Connie Li

11/10/18

Assignment 2- Text Mining

1. Project Overview

The main objective of this project was to conduct a sentiment analysis on Jon Bellion's most recently released album "Glory Sound Prep". This album was released yesterday, 11/9, and has been a hot topic of conversation amongst my friend group. My friends and I had very similar opinions on the album and I was curious to see whether Twitter users were vocalizing those same thoughts.

Sentiment analysis is commonly used among businesses. Marketing departments can use the data from sentiment analysis to gauge the consumers' overall feelings towards a product and gain insight on how consumers respond to certain marketing campaigns.

After using the "#GlorySoundPrep" to search 200 tweets, with the installation of Tweepy and TextBlog I was able to mine the text into three categories: positive tweets, neutral tweets, and negative tweets.

2. Implementation

First I began by installing Tweepy, which is a python library to access the Twitter API. This will allow us to gather the tweets that are needed to conduct the sentiment analysis. Next I installed Textblob, which is also a python library but the difference is that it is used to analyze textual data. Textblob provides a channel to execute common natural language processing tasks like sentiment analysis.

Next I created a class that would then be able to call multiple methods with different functions. First I began by simply just retrieving and printing 5 tweets that contained '#GlorySoundPrep'. Once I was able to retrieve these tweets, I could then begin the classification process by using TextBlob. If the polarity of the tweet was greater than zero, it was labeled as 'positive', if the polarity of the tweet was equivalent to zero, it was labeled as 'neutral', anything else was returned as 'negative'.

After the tweets were being retrieved and classified into their respective categories, I had to present the data and information in an effective way. In terms of design, there were many options that could have been explored. Ultimately, I ended up calculating the percentages of positive and negative tweets and then displayed the first five of each category. Another effective method

would have been presenting the most popular, whether that be the most retweets or likes, tweets. Popular tweets are a good representation of what the general population believes. Another alternative would have been to present tweets with the most replies that were opposite in polarity. Controversial tweets could provide interesting insights.

3. Results

After analyzing 200 tweets containing '#GlorySoundPrep', I found the following percentages:

```
Positive tweets percentage: 49.12280701754386 %
Negative tweets percentage: 8.771929824561404 %
```

About 50% of the tweets analyzed were considered to be positive, while about 9% were considered to be negative. This would leave the remainder of the tweets to be marked as neutral. Some of the positive tweets include:

```
I wish I had something clever to say about @jonbellion new album #GlorySoundPrep but I don't. It is beautiful
Jon Bellion is one of God's special gifts to the world!
Kept finding more things to like at every listen…
Want to have a good weekend? Go listen to @jonbellion 's new album #GlorySoundPrep
When someone says Blu isn't the greatest thing to ever have blessed my speakers #GSP #GlorySoundPrep
```

Unsurprisingly, all of the returned tweets were very positive and very much praised the newly released album. The program takes an interesting turn when presenting the negative tweets. Here are some of the negative tweets that were included in the 8.77%:

```
Who told @jonbellion to hit us so hard with this album? Please tell them thank you for me #GlorySoundPrep
Jon Bellion's new album is RIDICULOUS #GlorySoundPrep
The Human Condition almost made us forget that @jonbellion, on some days or every other day, considers himself a rapper. #GlorySoundPrep
that last beat, thats only played for a few seconds before it changes on "Adult Swim", that shit sounds like it was produced in orbit
My personal favorites from #GlorySoundPrep have to be: Let's Begin, Stupid Deep, The Internet, and Adult Swim!
```

The program actually categorized these tweets as negative, when in fact they are actually positive. This then forces me to question whether or not the percentages calculated are accurate.

4. Reflection

I am unsure as to how accurate my sentiment analysis is because the "negative" tweets that were returned were actually all positive, but my code perceived these tweets as negative. Here is an example of a positive tweet that the program labeled as negative:

Jon Bellion is not a human. 😭😭 #GlorySoundPrep

The program interpreted this tweet negatively or as an insult, but in reality the user intended this comment to fall into a category of high praise. This tweet implies that the user believes the artist to be so talented that he could not be on a normal human level, but because the program simply analyses the text, intentions, sarcasm, metaphors, etc. will not be taken into consideration. This program would have to be strictly used to analyze simple, black and white tweets. It would be interesting to program machines to understand human language beyond the surface level.