Political Stance Detection on Twitter

Thomas Lam, Grant Park, Fanhao Yang

Introduction

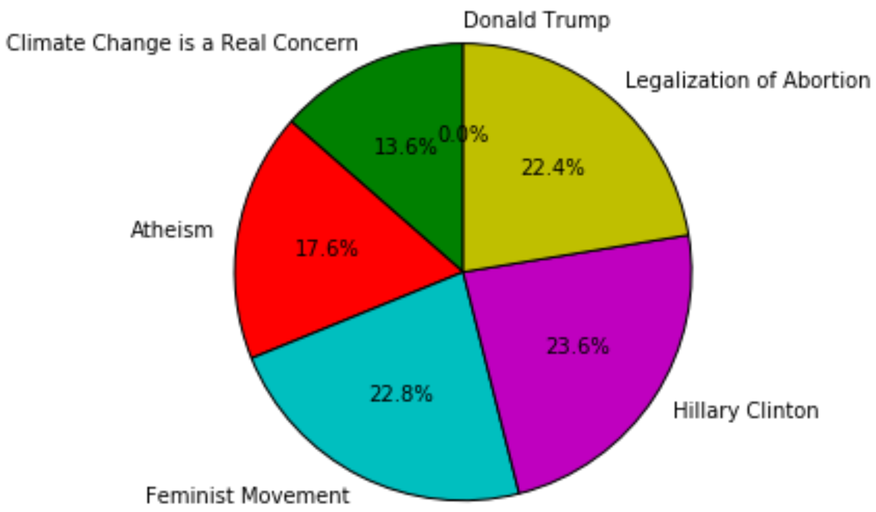
Problem: Detect a person's political stance on a given subject - more generally stance detection in NLP, i.e. determine from text whether the author of that text is in favor of, against or neutral towards a given subject.

Objective: What is the target subject that is being discussed in this tweet?

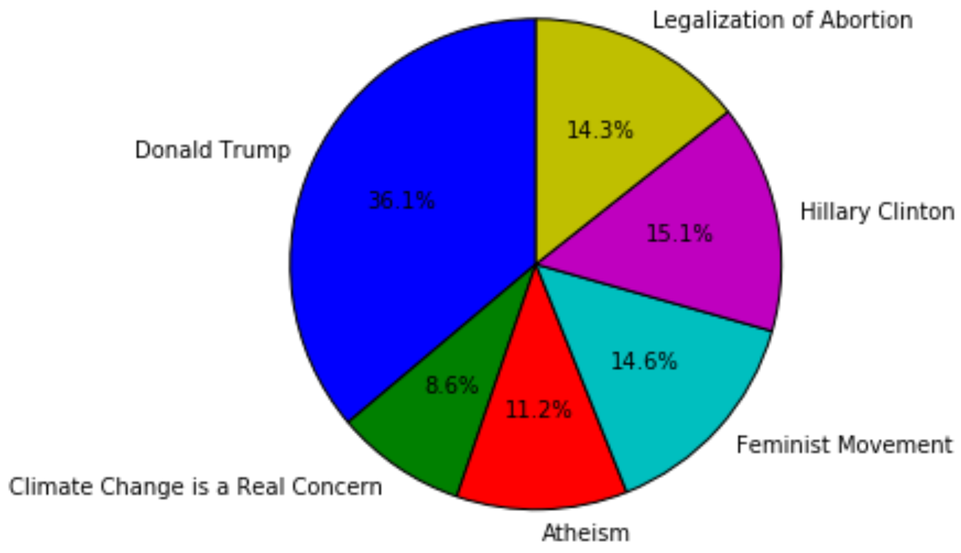
What is the person's opinion towards that subject?

Dataset

In this dataset, we are only interested in the target and stance labels, so we will be ignoring the other annotations labels for each Tweet.



The **training** dataset includes 2,914 tweets of the subject categories shown in the pie chart below. Notice that there are no tweets in this dataset which are annotated as being under "Donald Trump."



The **testing** dataset includes 1,956 different tweets of the subject categories.

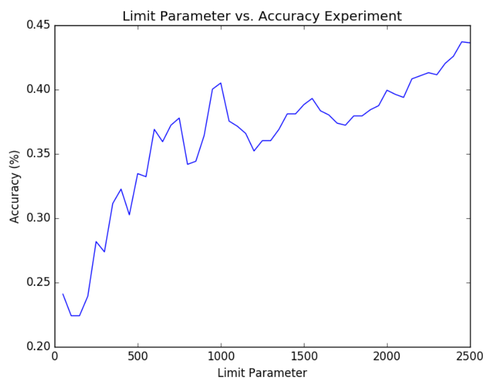
Methodology

We each implement different classification algorithms either from scratch or adapt existing library implementations into our project setting. The algorithms used are: Perceptron, Naive Bayes classification, Logistic Regression

For our baseline, we use majority class

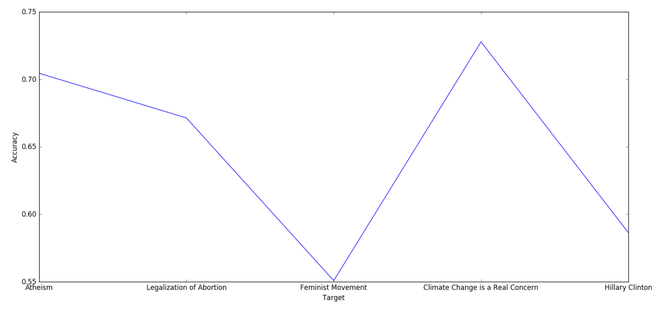
Big assumption: we assume detecting stance is a classification problem, i.e. similar to categorizing tweets into topic. In other words, for each classification algorithm, we perform on the dataset twice, once for identifying target and then for detecting stance

Discussion

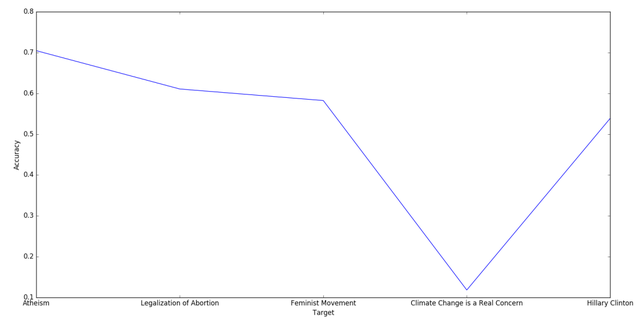


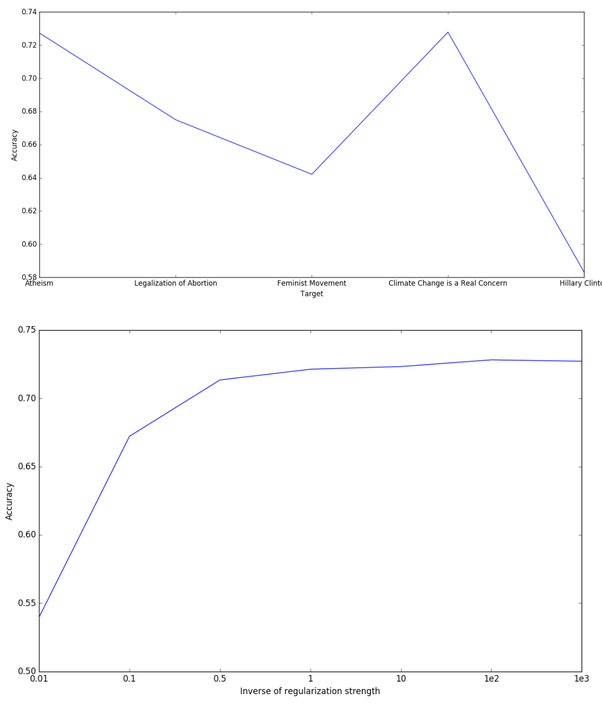
For Naive Bayes, we achieved a 66.37% accuracy on our test data with having trained on approximately 3000 tweets. This can be increased with more training samples as shown, where "limit" represents the number of tweets the classifier is trained on.

Using logistic regression, every target trained with its own model and applied to the test set



Logistic regression accuracy on target classification using improved tokenization





Applying the model trained using the tweets on "atheism" to the all the test sets

Using majority class classifier, every target trained with its own model and applied to its own test sest

Results

Among all three algorithms, logistic regression works best compared to Naive Bayes and Perceptron. Better tokenization increases the performance of logistic regression by 8%. Because the data is skewed, in stance detection, the majority class sometimes performs better.

References

[1]. Mohammad, Saif M., Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. "Semeval-2016 task 6: Detecting stance in tweets." InProceedings of the International Workshop on Semantic Evaluation, SemEval, vol. 16. 2016.

[2]. Sobhani, Parinaz, Saif M. Mohammad, and Svetlana Kiritchenko. "Detecting Stance in Tweets And Analyzing its Interaction with Sentiment." cc2016 The* SEM 2016 Organizing Committee. All papers cc2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: http:// creativecommons. org/licenses/by/4.0 (2016): 159.