

### Progress Report:

- We've changed our idea for our project since we submitted our proposal. Originally we planned to collect tweets with hashtags supporting either candidate from a wide date range as training data for a political sentiment classifier. We would then use tweets with the same hashtags (removed in processing) to test the accuracy of our classifier.
- Now, we want to collect tweets with hashtags supporting either candidate from distinct time periods as training data, then test against tweets mentioning the relevant topic from that time period (i.e. debate, scandals, interviews, Saturday Night Live, etc) to see if our classifier produces results similar to the polling data at that time.
- Besides doing more research and changing our direction, we've made a lot of progress on building the classifier. We've written all of the code for a political sentiment classifier given our expected input. However, we haven't been able to test it yet since we hit an obstacle with collecting data.
- At first, we tried to use the Twitter API, but found that it only returns tweets from the last week. We then looked into web-scraping a page from Twitter's Advanced Search, but the search produces all matching tweets in chronological order, making it difficult to collect a sampling from a wide date range.
- Finally, we found a dataset of 30,000+ tweets from Oct 3-4 mentioning Donald Trump or Hillary Clinton. These tweets were collected using web-scraping on Twitter's Advanced Search. However, these are not tagged with which candidate they support.
- This is how we came to the idea of collecting tweets from multiple different 48-hour time ranges. If we can only find data from specific narrow date ranges, we could collect from only the 48 hour time ranges we chose, and divide tweets from then into separate training and testing corpora.
- We plan to collect tweets from significant 48 hour time periods this week as well as polling data and work on testing our classifier.
- We will run our classifier and analyze results over the weekend.
- We don't have a preference for presenting but would prefer not to go first.