IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

Depression Detection Using Natural Language Processing For YouTubeData Users

Aftab Pathan[1], Seenin Sayyad[2], Musarrat Bakali[3]

[1,2,3] Student of Computer Engineering, Trinity College of Engineering and Research, India

Abstract: Depression is a mental illness that affects an individual negatively. It is considered as a serious disease by mental health care professionals. Depression detection is important to avoid unwanted consequences of not acknowledging the disease. A research was carried out in 2012 and an estimate was found. It was observed that there were roughly 258000 suicides. Further, it was observed that the age group that was mostly affected was between 15-49 years of age [1]. This estimate informs us that the aforesaid age group is prone to depression. This age bracket spends maximum time on social media and shares their view on it. It reflects their mental condition. This fact encourages us to develop a system to detect the depression level of the users and provide necessary information to the guardians to enable the guardian to take appropriate actions. The system is beneficial in informing the user and their guardian to prevent self harming or worsening of the condition. The death rate will significantly reduce if the user and the guardian are aware of the mental state of a user. The system is expected to be beneficial to reduce the percentage of death due to depression. It'll provide awareness to users and their guardians by automatically detecting depression [3]. This approach will utilize the emotions of the user detected from videos watched by the user. The title of the video indicates the content or category of the video. This enables us to get an insight to the user's inclination towards negative polarity.

Keywords: Depression detection, NLP.

I. INTRODUCTION

Depression has proved to be a mental illness that has taken away the mental well-being of those affected with it. Depression is classified as an affective disorder. An affective disorder is a mood disorder. The 10th Revision of International Classification of Diseases ICD-10 which is considered to be the basis for diagnosis of mental disorders in the Czech Republic classifies depression as a mood disorder. The forms of the disease may vary and depression is likely to have three main forms i.e. mild, moderate and severe forms. Initially the symptoms are more of negative polarity in their mood. Negative polarity is characterized by feelings of sadness and needlessness. Negative mindset and feelings are major symptoms of depression. The research was carried out in2012 and an estimate was found. It was observed that there are roughly 2,58,000 suicides [1]. Further, it was observed that the age group that was mostly affected was between 15- 49 years of age. This estimate informs us about the aforesaid age group to be prone to depression. This particular age group requires more attention and care as far as mental health is concerned. Techniques which are automated are helpful in understanding the feelings of an individual based on videos on the web which are user generated and are usable for many applications [7]. This function can be used by the Government to comprehend and analyze the reaction and get clarity of their feedback. This paper introduces an approach where the comprehensive computation is used. The comprehensive computation enables us to comprehend the emotions based on the analysis of the video titles. Prominent research on emotion analysis of images has been done (Joshi et al. 2011). However, in case of emotion analysis of video such research has only been carried out on movie data (Wang and Cheong 2006). As per our knowledge, there is no current work based on videos generated by users and are diverse in content. The evaluation of a large set of attribute features and computation of semantic attributes is introduced. Multiple significant informational pieces have been obtained which has formed the base of the upcoming research of this challenging problem. The emotions associated with a video may not be identical to the emotions associated with a person. However, there are possible computational solutions for prediction.

1.1 Motivation

- Over the past few years the analysis states that the single largest contributor to Global Morbidity is depression. WHO too has analysed the issue. It is a significant issue that's found in individuals aged between 15-49 years old as estimated in 2012[1]. There is an immediate need to find a solution to the problem as the lives of individuals can be safeguarded. The tracking of depression level will give an insight to the mental health of individuals. Therefore, a system that caters to such a need is required.
- There are various surveys carried out and one such survey held by the National Mental Health Survey in 2015-16 estimates that 15 percent of Indian adults are in need of guidance in taking care of their mental health [1].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

1.2 Problem Definition

In this research, a machine learning approach is used to detect depression level by analyzing the social media posts of user There are lots of parameters to be acknowledged to indicate depression of a user. Most of the users express their emotional state through posts and tweets and watch video on YouTube, videos in the proposed model, at first Beautiful Soup is applied to collect tweets. Facebook posts are collected manually with the permission of some users. collected data are processed and read into the machine learning model. Then collected data are uncluttered by using NLP

1.3 Objective

- To analyze the depression level of a user
- To make the user and their guardian aware of their deteriorating mental health
- To prevent consequences of depression, especially the ones that involve self-harming

1.4 Statement of scope

The scope of our project is to take a part of the whole depression level detection process. From a technical point of view it focuses on below points:

1. Registration and login of the user.
2. Link YouTube API with the application
3. Shows the video as per search keyword by user
4. Save history about keywords search and watch video
5. Apply an algorithm on history with the help of API
6. Find depression level
7. Show depression level to user
8. If depression is high send notification mail on guardian mail ID.

II. LITERATURE SURVEY

Title: Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences
Author: Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich
Summary: Depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems. Study addresses the early detection of depression using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embeddings is evaluated and compared to a classification based on user-level linguisticmetadata.

Title: Young adults with mental health conditions and social networking websites: seeking tools to build community.
Author: Gowen K1, Deschaine M, Gruttadara D, Markey D.
Summary: Young adults living with mental illnesses are currently using social networking sites and express high interest in a social networking site specifically tailored to their population with specific tools designed to decrease social isolation and help them live more independently.These resultsindicate that practitionersshould themselves be aware of the different social networking sites frequented by their young adult clients, ask clients about their use of social networking, and encourage safe and responsible online behaviors.

1.1 Machine learning for emotion analysis

A Machine Learning-based human emotion examination approach is represented by Riyadh. During this research work, the authors use sadness, happiness, disgust, and surprise for his or her distributed task. They collected tweets from Sentiment140, labeled them manually, abolished tweets with no emotion, and created a balanced dataset to accommodate 3,750 tweets [4]. 3,500 tweets were selected because of their training dataset and 250 tweets as the testing dataset. For feature extraction, the Unigram model and therefore the Unigram model with POS tagging were used. The authors use the frequency of the Bag of Words model as a feature to coach their classifier.

1.2 RNN for Depression Forecasting

A novel approach for depression forecasting was initiated by Suhara, using RNN. The authors design the LSTM RNN based deep learning algorithm. They used their model to elaborate embedding layers regarding every absolute parameter, which also assimilated a day-of-the-week variable to work out the day-of-the-week consequences in their imitation [6]. They collected depressing data from 2,382 self-declared depressed persons, covering 22 months' time span,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

via an android application. Their technology was successfully ready to forecast 84.6%, 82.1%, and 80.0% severe depression instances in 1, 3, and 7 days beforehand, respectively.

1.3 Machine Learning for Depression Analysis

Wang et al. conducted an investigation on Sina microblog, a Chinese micro-blog, which is one among the foremost influential social media services in China [2]. They integrate both Psychological and Machine Learning knowledge for their evaluation. From the technical perspective, Machine Learning techniques, like Decision Tree, Naive Bayes, and Rule-based classifiers were used. Their described method contained mainly three types, namely, polarity calculation of sub-sentences, sentence and word segmentation, and polarity calculation of sentences. Their model was ready to achieve80% precision.

III. PROPOSED SYSTEM

Our proposed system is aimed at individuals who require mental health support. With the rapid increase in digitalization, individuals are highly connected to social platforms. We have leveraged a part of the existing YouTube platform to aid in the process of depression detection. Apart from keywords in the search bar, facial expressions, visual-audio features are also used to identify the level of depression based on a scale. With the help of these identifiers, the system will be able gauge the individual's mental condition and mood while using the application. Videos are sourced through the official YouTube API. A depression factor is assigned to each video as per its title to build a database that will be referred to while gauging the individual. To reduce the percentage of death due to depression the system will be beneficial and it will provide social support to users by automatically detecting depression. This system will use emotions of the user recognized from videos watched by the user. The title of the video describes the content or category of the video with the help of this we can find mood and depression state of the user to help them track their depression level. The proposed methodology is shown in fig. 1: Proposed system architecture.

The first step requires the user to register and login to the android application. In a format similar to YouTube, the user will be able to stream videos of their choice with the help of the search bar. Videos that match the keywords mentioned will be displayed. As the user browses through the application, a history will be generated of the videos watched and keywords searched. On the basis of this accumulated data, natural language processing will be used to calculate the level of depression as per the scale previously designed.

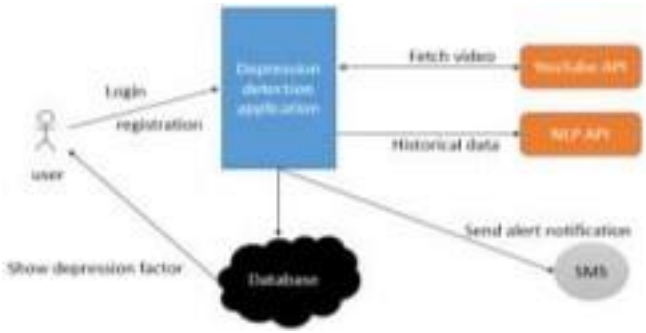

Fig. 1: Proposed system architecture.

- **User Login/Registration:**

The first step is to register on the android application and login to avail the services and features.

- **You Tube Interface:**

The interface is designed with the help of the official YouTube API. It stores the details of the video in the database.

- **Historical Data**

System user Historical data i.e. data search by user for watch video to calculate the depression factor using sentiment analysis as a NLP

- **Result Generation:**

The dashboard displays the results of depression level calculated by the NLP algorithm based on the user history.

- **Database**

The database stores login credentials and history of each YouTube video search title.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

IV. USAGE SCENARIO

This section provides various usage scenarios for the system to be developed. We have seen that many people feel great when applications take the minimum time for processing and also the battery usage is also minimum as compared to localsystems.

1 User Profile
This system will have one user/actors. He/she will be a user of a Smart-phone.
2 Use-cases
A Use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. The use cases are represented by a circle or ellipse. A key concept of use case modeling is that it helps us design a system from the end user's perspective.

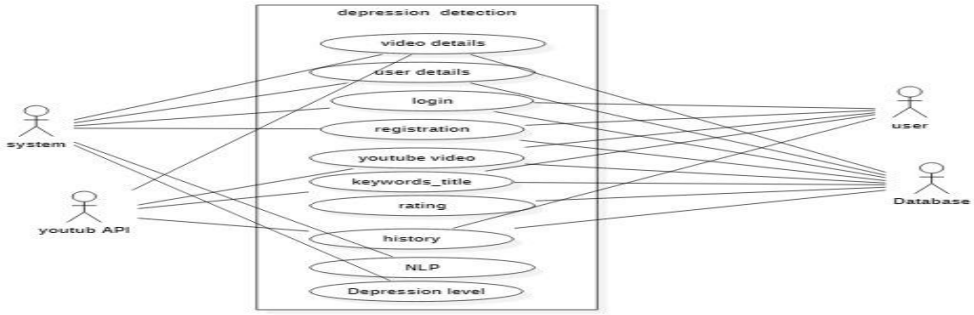

Fig 2: use case diagram

V.DATA MODEL AND DESCRIPTION

1 Data Description
Data objects that will be managed/manipulated by the software are described in this section. The database entities or files or data structures required to be described. For data objects details can be given as below

2 Data object and Relationships
Data objects and their major attributes and relationships among data objects are described using an ERD- like form.

VI. FUNCTIONAL MODEL AND DESCRIPTION

A description of each major software function, along with data flow (structured analysis) or class hierarchy (Analysis Class diagram with class description for object oriented system) is presented.

1 Data Flow Diagram
A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. DFDs can also be used for the visualization of data processing.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

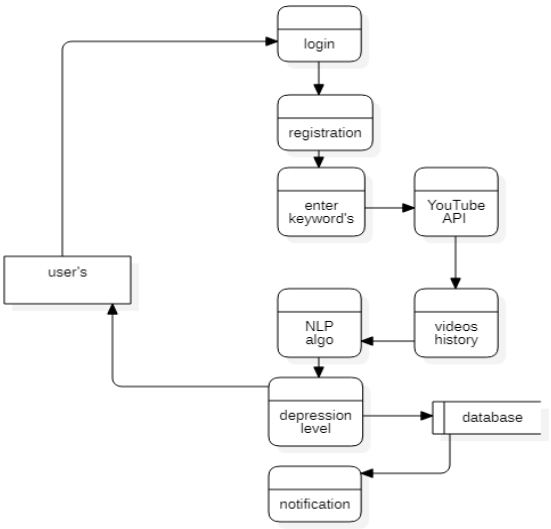

Fig 3: DFD (data flow diagram)

Activity Diagram

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The main element of an activity diagram is the activity itself. An activity is a function performed by the system. After identifying the activities, we need to understand how they are associated with constraints and conditions.

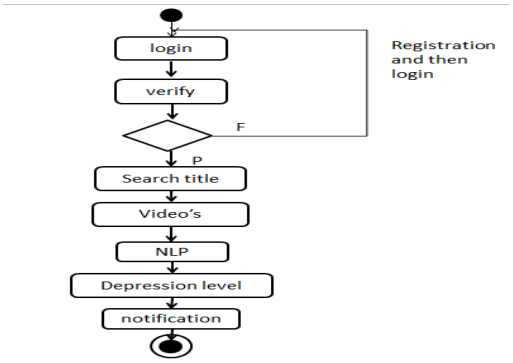

Fig 4: Activity diagram

VII. NON FUNCTIONAL REQUIREMENTS

Interface Requirements 1. Interface is not a major issue for this project because only the mobile phone directly interfacing with the inbuilt database Performance Requirements Considering the interactive task the system must have following characteristics:

1. Minimum Response Time
2. Efficient CPU Utilization
3. Less Memory Space
4. High Reliability
5. User Friendly

Non Functional Requirements defines some system properties and constraints it arises through the users need due to external factors such as safety regulation, privacy registration and so on, Non-Functional Requirements are:

1. Security
2. Reliability
3. Maintainability
4. Portability
5. Application

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

Compatibility Software quality attributes such as availability [related to Reliability], modifiability [includes portability, reusability, scalability], performance, security, testability and usability [includes self-adaptability and user adaptability

VIII. COMPONENT DESIGN

Class Diagram
The number of classes are identified and grouped together in a class diagram that helps to determine the static relations between them. With detailed modeling, the classes of the conceptual design are often split into a number of subclasses.

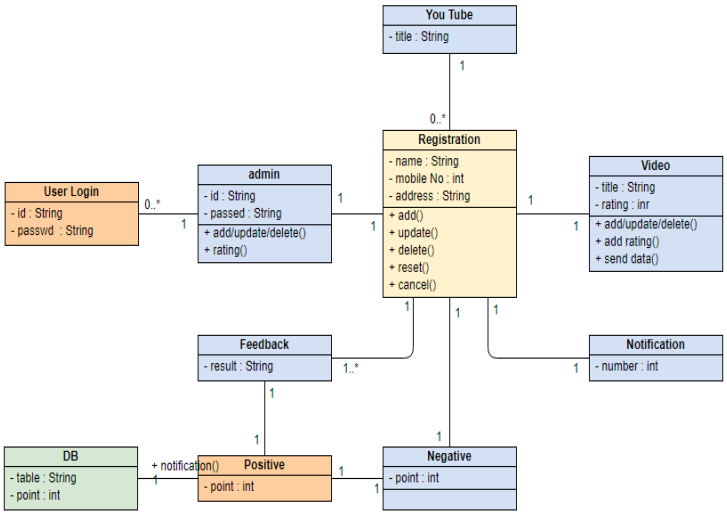

Fig 5: Class diagram

IX. COMPARING DEPRESSION MODULE

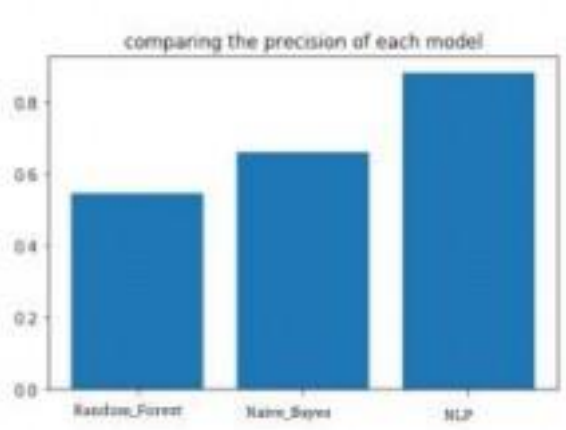

Fig 6: Precision of each module

As comparative study for our project we calculate polarity or word classification using three algorithm as Random forest work on multi output classification module when system have small training dataset. Naïve Bayes will perform better when system have large data set but system take much time to compare positive and negative polarity. The more accurate and precise algorithm in machine learning approach is natural language processing (NLP) [7]. System shows positive neutral and negative keywords using API. As a real time working we uses or preferred NLP algorithm for word classification.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10662

X. CONCLUSION

As per studies, depression if unaddressed can negatively impair an individual's ability to regulate their thoughts, behaviors, and emotions. This research demonstrated the use of machine learning to detect depression among individuals. The proposed system presents a comprehensive framework for depression detection from video emotion based on video title analysis. The natural language processing based technique analyses the user's watch and search history to determine their depression level. By providing social support to the user we can aid in receiving timely medical care. The features are also highly complementary, combining attributes with the title features shows very promising results. The future scope could require a focus on building a system that would limit the opportunities of exploitation of an individual's privacy. One such approach could be compliance with data protection regulations. Provision of transparency about the quantity as well as quality of data collected and to make sure that information collected for one purpose is not used anywhere else.

REFERENCES

1]"Depression," World Health Organization, 04-Jul 2017. [Online]. Available: https://www.who.int/mental_health/management/ depression/en/. [Accessed: 21-Dec-2018].

2]X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," in PacificAsia Conference on Knowledge Discovery and Data Mining, pp. 201–213, Springer, 2013.

3]"Languages of the World," Ethnologue. [Online]. Available: https://www.ethnologue.com/. [Accessed: 02-Feb-2019].

4]A.Z.Riyadh, N. Alvi, and K.H.Talukder, "Exploring human emotion via twitter," in 2017 20th International Conference of Computer and Information Technology (ICCIT), pp. 1–5, IEEE, 2017.

5]S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in bangla microblog posts," in 2014 International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1–6, IEEE, 2014.

6]Y. Suhara, Y. Xu, and A. Pentland, "Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks," in Proceedings of the 26th International Conf on World Wide Web, pp. 715–724, International World Wide Web Conferences Steering Committee, 2017.

7]B. A. Primack et al., "Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults," Comput. Hum. Behav., vol. 69, pp. 1–9, Apr. 2017.

8]F. Sadeque, T. Pedersen, T. Solorio, P. Shrestha, N. Rey-Villamiza, and S. Bethard, "Why do they leave: Modeling participation in online depression forums," in Proc. 4th Int. Workshop Natural Lang. Process. Social Media, 2016, pp. 14–19.