CS 141 Fall 2023

Final Project

This project is an opportunity for you to apply and expand upon some of the concepts that you learned in the second part of the course.

Project Phases:

1. **Preparation**
2. **Writing a project proposal**
3. **Data collection and data cleaning**
4. **EDA**
5. **Completing the main part**
6. **Writing a report**
7. **Presentation**
8. **Writing Questions**
9. **Answering Questions**

1. Preparation
Join (create) a group of 4-5 students:
https://docs.google.com/spreadsheets/d/1hPgfmOohmHFyu7FwOketWSb8QQRHJSnm/edit#gid=29070991
Find a topic that interests you (disease prediction, text processing, image compression, recommender system development, ...). Be creative, avoid popular ideas. Choose a topic that you care about.
- State a question(s) or set a goal. Think about how you can answer your questions or achieve your goal.
- Identify the data you need. Try to find a sufficiently large dataset(s) online. (Do not use Kaggle.)
- Explore methods, techniques, and algorithms that you can employ to answer your questions.

2. Writing a project proposal (approximately 1 page)
Introduce your chosen topic, provide a brief description of your project, outline the technique(s) you plan to use (feel free to submit an informal proposal via our shared Google Drive document if you would like to receive feedback), and consider including any hypotheses you may have. Once finalized, submit your proposal as a PDF on Gradescope. Additionally, prepare a few slides for presenting it in class.

3. Data collection and data cleaning
Depends on your project and data that you found.

4. EDA
Perform EDA on your dataset to gain a deeper understanding of the data. Decide, what needs to be done to capture pertinent information related to your topic. Present your findings using visualizations.

4. Completing the main part
What methods/techniques/algorithms can you use? Select one supervised learning technique and one unsupervised learning technique, excluding linear regression, as it was covered in detail in the lab. (For projects on text mining, image compression or recommender systems, the requirements may be slightly

different, and can be discussed during the meeting.) Apply these techniques to your data, and thoroughly analyze the results.

5. Project report
Provide a detailed description of all the work you performed on the project, following the order outlined in Project Phases 2-5. Begin with the project description, which is an edited version of your project proposal. Make sure to discuss and analyze the results. Complete the report before preparing your presentation. Include each member's contribution at the end. Upload the report to Grdescope.

6. Presentation
Prepare slides for your presentation.

7. Writing questions
Write three questions about your project and include them into your ppt. All students should be able to answer your questions after watching your presentation (even if they were not very familiar with the topic before).

8. Recording
For this part, you need to record a 8 –10-minute presentation covering all aspects of your project.

9. Answering questions (to be completed individually and uploaded to Gradescope).
Watch 10 presentations prepared by other groups and respond to at least one question from each presentation. Upload your answers to Gradescope.

What to submit:
1. Project proposal and techniques that you are planning to use (Gradescope/Project Proposal)
2. Clean dataset (include the source) or a link to the dataset (Google Drive, your team folder)
3. Jupyter notebook (Google Drive, your team folder)
4. Report, pdf (Gradescope/Project Report)
5. Video recording and PPT (Google Drive, your team folder)
6. Answered questions (Gradescope/Answering Project Questions).

Grading:

Project description – 10 points
Data preparation and EDA – 10 points
Main part and conclusion – 65 points (15p./65p. for difficulty/creativity)
Recorded presentation (including PPT) – 10 points
Questions (answering) – 5 points

Timeline:

Forming groups, choosing and finalizing the project idea, submitting the proposal – **November 30, 11:59pm**; late submission – December 1, 5% off of the project score, no late passes.

Uploading videos, slides, report – **December 17, 9am.**

Answering questions – **December 17, 11:59pm.**