Course Projects

Instructor: Alan Ritter

Final Project
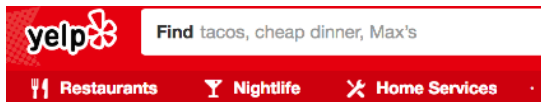
- Groups of 3-4
- Sign up here:
  - https://goo.gl/forms/FK6Yv84ac2fhFyH42
- **Highly Recommended:** stop by the office hours to chat and get feedback on your project.
- **Scope:** on the order of one of the programming assignments
  - But, need to define the problem, come up with right feature representation, write up results in a formal report.
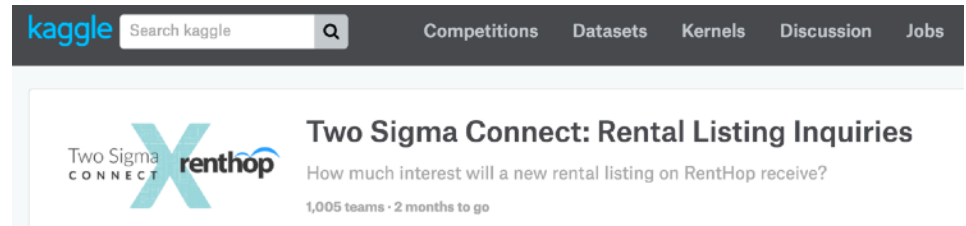
Selecting a Topic

- Part of your thesis? Great!
- Find a problem you are interested in where you think ML can help.
- Experiment with one of the algorithms we discussed about in class.
- **First question:** what is the dataset?

Datasets

- Netflix Problem
- Yelp Dataset Challenge
  - https://www.yelp.com/dataset_challenge
- Various Kaggle Competitions
  - https://www.kaggle.com/competitions
- Semeval Tasks:
  - http://alt.qcri.org/semeval2018/index.php?id=tasks

Yelp Dataset Challenge

Requirements

- 4 Page Report
  - Due December 10
- Include empirical analysis of your approach
  - Report performance on dev / test set
  - Compare against some reasonable baseline method.
- In class presentation during scheduled Final Exam time

Advice

- **First question:** is the data available?
- Try to get a simple baseline working as early as possible to determine whether your project idea is feasible.
- Start with a manageable-sized dataset
  - Then scale up…