# CS 272: Statistical NLP: Winter 2020

# Project Pitch

Sameer Singh
https://canvas.eee.uci.edu/courses/22668

One of the biggest indicators of whether a course project will be enjoyable and successful is how early do the students start working on it. Thus, I want to encourage you to start forming groups, discuss project ideas, and get feedback from me as early in the quarter as possible. To facilitate this, you will be submitting a short *pitch* of your project. Each pitch (one per group) should be uploaded to Canvas by **January 23, 2020**, containing the following sections.

## 1 What to Submit?

Each project pitch (one per group) will be a **one page** PDF, plus references, if any. It is mandatory to use the ACL style files for this report. Although the style files are both for LaTeX and Microsoft Word, I suggest the use of the LaTeX template, mostly for your ease of use, but you are free to use any. The style files are available here: https://github.com/acl-org/acl-pub/tree/gh-pages/paper_styles. Please use `\aclfinalcopy` option, as described in the example.

### 1.1 Team Details

Your pitch should contain the following details about the group:

1. *Team Name:* A short, memorable name for your group. If you are struggling with this, just find an interesting word that uses all your initials, or simpler still, just pick a random food item :)
2. *Members:* Names and UCINetIDs of all the group members (do NOT include your Student ID number). Add yourself to a "group" on Canvas (see instructions below). For < 3 members, email me on/before **January 16, 2019**.
3. *Division of Labor:* In a single sentence for each member, describe the relevant background, and the part of the project are they likely to contribute to.
4. *Diversity:* Add a single sentence to describe what you think makes your group diverse.

#### Joining a Group on Canvas

1. Login to your canvas account, and navigate to the CS 272 page.
2. Click 'People' on the left navigation bar.
3. Click the 'Groups' tab. You should see a list of potential groups to join.
4. If your team has not already claimed one, join one of the empty "Project Group X" groups. Otherwise, join the same group as your teammates. *Warning:* Do NOT use the "+GROUP" button to create a new group!
5. Change the group name from "Project Group X" to your team name. To do this, navigate to your group's homepage and click the "Edit Group" button. *Note:* You will only be able to do this if you are the first member of your team to join the group.

### 1.2 Problem Setup and Motivation

In a paragraph or so, describe the problem that your contribution will be addressing. This should include, at the very least, what the formal problem set up is (what is the input and the output?) and why is it important (potential applications?). A brief understanding of related work should be included to support your project, such as the similar tasks that have been studied in the literature so far. For **replication** project, provide name/title of the paper you picked, describe briefly the problem, and why you picked it.

### 1.3 Proposed Approach

In few sentences, describe what strategy you envision taking to solve the problem you proposed above. I do not need a detailed solution, but instead a very high-level understanding of how you intend to bridge the gap between

the input and the expected outputs. Of course, this is not set in stone, and I expect it to change considerably as you learn more about NLP and your project, however this section is where you will argue why it is an achievable goal for the quarter. For **replication** project, describe whether the dataset is available and accessible, and describe the code, if any, that is available for the approach.

### 1.4 Evaluation Plan

Describe, in a paragraph, how you intend to evaluate whether your project was successful. Although you may include a brief description of the qualitative results (such as examples, important features, etc.), you should focus on empirical evaluation. What is the dataset you will evaluate on? If you do not have any, how will you collect it? What is the evaluation metric? What is the baseline you definitely want to beat, or alternatively, what is the performance you would be happy with (e.g. $> 80\%$ accuracy)?

### 1.5 Computational Requirements

In a single sentence or two, describe your estimated computational requirements (storage, GPU, compute, etc.) and the resources you are relying on to address this. You can include the Google Compute credits that will be provided to you.

## 2 Tips and Suggestions

Here are some suggestions as you think about your project ideas and groups:

- *Play to your strengths.* I want you to work on something you like doing, and have some expertise in, as much as possible. If you are a machine learner, think about what NLP tasks can your methods be applied to. If you have some problem you care about, propose a novel task or dataset for others to collaborate on. If you have some ongoing research that is relevant, identify a small, independent research question, and propose it.
- *Skim papers.* Go through the list of papers on the course website (I will be adding papers soon). Read their abstracts, and see if any of them get you excited. Browse titles on recent NLP and ML conferences (ICML, ICLR, NIPS, AAAI, ACL, NAACL, EMNLP, and KDD), see if you can find their datasets or codebases. For even more "niche" topics, look at conference workshops to see if you like something, e.g. `http://naacl.org/naacl-hlt-2016/workshops.html`.
- *Use Github.* This may be obvious to many of you, but needs to be said. Learn `git`, and use Github to share as much as possible with your group: reports, code, data, documentation, etc. You may also want to consider using other features like issues, website (Github Pages), and the wiki. I also encourage you to make your repositories public for open source access (maybe use a license like Apache), but you are free to keep it private if you like (Github provides free private repos). Feel free to include the repo URL in the pitch if you decide to make it public, or want to share the private repo with me (Github username: `sameersingh`).
- *Use Google Sheet on Campuswire.* There is a shared Google sheet posted on Campuswire that can be useful to find classmates based on expertise and brainstorm different ideas. For example, if you have an idea but not a team, propose it to see if someone else is interested. You can also advertise yourself ("Can speak fluent Tensorflow!") and see if someone needs you. I encourage you to continue using Campuswire even after the teams are finalized, such as to compare your results to others working on the same/similar data, or to solicit feedback/suggestions from others on software libraries.
- *Start work early.* I really cannot stress this enough.