

Project Progress Report: Classifying Party of Political Tweets

Melissa Lynn

For my project, I will scrape political tweets from Twitter, build a dataset of tweets labeled by political party, and train a classification algorithm to classify tweets by political party.

1. Which tasks have been completed?

I have assembled a dataset (in a `pandas` dataframe) containing a record for every member of the United States Congress, containing the following information:

- (1) Their first and last names.
- (2) The state and/or district they represent.
- (3) Whether they are in the Senate or the House of Representatives.
- (4) Their political party.
- (5) Their Twitter Handle.

I decided to use `Tweepy` to scrape Tweets from Twitter. In order to use `Tweepy`, I need authorization from Twitter, so I applied for a developer account. Hopefully that will be approved within the next few days.

In the meantime, I have written code that should scrape the tweets and store them in a dataframe, along with the additional context of the dataframe mentioned above. I have partially tested this code, and will start scraping tweets as soon as my account is approved by Twitter.

2. Which tasks are pending?

Once my account is approved by Twitter, I will scrape tweets to create my dataset. Once I have a dataset of tweets, I will do the following:

- Do some data preprocessing on the tweets, to create a vector as in the bag-of-words model. I would like to experiment with different options here, to see how they affect the model I train.
- Exploratory data analysis. Since I think that the dataset itself will be interesting, I want to dedicate significant time to exploring the data, and trying to identify any interesting patterns.
- Training classification algorithms. I will experiment with different classification algorithms and parameters, to try to produce the most accurate classifiers that I can. I think it would be particularly interesting to look at decision trees and linear models, where I can easily interpret how the models are determining a classification.

Note: the code that I've written will automatically label the tweets with the political party, so I don't need to do any extra labeling of the data.

3. Are you facing any challenges?

Just waiting for my authorization for a Twitter developer account to be approved.