Math 4750 / 6920
Final Project Instructions

Overview

The goal of this class project is threefold: Firstly, to solidify your understanding of data mining and predictive modeling by applying these concepts practically. Secondly, to prepare you for real-world collaborative analytics projects through team-based design and execution. Lastly, to enhance your portfolio or resume with a demonstrable application of your skills in data analytics, positioning you favorably for future opportunities. Choose a dataset that interests you. This can be from any domain such as sports, economics, health, technology, entertainment, etc. Ensure the dataset is appropriate for class room use and substantial enough to allow for in-depth analysis. Work in groups of two. Begin early, especially in data procurement and background research. Utilize data mining techniques such as supervised, and unsupervised techniques learned throughout the course. The project is divided into three main components:

1. A project proposal
2. A short project presentation
3. A project report

The expectations for each component are detailed in the sections that follow.

Due Dates

Project Proposal: Due Friday, April 5, 2024.
Completed Project: Due Monday, April 22, submit report.

1 Project Proposal

The primary goal of the project proposal is to ensure your chosen project is feasible and aligns with the course's scope, providing an opportunity for feedback and guidance. You'll work in pairs for this project. You are encouraged to form groups on your own, as discussed in class. If you cannot find group members, I will assign you to a group randomly. Should you have any concerns about working with someone in your group, please speak to me for accommodations.

Submit a one-page proposal by the due date. Include all group members' names and section leaders. Address the following:

- **Proposal Format:**
  - The project proposal is a 1-page document, excluding references.
  - You are encouraged (not required) to use 1-2 figures to illustrate technical concepts.
  - The proposal must be submitted as a PDF document.
- **Introduction:**
  - Describe what you are planning to do.
- **Motivation:**

  - Describe why your project is exciting. For example, you can describe why your project could have a broader societal impact. Alternatively, you may describe the motivation from a personal learning perspective.

- **Resources:**
  - What resources are you going to use (datasets, computational tools like sklearn, etc.)?

1.1 Project Proposal Assessment

The proposal will be graded based on completeness of each of the 3 sections (Introduction, Motivation, and Resources). For each section, you can receive a maximum of 5 points, totaling 15 pts for the proposal overall.

2 Presentation

Prepare a 3 minutes presentation of your final project to share with the class. The presentation is "free form" but may cover the following:

- introduce the topic to a general audience;
- summarize the main approach;
- highlight the outcomes of your work.

One member of the group should participate in the presentation. The presentation will be all virtual and be submitted via Canvas. I will then upload the videos on Canvas so that the other students can watch them and make comments.

2.1 Project Presentation Assessment

The rubric for assigning points (out of 40) for the presentation is provided below:

- **5 pts:** Is there a motivation for the project given?
- **10 pts:** Is the project described well enough that a general audience, familiar with data science, can understand the project?
- **5 pts:** Are all figures legible and explained well?
- **20 pts:** Are the results presented and adequately discussed?

3 Project Report

The project report is expected to be 4-8 pages long (excluding references) and should include the following sections:

1. Abstract
2. Introduction
3. Proposed Method
4. Experiments
5. Results and Discussion

6. Conclusions
7. Contributions

More details are provided in the report template, which can be found at Canvas file section.

Please follow the report template file for writing your report. Additionally, you are required to submit all the code, computations, and experiments developed and conducted for this project. The quality of the code will not influence your grade but will verify the originality and authenticity of the reported results.

4 Project Report Assessment

The rubric for grading the project reports is outlined below.

4.1 Abstract (5 pts)

- Is enough information provided to get a clear idea about the project?
- Does the abstract convey the findings?

4.2 Introduction (5 pts)

- Does the introduction provide the necessary background information?
- Is there a motivation explaining the project's relevance and importance?

4.3 Proposed Method (10 pts)

- Are descriptions of symbols and mathematical notations included?
- Are the main algorithms described comprehensively?

4.4 Experiments (10 pts)

- Is the methodology described clearly?
- If datasets are used, are they referenced properly?

4.5 Results and Discussion (20 pts)

- Are results described clearly and logically analyzed?
- Are all figures clear and labeled correctly?
- Is the discussion critical and honest, including potential weaknesses?

4.6 Conclusions (5 pts)

- Do the conclusions reflect on the achievement of the project's goals?
- Are future directions suggested?

4.7 Contributions (5 pts)

- Are the contributions of each team member listed clearly?
- Did each member contribute approximately equally?

Group Dynamics

Ensure equitable distribution of work among group members.