IDENTIFICATION OF MENTAL HEALTH RELATED ISSUES FROM SOCIAL MEDIA USING NATURAL LANGUAGE PROCESSING

Midhun VM[1], Sharanya Menon[2], Akash Patil[3], Prof. Dhiraj Amin[4]

[1,2,3]B.E Student, Dept. of Information Technology, Pillai College of Engineering, Navi Mumbai, India - 410206
[4]Faculty, Dept. of Information Technology, Pillai College of Engineering, Navi Mumbai, India - 410206

Abstract - *Mental illness is one of the most pressing public health issues of our time. While counselling and psychotherapy can be effective treatments, our knowledge about how to conduct successful counselling conversations has been limited due to lack of large-scale data with labelled outcomes of the conversations. Social Media posts contain various types of topics in our daily life, which include health-related topics. Analysis of health-related social media posts would help us understand health conditions and mental health issues encountered by people. The approach is to extract casualties from social media platforms like twitter using natural language processing (NLP) techniques. The basic idea is to create a system that can analyse the data extracted from social media platforms to interpret a person's mental health. This data can be extracted from as many social media platforms as possible to increase the chances of getting an accurate result. Natural Language Processing will be used to analyse the words used in each post. Based on the result the person can be advised as to what must be their next approach to better mental health condition. The system will try to find certain keywords like 'stress', 'melancholy', 'insomnia', 'sad', etc., in the social media posts and try to identify the mental health condition of the user and the help he or she needs.*

***Key Words*: Mental Illness. Natural Language Processing, Mental Health, Social Media, Text Mining, Artificial Intelligence**.

1. INTRODUCTION

Mental health includes our emotional, psychological, and social well-being. It affects how we think, feel, and act. It also helps determine how we handle stress, relate to others, and make choices. Mental health is important at every stage of life, from childhood and adolescence through adulthood.

Over the course of your life, if you experience mental health problems, your thinking, mood, and behavior could be affected. Many factors contribute to mental health problems, including:

- Biological factors, such as genes or brain chemistry
- Life experiences, such as trauma or abuse
- Family history of mental health problems

Everyone has some risk of developing a mental health disorder, no matter their age, sex, income, or ethnicity. Social and financial circumstances, biological factors, and lifestyle choices can all shape a person's mental health. A large proportion of people with a mental health disorder have more than one condition at a time. It is important to note that good mental health depends on a delicate balance of factors and that several elements of life and the world at large can work together to contribute to disorders.

According to the World Health Organization (WHO): "Mental health is a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to his or her community." People tend to forget the fact that mental health is equally important to physical health. Studies show that mental health can directly affect the physical health of a person as well. The mental health of a person can be suggestive of their immunitive. High stress can lead to poor immunity.

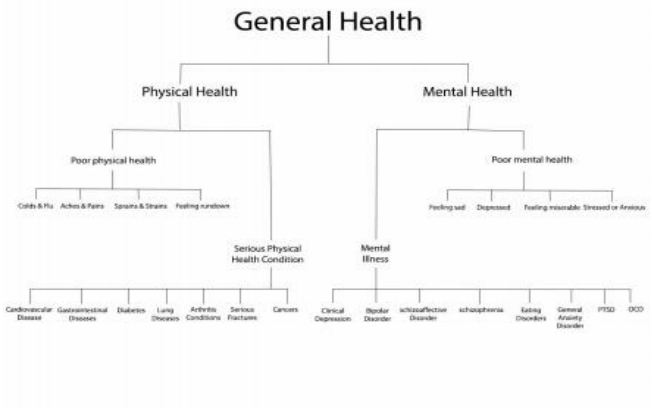


Fig 1. General Health spectrum of a person

The figure 1.1 shows the General Health spectrum of a human being and as we can see half of it focuses on mental health. This is plenty of evidence proving the fact that mental health is equally important.

2. RELATED WORK

Extract health-related causality from Twitter messages using Natural Language Processing: Son Doan, Elly W.Yang, Sameer S.Tilak, Peter W. Li, Daniel S. Zisook and Manabu Torii (April 2019)

In this report an NLP approach was made to extract causality from Twitter messages. They collected the results worth four months of twitter posts related to health and used them to analyse the topics of interests in health. Specifically, the number of matched sentences was 501 out of 29,705 for stress(1.6%), 72 out of 3827 for insomnia(1.8%), 94 out of 11,252 for headache(0.8%). The final casualties extracted were 41 for insomnia, 98 for stress, and 42 for headache. The study had a few limitations as well. The number of rules and patterns are small and they may miss some cause-effect relations in sentences. This study only considered simple cases explicitly reported in single sentences.

A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining: Hong-Han Shuai , Chih-Ya Shen, De-Nian Yang , Senior Member, IEEE, Yi-Feng Carol Lan, Wang-Chien Lee, Member, IEEE, Philip S. Yu, Fellow, IEEE, and Ming-Syan Chen (July 2018)

In this report,online mining social behavior provides an opportunity to actively identify SNMDs at an early stage. Instead, a machine learning framework, namely, Social Network Mental Disorder Detection (SNMDD), that exploits features extracted from social network data to accurately identify potential cases of SNMDs. It also exploits multi-source learning in SNMDD and proposes a new SNMD-based Tensor Model (STM) to improve the accuracy. Our framework is evaluated via a user study with 3,126 online social network users. We conduct a feature analysis, and also apply SNMDD on large-scale datasets and analyze the characteristics of the three SNMD types. The results manifest that SNMDD is promising for identifying online social network users with potential SNMDs.

Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project: Richard G Jackson, Rashmi Patel, Nishamali Jayatilleke, Anna Kolliakou, Michael Ball, Genevieve Gorrell, Angus Roberts, Richard J Dobson, Robert Stewart (Jan 2017)

In this report, NLP is used to create a language model to capture key symptoms of Severe Mental Illness(SMI) from clinical texts. This is done by development and validation of information extraction applications to make sure of SMI symptoms. The distribution of derived symptoms was described in 23 128 discharge summaries from 7962 patients who had received an SMI diagnosis, and 13 496 discharge summaries from 7575 patients who had received a non-SMI diagnosis. Data extracted from 46 symptoms with median score of 0.88, four poorly performed models were excluded.

Detecting depression and mental illness on social media: an integrative review: Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar and Johannes C Eichstaedt (2017)

This paper reviews the recent studies on predicting mental illness through social media. With an improved diagnosis rate of mental illness, some cases might remain undetected. These cases can be found on social media such as Twitter and Facebook. These cases are identified based on their activities on various social platforms, online forum membership, screening surveys and patterns in their language. Large-scale monitoring of social media through automated detection methods also helps to identify depression and other mental illnesses.

Natural language processing in mental health applications using non-clinical texts: Rafael A Calvo, David Nicolas Milne, M. Sazzad Hussain, Helen Christensen (Nov 2016)

This paper gives a taxonomy of data sources and techniques used for mental health support and intervention. Using Facebook, Twitter and other social media to create an online pathway to direct people to health information. It also aims in providing mental health assistance and to generate personalised intervention. Here, the use of social media as a data source which in turn is used to detect emotions and identify people in need of psychological assistance is taken into consideration. Various techniques that are used in labelling and diagnosis is mentioned. Paper also includes wayse aligned to generate and personalise mental health interventions along with the aim to develop a common language that helps to deal with mental health without regional boundaries.

3. PROPOSED SYSTEM

The architecture of the project that we propose is as follows:

Here, the process begins with the collection of raw data from various social networking sites. This data is then pre-processed. This includes cleaning the data that is acquired from the sources and to remove errors from the data. The irrelevant data collected is also removed during this process. It is done to identify and remove missing data and to reduce the original data so as to use the meaningful and reliable data set (Feature Extraction) for better results.

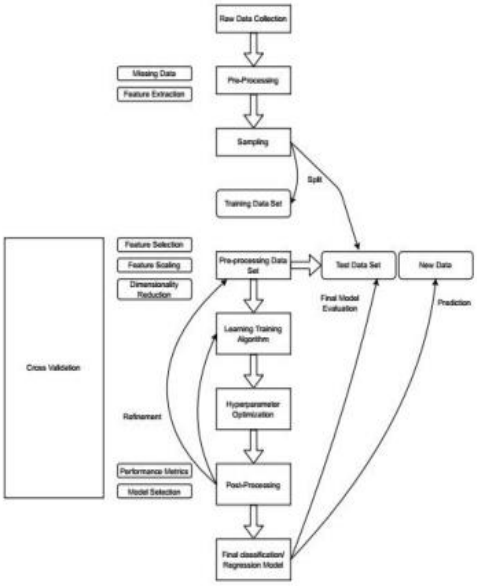


Fig 2. Proposed system architecture

Some common feature extraction techniques are:

- Principal Components Analysis (PCA) :- PCA is one of the most used linear dimensionality reduction techniques. When using PCA, we take as input our original data and try to find a combination of the input features which can best summarize the original data distribution so that to reduce its original dimensions. In PCA, our original data is projected into a set of orthogonal axes and each of the axes gets ranked in order of importance.
- T-distributed Stochastic Neighbour Embedding (t-SNE) :- t-SNE is non-linear dimensionality reduction technique which is typically used to visualize high dimensional datasets. It is extensively applied in image processing, NLP, genomic data and speech processing. It minimizes the divergence between two distributions: a distribution that measures pairwise similarities of the input objects and a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding.

This pre-processed data is then subjected to sampling after which it is split into training and testing data sets. Training data set is again pre-processed to perform:

- Feature Selection :- It is a process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model.
- Feature Scaling :- It is a method used to normalize the range of independent variables or features of data.
- Dimensionality Reduction :- It reduces the time and storage space required. Removal of multicollinearity improves the interpretation of the parameters of the machine learning model. It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D.

This data set is then subjected to the learning algorithm. Hyperparameter optimization is done, that is, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process.

- Training Loss : Training loss or Loss is the error on the training set of data.
- Validation Loss: Validation Loss is the error after running the validation set of data through the trained network. Train/valid is the ratio between the two. Unexpectedly, as the epochs increase both validation and training error drop.

Given below is the graphical representation of training loss and validation loss of the model.

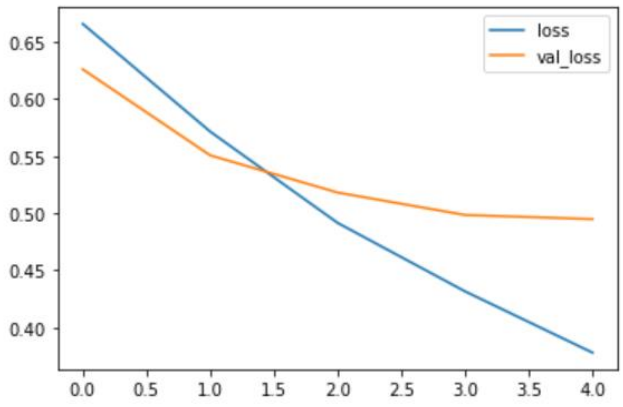
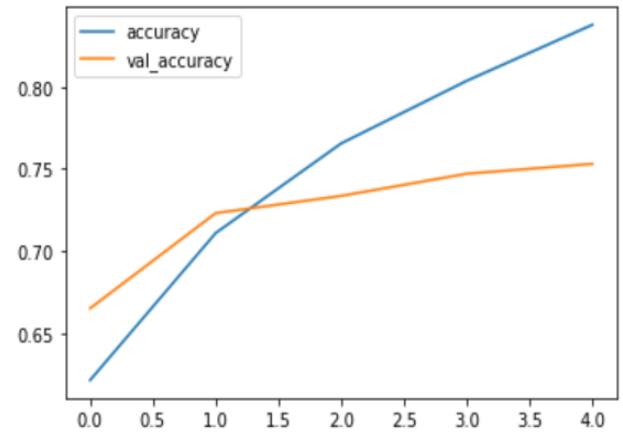


Fig 3. Loss and Validation Loss

Explanation of the terms associated with confusion matrix are as follows –

1. True Positives (TP) – It is the case when both actual class & predicted class of data point is 1.
2. True Negatives (TN) – It is the case when both actual class & predicted class of data point is 0.
3. False Positives (FP) – It is the case when the actual class of data point is 0 & predicted class of data point is 1.
4. False Negatives (FN) – It is the case when the actual class of data point is 1 & predicted class of data point is 0.0

- Accuracy :- Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made.



- Precision :- It is defined as the number of correct documents returned by our ML model.

The testing data set is used to test the model and predict the accuracy.

Given below is the snapshot of the application in work. It analyses the text entered by the user and detects whether the user is suffering from anxiety.

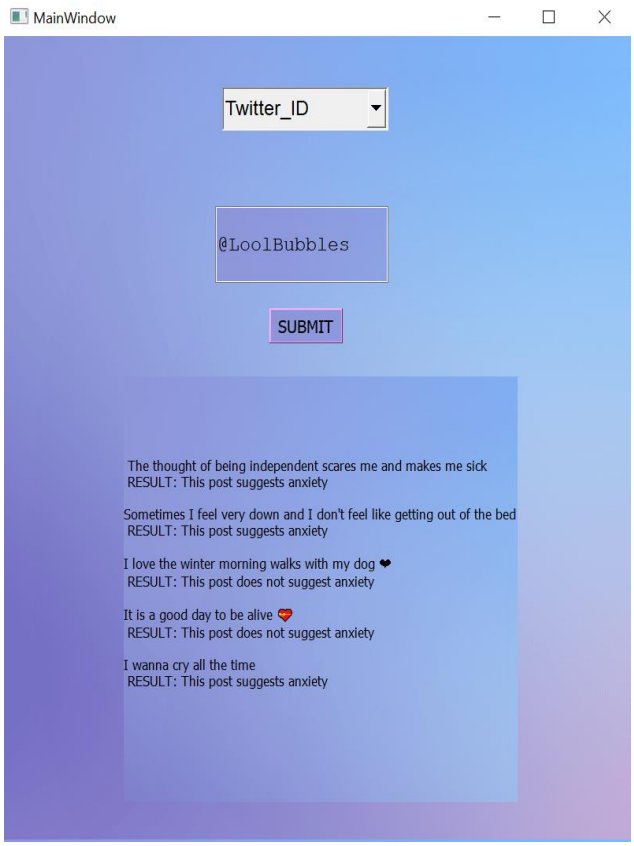


Fig 5. Working of the application

4. APPLICATIONS

The application is capable of analysing texts and tweets that the user posts and utilising it to predict whether the user has anxiety and needs help alleviate the same. It can be equipped to detect similar mental health problems from the emoticons that the user frequently uses since it has become a huge part of speech while expressing oneself on social media platforms. The application can be improved by inculcating audio analysis of the user to better understand the state of mind of the users from their voice. It can also be equipped to perform identification of complex facial features and expressions using emotion detection.

5. SUMMARY

Here, we have briefly presented an application for Identifying Mental Health related issues from social media using NLP, we have explained the objectives of the proposed system, we have conducted a literature survey of previous works and at the end of chapter 2 we have summarised the literature survey and listed the the advantages and disadvantages of each research paper. We have briefly explained the existing system architecture and

the proposed architecture. Further, the report also explains all the tools and technologies implemented in the proposed system such as PyQT and Tensorflow.

ACKNOWLEDGEMENT

We would like to extend our deepest gratitude to our Project Guide, Prof Dhiraj Amin, who guided us and provided us with his valuable knowledge and suggestions on this project and helped us improve our project beyond our limits. Secondly, we would like to thank our Project Coordinator, Prof Krishnendu Nair, who helped us finalize this project within the limited time frame and by constantly supporting us. We would also like to express our heartfelt thanks to our Head of Department, Dr. Satishkumar Verma, for providing us with a platform where we can try to work on developing projects and demonstrate the practical applications of our academic curriculum. We would like to express our gratitude to our Principal, Dr. Sandeep Joshi, who gave us a golden opportunity to do this wonderful project on the topic of 'Identification of mental health related issues from social media using Natural Language Processing', which has also helped us in doing a lot of research and learning their implementation.

REFERENCES

[1]Government of India Mental Health Article, November 2020

[2] Medical News today article on Risk Factors for Mental Health conditions, October 2020

[3] Son Doan, Elly W.Yang, Sameer S.Tilak, Peter W. Li, Daniel S. Zisook and Manabu Torii, "Extract health-related causality from Twitter messages using Natural Language Processing", BMC Medical Informatics and Decision Making, Article number: 79, April 2019.

[4] Richard G Jackson, Rashmi Patel, Nishamali Jayatilleke, Anna Kolliakou, Michael Ball, Genevieve Gorrell, Angus Roberts, Richard J Dobson, Robert Stewart, "Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project", Jan 2017.

[5] Hong-Han Shuai , Chih-Ya Shen, De-Nian Yang , Senior Member, IEEE, Yi-Feng Carol Lan, Wang-Chien Lee, Member, IEEE, Philip S. Yu, Fellow, IEEE, and Ming-Syan Chen, "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining", vol 30, no. 7, July 2018.

[6] Rafael A Calvo, David Nicolas Milne, M. Sazzad Hussain, Helen Christensen, "Natural language processing in mental health applications using non-clinical texts", received Jan 2016, revised Nov 2016, accepted Nov 2016.

[7] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar and Johannes C Eichstaedt, "Detecting depression and mental illness on social media: an integrative review", 2017.

BIOGRAPHIES

Midhun VM,
Student of Pillai College of Engineering, New Panvel.
Pursuing Bachelor's of Engineering Degree in IT Engineering from University of Mumbai.

Sharanya Menon,
Student of Pillai College of Engineering, New Panvel.
Pursuing Bachelor's of Engineering Degree in IT Engineering from University of Mumbai.

Akash Patil,
Student of Pillai College of Engineering, New Panvel.
Pursuing Bachelor's of Engineering Degree in IT Engineering from University of Mumbai.

Dr Dhiraj Amin,
Professor at Pillai College of Engineering, New Panvel,
Department of IT Engineering.