentor and teammates ahead of time, preferably before the work week starts.

Communication

Teams will be used as the central channel for course collaboration. Please use Teams to post questions regarding project problems, objective identification, data collection and/or compilation, data processing, data structure and management, and exploratory data analysis. We'll use this for current updates for dates/times…etc.

Grading

Participation	10%
Progress Report I	15%
Progress Report II	15%
Paper	35%
Presentation	15%
Code Repository	10%

Course GitHub Repo

The course has a GitHub repo saved here: https://github.com/UVADS/ds6015/

University of Virginia Honor System

All work should be pledged in the spirit of the Honor System at the University of Virginia. The instructor will indicate which assignments and activities are to be done individually and which permit collaboration. The following pledge should be written out at the end of all quizzes, examinations, individual assignments and papers: "I pledge that I have neither given nor received help on this examination (quiz, assignment, etc.)." The pledge must be signed by the student. For more information, visit www.virginia.edu/honor.

Special Needs

It is our goal to create a learning experience that is as accessible as possible. If you anticipate any issues related to the format, materials, or requirements of this course, please meet with your faculty mentor outside of class to explore potential options. Students with disabilities may also wish to work with the Student Disability Access Center to discuss a range of options to removing barriers in this course, including official accommodations. Please visit their website for information on this process and to apply for services online: sdac.studenthealth.virginia.edu. If you have already been approved for accommodations through SDAC, please share your accommodation letter with your faculty mentor.