

MATHEMATICS DEPARTMENT, UNIVERSITY OF MASSACHUSETTS DARTMOUTH  
**Data Science Capstone Project I/II**  
**DSC (498) 499 — Spring 2021**  
**Course Information and Syllabus**  
Updated January 17, 2021

---

<b>Instructor:</b>	Scott Field
<b>Instructor email:</b>	sfield@umassd.edu
<b>Instructor phone:</b>	508-999-8318
<b>Office:</b>	LARTS 394C
<b>Office hours:</b>	Tues 2:15pm - 3:15pm, Tues 5:00pm - 6:00pm, Wed 12:00pm - 1:00pm, Thurs 5:00pm - 6:00pm
<b>Office hours location:</b>	Zoom (see MyCourses for the link)
<b>Class time and location:</b>	Wednesday, 3:00pm-5:00pm, Zoom (see MyCourses for the link)

**Course Websites** Our course website will be hosted on MyCourses. For communication and coordination we will use Slack. Each student project will be hosted on bitbucket, github, gitlab, or any other git-based server of your choosing.

**Prerequisites** As this course is required for all data science seniors, I will assume you have taken the standard sequence of data science courses covering freshman, sophomore, and junior years.

**Textbooks and other resources** There will be no required textbook for this class. The material will be based on lectures, discussions, and online resources. All materials will be passed out in class or by email.

**Course description** A highlight of the data science degree is the capstone project, where, over two semesters, you will have an opportunity to tackle a real-world project of your choosing. This project-based capstone course provides students with a framework in which students gain both understanding and insight into the application of knowledge discovery tools and principles on data within the student's area of interest.

Each student is expected to design their own project, including the deliverables, goals, and analysis tools. Plans must state how they will utilize core data science skills (for example, statistics, programming, data mining, etc.) throughout the project. In keeping with the Capstone goal of engagement, successful projects must engage with people or groups outside of the UMassD campus. This could come in many forms including email lists, interacting with public github projects, a blog, or the project itself might include external collaborators.

Projects should make use of the data science “workflow”: (1) data collection such as data wrangling, cleaning, and sampling to get a suitable data set, (2) data management such as accessing data quickly and reliably, (3) exploratory data analysis such as generating hypotheses and building intuition, (4) prediction or statistical learning such as model building, and (5) communication such as summarizing results through visualization, stories, and interpretable summaries.

This course is intended for data science majors only.

**Learning Outcomes** By the end of the course, students will be able to

- use data science skills and techniques acquired over the past three years to complete a complex, real-world project that they design
- gain insights, build models, and answer questions with application-specific datasets
- communicate their findings through oral and written formats
- leverage their capstone experience to continue their careers after graduation

**Year-long timeline** By the end of this first semester (DSC 498 – a 3 credit class), students are expected to have made significant progress on their project. This includes writing a clear, thoughtful project plan as well as having preliminary results. Students will give regular updates on their project, including a short final report summarizing the work accomplished in the first semester.

During the second semester (DSC 499 – a 2 credit class), students will complete their project, write a final report, and give a public talk open to the entire UMassD community.

**Data Science Capstone Day** Each year, data science seniors (you!) and masters students embark on a capstone project. At the end of the semester, all students will present their work to the public. This is your final presentation. The date, time, and format will be described later in the semester.

**Classroom time** Classroom time will be used to discuss and work on projects. We may also discuss best practices, tips for writing scientific reports, giving presentations, and other skills necessary to carry out a successful project. While attendance does not directly factor into your grade, you are expected to give weekly project updates and these will be graded.

Most weeks will also feature a short mini-tutorial led by a student. These are referred to as topic talks below.

**Outside-classroom time** Please plan to spend about 5 hours per week on your project. To discuss technical issues, you are encouraged to stop by my office, send me an email, or find me on Slack.

**Grading Policy** Your grade in DSC 499 will be determined by tracking and documenting all code on github, bitbucket, or gitlab such that the projects are self-contained and usable by others (15%), beginning-of-the-semester report summarizing where your project stands and what needs to be accomplished over the Spring (10%), final end-of-year talk open to the entire campus (25%), final end-of-year report summarizing your results (25%), weekly project updates (15%), special topic talks (10%). When computing your grade, please keep in mind how rounding will be carried out: after a grade computation I will apply standard rounding rules (e.g.  $80.5 \rightarrow 81$  and  $80.4999999 \rightarrow 80$ ).

A	93% - 100%	B-	80% - 82%	D+	67% - 69%
A-	90% - 92%	C+	77% - 79%	D	63% - 66%
B+	87% - 89%	C	73% - 76%	D-	60% - 62%
B	83% - 86%	C-	70% - 72%	F	0% - 59%

A grade of an A+ will be given for exceptional performance.

**Topic talks** Each student will lead an approximately 20 minute “seminar” on topics of interest to them. The format and topic will be up to you to choose. You could run a hands-on tutorial on Tableau or tensorflow, for example. You could give a lecture on important probabilistic distributions and how you are using them in your project. You could provide an overview of tools and tips on how to use github to get a job. The important thing is that the topic is of wide interest to the class and relates to data science tools, skills, or post-graduation goals. You should also choose a topic that you are an expert in – this project is not meant to learn something new, but rather share your knowledge with the class. For example, if you have been using the Selenium package to webscrape, you could make this your topic.

You should propose a topic and format for approval before starting work on it.

**Weekly project updates** Each class you will give a short update (5 to 10 minutes) on what you’ve been up to over the past week. Do not summarize your project or what was done in past weeks. Your updates can be entirely verbal, you can make slides, you can screenshare your code, or anything else so long as it makes it clear what work was done. This is also the time to ask for feedback and help from the class and is meant to be collaborative. Please also describe what you will be working on over the next week. Be specific! Good: “I will finish webscraping historical datasets from 2018, and will manually compare the scraped data to whats on the website. Once this has been checked for correctness, I’ll have a CSV file to carry out exploratory analysis.” Bad: “I will get data for my project.”

Weekly-update grades: 0 (no update), 1 (technically an update was given, but it wasn’t clear what progress has been made or what will be done next), 2 (update was OK, but lacked in either progress or goals), 3 (nice update – there was good progress and what’s next is clear). These will be converted to a grade; roughly speaking 0=F, 1=C, 2=B, 3=A.

**Making your results reproducible (code hosting)** Data science is a science. Science, and therefore your project, should be quantitative, state testable conclusions, and be reproducible by other researchers. Your project must pass the reproducible test: if you gave another student (or the instructor) your report and access to your code, can they reproduce all of the results from that report?

By the end of the Spring, please make sure all of your code and data are accessible to the instructor along with any extra information needed to reproduce your work. Ideally this will be hosted on github, bitbucket, gitlab, or a similar code-hosting service. If you are looking to get a technical job after graduating, you should attempt to follow git best practices for developing and sharing code.

**Attendance** Attendance at class meetings is strongly encouraged. This is the best time to tackle any obstacles, get clarifications, and otherwise make rapid progress. Additionally, you are required to give weekly updates that will be graded. If you are absent from class without reason, points will be deducted from your weekly update grade.

If you believe you will be absent for an extended period of time due to illness or other emergency, notify me immediately to arrange make-up of any assignments that will be missed. No assignments can be “made up” after an absence, except in extenuating circumstances.

**Incomplete grades** An incomplete grade will be given only in exceptional circumstances at my discretion. You must be passing the class at the time of the request for an incomplete grade, or be sufficiently close to passing. If the work is not completed within one year of the recording of the incomplete grade, the grade will become an F(I). The incomplete policy for this course is that at least 70% of the course must be already completed and an exceptional circumstance (i.e. medical issue) must exist. If you feel you require an incomplete for an exceptional reason, you need to email me and state your reasons for the incomplete in writing. We will then schedule a meeting and decide how to proceed.

**Students with disabilities** If you have a documented disability and require accommodations to obtain equal access in this course, please meet with me at the beginning of the semester and provide the appropriate paperwork from the Center for Access and Success. The necessary paperwork is obtained when you bring proper documentation to the Center for Access and Success, which is located in Woodland Common, Room 111. Tel: 508-999-8711.

### Important dates

Tuesday, January 19, 2021	Classes begin
Tuesday, January 26, 2021	Last day to add, drop, or audit
Wednesday, March 10, 2021	Spring recess ( <b>no class</b> )
Friday, April 9, 2021	Last day to withdraw from a class
Wednesday, April 21, 2021	Follow Monday’s class schedule ( <b>no class</b> )
Wednesday, April 28, 2021	Last class
TBD	DSC Capstone Day
Wednesday, May 12, 2021	Last Day to Select or Revoke Pass/No Credit