# Sentiment Analysis of Twitter and How it Relates to Illini Football

STAT 385 FA2018 - Team Poisseidon

*Joe Broton - jbroton2*
*Braden Gebavi - gebavi2*
*Brad Gibbons - btgibbo2*
*Tim Soo - soo3*

*November 13, 2018*

**Abstract**

The purpose of this project is to explore sentiment analysis of twitter and how it relates to the illini football team. We will be analyzing the tweets regarding the last game against Northwestern on 11/24. This is in order to see how fans and students view the illini football team and their opinions on Lovie Smith. Lovie Smith has a 6-year contract with the Illini and after 3 years we have failed to make a bowl game.

# Contents

# 1  Introduction

The topic we are addressing is sentiment analysis on live Fighting Illini Tweets. The idea is to analyze live tweets with sentiment analysis in order to further understand people's reactions to the Fighting Illini and Coach Lovie's performance.

Lovie Smith has previous ties to the state of Illinois from his head coaching tenure for the Chicago Bears; he was a very splashy hire back in 2016 and expectations were high coming in. So far, the Illini football program has found little success and the sentiment around Lovie seems to be changing. There are articles written about this topic including Chicago Sun Times publishing an article that talks about Lovie Smith's Job security. People are speculating that he will be done after a 4th season if there isn't improvements as there hasn't been any in the past 3 seasons. The current generation of students at Illinois probably remembers Lovie for being the coach that led the Bears to their second ever Super Bowl in 2006 so it is interesting see what kind of reactions fans are tweeting about him now that he is struggling to revive the university's football program. If the struggles continue, there is an expectation that he will be replaced.

The topic came from just naturally being in our school and personal environments. Whether it is going to class or watching the game with friends, students and fans do not wish our football team was this bad. Given his previous success with the Chicago Bears up north, we hoped Lovie Smith would be the change needed. After seeing little to no improvement, we wanted to confirm our suspicions that students and fans are as displeased with the team on twitter as they are in the conversations we hear. There seemed to be a buffer period in his first few seasons where fans gave him the benefit of the doubt, as he was getting used to the college scene, but we have a feeling that period may be over and fans' patience are wearing thin.

We plan on extracting live tweets and matching words/portions with positive/negative/neutral words. From this, we are able to determine if if each individual extracted tweet is overall positive/negative/neutral, by assigning positive words ranging from 1 to 5, negative words ranging from -1 to -5, and neutral words to 0, and finding the overall sum. Ultimately, we want to apply the extracted tweet, the sentiment analysis, as well as diagnostic plots into an interactable shiny app to determine overall reactions to the team and coach.

Our ultimate goal is to determine the overall fanbase opinion on the team and Lovie. We plan on using bar plots that highlight the occurrences of certain keywords and/or feelings. We are also considering creating a wordgraph that increases/decreases the size of particular words based on occurence. At the end, we plan on creating an RShiny app that will be interactive, allowing the user to look at requested graphs with drop down boxes that allow for variation when choosing which words to plot on our graphs, as well as which graphs we want to produce or hide.

Upon further discussion, we determined one of the better methods of implementing our findings into a Shiny app, would be to split the data into two columns, one for NAME, and the other for the associated SCORE. We noticed that problems arise as the values in our package, were predefined by some person, or some group of people. However, everyone is different, and feel different about certain words. For example, if we have 'decent' as one of the words, some may associate it as being positive, while others may associate it as being negative. To fix this, we came up with a couple methods: 1.Implement a drop down menu, set the default scores as the scores from the package, but offer a user input that allows the user to select certain words, and assign different values to them (that differ from whatever the package scores list them as). 2.Provide the top 50 words from the package dictionary, but only display 10 at a time, with 'next' button that allows the user to move from page to page in order to avoid clutter in the drop down menu.

Sentiment Analysis requires an understanding of Regex. The reason for this, is to identify patterns, and using these patterns to extract specific portions of tweets that we would like to analyze. Furthermore, Regex helps us also to split and mutate certain tweets, that can further be compared to our positive, negative, and neutral words.

We also will be applying more recent topics, such as RShiny, that allows for interactive sliders and drop down menus that will assist the user in choosing exactly what they want to display. In our case, we will have options to graph specific plots (i.e bar plots) as well as specific words that we would want to appear along the x and/or y axes of our plots.

## 2  Related Work

We plan on using the information within the Text Mining with R book written by Julia Silge and David Robinson. In the book they show how to use the tidytext package which contains several sentiment lexicons within the 'sentiments' dataset. There are three general-purpose lexicons contained in the 'sentiments' package which are 'AFINN', 'bing' and 'nrc'. We plan on using 'AFINN' and 'nrc' of the lexicons as 'bing' can be considered a more specific lexicon of 'AFINN'. This will be necessary as these are essentially dictionaries of words pre-categorized which will lead to more precise data than if we were to create a dictionary on our own.

## 3  Methods

We plan on using the packages listed: "Tidytext" "ggplot2" "Rvest" "Tidyr" "dplyr" "twitteR" "stringr"

In order to run sentiment analysis, we will first need to collect the data set. This will involve using the twitter API along with the "twitteR" package to web scrape the data. Now, we will have a string of text containing what the tweet was. From there, we can use the "stringr" package and split the tweet into individual words, along with omitting some of the data like special characters, emojis, and hashtags. After that has been completed, we will need to organize it with "Tidyr" and "dplyr" in order to begin using the "Tidytext" package to run the sentiment analysis. Finally, we will create an RShiny app to display our graphs made with "ggplot2".

To create the RShiny App, we will need two main tabs. One will serve as a "Data" section, and the other will be "Visualizations/Graphs". The "Data" section can be referenced in Figure 1 in the appendix. It will have a menu on the left that will allow a user to search for a word and assign its value if the users' opinion differs from the lexicon. The words that the user has reassigned its values will be displayed to the right of the section, along with a reset button to erase user input. Below that will be a "Filter" section that for right now will only filter the most frequent words. That information will be displayed to the right of the filter section. You can divide the data even further than that by displaying positive, neutral, and negative words (default is all 3). We are considering implementing a slider which you can then further filter words, i.e. instead of positive values, you can now only show positive values 3 to 5.

In the "Visualization/Graphs" section users will be able select a graph to display from the dropdown menu. The graph will display to the right. Possible graphs include emotions in tweets, contribution of sentiment, and sentiment scores of tweets.

## 4  Feasibility

In order for this to be completed by the deadline 12/18 this is how the work must be split up and be completed by.

Joe - Web Scraping the Tweets after the Illini Football Game - by 11/27/18
Braden, Brad - Organizing the data - by 11/30/18
Joe, Tim -Running sentiment analysis on it - by 12/5/18
Brad, Braden - Creating the plots for our R Shiny App - by 12/6/18
Joe, Tim- Create an R Shiny App that displays the information we have gathered - By 12/15/18
Joe, Tim - Create a demo video - 12/16/18
All - Finalizing the report - 12/17/18

# 5  Conclusion

To summarize, our objective is to use sentiment analysis of UIUC football tweets, to determine the overall reactions from the fan base over the performance of the Fighting Illini and Coach Lovie Smith. Because our project is an implementation of sentiment analysis, our goal is to simply analyze fan base reactions, determining if people are feeling positive or negative about the overall football season, rather than identifying a problem. Specifically, our project helps to identify how the Fighting Illini and Coach Lovie are performing by analyzing a score that indicates positive or negative reactions. That being said, if required to identify a problem, it would be that the Fighting Illini/Coach Lovie have been underperforming. Despite Lovie's NFL coaching experience and previous ties to the state of Illinois, the solution may be to potentially release Coach Lovie Smith.

# 6 Appendix

## 6.1 Example Data

```
##  [1] "I know it hurts me....everything that was told to me was definitely not true. Not to mention t
##  [2] "The Citadel had a better performance away against #1 Alabama than @IlliniFootball is having ag
##  [3] "Hey @IlliniFootball & @LovieSmith get a real defensive coordinator and special teams instead o
##  [4] "I'm a diehard @IlliniFootball fan and I don't think I can watch this garbage @LovieSmith #wesu
##  [5] "Will  #Illini fire Lovie Smith at the end of the Season? @IlliniFootball"
##  [6] "And that it. The worst loss in @IlliniFootball history.\n\nUnforgivable."
##  [7] "PICK!\n\n@IlliniFootball starts fast with a Del'Shawn Phillips interception: pic.twitter.com/7
##  [8] "13:42 Q2 | Touchdown Iowa. Fumble recovered by AJ Epenesa and returned for the TD. PAT good. \
##  [9] "13:50 Q2 | 3-yard TD run by Sargent for Iowa. PAT good. \n\n@HawkeyeFootball - 14\n@IlliniFoot
## [10] "Do I have to have another tweet rant about why we shouldnt be punting or do we get it now? @Lo
## [11] "END Q1\n\n@HawkeyeFootball - 7\n@IlliniFootball - 0\n\n#Illini"
## [12] "HALFTIME\n\n@HawkeyeFootball - 35\n@IlliniFootball - 0\n\n#Illini"
## [13] "8:11 Q3 | Touchdown Iowa. 5-yard TD run by Toren Young. PAT good.\n\n@HawkeyeFootball - 49\n@Il
## [14] "12:19 Q1 | AJ Bush Jr. sacked at the Iowa 44. Turnover on downs. \n\n@HawkeyeFootball - 0\n@Ill
## [15] "I was pretty interested in finding some good football to watch, until I remembered that it's C
## [16] "Freakin @IlliniFootball has given up on @LovieSmith #embarrassing"
## [17] "14:19 Q1 | @IlliniFootball's Del'Shawn Phillips intercepts an Iowa pass at the Iowa 41. 2-yard
## [18] "13:26 Q3 | Touchdown Iowa. Illinois punt returned 61 yards by Kyle Groeneweg for the TD. PAT i
## [19] "#ILLINI @IlliniFootball 63 to 0. Coaches need to take ownership, but more importantly ADs need
## [20] "#illni something has to give.  You can't keep giving up 40-60 points per game.  Where is the d
```

These were some of the tweets regarding the last game against Iowa where the Illini lost 63 - 0. As you can see some of the tweets already show disdain against Lovie Smith and the team. Fans are embarrassed and some are saying they will not renew season tickets until the defensive coordinator is fired. We will categorize the data for sentiment analysis putting the words in the tweets into categories such as "anger", "negative", "positive", "fear", "trust", etc. By breaking down the tweets like this we will get a better understanding of the meaning of each tweet without having to manually go through each one and decipher them.
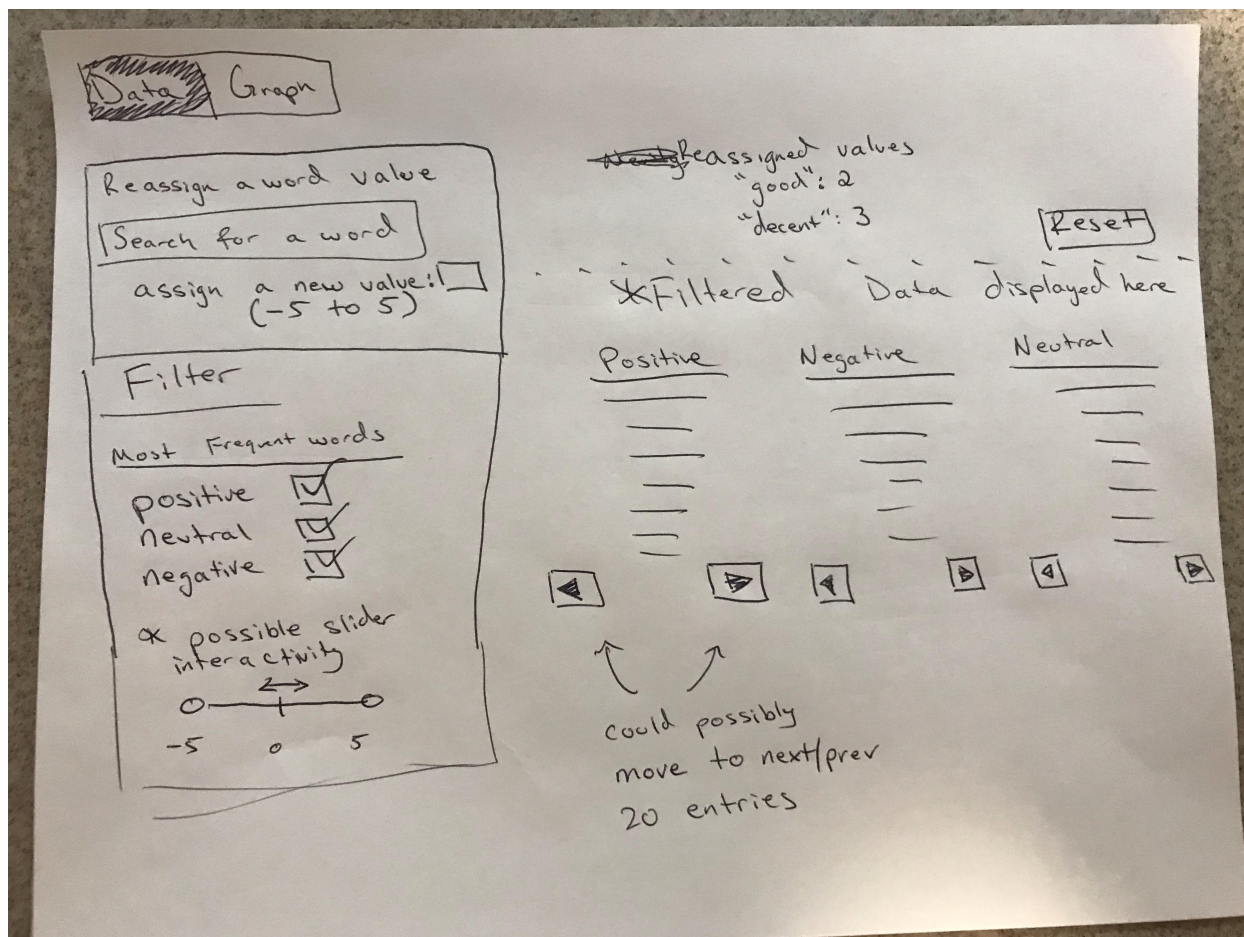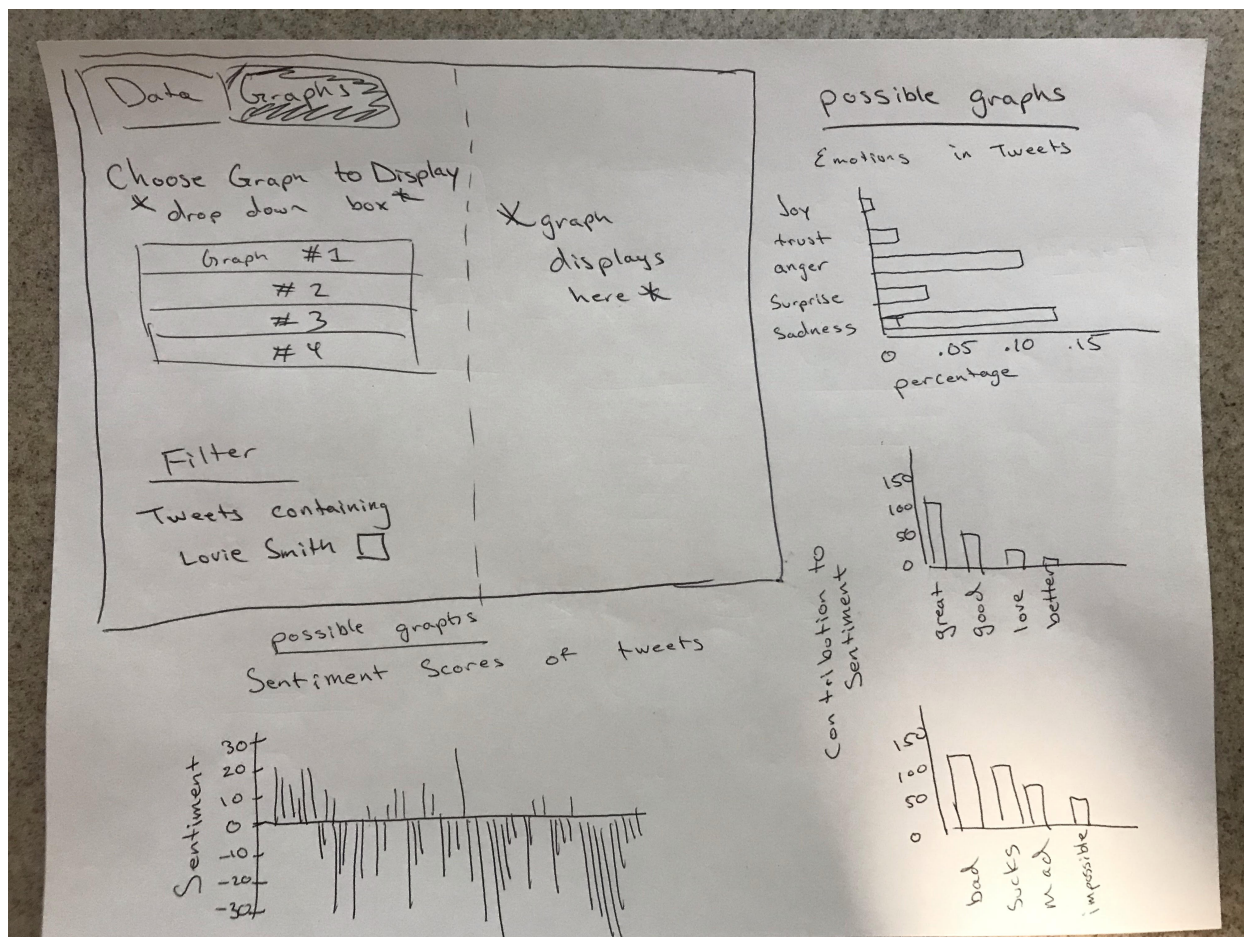
Figure 1: Example RShiny App; Data Section

Figure 2: Example RShiny App; Graph Section

```
##
## To cite ggplot2 in publications, please use:
##
##   H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
##   Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     author = {Hadley Wickham},
##     title = {ggplot2: Elegant Graphics for Data Analysis},
##     publisher = {Springer-Verlag New York},
##     year = {2016},
##     isbn = {978-3-319-24277-4},
##     url = {http://ggplot2.org},
##   }
```

# 7   References

https://www.tidytextmining.com/sentiment.html

http://bogdanrau.com/blog/collecting-tweets-using-r-and-the-twitter-search-api/

http://dataaspirant.com/2018/03/22/twitter-sentiment-analysis-using-r/

https://chicago.suntimes.com/sports/lovie-smith-illinois-illini-northwestern-wildcats/

https://newsok.com/article/5613773/smith-an-nfl-winner-hasnt-won-at-illinois

Silge J, Robinson D (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *JOSS*, *1*(3). doi: 10.21105/joss.00037 (URL: http://doi.org/10.21105/joss.00037), <URL: http://dx.doi.org/10.21105/joss.00037>.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham and Lionel Henry (2018). tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. http://tidyr.tidyverse.org, https://github.com/tidyverse/tidyr.

Jeff Gentry (2015). twitteR: R Based Twitter Client. R package version 1.1.9. https://CRAN.R-project.org/package=twitteR

Hadley Wickham (2018). stringr: Simple, Consistent Wrappers for Common String Operations. http://stringr.tidyverse.org, https://github.com/tidyverse/stringr.