

Project 3 Group 6

What does Twitter think about Ted Cruz?

Introduction

This project will examine tweets and their respective hashtags from Twitter on Senator Ted Cruz. The domain of this data was found via Twitter and the dataset was collected from Twitter's Standard API and was then analyzed through NLTK's NaiveBayesClassifier. The code used for analyzing the data can be found here: <https://github.com/jebidiah252/project3>. The slides for this presentation can be found here:

<https://docs.google.com/presentation/d/17UXnSnBEPKhRGXGp6JicTfJWd6gisko-zDuPz6rgFAG/edit?usp=sharing>.

Dataset

The dataset that we decided to observe was tweets that mentioned Texas Senator, Ted Cruz, from our domain of Twitter. We collected this data through Twitter's Standard API, version 1.1. What we did was we searched for tweets that contained the specific phrase, "Ted Cruz" and then tracked all tweets that were unique. We added the uniqueness to the tweets that we were getting, as having duplicate data was causing us to have skewed data when we started analyzing it.

Technique

Our techniques on this data were quite simple. We decided to first analyze all of the most popular hashtags from tweets that contained the phrase, "cruz". We figured that this would be enough to search for, as Ted Cruz was trending for a small amount of time. After we got the most popular hashtags from Twitter, we then decided to find the words that were recurring the

most in all of the tweets that we fetched. Our next analysis was to find the most popular hashtags by day when searching for tweets around Ted Cruz. The last analysis that we added

was a supervised

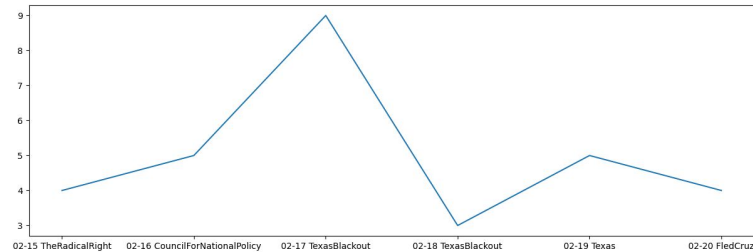
Machine Learning

model. We created

this model by giving it

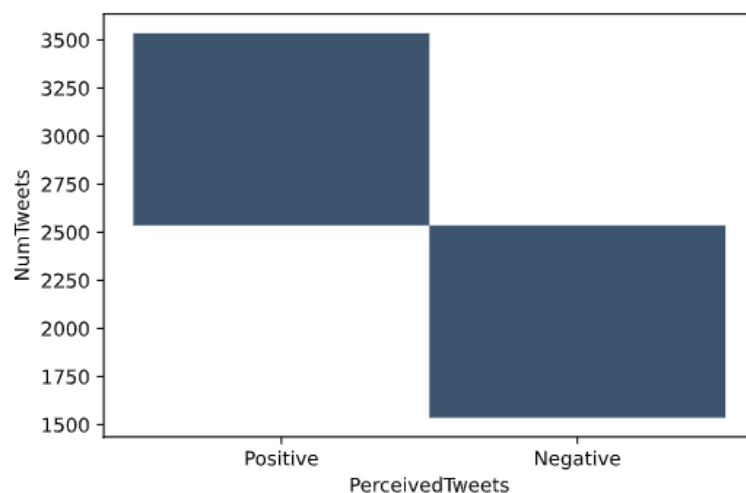
example tweets from

NLTK and then having it look over the Ted Cruz tweets. When setting up the model, we were seeing that the model had around 99% accuracy.



Results

The results that we found were interesting. We found that by visually looking at the occurrences of tweets and hashtags produced from Twitter that lots of Twitter users have a very great distaste for Ted Cruz. Along with the hatred towards Ted Cruz, it also looked like that anytime Ted was mentioned, there was a possibility of seeing his name with Donald Trump. What surprised us the most was that after looking at the data visually, the Machine Learning



model found that there were a greater number of tweets that were perceived as positive as compared to the very negative hashtags. What this could mean is that there were not enough negative tweets from the supervised set of data to formally

say that some of the tweets directed at Ted Cruz were negative, or people think Ted Cruz is a great guy. We're more likely to think the former.