2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) | 978-1-6654-5635-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICERECT56837.2022.10059773

Depression detection from Twitter posts using NLP and Machine learning techniques

Shreyas S Korti
KLE Technological University
Hubballi-580021
kortishreyas05@gmail.com

Suvarna G Kanakaraddi
KLE Technological University
Hubballi-580021
suvarna_gk@kletech.ac.in

***Abstract*—Depression is the one of the most seviour mental issue that the people of world-wide are irrelevant of their ages gender caste and races..etc. In this modern communication world peoples are more comport to express their thoughts in front of social media almost every day. The main agenda of this paper is to propose the data-analytics based model to detect depressed tweeter tweets of the peoples. In this paper then data is going to collect from different user's posted tweets from most popular social-media website like twitter. The depression level can be identified based on the tweets of the users in social-media. The standard methods to detect depression of the users via tweets which is in the form of structured, these methods needs a larger amount of the data from the users. Now a day's social media platform like twitter. Twitter has become more popular to express their views and their emotions in the form of tweets. The data screening can be done based on tweets it shows depressive symptoms of the users. By using machine learning technique we are going to do pre-processing of the data collected from the users. And even using Recurrent neural network (RNN) and NLP techniques, LSTM Deep-learning techniques to identify the depressed tweets in a more convenient manner.**

***Index Terms*—LSTM, RNN, NLP, Machine learning, CNN, SVM.**

I. Introduction

Current population of the world is 400 million, by this survey we come to know that 4.5 percent from the world population are suffering from depression they don't want to express their views and ideology they want to leave alone. In medical world we can say it's a chronic disease through this patient can undergo depression further it leads to suicide. As per survey we come to know that 800000 people suicide every year under the age of(15-29)years old. The main symptoms are collected through depressed tweets via social media i.e., Twitter. By taking user's tweets, we are going check the severity of the posted tweets. And by this we are going to identify the mood of the user by seeing his tweets. Being sadness is no-use its self-demotivating, about themselves these are all beginning stages of psychological symptoms. Symptoms can identified by the persons behavior way of expressing his views neglecting friends family members etc. Depression may leads to physical illness. And the user experiences a huge sudden changes in his body either through weight loss, nervous ness, hairless. Muscle pain. . . .etc. Basically it has good treatment rate nearly 67% of peoples undergone for proper treatment. The more convenient way to detect the depressed peoples via tweets. By predicting the level of risk of the users can be viewed. In this paper we are going to use machine learning techniques to analyze from the user tweeted posts via from twitter. Most of the users express their emotions via tweets. The main motive is to analyze the raw data collected from the twitter users is that first of all its very big platform where millions of peoples connect through this socialmedia by expressing their ideology. To understand the interactions of the people. Basically twitter is a best platform. twitter has nearly 321 million active users this users mainly like to express their lifestyles, thoughts either it may be happy or sad in social-media. Tweets have maximum 280 characters. by collecting the data from the users profile we can identify the behavior of that users. Whether this personality changes every day-to-day activity or not. This behavior of the users can be identified via their tweets. Here beautiful soup method is applied to collect the data from the twitter. Once the data is collected further processing can be done using machine learning techniques. The data is unclusterd format by using NLP classifier we can classify the data. First we have to pre-process by removing special characters and emojis, by applying various techniques such as (RNN), LSTM Deep-learning techniques to classify the tweets. Whether the level of the tweets are positive, negative or moderate.

II. Related Works

1) ***Early Detection of Depression****:* Social Network Analysis and Random Forest Techniques. Depression is one most major problem were we can't understand the mood of the user. We use to identify the users mood through tweets basically as mentioned in the above paper this psychological disorder is spreading like a virus to other people they are suffering and even suicide rate is more due to lack of treatment. These data are extracted from social media like twitter. Twitter user's tweets the data by keeping this in mind the author is using various techniques for early detection as MDD(major depressive order). Depression is like a symptoms effecting more than 250 millions of people across the world here they have used various machine learning techniques to classify the twitter data and some more techniques to classify the behavior the data i.e. 1.Random forest with 2 threshold functions and 2. Independent RF- classifiers and to detect depressed tweets and to classify non-depressed tweets. In both the Cases the features are defined in the form of text and symbolic similarity. the author address how this model is better than singleton model

978-1-6654-5635-7/22/$31.00 ©2022 IEEE

to improve the state-of-art. By this we say that this will helpful for finding new solution to deal with this problem. [1]

*2) **The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods**:* How Netherlands peoples are Undergoing depression in short i.e. (NESDA). Basically we are dealing with long-term anxiety disorder and they are dealing with some bio-logical problems like long term anxiety. There are some family which has a record of depression. It is transferred via through genes for upcoming generations of their family. As per survey of the author there is place in Netherlands where the population is 2981 person in that 1701 is diagnosed that they are in depression and anxiety, And even 907 person was declared permanently depressed and among that 373 peoples are healthy. Hence by keeping this in mind several samples has been collected from the depressed peoples information is gathered for better treatment of their lives. [2]

Fig. 1. Depression under NESDA peoples

*3) **Depression: Prevalence and Associated Risk Factors in the United Arab Emirates**:* Now a days depression has become more risk factor in the current world each and every person is undergoing depressed mode. The author trying to convince in this paper is that how UAE peoples are suffering from depression, Here they feel comfort for systematic way of approach, Searching some electronic databases to collect information of UAE populations relating to depression slowly. By this we come to know that not men's and women's even children's are also suffering from this diabetes depression and here the work load is more heavy. They can't get sufficient rest and due to lack support from society and mainly this leads to seviour chronic disorder such as Headache, weight loss....etc. And due to deficiency of vitamin-D that leads to risk factor suppose the family has history of depression. Here we can see low academic performance of the peoples. And in-taking of alcohol is very dangers to health to come out from the depression it's not only one solution to get out from that there are many solution like meditation can help to come out from that depression. [3]

*4) **Detecting suicidality on Twitter**:* The author tries to explain how peoples are committing suicide via social-media platforms and how twitter is taken for under investigation we will discuss in this paper. So, now a days peoples are

Fig. 2. Depression under UAE peoples

undergoing depression from past many years. Where peoples are increasing day-by-day and their level of problems like mental health illness....etc are also increased. Now a days peoples of the world don't want to exposure to new world they want to sit around four corners of the room and they don't want interact first of all and this is the reason they get affect mentally very badly. The people don't want to communicate with persons, But they like to interact with social-media platform like twitter .....etc. The user will post some depressed tweets, where whole world is watching online seeing that tweet also, some time they will think of that user, some of them get depressed and finally they suicide themselves. Here in this they are using machine learning techniques to detect that depressed similar tweets. [4]

Fig. 3. Depressed tweets

*5) **Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences**:* Depression is one of the major disease till today millions of people suffering can directly leads to death. Till today many peoples are suffering from depression due to lack of treatment and negligence of the people can lead to severely mental problems and all peoples can express their views and thoughts the most popular social media websites like twitter. users are spending lots of time in social media can impact severely

.By using machine learning techniques we can classify the social media tweets by applying suitable machine learning techniques such as CNN and deep-learning techniques can be done for classification based on user-level linguistic metadata. By inter-mixing both the above techniques it helps to detect early depression stage of the users. [5]

*6) **Depression detection in Tweets from Urban Cities of Malaysia Using Deep Learning**:* In this we are going to see how Malaysian peoples are suffered during pandemic time of covid-19. While every persons of the world are maintaining social distance and following covid-19 rules and regulations for betterment of peoples. Some of the people are sitting inside the house for self-precaution purpose those people lead to some physical and mental illness issues and slowly turned towards social media websites like twitter. A Twitter depression analysis main intention is to identify the depressed tweets in social media. So the peoples are addicted towards social media by posting some of the depressed tweets and many of them were where viewing that posts that leads to depression to other peoples. It's like chain process. Here the depressed tweets can be identified through some of the machine learning techniques (LSTM) Deep learning techniques and NLP by using this approach the author is going to solve in this paper. The main aim of this paper to identify the peoples who are suffering from depression there is necessity of early treatment. And by using above model they attained the accuracy of 94%. [6]

*7) **Depression Detection on Social Media Using Text Mining**:* Depression is the one of the seviour disease to the people, it is very difficult to identify the mood of the users, Basically Depressed tweets make some time people uncomfortable such that they take some wrong decisions that leads to death. Social media is a platform where millions of peoples are interacting with each other sharing their thoughts and opinions, the user who tweets depressed tweets can be easily identified based on the level of tweets that they have posted. Due to rapid increase of patient, handling has become more difficult. In this paper the author came up with some conclusion which helps to identify the depressed tweets using some machine learning techniques such as (CNN+LSTM) classifier and logistic regression for more accuracy. [7]

*8) **Depression Screening in Humans With AI and Deep Learning Techniques**:* Now a day social media platform has widely used mainly for communication purpose. Where peoples use to express their thoughts and views in front of the world and sharing their personal and family background information. These sensitive information can lead further too depression and finally commits suicide. First of all committing suicide is against the law. So here the author want to predict early detection of people such that if they get proper treatment in early stage there are more chance of early recovering. The author used machine learning techniques, NLP. Through this early detection tweets can be analyzed based on tweets. Treatment is provided for only concerned persons. Every individual person has different mindset the way of thinking addressing his thoughts differs from person-to-person. Here in this paper they have used AI and Deep learning techniques for identifying social-media tweets. Even they have used (CNN+LSTM). In this live data is collected from different social media like Facebook, Instagram YouTube....etc. And all the data is combined and by using above techniques the accuracy was attained. [8]

Fig. 4. Identification of depression

*9) **Sentiment Analysis on COVID Tweets**:* : The author wants to explain by taking sentimental tweets based on count-vectorizer and tf-idf approach. And depressed tweets using deep learning techniques. Due to high demand of social media such as twitter during covid-19 has become more popular in every sector. This is causing highly depressive and untolerable situation of the users around the world. Posting some depressed tweets leads to suicide. Here the author used different classifiers on covid-19 data from twitter for even training and using deep neural networking for attaining more accuracy rate using to most popular embedded techniques. They are 1.Count Vectorizer and 2.Term Frequency-Inverse hence they have compared the accuracy based on above techniques mentioned in this paper. And the data set was huge in volume in case of covid-19 tweets. The accuracy was quite compromising in this paper which has 10 efficiency. [9]

III. Models

*1) **RNN (Recurrent neural network)**:* It's a machine learning technique mainly used to understand the data. And it has its own memory to maintain the data by which it can keep track of both (Previous + Current) state. And it will handle sequential data too. It basically tries to maintain same function for every network data.

Here, Ht stands for requisite time stamp and and Function f(.....) stands for using has neural network techniques , ht-1 stands for previous state and Xt stands for current input state. We use formula as,

$$H(t) = f(h_{t-1}, X_t) \tag{1}$$

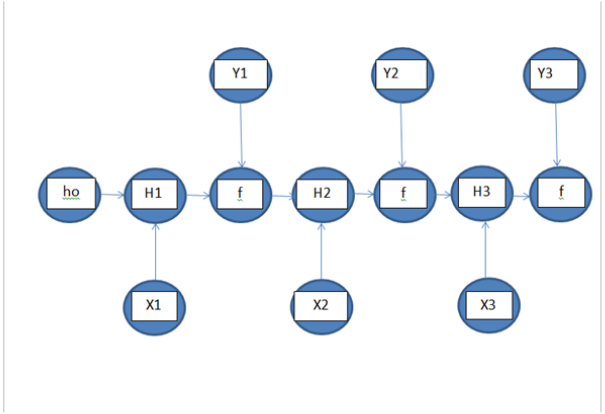

Fig. 5. Architecture diagram of RNN

Activation function Formula:-

$$h_t = tanh(w_{hh} * h_{t-1} + w_{xh} * x_t) \tag{2}$$

Here, h_t stands for requisite time stamp. Tanh function mainly used in hidden layers lies in between -1 and +1. W_{hh} Stands for weight of previous hidden state. H_{t-1} Stands for previous state. W_{xh} Stands for weight of current input state. X_t Stands for current input state. **For calculating output state**:-

$$Y_t = w_{hy} * h_t \tag{3}$$

*2) **LSTM (Long short term memory networks)**:* LSTM is a deep learning technique. Basically it carries some important data as input and computational is going to happen for output.

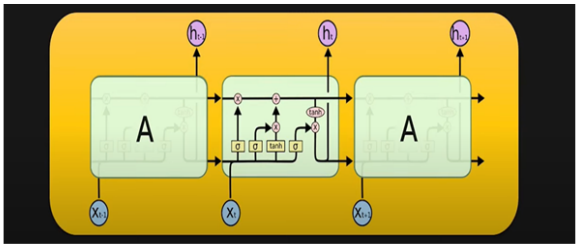

Fig. 6. Architecture diagram of LSTM

And output will be generated with the Knowledge of previous state or we can say it's a combination of (previous + current) state keywords matching is been done finally output will be generated. Input is taken from the users and sends it to neural networking by using Tanh(activation function) for computation such that output will be generated. It will directly store important keyword and it will be more helpful for model to predict the data. And at every state new input will be added it will extract important keywords from the data then It relates with upcoming text with the previous knowledge of keywords.

IV. Comparitive Analysis Based on Machine Learning Techniques

Methodology	Outcomes
Using machine earning techniques. [1]	40% of accuracy
Using machine earning techniques. [2]	43% of accuracy
Using machine learning techniques. [3]	45% of accuracy
SVM and Logistic Regression. [4]	There are mainly 2 accuracy 64% and 61%.
Hybrid models are (LSTM+NLP). [5]	76% of accuracy.
(LSTM Deep-learning + NLP). [6]	94% of accuracy.
Text-Mining,NLP,(CNN+LSTM). [7]	93% of accuracy
(CNN+LSTM) techniques. [8]	98.54% of accuracy
(Deep learning+ NLP + Naivebayes) [9]	84.5% of accuracy

V. Results and Discussion

*1) **Dataset**:* Basically Twitter data is taken from kaggle platform. The dataset is in the form of csv file format. It has mainly 7 columns and the data has nonce values like emojis and special characters. By applying NLP classifiers, we have to pre-process the data by removing the nonce value.

	tweet.id	created_a	text	location	retweet	favorite
0	1.45E+18	########	Open disc	Cheyenne	0	0
1	1.45E+18	########	Plenty of things are		0	0
2	1.45E+18	########	I feel a little hopeles		0	0
3	1.45E+18	########	Which is n	Denver, C	0	0
4	1.45E+18	########	So	Portland	0	2
5	1.45E+18	########	No parent	Bombay, I	1	4
6	1.45E+18	########	Being in	Alberta, C	7	17
7	1.45E+18	########	I am so		0	1
8	1.45E+18	########	Open disc	Cheyenne	0	0
9	1.45E+18	########	Plenty of things are		0	0
10	1.45E+18	########	I feel a little hopeles		0	0
11	1.45E+18	########	Which is n	Denver, C	0	0
12	1.45E+18	########	So	Portland	0	2
13	1.45E+18	########	No parent	Bombay, I	1	4
14	1.45E+18	########	Being in	Alberta, C	7	17
15	1.45E+18	########	I am so		0	1
16	1.45E+18	########	Open disc	Cheyenne	0	0
17	1.45E+18	########	Plenty of things are		0	0
18	1.45E+18	########	I feel a little hopeles		0	0
19	1.45E+18	########	Which is n	Denver, C	0	0
20	1.45E+18	########	So	Portland	0	2
21	1.45E+18	########	No parent	Bombay, I	1	4
22	1.45E+18	########	Being in	Alberta, C	7	17
23	1.45E+18	########	I am so		1	2

Fig. 7. Analysis of Twitter dataset

*2) **Results**:* LSTM performs better as campared to RNN. RNN is very much difficult to train And LSTM has a memory cell which maintains information in the memory for a long duration. LSTM is more powerfull as campared to RNN. LSTM is more faster than RNN. But By combining hybrid model like LSTM+RNN we can achieve better result as campared to LSTM and RNN individually by this we conclude that hybrid LSTM+RNN is best model.

TABLE I
ANALYSING THE RESULTS

Methodology	Outcomes
(RNN)	70% accuracy
(LSTM)	89% accuracy
(LSTM+RNN)	91% accuracy

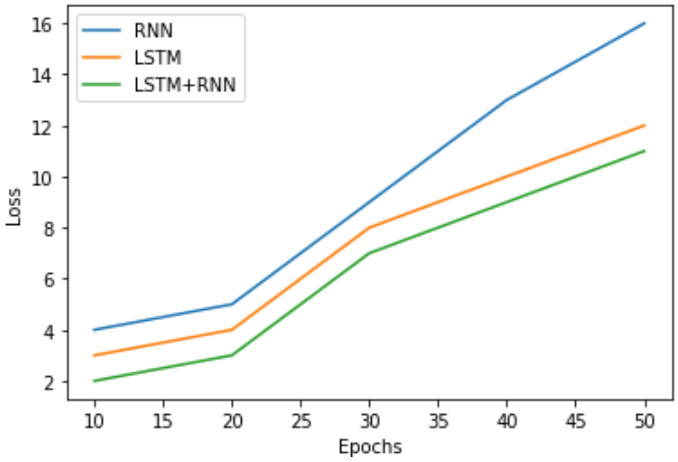

Fig. 8. Analysis of Epochs vs Loss Function

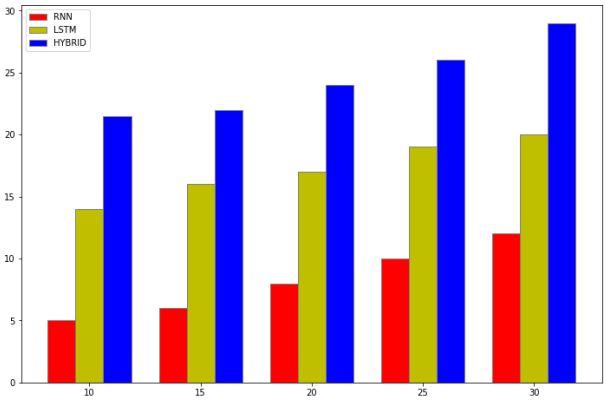

Fig. 9. Analysis Of model accuracy

VI. CONCLUSIONS

Depression can lead to some severe mental problem towards the path of the suicide. By approaching machine learning techniques and NLP classifier we can able to detect depressed tweets of the users in the social media. Twitter is a platform were people's communicate different parts of the world. But we see here misusing of the social media by tweeting negative tweets by reading it may affect even different person too. To classify the depressed tweets mainly we consider (positive, negative and moderate tweets). The main motive is to identify the user's depressed state via social media. Hence in fututre we will train model in a such a way that if Somebody tries to post the depressed tweets, it should be deleted /electronized in a such a way that no user can read that tweets by doing this there is a less chance of effecting other peoples of the world. And twitter has to make some policies such that a person who posts a depressed tweets he has pay fine in such a way that he has to re-think while posting a depressed tweets .

REFERENCES

[1] Early detection of depression: social network analysis and random forest techniques, Cacheda, Fidel and Fernandez, Diego and Novoa, Francisco J and Carneiro, Victor and others,2019.

[2] The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods, Penninx, Brenda WJH and Beekman, Aartjan TF and Smit, Johannes H and Zitman, Frans G and Nolen, Willem A and Spinhoven, Philip and Cuijpers, Pim and De Jong, Peter J and Van Marwijk, Harm WJ and Assendelft, Willem JJ and others.

[3] Depression: prevalence and associated risk factors in the United Arab Emirates, Razzak, Hira Abdul and Harbi, Alya and Ahli, Shaima ,Oman Medical Journal,2019.

[4] Detecting suicidality on Twitter, O'dea, Bridianne and Wan, Stephen and Batterham, Philip J and Calear, Alison L and Paris, Cecile and Christensen, Helen.

[5] Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences, Trotzek, Marcel and Koitka, Sven and Friedrich, Christoph M ,2018.

[6] Depression Detection in Tweets from Urban Cities of Malaysia using Deep Learning, author=Sri, EK Priya and Savita, KS and Zaffar, Maryam ,2021.

[7] Depression Detection on Social Media Using Text Mining, Dessai, Sukanya and Usgaonkar, Soniya Shakil,.

[8] Depression Screening in Humans With AI and Deep Learning Techniques, Wani, Mudasir Ahmad and ELAffendi, Mohammad A and Shakil, Kashish Ara and Imran, Ali Shariq and Abd El-Latif, Ahmed A ,2022.

[9] Global sentiment analysis of COVID-19 tweets over time, Mansoor, Muvazima and Gurumurthy, Kirthika and Prasad, VR and others,2020.

[10] CairoDep: Detecting depression in arabic posts using bert transformers, El-Ramly, Mohammad and Abu-Elyazid, Hager and Mo'men, Youseef and Alshaer, Gameel and Adib, Nardine and Eldeen, Kareem Alaa and El-Shazly, Mariam ,2021.

[11] Depression detection by analyzing social media posts of user, Al Asad, Nafiz and Pranto, Md Appel Mahmud and Afreen, Sadia and Islam, Md Maynul ,2019.

[12] Depression Detection in Tweets from Urban Cities of Malaysia using Deep Learning, Sri, EK Priya and Savita, KS and Zaffar, Maryam ,2021.

[13] Facebook social media for depression detection in the Thai community, Katchapakirin, Kantinee and Wongpatikaseree, Konlakorn and Yomaboot, Panida and Kaewpitakkun, Yongyos.

[14] Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models, Raza, Ghulam Musa and Butt, Zainab Saeed and Latif, Seemab andWahid, Abdul ,2021.

[15] Integrating deep learning for NLP in Romanian psychology, Schuszter, Ioan Cristian,2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC),2018.

[16] The detection of depression using multimodal models based on text and voice quality features, Solieman, Hanadi and Pustozerov, Evgenii A,2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus),2021.
[17] Use of Ensemble Machine Learning to Detect Depression in Social Media Posts, Jagtap, Nakshatra and Shukla, Hrushikesh and Shinde, Vaibhavi and Desai, Sharmishta and Kulkarni, Vrushali,2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC),2021.
[18] Usage of emotion recognition in military health care, Tokuno, Shinichi and Tsumatori, Gentaro and Shono, Satoshi and Takei, Eriko and Yamamoto, Taisuke and Suzuki, Go and Mituyoshi, Shunnji and Shimura, Makoto.
[19] Anxiety recognition of college students using a Takagi-Sugeno-Kang fuzzy system modeling method and deep features, Meng, Xiangmin and Zhang, Jie ,2020.
[20] Understanding Depression Detection Using Social Media, Latif, Aliza Abdul and Cob, Zaihisma Che and Drus, Sulfeeza Mohd and Anwar, Rina Md and Radzi, Husni Mohd,2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE),2021.
[21] S. G. Kanakaraddi, A. K. Chikaraddi, K. C. Gull and P. S. Hiremath, "Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 287-292
[22] Kanakaraddi, S.G., Chikaraddi, A.K., Aivalli, N., Maniyar, J., Singh, N. (2022). Sentiment Analysis of Covid-19 Tweets Using Machine Learning and Natural Language Processing. Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems. Advances in Intelligent Systems and Computing, vol 1415. Springer, Singapore.