

**DEPARTMENT OF ENGINEERING AND
TECHNOLOGY**



**BUSINESS ANALYTICS WITH R
CS2001**

ISIS TWEETER ACTIVITY

~SEMANTIC ANALYSIS/TEXT MINING~

Submitted by: Saumya Patel

Roll No.: 2019BtechCSE049

ABSTRACT

This project report contains the Exploratory Data Analysis of a dataset named “How ISIS uses twitter tweets” which includes about 11000 tweets from more than 100 tweeter users from all over the world .

INTRODUCTION

ISIS OVER SOCIAL MEDIA

The Islamic State, known as ISIS or ISIL, has exploited social media, most notoriously Twitter, to send its propaganda and messaging out to the world and to draw in people vulnerable to radicalization. By virtue of its large number of supporters and highly organized tactics, ISIS has been able to exert an outsized impact on how the world perceives it, by disseminating images of graphic violence (including the beheading of Western journalists and aid workers and more recently, the immolation of a Jordanian air force pilot), while using social media to attract new recruits and inspire lone actor attacks. Although much ink has been spilled on the topic of ISIS activity on Twitter, very basic questions remain unanswered, including such fundamental issues as how many Twitter users support ISIS, who they are, and how many of those supporters take part in its highly organized online activities. Previous efforts to answer these questions have relied on very small segments of the overall ISIS social network. Because of the small, cellular nature of that network, the examination of particular subsets such as foreign fighters in relatively small numbers, may create misleading conclusions. The information vacuum extends to—and is particularly acute within—the sometimes heated discussion of how the West should respond to this online campaign. While there are legitimate debates about the bounds of free speech and the complex relationship between private companies and the public interest, some have argued against suspending terrorist social media accounts on the basis that suspensions are not effective at impeding extremist activity online. These arguments that are usually predicated on very small samples of potentially misleading data, when data is proffered at all.

From 2013 to 2014, the organization primarily used mainstream platforms such as Twitter, Facebook, and YouTube. In 2014, these large social media platforms removed ISIS content. Since then, ISIS has chosen to utilize social media platforms that either protect their content or allow for content to quickly be reposted.

TWITTER

During 2014, there was an estimated 46,000 to 90,000 Twitter accounts that advocated for ISIS or were run by supporters of the group.^[11] In 2015, Twitter reported that it banned 125,000 ISIS sympathetic accounts.^[9] In 2016, it published an update of 325,000 deleted accounts.^[12]

Though many accounts have been suspended, ISIS supporters often create new accounts. Twitter defines those who recreate accounts as “resurgent’s” and explains

that these are often difficult accounts to remove completely, since they tend to pop back up in alternate forms. It is estimated that approximately 20% of all ISIS affiliated Twitter accounts can be traced back to fake accounts created by the same user. Many of these accounts are traced back to the “Baqiya family,” which is an online network of thousands of ISIS followers.^[10] Many of these accounts are active during important ISIS military victories. During the ISIS march on Mosul, there were about 42,000 tweets on Twitter supporting the invasion

SENTIMENT ANALYSIS

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. However, analysis of social media streams is usually restricted to just basic sentiment analysis and count based metrics. This is akin to just scratching the surface and missing out on those high value insights that are waiting to be discovered. Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative or neutral.

TEXT ANALYSIS

Text mining, also referred to as text data mining, similar to text analytics, is the process of deriving high-quality information from text. It involves the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources.

WORDCLOUD

A word cloud is an image made of words that together resemble a cloudy shape. The size of a word shows how important it is e.g. how often it appears in a text — its frequency.

People typically use word clouds to easily produce a summary of large documents (reports, speeches), to create art on a topic (gifts, displays) or to visualise data (tables, surveys).

ABOUT THE DATASET

Dataset named “How ISIS uses twitter tweets” which includes about 11000 tweets from more than 100 tweeter users from all over the world. This dataset was found on data.world (<https://data.world/socialmediadata/how-isis-uses-twitter/workspace/file?filename=how-isis-uses-twitter%2Ftweets.csv>).

The original csv file contains 8 columns of data with 11347 rows (excluding header row). The columns with and their data types are as follows:

* **`name`**: Looks like a combination of real and fake names of the tweet poster. Not sure if one can input fake names when register on Twitter. There are 112 unique names.

* **`username`**: Twitter usernames. Very straight forward. There are 112 unique usernames. Maybe matching names above?

* **`description`**: Seems to be the short paragraph located right beneath one's profile where he/she describes who he/she is, or the identity he/she would like to be viewed with. Interestingly, there are only 78 unique descriptions among these 112 users.

* **`location`**: This is the user's location where one can put on their tweeter profile. Interestingly, there are only 50 unique locations, with most being NA, and the second most popular being "Read my blog."

* **`followers`**: Number of followers the person had when the tweet was posted, with max being 34,690 and min being 16. No NA. Interesting.

* **`numberstatuses`**: Number of statuses the user had posted by the time this tweet was posted, with max being 33,091 and min being 1 => those who only tweets once and spotted as ISIS fanboy...

* **`time`**: Timestamp of the tweets in the dataset, in "Month/Day/Year Hour:Minute" format. Should be easy to parse.

* **`tweets`**: The meat in the dataset, 11 of which seems to be translated into English from the original version.

BUSINESS OBJECTIVE

Identifying violent extremists

BUSINESS CONSTRAINTS

No innocent user should be analysed wrongly

ANALYSIS

Our objective in this analysis is to analyse the dataset given, refine it and do text analysis on it. We tried using sentiment analysis over the tweets to identify or classify tweets as positive, negative or neutral. On that basis we tried to identify the real user for the given username. Using text analysis, we highlighted the important words used in the most with highest frequency, we plotted 2 bar graphs to analyse the frequency of those words. A word cloud is also plotted to show the most highlighted words in the tweet.

Due to time limitation, and unavailability of .txt files for ppositive, negative and neutral words, we were not able to identify the semantics of the tweets.

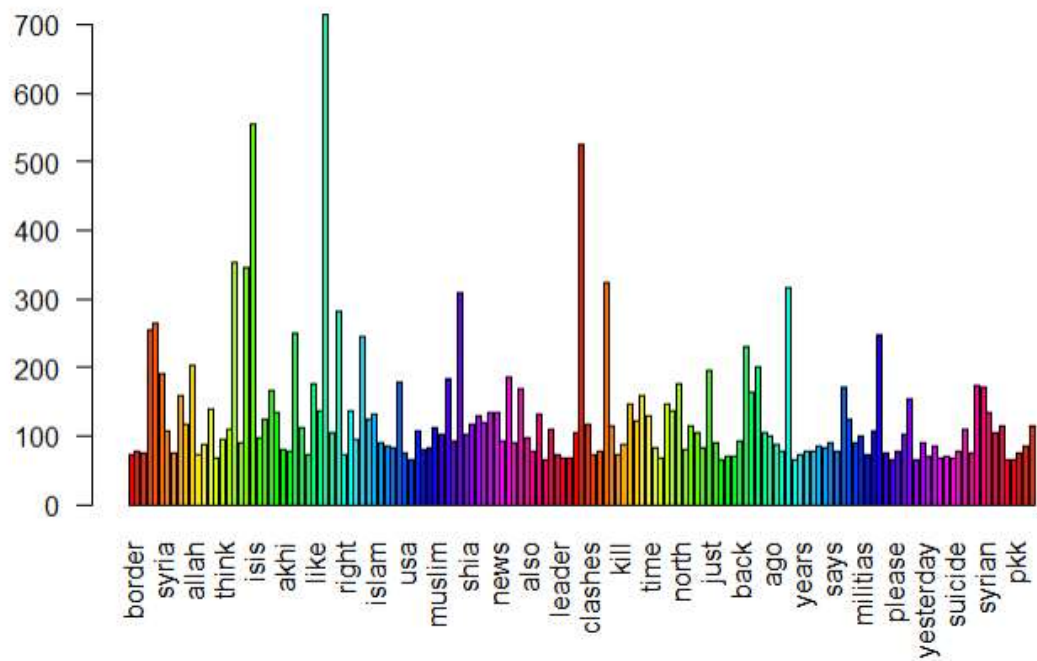


Figure 1 : Frequency of the all the words in the tweets

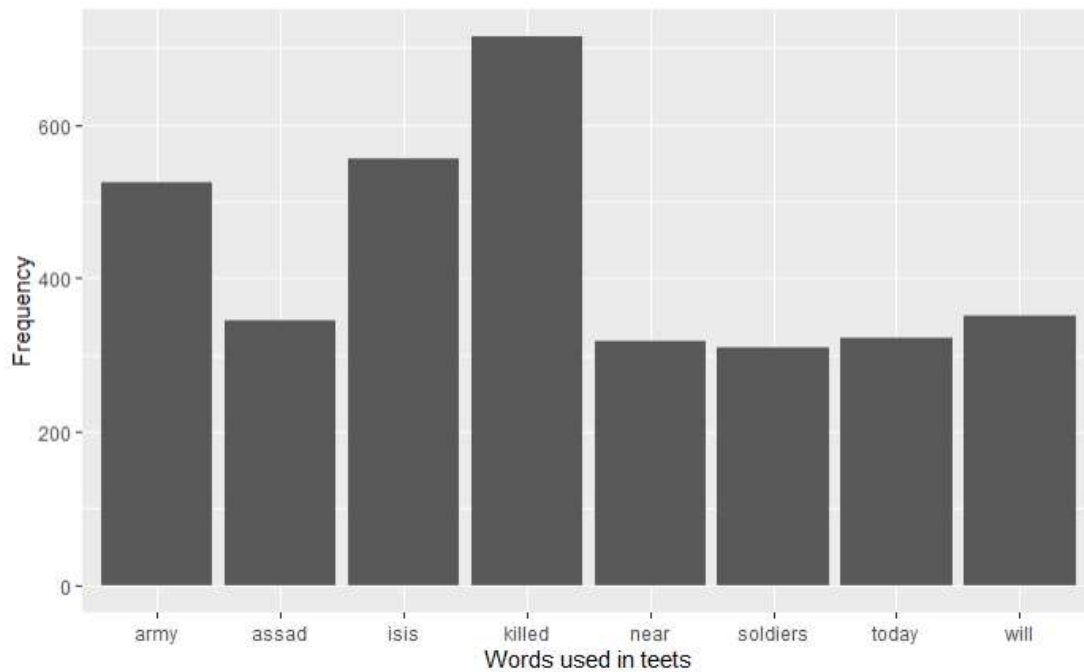


Figure 2 : Words whose lowest frequency is 300

