Final Project Grading

INFSCI-2160 Data Mining

Matt Berezo, University of Pittsburgh

The overall grade for the final project is 40% of the total grade. This can be done in a group of 2 – 3 students. The final project includes the following milestones due before the beginning of class on the respective weeks. Late milestones will decrease 10% on final project grade by each day late:

- Week of February 4, 2019: Abstract Due (5%)
    - Short paragraph describing: (1) Project name and team members, (2) problem you have chosen, (3) why the problem is interesting to you, (4) what kind of data you plan to use.
- Week of February 25, 2019: Proposal Due (5%)
    - Elaborated version (300 – 500 words in length) of abstract that includes: (1) Title and team members, (2) problem you plan to do, (3) why the problem is interesting, (4) approach/general project plan you intend to take, (5) what kind of data you plan to use and where/how you will get it
- Week of March 25, 2019: Progress Report Due (10%)
    - Same format as final report. Should include: (1) What has been accomplished, (2) preliminary description or analysis about your dataset(s), (3) description of data collection process, (4) any early numbers that you find notable in the report
- Week of April 15, 2019: Final project presentations
- Wee of April 22, 2019: Turn in final paper (4 – 8 pages in length) following the ACM SIG format (https://www.acm.org/publications/proceedings-template). Report should include:
    - Description of the problem
    - Goal of the study
    - Any benchmark models or notable research done by other groups on the problem
    - Dataset description and overview
    - Data analysis and predictive modeling approach
    - Evaluation of the model
    - Limitations and/or downfalls of the study
    - Recommendation/next steps
    - High-level description of how this could be deployed (if applicable)
    - Description of each member's role and work on the project
    - Final presentation and report will make up 80% of the final project grade

Useful places to find data:

www.google.com

https://toolbox.google.com/datasetsearch

Rubric

A: Technical Strength and Evaluation

Technical Strength: The goal of this project is for you to practice how to design, implement, and evaluate a data mining project in a practical, real-life scenario. To show the technical strength in your project, you need to provide sufficient details in your oral presentation and written report. Readers should easily understand:

- What are you trying to accomplish?
- What techniques did you use?
- What are the strengths and weaknesses of your project?
- The approach you took to achieve your goal

Experimental Evaluation: Describe your experimental results with systematic evaluation in your oral presentation and written report. Specifically, describe the evaluation criteria, including performance measures and baseline methods. Explicitly state the performance of your method and provide results and insight you gained from the evaluation.

B: Innovation and Challenge Level

Innovation: Use your project to do something creative and useful. You could demonstrate innovation by (i) applying existing data mining techniques to solve a new problem, (ii) incorporating new ideas to improve existing data mining techniques for solving an existing problem. If your project simply duplicates existing work, you will gain fewer points in this aspect. To get the points you deserve, present any innovative ideas included in your project, in your oral presentation, and written report.

Overall challenge level: Your project is expected to be reasonably challenging. This involves (i) non-trivial implementation effort, and/or (ii) innovative components. You should explicitly state why your problem is reasonably challenging in your oral presentation and written report. For groups with 2 or more members, the expected challenging level should be higher.

C: Presentation

You should clearly present required components as described above in both your oral presentation and written report. Present your project to the entire class and you will receive peer review from the class. In your written report, provide sufficient details within proper page length. Use the given template and follow the writing guideline as described in the course webpage "Final project guideline." In the written report, cite and provide reference for any source of information you have used in your project, including publications, online information, code or software packages.

Honor code: You may consult any papers, books, online references, or publicly available implementations for ideas and code that you may want to incorporate into your project, as long as you clearly cite the sources in your code and your written report.

D: Implementation

You need to submit R code for your project that is working and can produce the same experimental results as described in your final report. Hence, the experimental procedure and the testing data must be submitted together with necessary documentation describing how to run the experimentation.

E: Overall Project Management

You need to submit required materials to meet each milestone as described above. These milestones are to help you seek resources you may need, make sure you develop a concrete direction, and better manage your project. Before each milestone, if you find it difficult to meet the requirement, you should make an appointment and discuss with me in advance.

Group member evaluation:

To facilitate real collaboration within larger group, for a group with two or more members, I will ask each member to submit independent evaluation of member contribution for your group member(s), and adjust individual scores based on the member evaluation.