International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Mental Health Monitoring using Sentiment Analysis

Aagam Shah[1], Rohan Shah[2], Praneeta Desai[3], Chirag Desai[4]

[1]*Student, Dept. of Information Technology, K.J. Somaiya College of Engineering, Maharashtra, India*
[2]*Student, Dept. of Information Technology, K.J. Somaiya College of Engineering, Maharashtra, India*
[3]*Student, Dept. of Information Technology, K.J. Somaiya College of Engineering, Maharashtra, India*
[4]*Assistant Professor, Dept. of Information Technology, K.J. Somaiya College of Engineering, Maharashtra, India*

---***---

Abstract - *There is enough evidence to prove that overall physical, psychological and communal welfare of a human being is predominantly dependent on their mental health. Thus, early recognition and mediation in addressing issues regarding it is of utmost importance. Individualized and ubiquitous sensing technologies such as smartphones, smartwatches, activity trackers, etc. allow continuous tracking and gathering of data in an undisturbed and low-profile manner. Sentimental Analysis using Natural Language Processing has been proposed to be applied to the collected data to predict user information such as location, mood, activity, mental status, depression, anxiety, stress and soon. This paper surveys and proposes a methodology for data extraction and a model using Lambda Architecture to study the social media presence and other available data to predict the mental state of the user along with taking additional measures to maintain the secrecy of the user along with data privacy.*

Key Words: Mental Health, Natural Language Processing, Sentiment Analysis, Health Monitoring System, Mental Disorders, Smart Phones, Wearable Gadgets, Social Media, Lambda Architecture.

I. INTRODUCTION

Emotions are certain complicated neurological expression including three sections: a sentimental occurrence, a psychological and physical reaction, and a sociological or revealing reaction [6]. Mental State of a person is the perspective of that person and furthermore gives a sign of his/her general nature. Psychological sickness is a result of lopsided characteristics in cerebrum science. Depression is one of the most well-known and debilitating mental issues, and relevantly affects society. Despondency and schizophrenia are the primary purposes behind most suicides because of mental issues. Most people are subjected to pressure while many are influenced by depression because of numerous reasons. The assessment of mental well-being is important to comprehend and propose treatments for patients with a strayed mental conduct.

These days, a few applications have been proposed for depression and, by and large, for psychological wellness issues. These applications incorporate self-observing and mean to follow anxiety symptoms and to furnish unwinding activities to assist people with misery, uneasiness issues, outrage the executive's issues, and so on. They are not intended to supplant customary treatment, for example, Intellectual Behavioral Therapy but can be used to enhance it.

Self-destruction avoidance to mental turmoil people would thus be able to be of profiting the general public on the loose. The intricate issue is to recognize the person who has psychological instability. One in each four people are influenced by a psychological issue at some stage of their life. Considering enormous number of individuals that are legitimately and by implication experiencing psychological sickness, it is critical to contemplate the techniques that help in the distinguishing proof of psychological instability, follow and foresee the developing dysfunctional behavior procedures. Estimation and emotional detecting technology could assist with handling these goals by giving successful apparatuses and frameworks for target appraisal. Such devices and frameworks don't expect to supplant the clinician or therapist yet they could bolster their choices.

To provide an estimate of how prevalent the effects of mental illness are in society [1], consider the following statistics:

- According to recent estimates, about 20% of Americans, or about one in five people over the age of 18, suffer from profound mental disorder in a given year [1]
- Out of 10, 4 of the disabilities are Mental Illnesses such as Bipolar Disorder [4], Clinical Depression [4], Obsessive Compulsive Disorder [4] (OCD) and Schizophrenia [4].
- Around 3% of the population [1] suffers from more than one mental disorder at the same time.
- Approximately 5% of humans are so severely affected by mental disorders that it indirectly conflicts with their ability to work in society which include Schizophrenia, Bipolar Disorder, Depression, Panic Attacks, and Obsessive-Compulsive Disorder [1] (OCD).
- Around 20% of medical appointments are attributed to generalized anxiety disorder such as panic attacks [1].
- Every year 8,000,000 people are experiencing depressive episodes.
- 2 million Americans experience psychotic disorders, and three hundred thousand new cases are diagnosed yearly.

The next sections to be discussed are mentioned underneath. Literature Review is discussed in Section II. Section III explains the proposed methodology include the details regarding currently available data set and additional data to be gathered. It also describes the Sentimental Analysis Pipeline followed by procedure for data extraction and gathering. After which the paper explains the proposed architecture for the implementation of the system. We summarize and conclude in Section IV.

II. Literature Review

Discovering psychological resilience of Twitter User's by analyzing their Tweets:

In the observed work, they have captured the psychological resilience of a person, who is posting on twitter with the downturn establish tweets showed that a person, who is in distress and posting with one of the watchwords is tending to tweet with other despondent watchwords. They have gotten the keywords regarding tweets from twitter.com utilizing application program interface (API) [3]. They utilized keywords like depressed, defeat, dejected, anxious, uneasy, fatigued and undeserving were gathered from April to July 2018[3] on arbitrarily random dates [3]. They assembled more than 1.3 million tweets for [3] each of the above keywords or catchphrases to perform their examination.

Utilizing the "text-mate" bundle in R programming language, they evacuated all the numerals, different signs, blank spaces, and hyperlinks from the above tweets [3]. A spreadsheet is acquired and the word [3] occurrence table of each word [3] was established. They picked the best highest repeated expression from every watchword. They acquired comparative regularly utilized words from every one of the keyword independently and joined every one of these words set with their occurrences. They positioned with their recurrence of utilization, and it is seen that there are 202 unique expression in the assortment of expression sets of every single one of the keywords records [3].

Singular Value Decomposition is a critical instrument in the locale of Information Retrieval to handle the records that are generally correlated with a lot of catchphrases in a question. Relationship coefficients were acquired to these watchwords from the recurrence records and exercised the Singular Value Decomposition technique to recognize the quantity of limits that are adding to this methodology.

It was seen that individuals who posts with one discouraging catchphrase is subsequently using supplementary discouraged watchwords. The recurrence of expressions utilization delineated the expression utilization design with watchwords and came about that a person in a disconsolate phase may before long goes into another disconsolate phase. The outcomes acquired from the following approach brought about a comparative outcome demonstrating a commendation of 63 percent commitment to the initial three pivotal values. The Singular Value Decomposition approach likewise bolstered the outcomes, and it was observed that there is consistently an opportunity of individual tweeting in any event three of these keywords, in the event that he/she is mentally depressed [3].

Psychological well-being prediction using Data Mining:

They have planned a system with the essential objective of building up a site where clients are able to put in information in a form and acquire outcomes about current psychological instability reliant on their information. They have gathered a data-set that is accessible on the internet. The information assembled is analyzed and refined. The information carries various tags, for example, age group, sex, separation of work environment from home, past dysfunctional behavior, pedigree, and so forth [2]. They have mark encoded the information for better exactness and executed the Random Forest and choice Tree calculations for evaluating the information, also to locate the exact estimation of mental health status. They have applied the choice Tree grouping calculation for the arrangement of the information as it was seen as progressively precise. They broke down the information with the assistance of this calculation to discover different bits of knowledge that the information uncovered. They have structured a website where a client will sign in and top off a structure which has questions dependent on the informational collection accumulated. The client will respond to the inquiries and an outcome about his/her state of mind will be given on the website according to the sources of the given info.

They have structured the framework such that a likelihood of under 0.30 shows that the client doesn't experience any mental ill effects. Likelihood somewhere in the range of 0.3 and 0.63 demonstrates that the client may confront a psychological instability later on and a likelihood more noteworthy than 0.63 shows that he/she experiences dysfunctional behavior [2]. The precision they got with choice tree was 82% with 258 examples of information being characterized effectively out of 315 occasions [2].

III. Proposed Methodology

3.1. Data

Sentiment Analysis is currently implemented on Twitter tweets, Facebook messages and on Instagram Images. Sentiment140 Data set is used for sentiment analysis of twitter tweets.t contains 16 million tweets that have been extracted from the twitter API. The tweets are annotated a range of 0 to 4 which helps in determining the sentiment of the tweet. The data consist of 6 fields:

- target - It specifies the sentiment of the text
- id - Used for Identification of the data
- date - It consist of the date and time the post was uploaded
- user - Specifies the username of the twitter account
- text - Specifies the Tweet (Text) written by the user

These data help to determine the sentiment analysis. The Target field has 3 outputs {0,2,4} where 0 represents neutral, 2 represents positive and 4 represents negative.

Other Area where sentiment analysis is used is Facebook post. The Data set used for this area is available open source on Kaggle as Fb Sentiments. This data set contains only 3 fields of id, post and the date of the post based on these post's and using NLP inbuilt library of Nltk corpus, sentiment analysis is derived from this data.

For the Project Area of Mental Health most accurate data set should consider the following input attributes.

Table-1: Structure of Proposed Database

Attributes	Type
Name	String
Age	Number
Text Message (SMS)	Text
Social Media Images	Images
Social Media Emotions	Text
Social Media Location	Location
Heart Rate (Wearable Device)	Number
Steps (Wearable Device)	Number
Music	Audio
Active Screen Time	Number
Device Images	Images
Device Location	Location

The Table above contains many attributes that can be harnessed from the user's phone, Wearable devices and social media like Facebook, Twitter and Instagram. The data that is extracted from the device for processing always remains in device only in order to protect the privacy of the user. All the data is always encrypted using SHA-256 encryption ensuring complete privacy of user data.

3.2 Sentiment Analysis Pipeline

Sentiment Analysis is defined as the recognition of the text to be positive, negative or neutral. Sentiment Analysis is generally implemented using Machine Learning where the input text 'T' is given and a list of emotion types are given based on which the emotion is determined. The generalized pipeline of Sentiment Analysis could be broken down into 5 different steps:

- Input text:- The Input text could be in the form of Text, PDF, Audio Input and HTML Files.
- Pre-Processing:- Pre-Processing of input data is necessary in order to make the data in a standard format so that the optimized algorithm could process it and provide the most efficient result. Pre-process consist of process of Stemming, Lemmatization, Tokenization, Stop word removal, Removal of repeated characters and even spell check.
- Feature Extraction:- The process of Feature Extraction is used in the identification of parts of speech in the in input text. It helps for proper text formation.
- Feature Selection:- Feature Selection includes Information gain, selection based on Frequency, Point wise mutual information and gain ratio.
- Sentiment Classification:- Sentiment Classification can be achieved using multiple methodology which include Classification, Regression, Clustering and Association.

The Lexicon based approach uses a set of words to determine the sentiment of the text. The sentiment is divided into three main types Positive, Negative and Neutral. There are 2 main approach in the Lexicon based prediction Dictionary based and corpus based approach .The input text after preprocessing is tokenized and then compared with a dictionary set of values which determine the nature of the text, based on this approach the entire sentence is given a value and then its sentiment is calculated. In the corpus-based approach, using the seed list of the words and help from various semantic technique more context specific words are identified.

It is an iterative process that begins with a defined word collection but, by using multiple sources, broadens its quest range by using alternative synonyms, originating from the seed set of terms of opinions and using numerical and linguistic techniques, certain words of opinions belonging to a specific context are found in known corpus like dictionary and thesaurus.

3.3 Data Extraction or Gathering

The proposed model suggests gathering or mining of user data from various sources such as Mobile Phones (Smart Phones), Wearable Gadgets (Smart Watches) and Health Trackers (Fitness Tracker Bands). Using Smart Phones or phones in general is useful as the data captured or gathered may be shared or transmitted via various means such as SMS, MMS, USSD, Bluetooth, Internet (Wireless Mobile Data) or exchange of devices and hardware such as memory cards. Wearable Gadgets such as Smart Watches, Fitness Trackers, Connected Headsets, Smart Glasses, Wrist Bands, etc. are produced due to the development and advancement in mobile technologies and sudden keen interest of users in fitness and sports related activities. Social Media data is gathered information from social networks such as Twitter, Facebook, Linked In, Instagram, etc. showing user engagement, viewing, sharing and following with respect to content and profiles. It is metrics and statistics to determine the ups and downs of any given data to make sense out of it and translate the data and time into results.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

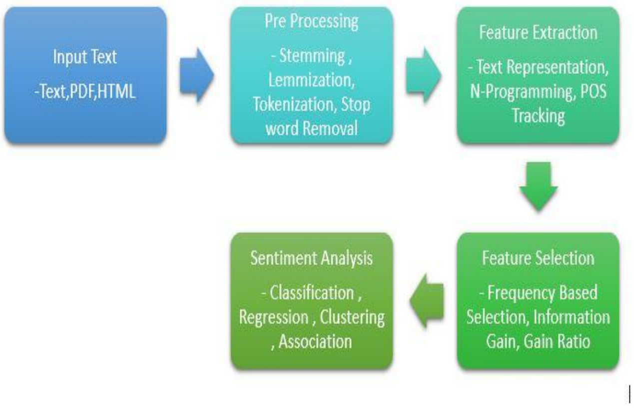

Figure-1: Pipeline for Sentiment Analysis

3.3.1. Mobile Devices

The Data that is extracted from the Smart Phones are the text messages sent and received, call recording of the user, images kept in the mobile gallery, active screen time while using the mobile phone, music preference and the location of the user over a period of time. All of this data is then encrypted and sent to the server for processing.

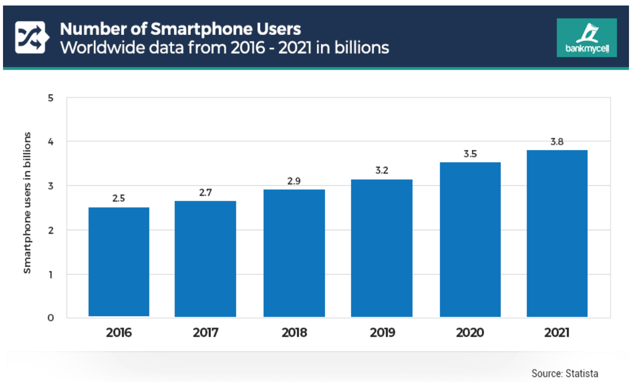

Chart-1: Number of Smartphone Users

3.3.2. Wearable Gadgets

The Data extracted from Wearable Devices include sensing mobility and activity through sensors, ecological momentary assessments (EMA), patient reported outcomes and Health events. The physical exertion, speed of walking, area-perimeter of travel may be recorded by itself using motion sensors. User may manually enter or record pain, mood, fatigue, disability, hurts, hospitalizations, activity restrictions, etc.

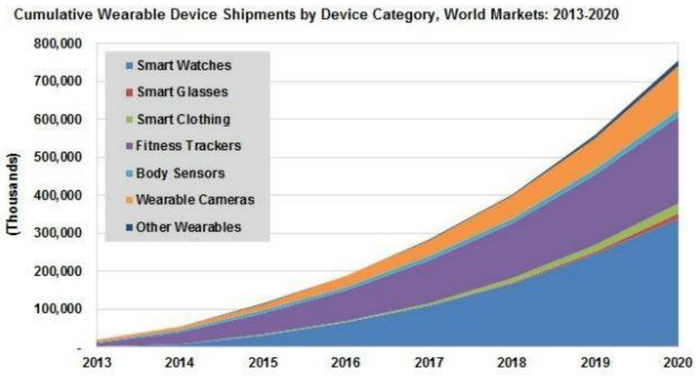

Chart-2: Number of Wearable Gadget

3.3.3. Social Media

Social Media data are mainly statistical data having numbers, percentages and values providing actionable insights into one's life and mental state. Some examples of data gathered from social media include posts or topic shares, liked posts, mentions of others users, impressions of pages, hashtags related to any trending topics, URL clicks on advertisements, keywords, followers of the user, following topics, comments on posts, interactions and so on.

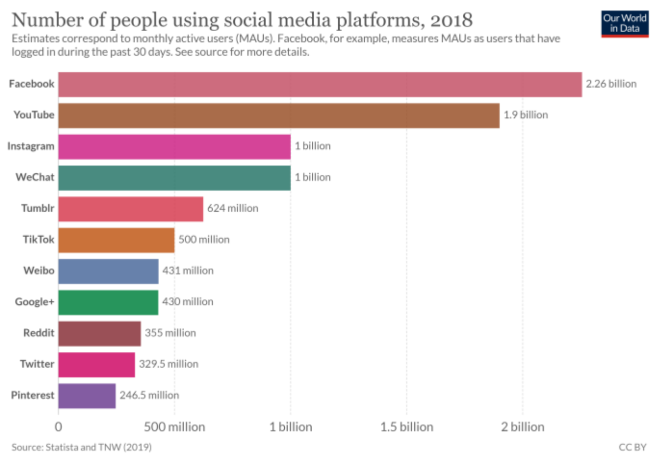

Chart-3: Number of people on social media

4. System Architecture

Massive quantities of data can be handled by using a data-processing architecture such as Lambda architecture by using both batch as well as stream-processing methods []. Lambda architecture uses 3 layers for more efficiency they are:

- Batch Layer
- Speed Layer
- Serving Layer

Batch Layer is used to optimize the precision of the result without considering time since it uses a distributed processing framework that has allowed the results to be calculated before handed and the result is then stored in internal read only database. Speed layer is the also called the real time layer its main work it to reduce the delay and generates result faster but this on a contrary causes a drop in precision. After both of the process when the data is in the usable format the result of the speed layer is matched with more attested tests, this leads to the third layer of functioning called the serving layer in Lambda Architecture. The figure below represents the Architecture

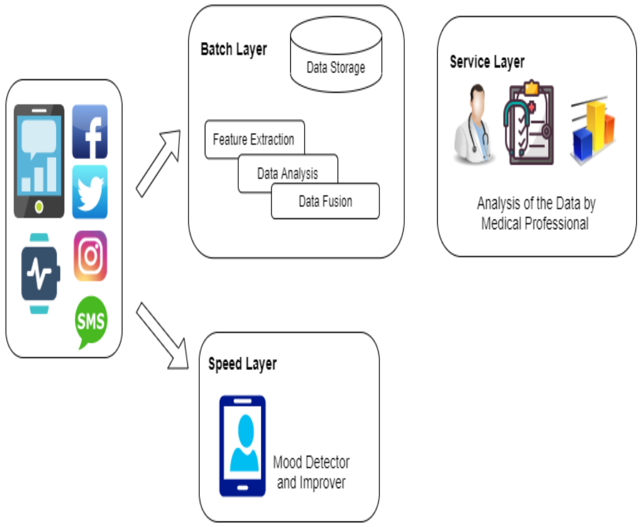

Figure-2: System Architecture

The figure 2 shows the architecture for the proposed architecture which is based on Lambda architecture. In this system the data is gathered from various sources and then processed in real time in order to detect the sentiment of the person. Offline processing is also possible due to Batch Processing in which all the features of text are extracted offline and they are also compared with the help of a set of corpus which is made offline beforehand, with this the entire data analysis is then connected to each other and a final report regarding the mental health is generated. Finally, these results are also merged with all the data of the user and the detailed report is generated which can be available in the main view of this system in service layer which can then be sent to medical professionals. In the speed layer the main aim is to immediately improve the mood of the person in case of any mental dis stress. This when the speed layer comes into existence the speed layer as the name suggests puts speed as its preferred choice over accuracy. This layer can potentially help save many lives which are facing mental dis stress and help improve their mood immediately.

IV. CONCLUSIONS

This paper clearly describes that healthy state of mind is predominantly a necessity for overall wellness of a person physically, psychologically and socially. Thus, this paper proposes a system for early recognition and intervention in addressing the problems faced in determining the mental state of a person. This paper discusses the use of mining or extraction of data from various sources such as Mobile Phones (Smart Phones), Wearable Gadgets (Smart Watches) and Health Trackers (Fitness Tracker Bands) and proposes a model using Lambda Architecture to analyses the extracted data to predict the mental state of the user maintaining user secrecy and data privacy. The paper suggests the use of Sentimental Analysis and Natural Language Processing on the gathered data to predict user information such as location, mood, activity, mental status, depression, anxiety, stress and other critical parameters.

REFERENCES

[1] National Center for Biotechnology Information.Information about Mental Illness and the Brain. https: / / www.ncbi.nlm.nih.gov/books/NBK20369/v. Last accessed 27July 2020. July 2020.

[2] V. Laijawala et al. "Classification Algorithms basedMental Health Prediction using Data Mining". In: 20205th International Conference on Communication andElectronics Systems (ICCES). 2020, pp. 1174–1178.

[3] S. T. Sadasivuni and Y. Zhang. "Analyzing Tweets toDiscover Twitter Users' Mental Health Status by a Word-Frequency Method". In: 2019 IEEE International Con-ference on Intelligent Systems and Green Technology(ICISGT). 2019, pp. 5–53

[4] WHO.Mental disorders affect one in four people. https://www.who.int/whr/2001/mediacentre/pressrelease/en/.Last accessed 27 July 2020. July 2020.

[5] Wikipedia.Lambda architecture. https : / / en . wikipedia .org/wiki/Lambdaarchitecture. Last edited on 28 February 2020. July 2020.

[6] C. Zucco, B. Calabrese, and M. Cannataro. "Sentiment analysis and affective computing for depression monitor-ing". In: 2017 IEEE International Conference on Bioin-formatics and Biomedicine (BIBM). 2017, pp. 1988–1995.

[7] H. Elmunsyah, R. Mu'awanah, T. Widiyaningtyas, I. A. E. Zaeni and F. A. Dwiyanto, "Classification of Employee Mental Health Disorder Treatment with K-Nearest Neighbor Algorithm," 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Denpasar, Bali, Indonesia, 2019, pp. 211-215, doi: 10.1109/ICEEIE47180.2019.8981418.

[8] V. Mody and V. Mody, "Mental Health Monitoring System using Artificial Intelligence: A Review," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-6, doi: 10.1109/I2CT45611.2019.9033652.

[9] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera and D. Manocha, "Learning Perceived Emotion Using Affective and Deep Features for Mental Health

Applications," 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Beijing, China, 2019, pp. 395-399, doi: 10.1109/ISMAR-Adjunct.2019.000-2.

[10] J. Qinghua, "Data Mining and Management System Design and Application for College Student Mental Health," 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, 2016, pp. 410-413, doi: 10.1109/ICITBS.2016.96.