Nov 29

* Web scraping: Kai and Bofei
  + **TODO:** find a way to scrape a random list of users
* Feature selection
  + **TODO:** research existing projects, decide most useful features
    - Based on available user and posting features
* Identifying fake users
  + Signs of fake users:
    - Significantly Lowered Costs
    - Carry similar items from the same brand (actual users would have shoes, pants, shirts, etc) but scam users only carry jackets for example
    - Zero Stars (especially with bad reviews)
    - Different People’s Listings are Used so Photo Backgrounds aren’t Homogenous
    - If we can scrape reviews, some reviews out the user as a scammer

**TODO** by next meeting:

* Scraping list of users - Bofei and Kai
* Choose important features - Serena, Esther, Kylie
* Identifying fake users (20 each) - Serena, Esther, Kylie
  + Only include listings where we are 100% certain about the label
  + Use keywords list and focus on one keyword each

Next meeting: Friday morning at 9 AM

Dec 1

* Presenting found fake users and their features
* Scraping has been run for 12+ hours

**TODO:**

* Finish annotating on Sheets - Kylie and Esther
* Annotate bofei’s user list - Serena
* Begin EDA and choosing a model - Esther, Serena, Bofei, Kai (We are prioritizing unsupervised models)
* Finish Scraping the current users, and start scraping the users everyone else looked at - Bofei, Kai
* Establish a pooling spreadsheet for all the users from everyone, and not try to look at more users than we already have (note)

Next Meeting: Saturday at 9 AM at location: Davis Hall Room 410 (meet outside)

Dec 4

**TODO**:

Run the rest of the scraping: Kylie

Test set: Kylie, Esther, Serena

Final Report editing: Esther

Presentation: Kylie, Esther, Serena

Tertiary annotation of Depop users: Esther

Running the pipeline again: Kai

Additional models: Bofei (NN-based, description CLS)

Dec 5

**TODO**:

Presentation (3 minutes each)

**Introduction**: Describe your research question(s) and discuss the project context and motivation (who are the stakeholder(s) and what are the benefits to them?)

* Kylie

**Dataset**: Describe the raw dataset: where or how did you collect it, descriptives, feature engineering, filtering, etc.

* Esther

**Methods**: Describe the machine learning methods you used to approach the problem at hand. You may use algorithms that we did not cover in class, but please give some background info if you do.

* Kai and Bofei

**Results and Conclusions**: Synthesize individual work and provide findings and conclusions. Do not forget to mention the implications of your project and its results. Would you say it's ready to be used in the real-world yet? If not, what needs to be solved before you would say it was ready?

* Serena
* Problem: model could be biased because we manually labeled the data
* Retrain with more data
* Potential to do more → computer vision to analyze photos