Sentiment Analysis of Tweets on Egypt in World Cup June 21 - June 25, 2018 before Saudi Arabia Match

Zeyad Kelani

June 26, 2018

***TEST PROJECT - written in LaTeX**

1 Disclaimer

This Demo project was made 7 hours before the Egypt Vs. Saudi Arabia game (PST) that took place on June 25, 2018 in the Russia 2018 World Cup. This project was complied just to test "R Codes" in extracting and cleaning tweets in English language.

2 Introduction

Collected random n=500 tweets using Standard Twitter API (https://developer.twitter.com) for the sake of analyzing sentiments based on two hashtags #Egypt and #World Cup starting June 21-25, 2018 in English language.API returned only randomly selected 500 tweets.

Performed data cleaning on tibble format using R(tiday), R(tm) and R(gsub) packages. Sentiment scores are calculated in two different ways: bing and nrc.

The bing lexicon categorizes words into a positive or negative categories; the nrc lexicon categorizes words into emotions like anger, disgust, fear, joy, ...etc.

3 General

Figures below show the most repeated words in returned tweets, word cloud based of frequency>15, world cloud for visualization, and most repeated unique words.

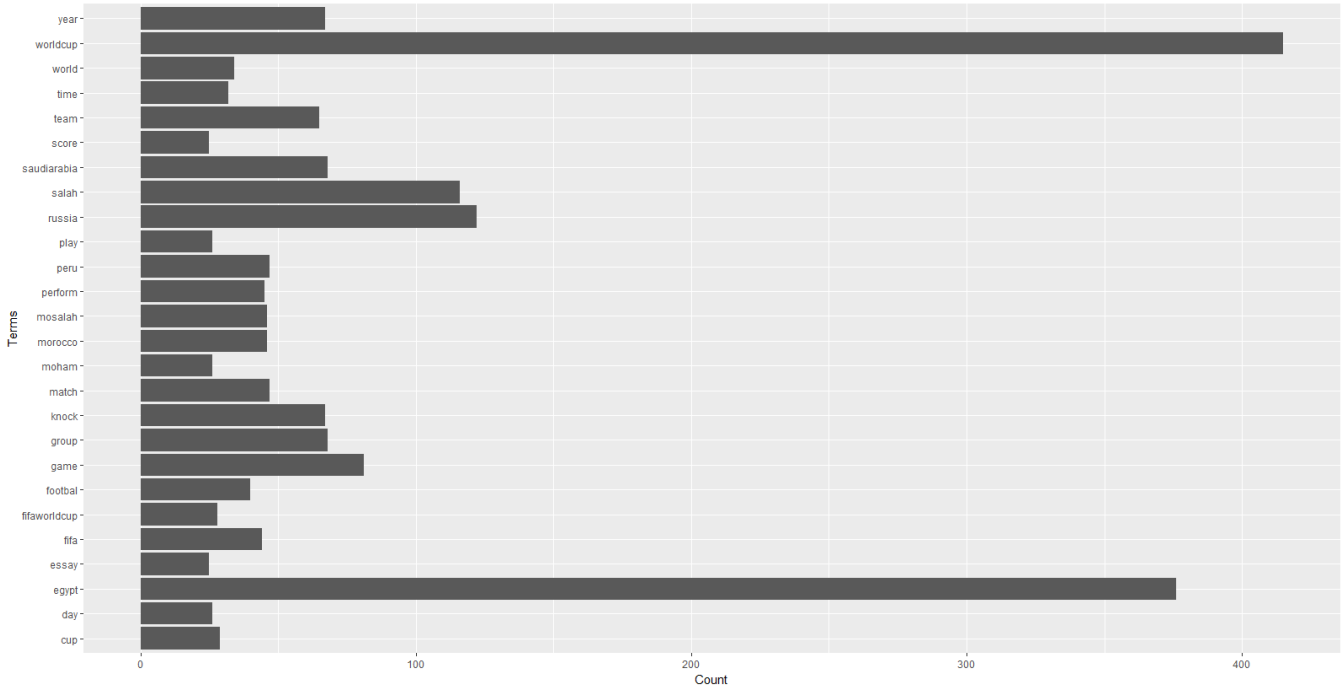

Figure 1: Most Repeated Words

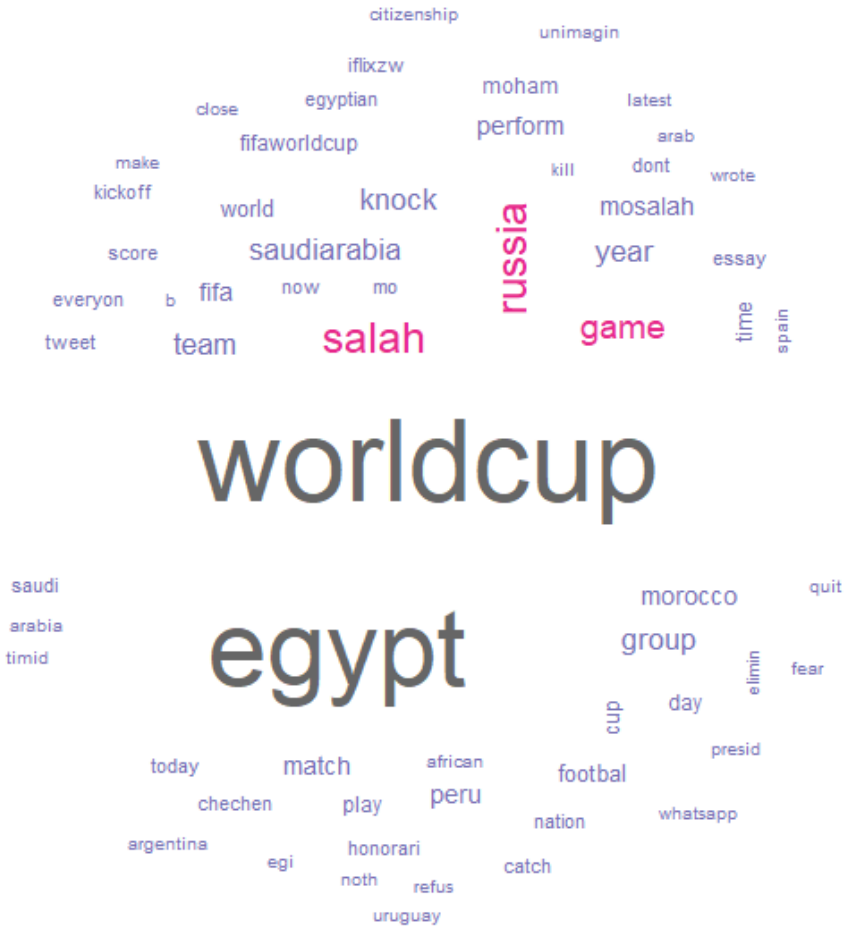

Figure 2: Word Cloud

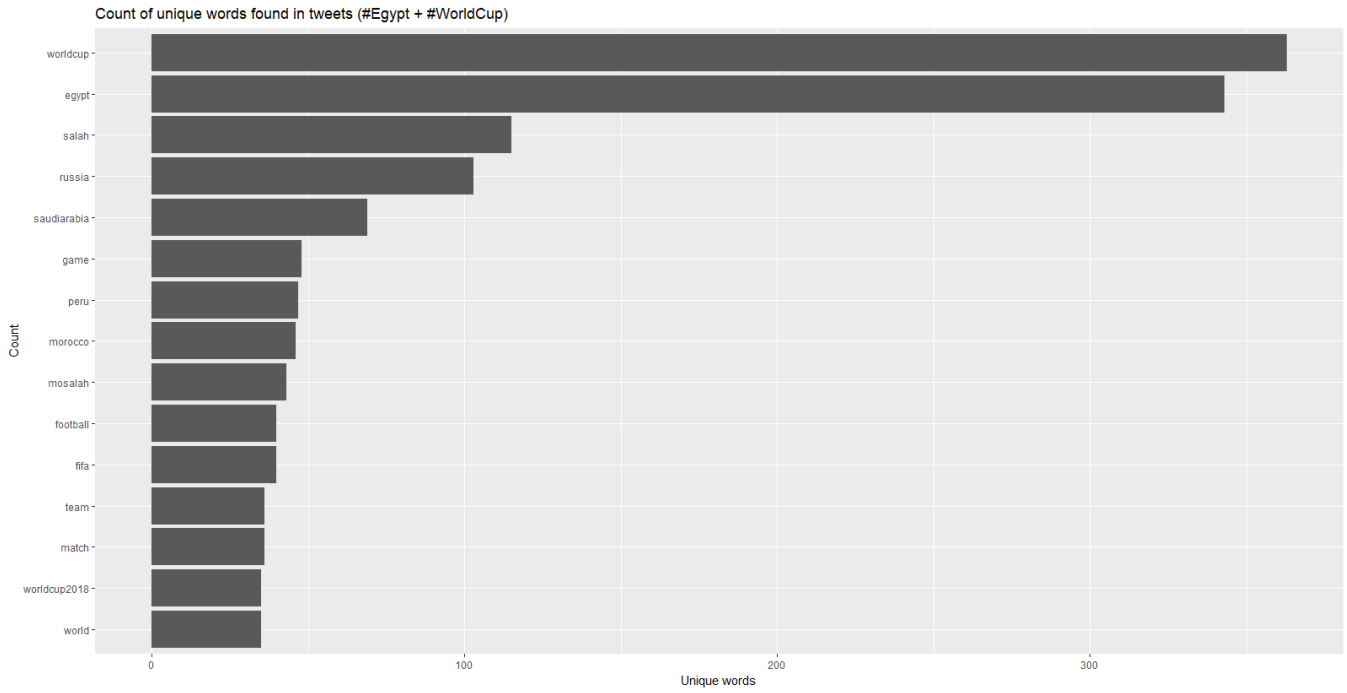

Figure 3: Most Repeated Unique Words

4 nrc

Figures below show topic classification based on latent Dirichlet Allocation and sentiment scores from nrc lexicon dictionary.

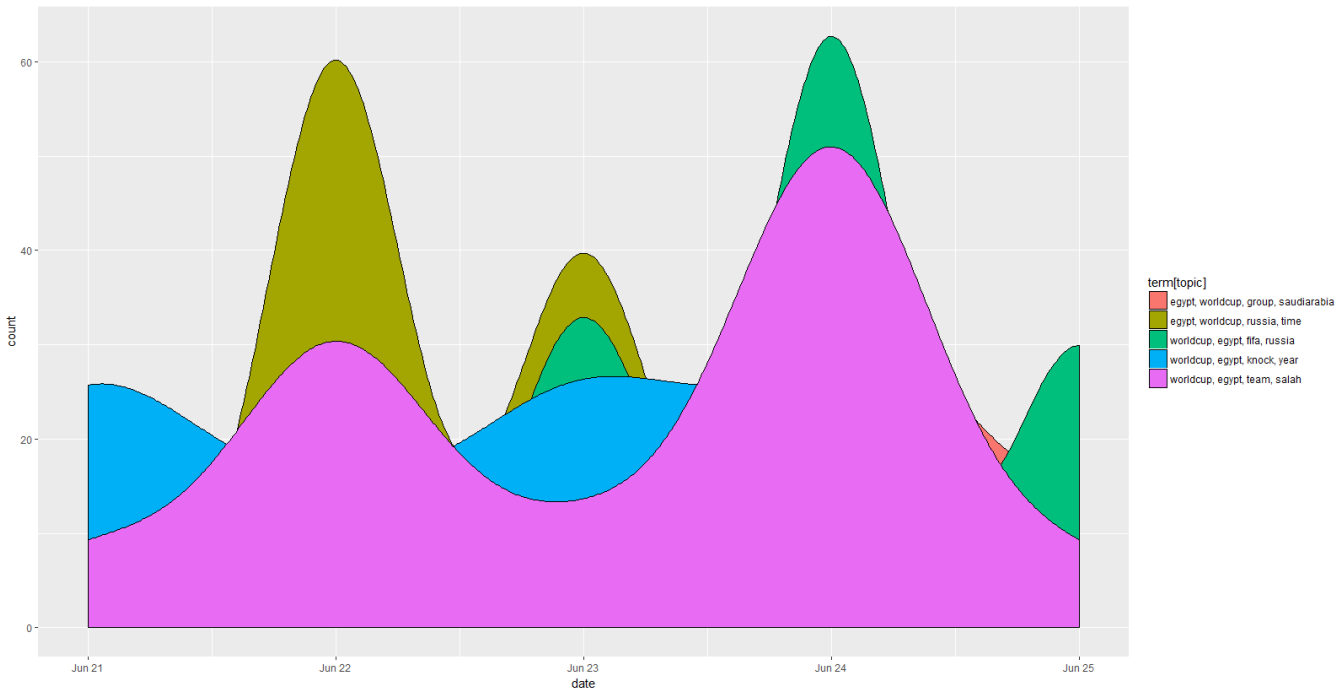

Figure 4: Topic modeling based on latent Dirichlet allocation

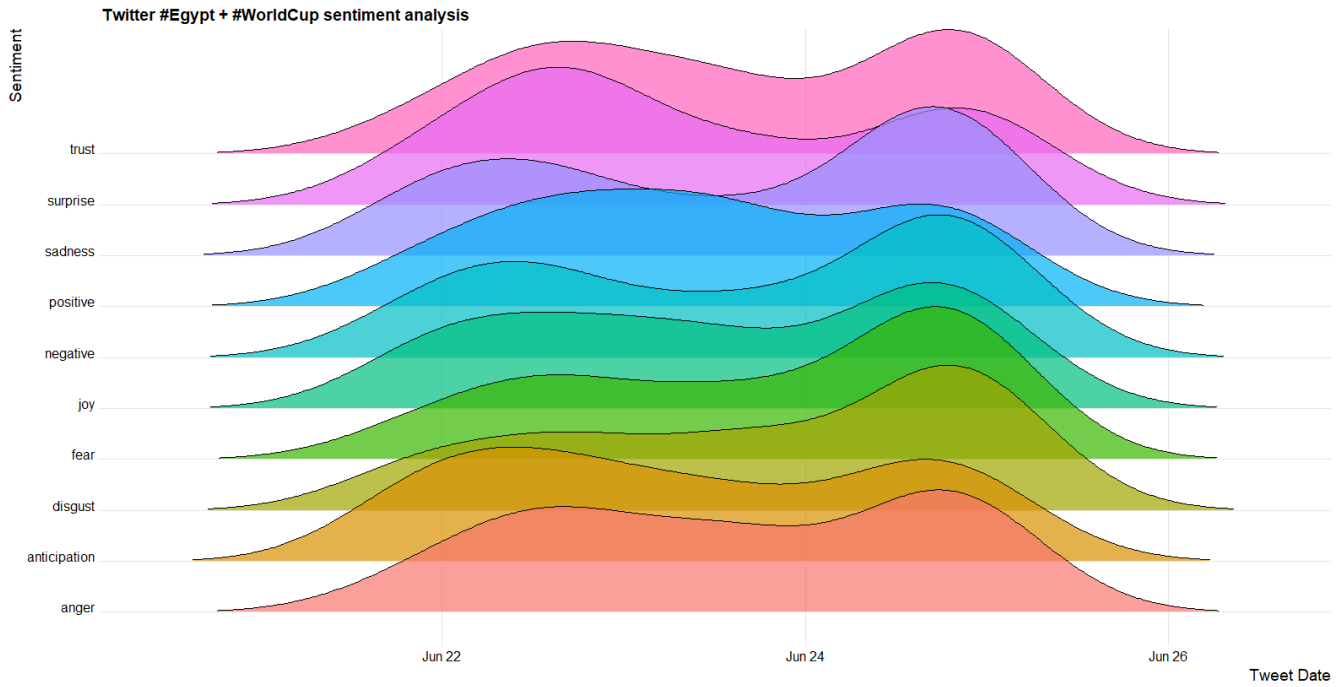

Figure 5: nrc Sentiment Emotions of Tweets over Time

5 bing

Figures below show bing scores for extracted tweets over time and words used in extracting such scores.

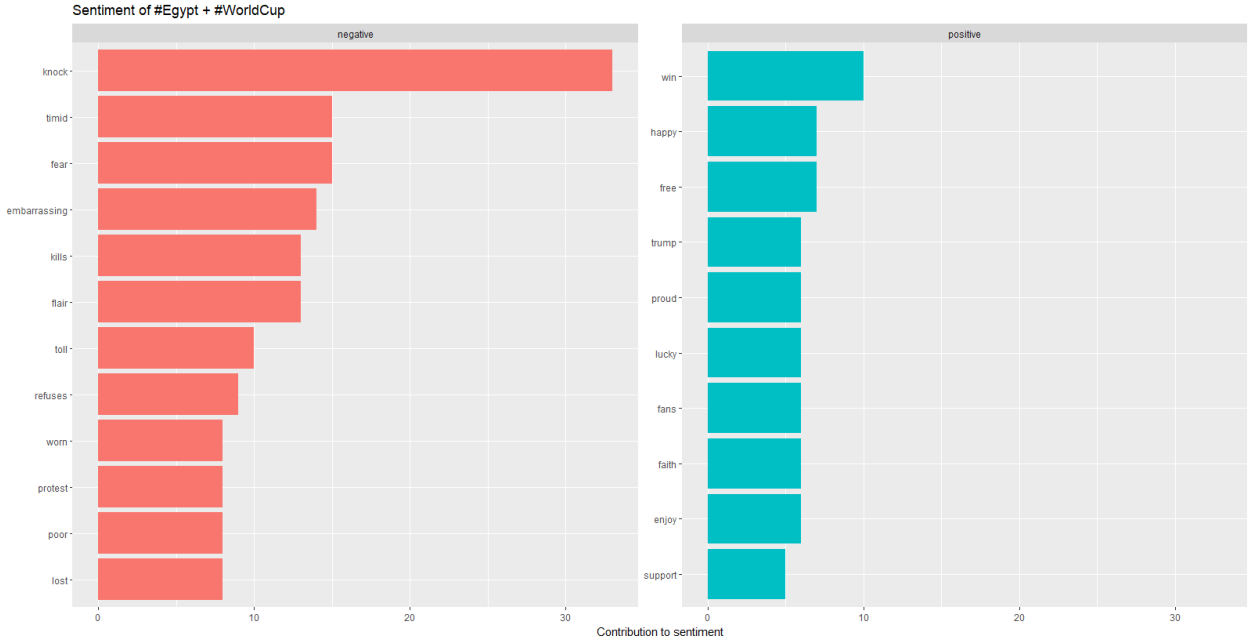

Figure 6: bing Classification of Words

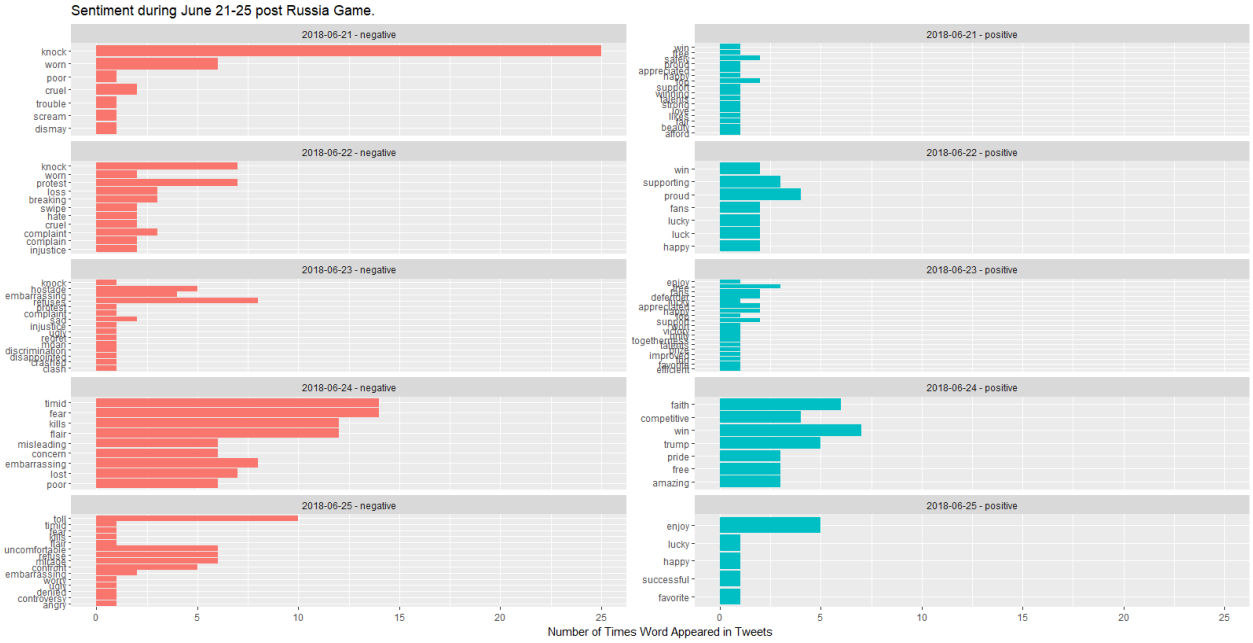

Figure 7: bing Scores by Day

6 Remarks

This project is limited in generalization since it tracks only tweets referencing couple of hashtags written in English, thus it does not reflect Egypt sentiments rather random collective reaction to Egypt's performance in world cup.

Next steps, test codes for analyzing Arabic Tweets using UTF-8 encoding, increase-N, assign ideal points estimations for users, predict topics based on latent variables.