Final project

Prof. Alexandra Chouldechova
95-791: Data Mining

Overview

- Work in teams of 2-4 students
  - You may assemble your own team
  - If you wish to be assigned to a team, please fill out the short final project survey

- Pick a project from the posted list
  - Multiple teams can work on the same project
  - Teams cannot collaborate with each other, especially if working on the same project

- Feel free to use any data mining methods we covered in this class, or any data mining methods you learn about in class or outside of class while working on the project

- **Deliverable**: an R Markdown-generated report addressing the key questions outlined in your project description

Deliverable: Final report

- Each team will submit 1 written report
  - Submit both your .Rmd file and .html output file

  - Report should have a References section clearly citing any non-course references that were used during the project

  - Got a great idea from reading a blog? Cite it.

  - Talked to a friend who has experience with that kind of data? Cite them.

  - Got an idea from something or someone who isn't a teammate?
    Cite. Cite. Cite.

Deliverable: Final report

- Each team will submit 1 written report
  - Submit both your .Rmd file and .html output file

  - Report should have a References section clearly citing any non-course references that were used during the project

  - Got a great idea from reading a blog? Cite it.

  - Talked to a friend who has experience with that kind of data? Cite them.

  - Got an idea from something or someone who isn't a teammate?
    Cite. Cite. Cite.

Due date: Friday, March 9, 11:59PM

Final report: Contents

- **Introduction:** Data and Problem summary
  - Begin with a section that clearly describes the problem(s) you are addressing and the data that you have available
- **Section 1:** Exploratory Data Analysis
- **Section 2:** Methodology
  - Give an overview of any Data Mining methods that you experimented with, along with a summary of how the methods performed
  - If you used both Descriptive and Predictive methods as part of your analysis, describe these in two separate subsections
  - Clearly state the final model(s) that your Findings are based on
  - Important: Clearly describe and provide supporting Figures/Tables of any and all Model selection and Model validation that led up to your final model
- **Section 3:** Key Findings, Main takeaways
  - Provide Figures, Tables, Nicely formatted model output, and accompanying discussion of your main findings
  - Be sure to clearly answer, in context-appropriate language, each and every question stated in your Project's description file

Final report: Grading

- Project is worth 30% of your Course grade

- Whether you are enrolled for a Letter Grade or Pass/Fail, you must achieve a passing score on the Final Project in order to pass the class

- Everyone on a team will receive the same score
  - You may contact me if you wish scores to be assigned to group members separately

  - If you do, I will implement a peer-review grade component to gauge individual contributions of the team members

  - Individual grading isn't guaranteed to result in a different (or better) score

Getting started

- Select your team by Friday, February 16.
  - A Google Form will be sent out to collect team information.
  - For those of you who wish to be assigned to a team, a survey will go out later today asking you for your project preferences.

  - Pick a team name

- Get started on your project!

- Your team will meet with the teaching staff on a weekly basis starting next week to discuss your progress