Descriptive Analysis of Suicide Ideation on Twitter

Lipisha Chaudhary, Vignesh Vijaykumar Nair, Ishika Prasad
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623
{lc2919;vn3246;ip1262}@rit.edu

ABSTRACT

The start of the online social networks have caused an spike in interactions between people. While this has a positive impact in society, it can also spread the ideations of suicide leading to the contagion effect.We used live twitter data for our database along with previously labelled twitter data and draw conclusions about the data to gain an insight on the possible behavior of users who are going through suicidal thoughts. We observed that people who are popular and very active on social media are likely to be depressed. We visualized a word cloud of the most popular word addressed by said depressed people. We also performed K means clustering on the tweets and obtained 6 clusters. We believe our work has helped gain some insight on suicide ideation on twitter.

1. INTRODUCTION

In today's world, social media is one of the platforms for personal expression of thoughts, ideas and to show our individuality. We post nearly all nitty-gritty details of our everyday life, from talking about what we eat to our opinions on political matters. Especially for young adults, social media is a form of catharsis. From these posts, we can assume human emotions and what a person goes through mentally. People who go through an unfortunate episode are likely to convey their feeling through social media indirectly[4]. In comparison with users who spent more time on social media like Twitter, Instagram, Facebook, and other platforms were shown to have a considerably higher rate of reported depression than those who spent less time, according to recent studies[2]. Recently a 16 year old girl committed suicide over Instagram poll results on users voting that she should die [3].There are several triggers and indications online which can theoretically determine if a person is likely to be suicidal or not. Our project focuses on identifying individuals who are suicidal based on the content that they post. We wish to contribute to the ongoing research on detection of suicide ideation and measure the performance of four machine classifiers in their accuracy of correctly distinguishing suicide related and non suicide related posts.

In terms of Data Science objectives, our project focuses on clustering and descriptive data mining. We want to cluster the available data, analyze the resultant cluster and compare the clusters using performance metrics. We will also perform data association to get an insight into what possible factors could have a strong link between suicidal thoughts and the persons' environment. Possibly from the correlation we observe, our list of deliverables may increase. We have decided to not perform predictive data mining, since, in order to label our dataset, we need strong domain knowledge of the nuances of human nature and to have a valid experiment we would need to label at least 50,000 records and we do not have enough resources to perform the same.

2. DATA SETS

We have the following sources used for our data mining process.

- **Twitter Live Data**: Since we are sticking to Twitter and we want to be able to get as much information as possible, we decided its best to have one source right from the horses' mouth. After successfully acquiring a key for developers, we gathered data live from twitter using Tweepy package in python. To be able to do this, we needed a list of search words which can indicate depressive episodes that was used as a query to obtain raw twitter data, so we came up with 20 terms taken from [6]. Since the package Tweepy takes these terms as tokens, we did not find the need to use Regular Expressions for filtering the terms. Following are some of the terms:
    1. Reference to death: "sleep forever", "want to die", "be dead", "better off without me", "better off dead","end my life", "never wake up", "die alone", "go to sleep forever".
    2. Reference to difficulty in living: "tired of living", "don't want to be here","can't go on", "not worth living".
    3. Direct reference to suicide: "suicidal", "suicide","my suicide note", "my suicide letter","ready to jump", "suicide plan".
- **Detect Depression In Twitter Posts**:[1]: This is a semi-processed dataset found on GitHub. It focuses specifically on suicide. We decided to take this dataset to balance the previously completely raw dataset, so

that we have a rough guideline of narrowing down our scope. This is an already labelled dataset. Each tweet is classified into 0,1,-1, representing neutral, positive and negative tweets respectively.

Both the datasets are stored and merged into PostgreSQL.

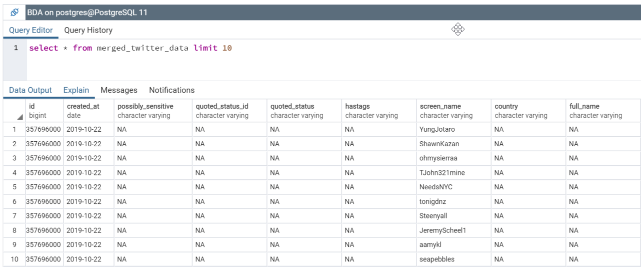
1. Format of data (Twitter Live Data): JSON
2. Format of data (Detect Depression In Twitter Posts): CSV
3. Format of Merged data: PostgreSQL

3. PREPROCESSING

Since the data from the first dataset is raw, there was a lot of cleaning required. Empty values were mostly discarded and some of the missing values were given a default while some like a missing tweet itself, was discarded. Both the datasets are merged and stored in PostgreSQL. There are 32 major attributes from the live twitter data [5]. This is available in the JSON format. Out of the 32 attributes, we decided to scrape the ones which we deemed ineffective. Below is a short description of the data that we did decide to keep:

1. **id**: 'id' gives a numeric id for any individual Twitter user.
2. **created_at**: 'created_id' is a date-time based attribute, gives information about when the tweet was tweeted by a particular user.
3. **text**: 'text' is the actual tweet tweeted by the user.
4. **user**: 'user' is the user_id of the person who started the particular tweet.
5. **source**: 'source' gives the utility where the tweet is posted.
6. **in_reply_to_status_id**: 'in_reply_to_status_id' is a integer type attribute which gives an integer value which indicates whether the tweet is replied to or not.
7. **country**: 'country' is a string type sub-attribute which gives the country from where the tweet originated from under the attribute Place.
8. **coordinates**: 'coordinates' is a integer-list type attribute which gives the coordinates of the country attribute .
9. **retweeted_status**: It is an attribute that contains a representation of the original Tweet that was retweeted
10. **truncated**: A Boolean value which indicates whether the value of the text parameter was truncated
11. **display_text_range**: Shows the text range of characters. It is also an additional indicator of tweet being truncated.
12. **Hashtags**: Lists the hashtags used.
13. **favorite_count**: 'favorite_count' is a integer type attribute which tells us how many times the tweet has been liked.
14. **Sentiment rank**: It ranks the sentiment into positive,neutral and negative.

Figures 1 and 2 provide a small preview of the merged data. The other snapshot are available in our sources folder, submitted separately.


BDA on postgres@PostgreSQL 11

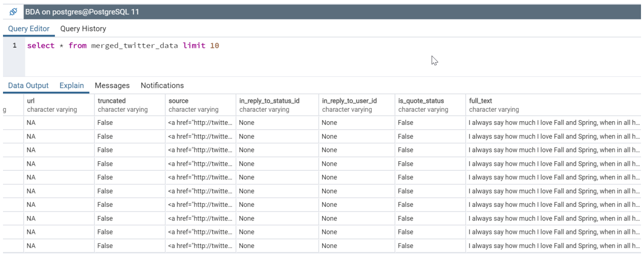
Query Editor Query History

```
1 select * from merged_twitter_data limit 10
```

Data Output Explain Messages Notifications

	id bigint	created_at date	possibly_sensitive character varying	quoted_status_id character varying	quoted_status character varying	hastags character varying	screen_name character varying	country character varying	full_name character varying
1	357696000	2019-10-22	NA	NA	NA	NA	YungJotaro	NA	NA
2	357696000	2019-10-22	NA	NA	NA	NA	ShawnKazan	NA	NA
3	357696000	2019-10-22	NA	NA	NA	NA	ohmysierraa	NA	NA
4	357696000	2019-10-22	NA	NA	NA	NA	TJohn321mine	NA	NA
5	357696000	2019-10-22	NA	NA	NA	NA	NeedsNYC	NA	NA
6	357696000	2019-10-22	NA	NA	NA	NA	tonigdnz	NA	NA
7	357696000	2019-10-22	NA	NA	NA	NA	Steenyall	NA	NA
8	357696000	2019-10-22	NA	NA	NA	NA	JeremyScheel1	NA	NA
9	357696000	2019-10-22	NA	NA	NA	NA	aamykl	NA	NA
10	357696000	2019-10-22	NA	NA	NA	NA	seapebbles	NA	NA

Figure 1: Merged Data Part 1


BDA on postgres@PostgreSQL 11

Query Editor Query History

```
1 select * from merged_twitter_data limit 10
```

Data Output Explain Messages Notifications

url character varying	truncated character varying	source character varying	in_reply_to_status_id character varying	in_reply_to_user_id character varying	is_quote_status character varying	full_text character varying
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...
NA	False	<a href="http://twitte...	None	None	False	I always say how much I love Fall and Spring, when in all h...

Figure 2: Merged Data Part 2

4. BASE STATISTICAL DESCRIPTION

We observed the distribution of the sentiments dataset in 3. We can see the majority of the dataset contains negative tweets. To understand what to do with the dataset, we decided a base statistical description would be helpful.Figure 4 provides base statistics of sentiment, count of friends, followers, and statuses. We can observe that most tweets found in this dataset was classified as negative, since the mode is -1. We also observe that standard deviation is relatively high,

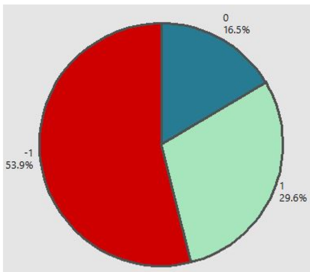


Figure 3: Division of sentiments

Variable	Mean	SE Mean	StDev	Minimum	Median	Range	Mode	N for Mode	Maximum
sentiment	-0.2424	0.0196	0.8815	-1.0000	-1.0000	2.0000	-1	1087	1.0000
friends.count	1898	242	10867	0	486	435544	0	16	435544
statuses_count	57832	3049	136948	1	18923	2095419	89277, 234595	6	2095420
followers_count	8100	3538	158887	0	582	6977554	495	9	6977554

Figure 4: Basic description

thus the differences are spread out, most values are not close to the mean. To understand more about the user, we compared the standard deviation of the followers, friends, and statuses count against grouped sentiments (positive negative and neutral). Figure 5 and 6 gives us that information. We can infer from the dataset, that the people who are de-

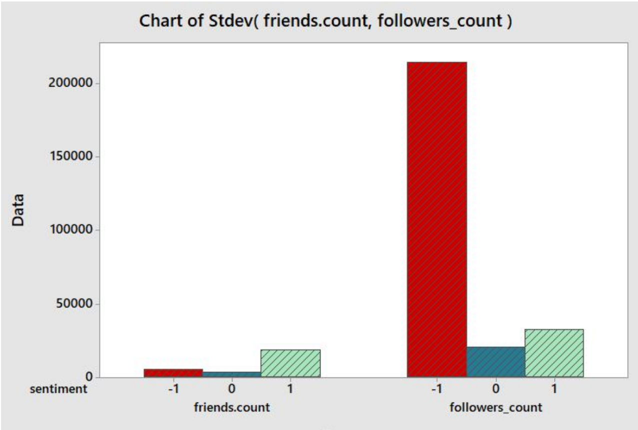


Figure 5: Sentiment vs friends and followers count

pressed or suicidal, are likely popular since they have a large followers count. However, we also see that they have lesser friends and they post more statuses than people who are non- suicidal. This conclusion supports the cases of [7, 9] which finds a link between ideations of suicide and social media usage.

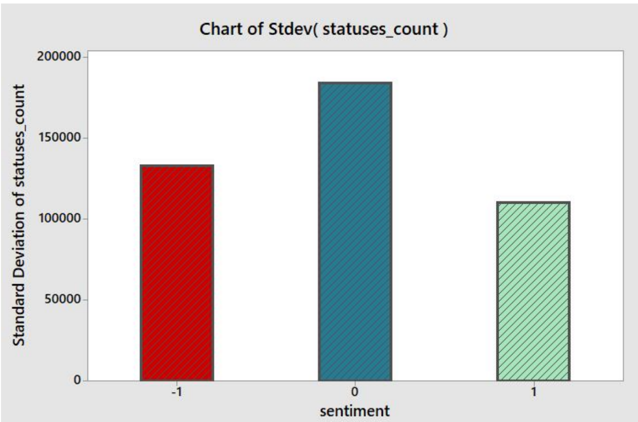


Figure 6: Sentiment vs statuses count

We tried to group tweets by location, trying to figure out which country tweeted the most depressive tweet but the problem with that was twitter API does not give coordinates for all the tweets, and the verbose description of location given is custom, which means users can name any location they want, so some of the locations were not even real which is why we could not do anything about it. This is a classic example of missing data and how it affects the accuracy of results.

4.1 Pairwise Comparison

In order to understand if there is a relationship between followers count,friends count and the number of times users

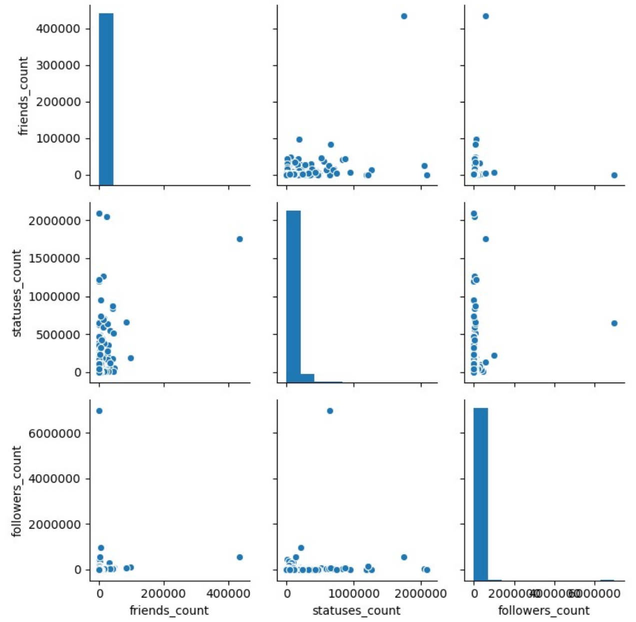


Figure 7: Pairwise Comparison

post statuses, we performed pairwise comparison. Figure 7 shows us that there is little to no correlation between the three of them. The Pearsons coefficient value between statuses vs friends count is 0.391(the highest amongst the three) which actuality is not high in value. Thus a regression line cannot be fitted over the data.

4.2 Word Cloud

Human behaviors are based on a pattern. This applies to the choice of words they apply to social media websites where they are directly or indirectly voicing their thoughts. Even if humans are trying to equivocate, like in case of taboo subjects like feeling to take one's own life, there is still some distinct pattern to those words. For this reason, we generated a word cloud based on the frequency of words used by tweets labeled as negative. Figure 8 is a visual representation of the most frequently used words. We can observe

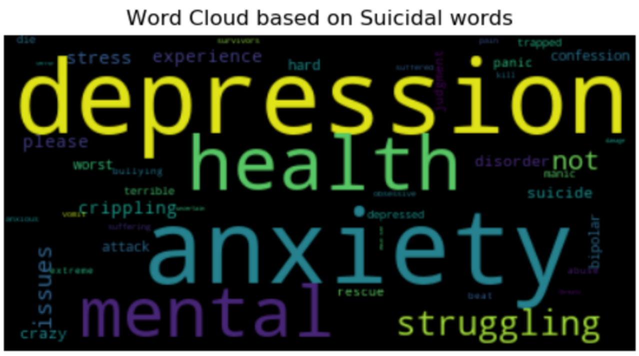


Figure 8: Wordcloud

that people going through suicidal thoughts frequently use the words 'anxiety', 'struggling' and 'depression'. There are also words like 'trapped' 'stress' and 'hard' which are likely used by users who do not directly want to admit or express that they may be going through depression.

5. K MEANS

Term Frequency(TF) corresponds to the number of times a word or a term appears in a text. The *Inverse Document Frequency(IDF)* is the logarithm of the number of documents (here, the number of tweets) divided by the number of documents in which a specific term appears, thus measuring the weight of a word in the review. At the word level, the TF-IDF[8] score is a matrix that represents TF-IDF scores for each word in the review text.

Algorithms have a hard time understanding text data, hence before any kind of computations can be performed with this data we need to transform it for our algorithm to understand and use it to draw meaningful conclusions. K means clustering algorithm is no exception to this condition, therefore to optimally run the k means algorithm on our processed Twitter data, TF-IDF was used to cluster similar types of tweets.

The main goal was to understand how many people tweeted have similar thought processes along with collecting these tweets into a single cluster. Although the K means clustering was performed on the tweet text, we used the Tweet ID to showcase similar tweets clustered together.

The total number of clusters created was 6, and each cluster contained tweets related to a particular keyword which we assumed would define a negative, positive or neutral tweet pertaining to suicidal thoughts. For example, the results achieved after running our k means algorithm, the first cluster created contained tweets having a very neutral context. On the other hand cluster 2, contained tweet IDs very specific to 'mental health' and cluster 3 had tweets that talked about 'anxiety'. Following are the examples of tweets based on each cluster:

Cluster 1:

- {870430852605333504: "Annual incomes and employment rates don't show the massive amount of variance in our monthly earnings. A #basicincome will abate insecurity. https://t.co/OKKlmDO7lW"}
- {870431472838168576: The wisdom. This is a word. https://t.co/mIGKcSQzk6}

Cluster 2:

- {870430814433144832: "It's time for physicists to talk about mental health https://t.co/R1bRZG9gpD"}
- {870431366474858496: "From @uoftmedicine @UofTMedDean: "Can we talk about physician mental health?" https://t.co/GpeQlXe5Im"}

Cluster 3:

- {870430762255953920: "Hey, look - I found my social anxiety again. Was wondering where that went."}
- {870430779439841280: "From @uoftmedicine @UofTMedDean: "@sabbunny I went there last Monday and I almost had an anxiety attack. there were so many people but we only spent 98 when we usually spend 23"}

Cluster 4:

- {870430799950004224: "According to studies, high-anxiety people are more likely to make bad decisions because they tend to catastrophize uncertain situations"}
- {870435830204055552: "I don't think I'm going to miss eighth grade. It's been a tough year. A lot of my friends are struggling with depression and self-harm..." https://t.co/gC2ldfT2db"}

Cluster 5:

- {870430820791726080: "it's fucking me https://t.co/80 8i3s7hmQ"}
- {870439964034682880: "I have terrible post-con depression help me"}

Cluster 6:

- {870430807344779264: "I am delighted to pledge support for people with mental health problems, in line with Mind's manifesto:... https://t.co/WWYmwnQO68"}
- {870430905529016320: "10 nutritional deficiencies that may cause depression https://t.co/NIfvlPlDQM #Mentalhealth #Nutrition"}

The most interesting cluster we observed was Cluster 5, it contained quite a few abusive words. Cluster 6 had all the tweets pertaining to people showing their support for combating suicides.

6. CONCLUSION

From observing the dataset, we can see many familiar patterns between people who are depressed and their usage of social media. We were able to find support to already published studies on social media use and depressive tweets. We observed that people who are popular and very active on social media are likely to be depressed. We visualized a word cloud of the most popular word addressed by said depressed people. We found cases of missing data and data which has little to no correlation. We also performed K means clustering on the tweets and obtained 6 clusters giving an insight to their usage of jargon's and thought processes. We believe our work has helped understand the concept, effect and influence of suicide on social media and vice versa.

7. REFERENCES

[1] Detect depression in twitter posts. `https://github.com/peijoy/DetectDepressionInTwitterPosts`. (Accessed on 09/23/2019).

[2] Does social media cause depression? | child mind institute. `https://childmind.org/article/is-social-media-use-causing-depression/`. (Accessed on 09/20/2019).

[3] Instagram urges users to report suicidal posts after 16-year-old kills herself over death or life poll results. `https://www.newsweek.com/instagram-urges-users-report-harmful-posts-after-16-year-old-kills-herself-146425`. (Accessed on 09/23/2019).

[4] These 5 social media habits are linked with depression | live science. `https://www.livescience.com/62718-social-media-habits-depression.html`. (Accessed on 09/20/2019).

[5] Tweet object — twitter developers. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object. (Accessed on 10/21/2019).

[6] P. Burnap, W. Colombo, and J. Scourfield. Machine classification and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, pages 75–84, New York, NY, USA, 2015. ACM.

[7] NCBI. Tracking suicide risk factors through twitter in the us. - pubmed - ncbi. https://www.ncbi.nlm.nih.gov/pubmed/24121153. (Accessed on 10/22/2019).

[8] towardsdatascience.com. Tf idf | tfidf python example. https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76.

[9] H.-H. Won, W. Myung, G.-Y. Song, W.-H. Lee, J.-W. Kim, B. J. Carroll, and D. K. Kim. Predicting national suicide numbers with social media data. *PLOS ONE*, 8(4):1–6, 04 2013.