**DATA 620 Final Project**

Deliverables Schedule

|  |  |  |
| --- | --- | --- |
| *Deliverable* | *Date* | *Points* |
| Final Project Proposal | Saturday July 13 | 25 |
| Final Project | Thursday July 18 | 150 |
| Final Project Presentation | By Video | 50 |

Policy on Collaboration

You may work in a team of up to three people. Each project team member is responsible for understanding and being able to explain *all* of the submitted project code. Remember that you can take work that you find elsewhere as a base to build on, but you need to acknowledge the source, so that I base your grade on what you contributed, not on what you started with!

Final Project Proposal

Your proposal should describe at a high level what you’re seeking to accomplish, and your motivation for performing this analysis. A guiding question or hypothesis to test is one good way to start. If you are going to work in a small group (encouraged!), you should also list your partners’ names.

You should briefly describe your data sources:

[20 Open Datasets](https://medium.com/@ODSC/20-open-datasets-for-natural-language-processing-538fbfaf8e38)

<http://cseweb.ucsd.edu/~jmcauley/datasets.html>

[How to Detect Fake Reviews](https://qz.com/1758846/how-to-detect-fake-reviews-while-shopping-online/)

<https://shopping.com/>

If one reviewer trusts another reviewer and both reviewers write a number of “questionable” reviews – flag them

Is their review history filled with similar five-star critiques, singing the praises of random products such as laptop stands and off-brand iPhone chargers? Chances are you’ve caught a professional fake reviewer.

If a large number of users who reviewed the item only have a single review in their history, they may be fake accounts.

**IDEA:**

Use NLP to look for Fake Reviews (criteria to be determined)

Then leverage the Social Network to try to identify who else trusted the fake reviewers, to find potential other fake reviewers

Combine the results of the above into a list of problem fake reviewers

Bonus: build a model that attempts to identify future fake reviewers

plan for doing the work, and up front concerns. If you are working in a group, please describe the roles and responsibilities of each group member.

We’ll treat this proposal as a planning document, not a blue print containing “firm, fixed requirements.”

Final Project

Your project should incorporate one or both of the two main themes of this course: network analysis and text processing. You need to show all of your work in a coherent workflow, and in a reproducible format, such as an IPython Notebook or an R Markdown document. If you are building a model or models, explain how you evaluate the “goodness” of the chosen model and parameters.

Final Project Presentation

We’ll schedule a short presentation for each team, either in our last scheduled meet-up or in additional office hours to be scheduled during the last week of classes.