CSCI 567, Spring 2023
Prof. Yan Liu

Course Project

Due: 11:59 pm, Apr 21, 2023 (Submission)
5:00 pm, May 5, 2023 (Report)

1 Instructions

In the project, you will have the chance to explore an interesting machine learning problem by participating in a DrivenData competition. In this competition, you will predict the level of damage to buildings caused by the 2015 Gorkha earthquake in Nepal. Specifically, given information about the buildings hit by Gorkha earthquake, such as the buildings' structure and their legal ownership, your challenge is to predict level of damage to these buildings. For more details, please refer to official competition website on DrivenData (Richter's Predictor: Modeling Earthquake Damage).[1]

2 Teams

- The project should be done in teams comprising of **2-3 members**.
- Fill out your team information in this google spreadsheet with your team details. Please log in with your USC email. You can find your team ID in column A within the spreadsheet (Note: please don't change the pre-specified team IDs in the first column).
- Form a team of up to 3 students by **11:59 pm, Mar 20, 2023** and create your accounts on the official competition website. Make sure that all members of your team are registered under a single team name which begins with CSCI567_id[TEAM ID], for example, CSCI567_id16. The member who creates the team should join the CSCI567 community: go to your profile page and enter CSCI567 under MY COMMUNITIES. You will see the class community appear with a link to the community leaderboard. As you submit, you'll be able to see the other participants and your performance within the community.
- A team can have members from different sections of the class (offline, online, DEN). Please feel free to use Piazza to arrange your team. You can post on Piazza to look for teammates (make sure to add the "project" tag).

3 Grading

- Your grade is bifurcated as follows: 1) your team's relative rank and score on the leaderboard (50%), and 2) final presentation and the project report (50%). The total base score is 100. The project is worth 20% of the final grade for the class.
- Note that the public leaderboard shows the ranking of all teams, not just those in this class. We will consider the final ranking on the public leaderboard as of **Apr 21, 2023, 11:59 pm, PT** for grading.
- Members of the same team will receive the same scores.

4 Final Presentation

The final presentation is on Apr 25 (the last class) and will be held online. The detailed logistic will be announced later.

[1] https://www.drivendata.org/competitions/57/nepal-earthquake/page/134/

5 Deliverables

- Each team needs to write the project report in NeurIPS format. (6 pages maximum, including references; this page limit is strict). The NeurIPS LATEX format can be found here: https://neurips.cc/Conferences/2022/PaperInformation/StyleFiles

  Not following the format will lower your grade.

- In your report, you should cover the details of your solutions, including the general ideas, the way of data processing and cleaning, the learning algorithms and models you have tried, the results you obtained, and any other insightful thoughts during the competition. You should also describe how to run your code to get the results.

- All teams are advised to use Python as the programming language for their code. Training should be doable solely on your laptop, but you can use the Kaggle Notebook, Google Cloud Platform, or Google Colab for running your codes, if needed.

6 Deadlines

- We require teams from this class to submit on DrivenData by **Apr 21, 2023, 11:59 pm, PT**. Please make your final submissions on time. We will grade you based on your final released ranking on the online public leaderboard as of the deadline.

- Each team also needs to submit one PDF copy of the project report and all the code via Gradescope. This is due by **May 5, 2023, 5:00 pm, PT**.

7 Policy on collaboration

In line with the rules of the competition, you are only allowed to share code within your own team. Your code will be analyzed to reproduce the results and compare its similarity to code from other teams. Any violation of the USC Integrity and Plagiarism policy will lead to an immediate "F" grade in this class and you might be subject to harsher penalties. However, discussion about approaches between each team members and cross-teams are allowed and we encourage you to actively engage in forums, piazza, and discussion with the DrivenData's community. This provides a great learning opportunity for you to explore how to collaborate and research new approaches to tackle hard challenges.

Using resources from the internet is acceptable, but you must cite and acknowledge any resources you used, and should understand and analyze everything you use as part of your report.