

Natural Language Processing for the Social Sciences (GR5067)

Class Project

Description:

Your team of 4-5 members was recently funded by [Sequoia Capital](#); one of Silicon Valley's premier Venture Capital firms. Your job is to conceptualize an idea requiring NLP practices and prototype/develop it in Python.

Each team needs to come up with a topic (some examples but not limited to are below). Requirement is to provide bi-weekly progress updates, a mid-semester presentation and end of semester presentation, demo and research report is required.

- A. Leverage New York Times API to extract key words, like 'trade war' and plotting against stock market movements
- B. Compare corpuses extracted between Republican vs Democratic speeches
- C. Detecting sentiment changes for geopolitical policy changes
- D. Streaming twitter sentiment engine, great for track real-time sentiment during debates!

Again these are just 'examples' and you are encouraged to come up with your own topics. Just make sure you pick a topic supported by existing data. I strongly recommend carefully thinking out your project as executed projects can very well turn into the next [Unicorn](#).

[Kaggle](#) is a great resource for projects ideas and to get access to data

Presentation DUE *last day of class*:

The final class each team will present for ~15 minutes.

As for the presentation PDF:

'Suggested' Slides (and you can be creative here):

- Intro
- Quick overview of data used
- Quick overview of methodology and validation
 - o Walk through of overall architecture
- Performance Measures
- What did you like least and best about the project?
- Conclusion

Final Research Report DUE a week after the last day of class by 11:59PM EST:

A final research report (double-spaced in Arial 12 font) is required and needs to include the following:

Abstract:

Paragraph synopsis of research

Introduction:

Description of your project

Data:

Description of the data used

Methodology:

Description of technology used

Results:

Summary of key findings

Conclusion:

Restate what problem or question your project addresses

References:

Citations, *APA style*

Team lead must submit to courseworks the following, by **11:59PM EST 1 week after the last day of class** in a single zip file containing the following:

1. Presentation .pdf
2. Research paper .pdf format

**** Please do NOT submit the data you used nor code, just submit what is listed above.**

Project Grading Rubric:

Project	Presentation	Knowledge of Material	Presenter was fully prepared with understanding of the parameters.	Presenter was unprepared for the talk, demonstrating little understanding of how the topic relates to the Missing backup material where conclusions were inferred	Did not show up for presentation without valid excuse	10
		Content	Excellent visuals		Did not show up for presentation without valid excuse	10
	Paper		An accurate and complete explanation of key concepts and theories is made. Enough detail is presented to allow the reader to understand the content and make judgments about it. Quality of paper is submission worthy to a journal.	The explanation is sufficiently inaccurate, incomplete, or confusing that the reader gains little information from the report. It appears that little attempt has been made to help the reader understand the material.		
		Content			Did not submit a report	30