Data Surveillance

Mayank Maroti1, Rishabh Singh Rajput2, Ashutosh Devpura3, Sathybhama4
Student1,2,3, Assistant Professor4
Department of Information and Technology
SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

Abstract:
For the project we have developed a graphical Java software application called The Twitter Mapper. This application allows the user to specify the search frame of context, and monitors tweets from a service that Twitter in real time interest. It also plots the location of the user on map of the world. To do this, the application makes use of an existing geo mapping library and the Twitter API. Our project is social media Analytics and Keyword sorting platform providing peculiar intelligence to engage twitter user or younger demographics based on their location, what they say, and who they are. Coupled with analysis of accounts, as well as the ability to research any hash tag and topic to examine keyword which can help in our analysis of that tweet. There are two parts of the project. First that it connects to the Twitter service to receive tweets, and selects those tweets that match the given pattern by the user. The next part of the project is to display a marker pointing out to the location in the world of the feeder that gave up the tweet, on the tips that show the text of each tweet when the user places the mouse over the pointer.

Keywords: data analysis, cyber bullying.

I. INTRODUCTION

The use of social media platforms has steadily risen over the past decade, and recent surveys estimate that 90% of online adolescents and young adults use some kind of social networking site. One of the most popular social networking sites is Twitter. In fact, 35% of internet users of ages 16-24 use Twitter. Twitter is a micro blogging platform where users create short updates that are less than 140 characters. Many teens facing depression or depression are using social media like Twitter to express them-self. Because of the ease with which Twitter allows users to connect with a large audience of acquaintances and strangers, its popularity has grown, especially with teens and young adults. Twitter users can feel free to openly communicate with friends and acquaintances about virtually anything, even topics as traditionally private as mental illness. In a related study, researchers examined the keyword “depression” on Twitter for two months in yielded 20,000 tweets; the researchers found initial evidence that individuals tweet about their depression and even disclosed updates about their mental health treatment on Twitter. Likewise, a case study of tweets posted by a Twitter user prior to committing suicide found that suggestions of suicide were noted in the individual’s tweets immediately prior to the suicide occurring. Another related study found associations between clusters of at-risk suicide Twitter conversations and increased prevalence of geographic-specific suicide rates reported in traditional data sources. Lastly, a study of Tweeters in Japan found that self-reported lifetime suicide attempts were associated with tweets expressing suicidality. Mentally ill users have been spotted using test surveys, their public sharing of a scan on Twitter, or by their accountantships in an online forum, and they were differentiable from control users by symptoms of core deficient. Automated detection methods may help to identify depressed or otherwise at-risk beings through the large-scale monitoring of social media, and in the future may appreciate existing screening procedures. We have examined depression-related tweets on Twitter to get insight into social networking about mental health. We assessed for random sample of depression-related tweets, Tweets were coded for little hints of DSM-5 symptoms for Major Depressive Disfunction. Supportive or helpful tweets about depression was the most common theme, closely followed by disclosing feelings of depression. more than 60% of tweets revealed one or more symptoms for MDD and communicated thoughts or ideas that were consistent with struggles with depression after making an account of them they showed a vast deviation from the happy mortal human world. Health professionals can use our findings to stickup a bolt data to give an support-structure to the riding and nature the creation with us.

II. RELATED WORKS

These are the key drives the work along the project
1. Find markers of weaker mental health
2. Find factors that lead to weak mental health

Node.js is built on Google Chrome's JavaScript Engine. MongoDB is the NoSQL database, designed for the running application in todays world A classless environment, testing framework and asset pipeline structure for blockchains using Ethereum Virtual Machine (EVM), this is to make a developer life much more simpler.

III. PROPOSED SYSTEM

The proposed system should have the following features. The transactions should happen in secure room between clients in the networking system.

it provides agility to the user to transfer the data through the network very easily by compressing the large file. It should
also identify the user and give a communication security with transfer of file requested and run the required process at the system if needed. In this system the data will be send in the form of audio file. The user who received the file will do embedding, decryption, and decompress in their levels etc.

1)Web Frontend-A HTML, CSS, 2)Backend-Node.js, MongoDB 3)FileSharig Protocol-IPFS 4)Twitter APIs

IV. RESULTS AND DISCUSSION

These are the fundamental aspects that enhance and support the main frame of our project, this inhabits the total working and fundamental theory of our project.
1. Faster and accurate finds
2. Built in Geographical locator
3. Seamless Integration with More complex “Views”

V. CONCLUSION

Social media can prove to be an extremely handy tool in detecting and preventing mental health issues in a quick, effective and economic way, or so they thought. While many for all intents and purposes have mostly attempted this approach a unified solution to the problems faced for the most part is required. The data that actually is collected for research purposes must generally be from kind of more diverse sources for effective conclusions to specifically be drawn. Social media can prove to be an extremely handy tool in detecting and preventing mental health issues in a quick, effective and economic way. While many have attempted this approach a unified solution to the problems faced is required. The data that is collected for research purposes must be from more diverse sources for effective conclusions to be drawn. Some extra features were added to Java framework. The regular expression rules for extracting the hash tag, mention and URL were changed for our needs. The sequence of those processing was also modified. One of the main contributions in tweet processing in this term is the inclusion of Named Entity Recognition and Parts of Speech tags in the data set. We used existing Spark APIs for the Stanford Core-NLP framework for that purpose. There were a limited number of stop and profanity words, and those were kept in a list inside the code. Instead we used the standard list of stop-words provided by Spark. Data collected from social media provides immense insight into the psychological profile of a person. While this has been misused to a certain extent responsible handling of this data could prove to be a huge boon to society. Writing styles, patterns can not only provide insight into people who are suffering from PTSD, depression or any other problem but can also help find instances of online abuse. Things once spoken stay in the memory and eventually fade with time but things written on the internet have a dangerously perennial nature.

VI. FUTURE SCOPE

It could be used to estimate the views of a population regarding any person or socio-political event. It could help predict the onset of the next medical pandemic by analysing the people at the center of large social circle. For the tweet cleaning task, several avenue of improvement is possible. For the named entities, addition tags can be added for twitter specific elements like emojis. Another improvement is necessary for the memory allocation of each executor in Spark. For the NER and POS tag processing, the library that is loaded into memory is too large. The profanity list requires further memory. So optimising the number of executors per node is necessary which could speed up the processing. Another optimisation can be done by merging data transferring and data cleaning steps into a single step task. For the social network, our plan is to parallelise the conversion of RDF N-triple format using Spark and instead of using two intermediary step conversions to CSV from JSON and then converting to N-Triple, we plan to parse the JSON file directly to N-Triple and store it on the server. We also aim to store more tweet relationships and along with user ID, we plan to store user names, user’s top N friends, followers and top N tweets posted by the user based on highest favourite counts.

VII. REFERENCES:


[6].Logistic Regression Detailed OverviewSaishruthi Swamina