

## Event detection using twitter sentiment analysis

People express joy, happiness, frustration, anger through tweets. The sentiments of the users are reflected by the words in their tweets. Twitter sentiment analysis performs an in-depth analysis of the words in the tweet to predict the sentiment of the user posting the tweet. The different attributes of the words in the tweets acts as features, and they can be used to classify a tweet.

People are very passionate about sports and their sentiments can be analyzed by performing an in-depth analysis of their tweets about sports. This can provide us with a good platform to perform sentiment analysis.

### Problem:

To detect the major events in the superbowl 49 game between Seattle Seahawks and New England Patriots.

### Methodology:

The steps are listed below:

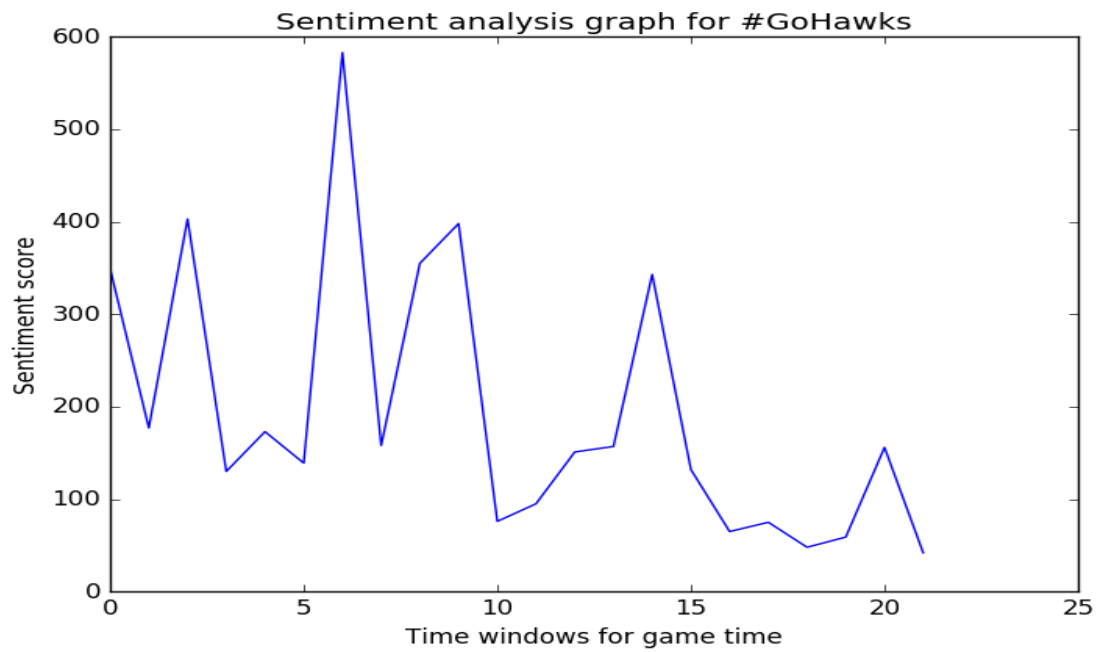
- Consider only the tweet posted for the #GoHawks and #GoPatriots data set in the window of the game time.
- Next, we use an open source dictionary of positive and negative words to assign each tweet a positive score and negative score. This score indicates the number of each kind of words which show up in the tweet. I used a “dictionary” which I had downloaded couple of months back.
- After this each tweet is assigned a class by comparing the tweet’s positive score to its negative score. If a tweet has more positive words than negative, it is assigned a sentiment score of +1 indicating a positive tweet. If the tweet has more negative than positive words, it is assigned a sentiment score of -1 indicating a negative tweet. If the positive and negative score for a tweet are equal, then it is assigned a sentiment score of 0 indicating it is a neutral tweet.
- Next, we aggregate the scores into bins of 10 minutes for each dataset over the duration of the game.

The table below shows shows the time during the game when each team scored a touchdown. These are the events that we try to detect or label using our sentiment score plots.

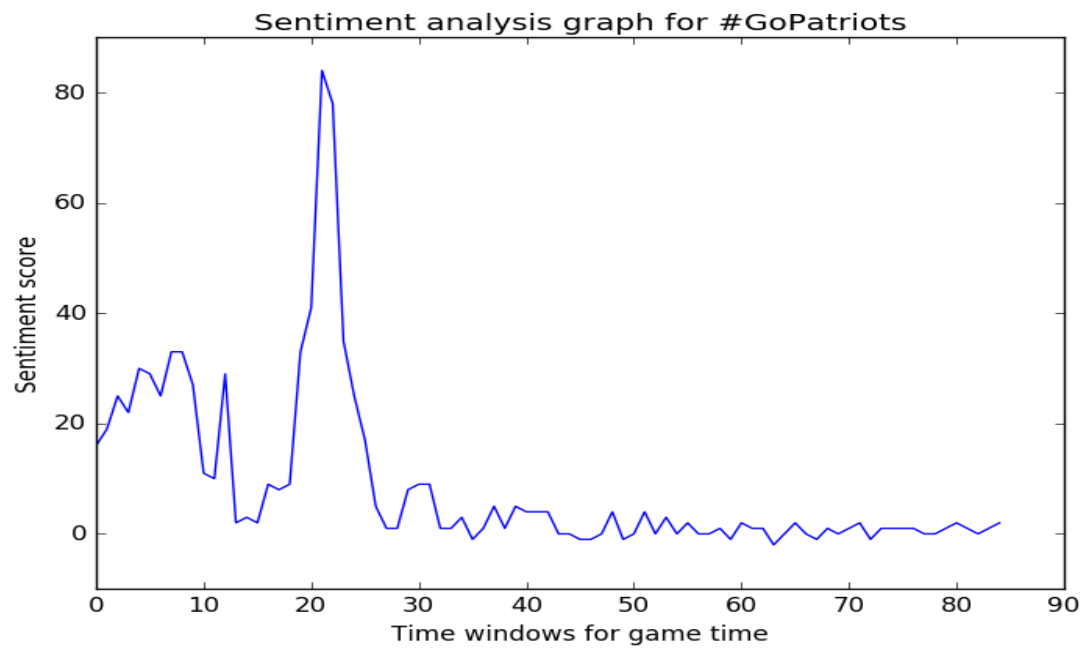
Hawks touchdown times	Patriots touchdown times	Corresponding 10 minute window
60 min	40 min	4 <sup>th</sup>
85 min	75 min	7-8 <sup>th</sup>
145 min	175 min	17-18 <sup>th</sup>
	195 min	19-20 <sup>th</sup>

## Preliminary Results

- #GoHawks



- #GoPatriots



## **Conclusions:**

### **1) Sentiment analysis for #GoHawks:**

At the 4th time window there is a drop in the sentiment score, which matches up with the time when the Patriots score a touchdown. This is followed by peaks at the 6th, 8-9th and 14-15th window indicating touchdowns scored by the SeaHawks. Finally, the sentiment score drops sharply towards the end of the game signifying the final touchdown scored by the Patriots and the SeaHawks losing the superbowl.

### **2) Sentiment analysis for #GoPatriots:**

We can see that the sentiment score increases at the 4th time window and window 7-8 which indicates the times when the Patriots score touchdowns. We can also see that there is a sharp decrease in the sentiment score during window 8-9 and window 14-15 which correspond to times when the SeaHawks score touchdowns. Finally, we see a tremendous spike in the sentiment score in the 19-21 window indicating the final touchdown scored by the Patriots and the Patriots winning the superbowl which is obviously marked with a sizable number of positive tweets by Patriots supporters.

Sentiment analysis or Natural Language Processing is a key research topic for most Data-Science jobs. Over the years, tweet blogspot website has posted hashtag analysis for elections and general opinion on topics like Climate Change. I would like to work on project like this analysis. In general develop a business or economic metric on rich data.

## **Dictionary Reference**

1) Bing Liu, Minqing Hu and Junsheng Cheng. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan.