INFO 490MH2/427 Final Project Outline and Guidelines

A large part of this class is dedicated to having you complete a project that allows you to explore a data science technique. This document helps outline some of the guidelines and grading policies we will use.

The main idea is to get you thinking and researching a problem to investigate, analyze and report on.

An important component of this project is for you to learn and research the machine learning algorithm (or statistical analysis). Most likely, we will not be able to cover the technique that's needed -- due to time constraints.

Overall Grading

Overall Grading:

- 10% Project Milestones
- 40% Project Notebook
- 20% Project Video
- 30% Peer Reviews (3-4)

Project Milestones

This is simply uploading to Moodle your project milestones. Specific instructions will come later. But it will involve uploading the same project notebook to Moodle each time. A snapshot of the notebook will be kept at each milestone.

Project Notebook submission

You will make a final Moodle submission of the project notebook that will contain the full story of your project. Your project is considered completed. It will be graded by both the staff and peers in the class. The grading rubric for the notebook is given below.

Project Video submission

You will also produce a video (e.g. screen cast, slide show) of you presenting your work. You do not have to be on-camera. Voice only is fine (with visuals). The video will also be graded by peers and staff graded (see rubric).

- You can use any video hosting service (e.g. Google Drive works).
- You can use your phone, laptop, etc.
- You must include the link to the video and include it in your notebook.
- The project template has a placeholder for your video.
- Your presentation should NOT be you simply running your notebook.

Peer Grading

Each of you will receive a 3-4 other projects (you can do more if you want). There will be a simple grading rubric for you to use for grading the notebook and the accompanying video. The grading must

be done in a professional manner and submitted on time. The reviews will also be reviewed by the staff for completeness and fairness.

Milestones

Milestones:

Week 8: Release of initial guidelines, rubrics, etc (this document) and release of project template file. The template file will be your project notebook – you will keep updating this as you keep working on your project and building it. At every milestone, you will submit an updated version of your project notebook with that milestone's requirements.

Milestone	What	Due Date End of week
M0	Submit ID of project notebook	8
M1	Submit identification of problem and accompany dataset(s) You will get 3 (weeks 8,9,10) weeks of research for this	10
M2	Submit analysis selected; Data cleaned and ready	11
M3	Weeks 12, 13, 14 you just work!! And Submit notebook with video link You will also receive URLs of peer notebooks to be reviewed	15
M4	Submit peer reviews (3-4)	16*

*may depend on finals schedule

MS0: Submit the ID of your project notebook

- The specific details are given in the Lesson section.
    - Copy the project template (given in the Lesson component) to your Google drive
- Note that the URL you share will be made available to others in the class for the peer reviews.
- Do not include any personal information in your notebook.
- If the staff needs to make updates to the template, we will announce them on Piazza

The template only contains the general format of the outline and a video placeholder. You will need to add additional text and code cells as necessary.

MS1: Project description and dataset

Identification of problem/challenge/question

This milestone will be completed by submitting your notebook to Moodle.

- write the introduction to your notebook in a markup cell in your notebook.
- identify the issue/problem/challenge/question you are addressing
- identify what you hope to show (e.g. your hypothesis, etc)
- colab's markup guide
- text cells can also include latex

Data requirements

- at least 1000 instances (rows)
- under 1 MB
- load within 3 minutes within a notebook
- the dataset has to be publicly available (available via http)
- if you have a personal dataset, you can host it on google drive and upload it into the notebook
- the notebook must fetch the data via
    - pandas open_csv with a URL (instead of a filename)
    - python requests library (do not use wget)

MS2: Identification of analysis, data cleaning/normalization techniques

You will write (in another text cell) what analysis you chose. This milestone will be completed by submitting your notebook to Moodle.

- discuss the necessary data cleaning/normalization techniques.
- discuss why you selected the analysis technique

MS3: Notebook and video

This milestone will be completed by submitting your notebook to Moodle.

You **CANNOT** be late for this milestone, otherwise the peers who view your notebook will grade it poorly (because it's not complete).

- At this point, you are done with both the notebook and video.
- You will receive the URLS of peer notebooks to be graded.

Notebook requirements

- a working notebook that tells the story of your analysis
- you can certainly add as many text/code cells as necessary
- the notebook must function properly (i.e. the code cells run)
- be creative!

Video requirements

- a 7-12 minute video presenting the problem, analysis done, results
- at least 3 minutes must be spent on explaining the techniques you did
- you are also helping teach the viewer

Grading Rubrics

Grading Rubrics

Part of MS4 milestone requirements

The following Score Card(s) will be used by both staff and peers (subject to change):

Overall Project

Score	Why
[0]	this work is not an original analysis (see comments)
[1]	little to no effort; not organized; incomplete; did not understand what was happening
[2]	not enough details given; not coherent story; distracting grammar/spelling issues
[3]	solid effort; well done; coherent story
[4]	outstanding analysis; presentation; conclusion

Notebook

Your notebook will be grading on the following components:

Introduction/Project Scope

Score	Why
[1]	missing; poorly written
[2]	basic introduction; did not understood what lies ahead
[3]	put the problem in a context I understood, I wanted to read on

Depth of Analysis

Score	Why
[1]	none; simple; only visualizations
[2]	basic statistics
[3]	machine learning, advanced statistics, NLP

Use of visualizations/images

Score	Why
[1]	little to none
[2]	visualizations were good, but needed more
[3]	great use of visualizations to add clarity

Performance:

Score	Why
[1]	some cells didn't run; the story was incomplete
[2]	took over 3 minutes; poor response time given the constraints
[3]	ran well, given the constraints of the dataset

Understandable/Usability

Score	Why
[1]	Poor explanation(s); I did not understand the analysis
[2]	Minimal explanation; I wasn't totally clear on what was happening
[3]	The author did a good job explaining what was happening

Code

Score	Why
[1]	poorly decomposed; poorly named variables, functions/classes;
[2]	well use of functions and/or classes, but poorly documented
[3]	well commented, understood the goals of each function/class

Overall Aesthetics

Score	Why
[1]	fundamental; not too much was done
[2]	illustrated (images, drawings, etc) but could use a bit more
[3]	well illustrated; illuminating, helped the understanding; well done

Video

Your video will be graded on the following components:

Content

Score	Why
[1]	lack of organization
[2]	focused subject matter, structured, scope of project (as told on video) is a bit unclear
[3]	creative, well-organized, well scoped

Presentation

Score	Why
[1]	distracting from the content (poor deliver, lack of effort)
[2]	smooth and effective, pace was adequate
[3]	engaging, well-rehearsed/scripted, good/great pacing

Production

Score	Why
[1]	distracting, inadequate; video poor; audio poor
[2]	good use of images, illustrations, examples; A/V non distracting
[3]	outstanding effort, very well done

How to grade "accurately"

Each of us will have our own definitions of what 'inadequate' or 'outstanding ' is. The important thing is to be consistent AND to grade each notebook independent of YOUR notebook and independent of others.

Don't grade relative to another person's project. It's unfair if your first notebook is so fantastic that you use it as the high bar. After each notebook you grade, reset your expectations, be fair, be consistent. Treat others' work with respect. It's okay to critique fairly.

Notices

Do I need permission?

You do not need approval to solve a specific problem -- use the grading rubrics and specified requirements help you decide. If you find it interesting, we will find it interesting. Let your interests drive the fun.

Good Coding Guidelines

Use good coding practices for all your code:

- ONLY use matplotlib's Object Oriented interface (fig and axes) for plotting. You can use seaborn (which we have not covered) but do not use Pandas to do your plotting.
- decompose your functions. Each function should be narrowly focused
- decompose your classes. You can do everything using Python classes (lesson on creating Python classes will be out soon)
- comment your code, so the reader at least knows what the function/class is responsible for
- use well named variables, functions, classes, CONSTANTS
- indent your code 4 spaces
- follow the pep style guide when in-doubt

ASK Questions (on Piazza)

Ask all questions regarding the project, under the Project Thread on Piazza so everyone can benefit.

Incremental Feedback

When you submit your notebook on Moodle for the milestones, the staff may not have time to review it for feedback. If you are stuck or have questions, you **MUST** ask on Piazza.

The staff will assume each submission you make meets the specific milestone requirements.

Integrity

Also, please don't copy any analysis that is available. Remember your work will be viewed by others and the staff. It's vital that all work you submit is from your own mind. It will serve no purpose if what you outline is just a copy&paste of someone else's blog or project.

Anonymous Feedback

All reviews will be done anonymously. When you grade someone else's work, be professional -- the staff will be able to see your comments and grading.

Good Luck!!