DTSC 691 Project Handbook

Table of Contents

Introduction

DTSC 691 is part two of the capstone in the Masters in Data Science. Students will complete a capstone project integrating their learning across courses. Students will complete a project proposal, including their data source, acquisition, cleaning, analysis, and presentation intentions. Can be taken jointly with DTSC 690.

Handbook Purpose

This project handbook is meant to provide students, mentors, and staff, with clear guidelines for the execution of the project. It is critical that students thoroughly read and understand all aspects of the handbook.

Student Responsibilities

Students are responsible for all policies and procedures both in the Handbook and the course syllabus.

Mentor Responsibilities

The primary responsibilities of mentors are (1) assist students in the formation and agreement of the project proposal and (2) guidance for students in the execution of the project.

Grade Scale

This course is graded pass/fail per the rubrics below.

Project overview

This section is meant to serve as a broad overview of the project process. Subsequent sections will give further detail.

Initial project consideration

Prior to the start of DTSC 691, students should spend a good amount of time considering the type of project they are interested in. While this is largely flexible, there are default project options students can consider:

- [Default databases project guidelines](#)
- [Data science for education project guidelines](#)
- [Default machine learning project](#)
- Custom machine learning project

It is strongly recommended that students thoroughly consider all options before the term begins. If students are interested in receiving feedback about a potential DTSC 691 project prior to the course beginning, please email dtsc_ga_691@eastern.edu.

Before-term inquiry

- Prior to and/or at the beginning of the term, students will be asked what type of project they are interested in in order to be matched with an appropriate mentor. Although students do not need to know the details of their project, we require students to indicate which general project they are interested in so they can be connected with a mentor.
- Students will be notified of their mentor assignments by, at the latest, the day after the add/drop deadline. Anytime prior to that, if you have not yet received notification of your mentor assignment, please be patient and in the meantime reach out to dtsc_ga_691@eastern.edu to receive project feedback. However, our goal is to have mentor assignments completed by the end of the day, Friday of the first week of class. Please fill out the Google form sent by your course instructor as soon as possible to assist us in meeting our goal. Students who have not filled out the google form by the add/drop deadline will be assigned to a mentor randomly.
- If a student will require a non-disclosure agreement for the completion of their project, they need to inform the IOR as soon as possible.

Optional pre-proposal submission

- You may submit to your mentor a brief, **<2** page document outlining your proposal for the purpose of general comments on the direction of the project. If a pre-proposal is >2 pages we will ask you to resubmit a shorter version. Please submit a Google Doc for this, clearly indicating this is your pre-proposal (as opposed to your proposal). This may be submitted up to 7 days into the semester. Share this google doc with your instructor and mentor as editors.
- Feedback will address whether the topic is appropriate, depth of the project, and more general feedback.

Course meeting

- We will have a class meeting in the first week of the course. An email will be sent prior to the beginning of the term with details, and additional information about time and Zoom location can be found in the syllabus.
- If you can not attend you will be able to view the recording following the meeting.

Project Proposal submission

- Project proposals are due two weeks after the beginning of the term
  - If the term starts Monday, 1/10, the proposal is due Monday, 1/24 at 11:59PM ET.

- ○ *However, we **strongly prefer and recommend** that you turn in your proposal one week after the beginning of the term.*
- Submit your project proposal through the [Google Doc](#) (using the appropriate topic template) as described in the 'Project proposal' section below. Clearly indicate that this is your proposal (as opposed to your pre-proposal). *You **must** upload a .pdf of the project proposal to Brightspace.*
- Expect feedback on your proposal within 48-72 hours after submission.
- You and your mentor will exchange feedback and come to an agreement, as outlined in the 'Project proposal' section below.

Project submission

- Submission of the project is due at 11:59PM ET on the final date of the term.
- Submit your project submission through corresponding assignment in BrightSpace and submit this [Google Doc](#) (using the appropriate topic template).
- You must also submit a video of the submission, as outlined in the 'Project walkthrough' section below.
- Expect feedback on your project within 1 week of submission.

Project proposal

Project proposal overview

Proposals serve as an agreement between mentor and student about the scope of the project. This assures that both you and the mentor agree to the expectations and requirements of the project and aren't surprised with any grade outcome. Once proposals are accepted, they form a contract such that if a student completes the outlined project in good faith, they will pass.

Project proposal details

Within the first two weeks of the course you will submit a project proposal (the sooner the better). Once the proposal has been received, mentors will review the proposals and either approve it or require revisions. If it is approved you will have the remainder of the term to complete your project. If it is not, it is expected that revisions will be completed in a timely manner, as indicated in the feedback. **If your proposal is not "approved", you are much more likely to fail the project.** Once the proposal has been approved, the mentor will email the student stating clearly, "Your proposal has been officially approved." and will cc the Instructor on the email.

You will submit your project proposal as a Google Doc link so we can work collaboratively. [You must use this template](#) (using the template for your corresponding topic). Once in the template, go to File → Make a copy and save your version. You must save this version using the format "<Your_Last_Name Project Proposal>" for the file name; for example,

"Morabito_Project_Proposal". To share, go to Share in the upper-right corner and add both your mentor and the course instructor as editors. Then, go to Brightspace and submit the link and a .pdf of the submission to the Project Proposal Assignment. **You must do both to get credit for the proposal**.

Optionally, students can submit a video that walks through their proposal. If you choose to record a video you may either share it with your mentor in a way they indicate they can receive it - perhaps through Google Drive or Brightspace.

Feedback to the proposal will be written in comments directly in the Google Doc.   Once the proposal has been submitted, students are expected to address all comments. It is critical that you thoroughly review all comments because you will be held responsible for them. Students should address the comments within the Google Doc itself. Do not "resolve" the comments, but rather, leave all comments from both students and mentors for documentation purposes.

Because your proposal is only a proposal, there is no guarantee that your final project will be successful; you could earn a pass on your proposal, but fail to complete what is necessary for a good grade. Alternatively, you could turn in a poor proposal, address all of our comments, but still not have a successful project. This is why it is critical that you communicate with your primary instructor throughout the process.

Feedback throughout the project will address issues such as depth and breadth of the work and topic, suggestions for improvement and additions, and other big-picture topics. However, mentors will not directly assist with issues such as cleaning data, debugging, model training, and other specifics of your project.

The project proposal should include the following sections:
- Goals of the project
- Data description
- Software
- Analysis plan
- Project communication

Details of each follow.

Goals of the project

500 word minimum overview of what the project hopes to accomplish. This should be at an introductory level; it should be able to be understood by anyone in our program and save complex details for later. Include key components such as: motivation, research question statement, hypotheses, materials and methods, deliverables, implementation strategy, etc. Feel free to supplement this textual overview with images or figures to aid in your proposal.

Data description

A detailed description of the data to be used. The most common problem students have in their proposal is in not providing enough detail about their data. You might feel like you are oversharing, but it is necessary for your mentor for your instructor to understand if your project, and especially the analyses, are adequate. Some suggestions:

- In all cases, describe the data in as much depth as you can
  - How many observations are there? How many variables?
    - Are you planning on subsetting the data in some way?
  - Include anything from descriptive statistics to associated manuals
  - What specific variables will be used?
- If the data have been previously collected, who collected it?
  - How were the data collected? When?
- Provide instructors a way to access data. If there are privacy concerns please contact us ASAP.
  - If you plan on collecting data, present a detailed plan of how the data will be acquired

Please take special note in regards to datasets that have associated notebooks (or other analyses) with them, such as those on Kaggle or GitHub. It is required that you disclose this information to your mentor by the project proposal at the latest. No work you find from these sites may be used in your analyses without your specific citation and permission from your mentor. If you do not cite the source you can face disciplinary action ranging from failing the course to dismissal from the program.

Software

What software will you use for your project? What will the software be used for? This will likely be only one paragraph or so.

Analysis plan

Include a detailed description of how you will complete your project. This will vary depending on the type of project you wish to complete, but should include enough detail that professors can understand the overall project architecture and implementation strategy.

In addition to the analysis description, you must also include weekly goals. We will not hold you to this timeline, but we want you to demonstrate you have given real thought to how you will complete a project of this size in a 7-week period.

Project communication

You will need to communicate the results of your project in some way. For some, a detailed presentation walking through a Jupyter Notebook could suffice. For others, a slideshow detailing their database construction. And for others still, an APA-style paper could be the best

way to communicate results. You should detail how your results will be communicated and the reasons for your choice.

Project

Submission Checklist

By the end of the term, you must submit the following to the "Project Submission" assignment in BrightSpace. **All of these pieces must be submitted to receive credit for your project:**

- **A link to a shared Google Drive folder with all your project files** *(change share settings to be shared with anyone at Eastern University with commenting privileges).* The folder should be organized and easy to navigate. Included in this folder (at minimum) should be:
  - **Project Submission - Google Doc** *(a link to this document – which is essentially your detailed and finalized proposal); see below for more details*
  - **Project Materials:** Any/all videos, notebooks, and other files related to your project; see more details below
  - **Video - Project Walkthrough**; *see more details below*
- **Single pdf of ALL project files merged together** *(including the Project Submission, and any other materials such as code, data description, Jupyter notebooks, html files, .sql files, other files, etc.) Please do not include your data as this will make the submission file too large.*

Project Submission - Google Doc

The project submission will give a broad overview of your project and will relate heavily to your project proposal. The purpose of the submission is for you to provide us with a clear understanding of each portion of your project.

You will submit your project submission as a Google Doc link (included in your shared Google Drive folder) so we can work collaboratively. You should use the project submission template (using the appropriate topic template). Once in the template, go to File → Make a copy and save your version.

The project submission Google Doc has four sections from the project proposal. Some sections might not change from the proposal to the submission, but none should be shorter; if there are changes they must be detailed. Please make sure to add as much detail as necessary for us to understand what your project entailed; if we do not know why changes were made we cannot adequately assess your project, and if we are not told what your project entailed we cannot give you proper credit.

Project Materials

All materials, such as code and data description, must be provided in some fashion, but the specifics will vary project to project. Please make sure that all aspects of your project work properly; if we cannot run the code or access data we can not adequately assess your project.

Materials should be submitted via Brightspace within the shared Google Drive folder; the link of which you will provide in the Project Submission assignment.

All materials (including your Project Submission - Google Doc) must also be merged and included in the single pdf turned into the Project Submission assignment in BrightSpace. This allows our plagiarism checking software to accurately assess all aspects of your project including your code.

Video - Project Walkthrough

You must submit a 30-minute (or less) video walkthrough of your project to help guide us through your work. The video may be no longer than 30 minutes; for each additional minute you will lose 10% of your grade. The video should be thorough and professional.

It is your responsibility to effectively manage your presentation time. Avoid excessive detail about every aspect of your project. Rather, you should aim to tell a well-constructed story about your project - paying attention to both the big picture and the relevant details.

Your presentation should be recorded via Zoom (or similar video recording software) and included in the Google Drive folder you submit to the Project Submission assignment in BrightSpace..

- [This video](#) will walk you through how to record in Zoom

The walkthrough video should be different than videos you might to fulfill the Project Communication portion of the proposal

Database Project Proposal Rubric

	Pass	Pass	Fail	Fail	Fail
Background and problem objective	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
ER diagram	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Software	Clear and concise description of what software will be used and how.	Incomplete description of software, but can largely be understood	Substantial lack of clarity surrounding software	Little to no description of software or unprofessionally done.	No section present
Analysis plan	Complete analysis plan and weekly goals. Reader can clearly understand what analyses will be conducted.	Incomplete plan and/or goals, but can be largely understood	Substantial lack of clarity surrounding analyses	Little to no plan on, unprofessional, or needing substantial clarification.	No section present
Presentation Plan	Complete and clear description of how project will be presented	Incomplete plan but can be largely understood	Substantial lack of clarity	Little to no plan, unprofessional, or needing clarification	No section present

Data Science for Education Project Proposal Rubric

	Pass	Pass	Fail	Fail	Fail
Introduction (Description of Topic)	Complete introduction to the project. Clear learning objectives presented. Easy to read and understand	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description	Description of goals minimal, unprofessional, or needing substantial clarification	No section present
Summary of Modules	Complete description of the modules, including videos and materials in each one.	Sufficient depth, but lacking in some details or not presented clearly enough	Substantial lack of clarity surrounding modules; did not fulfill requirements	Description of modules minimal, unprofessional, or needing substantial clarification	No section present
Data and Software Description	Clear and concise description of what data and software will be used and how	Incomplete description of data and software, but can largely be understood	Substantial lack of clarity surrounding data and software	Little to no description of data and software or unprofessionally done	No section present
Assignments and Quizzes	Brief description of all assessment materials. Relates each to learning objectives	Incomplete description, but mostly understandable	Description present but minimal	Description unprofessional or needing substantial clarification	No section present

Machine Learning Project Proposal Rubric

	Pass	Pass	Fail	Fail	Fail
Goals of the project	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
Data description	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Software	Clear and concise description of what software will be used and how.	Incomplete description of software, but can largely be understood	Substantial lack of clarity surrounding software	Little to no description of software or unprofessionally done.	No section present
Analysis plan and model specifications	Complete analysis plan and weekly goals. Reader can clearly understand what analyses will be conducted.	Incomplete plan and/or goals, but can be largely understood	Substantial lack of clarity surrounding analyses	Little to no plan on, unprofessional, or needing substantial clarification.	No section present
Project communication	Complete and clear description of how project will be communicated	Incomplete plan but can be largely understood	Substantial lack of clarity	Little to no plan, unprofessional, or needing clarification	No section present

Custom Project Proposal Rubric

	Pass	Pass	Fail	Fail	Fail
Goals of the project	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
Data description	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Software	Clear and concise description of what software will be used and how.	Incomplete description of software, but can largely be understood	Substantial lack of clarity surrounding software	Little to no description of software or unprofessionally done.	No section present
Analysis plan	Complete analysis plan and weekly goals. Reader can clearly understand what analyses will be conducted.	Incomplete plan and/or goals, but can be largely understood	Substantial lack of clarity surrounding analyses	Little to no plan on, unprofessional, or needing substantial clarification.	No section present
Project communication	Complete and clear description of how project will be communicated	Incomplete plan but can be largely understood	Substantial lack of clarity	Little to no plan, unprofessional, or needing clarification	No section present

Database Project Submission Rubric

	5	4	3	2	0
Background and problem objective	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
ER diagram	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Analyses	Complete, coherent, professional analyses.	Analyses incomplete or unclear	Substantial lack of clarity surrounding analyses	Unprofessional analyses	No section present
Presentation	Complete, coherent, and professional presentation.	Sufficient presentation but lacks details, preparedness, or clarity.	Substantial lack of clarity, details, or preparedness.	Unprofessional presentation	No section present

A combined score of 15 is required to pass

Data Science for Education Project Submission Rubric

	5	4	3	2	0
Lecture Materials	Lecture materials complete and professionally created	Lecture materials complete but not ready for presentation	Lecture materials need substantial revision	Little or unprofessionally done materials	No Lecture Materials present
Lecture Scripts/Notes	Scripts and notes completed professionally and with sufficient depth	Scripts and notes need editing or revision	Substantial lack of clarity in lectures and notes	Lectures and notes unprofessionally completed	No Lecture Scripts/Notes present
Quizzes and Assignments	Quizzes and assignments thoughtful, clear, and ready for use	Quizzes and assignments lack clarity or need revision	Quizzes and assignments need substantial overhaul	Unprofessional or minimal quizzes and assignments	No Quizzes or Assignments present

A combined score of 12 is required to pass

Machine Learning Project Submission Rubric

	5	4	3	2	0
Goals of the project	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
Data description	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Model Specifications	Complete, coherent, professional analyses.	Analyses incomplete or unclear	Substantial lack of clarity surrounding analyses	Unprofessional analyses	No section present
Presentation	Complete, coherent, and professional presentation.	Sufficient presentation but lacks details, preparedness, or clarity.	Substantial lack of clarity, details, or preparedness.	Unprofessional presentation	No section present

A combined score of 15 is required to pass

Custom Project Submission Rubric

	5	4	3	2	0
Goals of the project	Complete introduction to the project. Fulfills length requirement. Easy to read and understand.	Incomplete introduction, but project can be mostly understood	Introduction present but is an incomplete description.	Description of goals minimal, unprofessional, or needing substantial clarification.	No section present
Data description	Complete description of the data, fully answering questions posed in the syllabus. Reader can easily understand all parts	Sufficient depth, but lacking in some details or not presented clearly enough.	Substantial lack of clarity surrounding data; did not fulfill requirements.	Description of data minimal, unprofessional, or needing substantial clarification.	No section present
Analyses	Complete, coherent, professional analyses.	Analyses incomplete or unclear	Substantial lack of clarity surrounding analyses	Unprofessional analyses	No section present
Presentation	Complete, coherent, and professional presentation.	Sufficient presentation but lacks details, preparedness, or clarity.	Substantial lack of clarity, details, or preparedness.	Unprofessional presentation	No section present

A combined score of 15 is required to pass