Descriptive Analysis of Suicide Ideation on Twitter

Anonymous

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.

  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. **in_reply_to_screen_name**: If the represented Tweet is a reply, this field will contain the screen name of the original Tweet's author
14. **in_reply_to_user_id**: If the represented Tweet is a reply, this field will contain the integer representation of the original Tweet's author ID.
15. **favorite_count**: 'favorite_count' is a integer type attribute which tells us how many times the tweet has been liked.

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

4. LIST OF COMPLETED OBJECTIVES

1. Data gathering from live and existing datasets
2. Data merging
3. Data Storage in a DBMS
4. Data cleaning
5. Data pre-processing

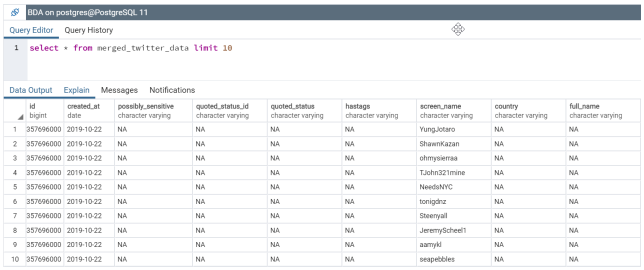

Figure 1: Merged Data Part 1

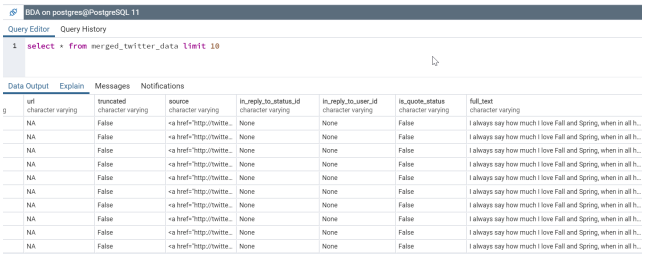

Figure 2: Merged Data Part 2

5. LIST OF FUTURE DELIVERABLES

1. Clustering using :
   (a) K Means
   (b) K medoid
   (c) DB-Scan
   (d) Expectation Maximization
2. Comparison of the performances of these algorithms using Silhouette index,Dunn index,R-squared index and Homogeneity index.
3. Visualization of the clusters and their performance.

6. 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` (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.