

ANALYSIS AND DETECTION OF DEPRESSION LEVEL USING SOCIAL MEDIA

Pragati J. Dhengale*¹, Nikita G. Zade*², Sakshi S Hedau*³, Pooja R. Sharma*⁴, Pallavi Lanjewar*⁵, Prof. Yogeshre D. Choudhary*⁶

*^{1,2,3,4,5}Department Of Information Technology, Karmaver Dadasaheb Kannamwar College of Engineering, Nagpur Maharashtra, India.

Guide, Department Of Information Technology, Karmaver Dadasaheb Kannamwar College of Engineering, Nagpur Maharashtra, India.

ABSTRACT

Detection of depression through messages sent by a user on social media is usually an elaborate task owing to their recognition and trends. In recent years, messages and social media has over up being an extremely shut illustration of a person's life and his standing. Usually|this can be} often a huge stockpile of knowledge on one or two of person's behaviour and could be used for detection of assorted mental sicknesses (depression in our case) exploitation tongue method and Deep Learning. This project is concerning constructing a deep learning model exploitation IP to predict such mental disorders. Remembering networks area unit well-suited to classifying, method and making predictions supported datum information since there are a unit usually lags of unknown length between necessary events throughout a datum.

Keywords— Depression, social media, mental illness, deep learning, NLP.

I. INTRODUCTION

Depression could be a typical mental state disorder that has long been denned mutually ill health with a gaggle of diagnostic criteria. it always co-occurs with anxiety or different psychological and physical disorders; and includes pertaining to the emotions and behaviour of the affected individuals. Keep with the globe Health Organization study, there square measure 322 million of us estimated to suffer from depression, care for four.4% of the worldwide population. In today's world, communication through social media is rising as a vast deal. They're willing to share their thoughts, stories and their personal feelings, mental states, wants on social network sites, blogging platforms etc. Receivers use the manuscripts from emails and different kinds of social media comments to create correct reasoning and to correct mistakes. Once we write digitally on social media, their texts square measure processed automatically. Language method techniques square measure acquainted with infer people's mental behaviour.

According to World Health Organization, depression might be a standard worldwide folio that affects an enormous amount of individuals in spite of their age. There square measure multiple factors that interfere with the amount of your time detection and treatment like lack of consummate specialists, social shaming, and improper identification therefore on. The ever-lasting depression disorder could lead to suicide if the depressed individuals do not appear to be equipped with correct practice, instant facilitate and would possibly to boot suffer from anxiety. This work is targeted at the detection of depression and anxiety from tweets. The experiment conducted throughout this work wants the text data that the chosen data provide is Twitter where of us tweet relating to our feelings, hopes, desires, thoughts, stories and mental states.

The goals of our analysis are: collect the general public on the market media messages of healthy and self-diagnosed those that contains mixed emotions thus live the extracted Twitter data and apply NLTK and deep learning classifiers like LSTM-RNN to predict depressive and anxiety tweets. we tend to square measure able to explore for a solution to a performance increase through an accurate choices alternative and their multiple feature mixtures. First, we tend to decide the foremost useful linguistic choices applied for depression identification to characterize the content of the posts. Second, we've got an inclination to investigate the correlation significance, hidden topics and word frequency extracted from the text. we've got an inclination to match the performance results supported by three single feature sets and their multiple feature mixtures. In our experiment, we've got an inclination to use the data collected from the Reddit social media platform.

II. AIMS AND OBJECTIVE

- Identify the foremost effective deep neural spec among some of the elect architectures that were with success utilized in tongue process tasks.
- The architectures square measure wont to notice users with signs of mental diseases (depression in our case) given restricted unstructured text knowledge extracted from the Twitter social media platform.
- To investigate the result of depression detection, we have a tendency to propose the Deep learning technique as an associated economical and ascendable technique.
- The main contribution of this study lies in exploiting a chic, diverse, and discriminating feature set that contains each tweet text and behavioural trends of various users.
- This study is often extended within the future by considering additional deciliter models that live} extremely unlikely to over-fit the used knowledge and notice an additional dependable thanks to measuring the features' impact.

III. LITERATURE SURVEY

MICHAEL M. TADESSE , HONGFEI statue maker , BO XU , AND LIANG principle : we are able to significantly improve performance accuracy. the only single feature is a word with the Support Vector Machine (SVM. The strength and effectiveness of the combined choices are most successfully incontestable with the Multilayer Perceptron (MLP) classifier resulting in the very best performance for depression detection reaching cardinal accuracy.

P.V. Rajaraman AsimNath, Akshaya.P.R, Chatur Bhuja.G : Messages and social media has all over up being a really shut illustration of a person's life and condition. This is often an enormous stockpile of knowledge of a couple of person's behaviour and might be used for the detection of assorted mental diseases (depression in our case) victimization language process and Deep Learning.

Akshi Kumara, Aditi Sharmab, Anshika Arorac : This mixed anxiety-depressive disorder is predominantly associated with erratic thought technique, restlessness and wakefulness. supported the linguistic cues and user posting patterns, the feature set is printed using a 5-tuple vector. Associate in Nursing anxiety-related lexicon is built to look at the presence of tension indicators.

HoyunSong, Jinseon You, Jin-Woo Chung Jong C. Park : we tend to propose a Feature Attention Network (FAN), galvanized by the tactic of designation depression by associate knowledgeable global organization agency that features a background regarding depression. we tend to appraise the performance of our model on the associate outsized scale general forum (Reddit Self-reported Depression Diagnosis) dataset. Experimental results demonstrate that FAN shows good performance with high interpretability despite a smaller kind of posts in work information. we tend to analyze whole totally different aspects of posts by depressed users through four feature networks designed upon psychological studies, which might facilitate researchers to investigate social media posts to look out a useful proof for depressive symptoms.

IV. DESIGN AND IMPLEMENTATION

A. Proposed System

Mental illness detection in social media is going to be thought of as a complicated task, primarily as a result of the troublesome nature of mental disorders. In recent years, this analysis area has begun to evolve with the continual increase within the quality of social media platforms that became academic degree integral a locality of people's life. This shut relationship between social media platforms and their users has created these platforms to duplicate the users' personal life with fully completely different limitations. In such academic degree surroundings, researchers area unit bestowed with a wealth of data concerning one's life. to boot to the quantity of quality in identifying mental sicknesses through social media platforms, adopting supervised machine learning approaches like deep neural networks haven't been widely accepted as a result of the difficulties in obtaining good amounts of annotated employment data. as a result of these reasons, we have a tendency to tend to try to determine the foremost effective deep neural specification among a variety of selected architectures that were successfully utilised in tongue method tasks. we have a tendency to tend to gift the latest model NLTK with LSTM. The chosen architectures area unit accustomed observe users with signs of

mental sicknesses (depression in our case) given restricted unstructured text data extracted from the Twitter social media platform.

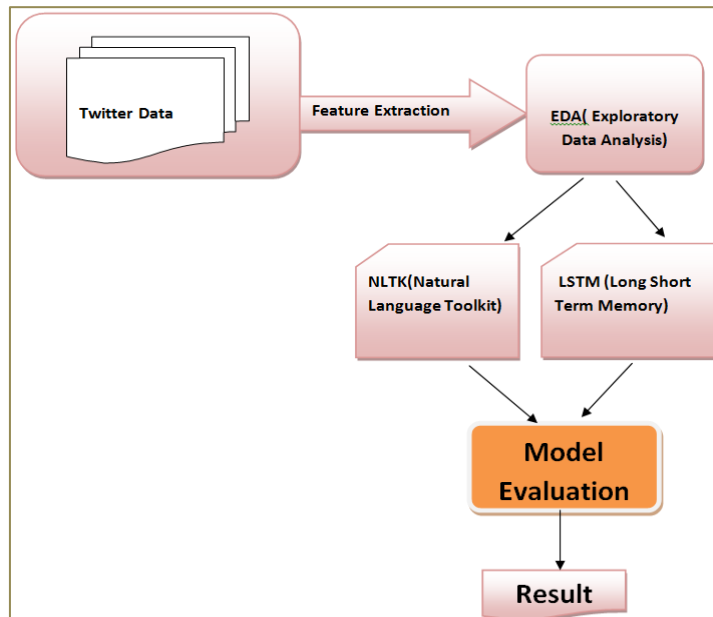


Fig. 1 System flow Diagram

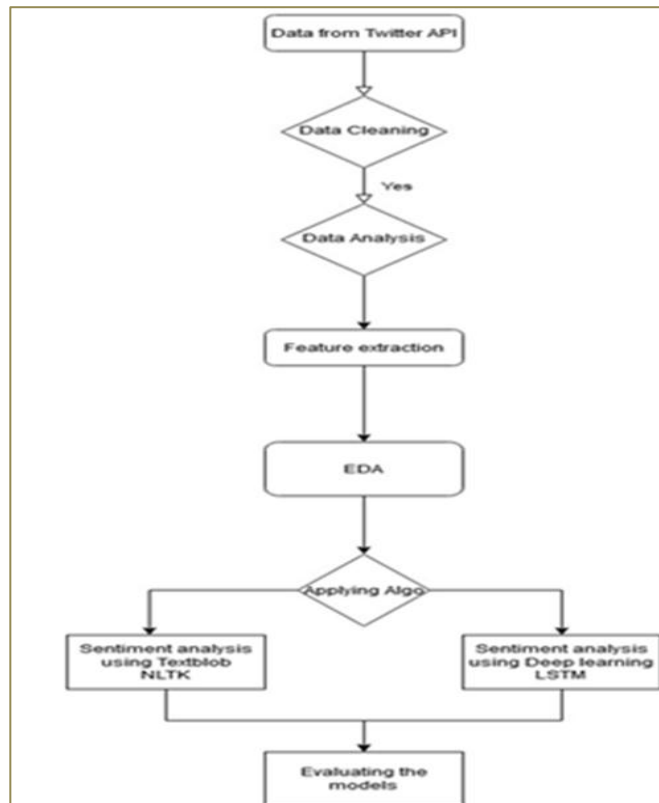


Fig. 2 Data Flow Diagram

Data for the model square measure extracted from the Twitter website. info is in unstructured format and it will have uncountable useless info. so we'll initial do info improvement. Once we've got a bent to wash the data we'll apply EDA for locating the polarity of the data like positive, neutral, or negative automatically or many difficult sentiments like happiness, sadness, anger, joy, etc. Once EDA is finished will pass the data to the NLTK model and LSTM model which we are able to opt for the model giving the only accuracy.

Modules

1. data selection: a training set is utilized to teach the machine learning technique to grasp the potential relationship between the informative variables and target variable.
2. data Cleaning: data improvement suggests that filtering and modifying your data specified it's easier to explore, understand, and model. Filtering out the weather you are doing not want or would really like so as that you just don't get to take into account or technique them.
3. data imputation: Machine learning algorithms would like numeric input values, and price to be present for each row and column in an exceedingly} very dataset. As such, it's normal to identify missing costs in an exceedingly} very dataset and replace them with a numeric worth.
4. data Analysis: the simplest way of information analysis that automates analytical model building. it is a branch of computing that supported the thought that systems can learn from data, verify patterns and make choices with the smallest amount of human intervention.
5. data image: data mental image is that the illustration of or data in an exceedingly} every graph, chart, or different visual format. Machine learning makes it easier to conduct analyses like revelatory analysis, which could then operate helpful visualizations to gift.
6. coaching: coaching data is that the data you utilize to teach associate formula or machine learning model to predict the result you vogue your model to predict.
7. Testing: A take a glance at the dataset is also a dataset that is freelance of the working dataset, but that follows identical chance distribution attributable to the working dataset.
8. formula selection: Machine learning algorithms square measure the engines of machine learning, which suggests it is the algorithms that flip a chunk of data set into a model.

B. planned Methodology

1) Linguistic communication Toolkit (NLTK): The linguistic communication Toolkit, or a lot of usually NLTK, maybe a suite of libraries and programs for symbolic and applied mathematics linguistic communication process (NLP) for English written within the Python programing language. it had been developed by Steven Bird and Edward Loper within the Department of pc and Data Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample information. it's in the middle of a book that explains the underlying ideas behind the language process tasks supported by the toolkit, and reference work. NLTK is meant to support analysis and teaching in information science or closely connected areas, together with empirical linguistics, scientific discipline, computer science, info retrieval, and machine learning. NLTK has been used with success as a teaching tool, as a private study tool, and as a platform for prototyping and building analysis systems.

2) Long Short Term Memory (LSTM): The central role of the associate degree LSTM model is controlled by a memory cell referred to as a 'cell state' that maintains its state over time. The cell state is that the horizontal line that runs through the highest of the below diagram. It is often envisioned as a transporter through that info simply flows, unchanged. info is often additional to or aloof from the cell state in LSTM and is regulated by gates. These gates optionally let the data flow in and out of the cell. It contains a point-wise multiplication operation and a sigmoid neural internet layer that assist the mechanism.

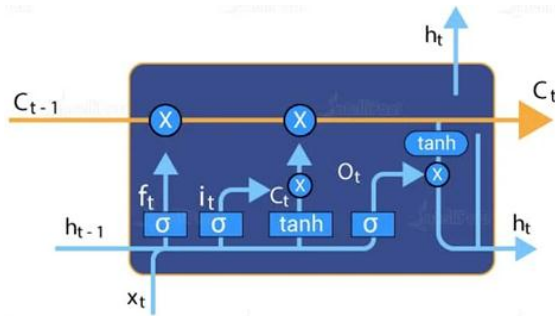


Fig. 3 LSTM Logic

V. RESULT AND DISCUSSION

```

1 Tweets,User,User_statuses_count,user_followers,User_location,User_verified,fav_count,rt_count,tweet_date
2 @gurustulyahesh @glpustolyevs @sashitika @ivekkrishnani @majorthefilm @prakashraaj,Salman
3 Khan,42942,42411934,MUMBAI,True,14885,3255,2021-04-12 10:51:45
4 1,Isse kehte hain dhamakedaar teaser! Really happy and proud to launch this. Congratulations to the team.. and salute..
5 https://t.co/tw3kaPBKri,Salman Khan,42942,42411934,MUMBAI,True,28723,6128,2021-04-12 10:51:45
6 2,"Happy birthday brother, wish u all the health, happiness, success.. stay blessed stay safe . @jaydevgn https://t.co/4x8ty23hbB",Salman
7 Khan,42942,42411934,MUMBAI,True,79460,5631,2021-04-02 17:49:49
8 3,"Karo apne aap ko aur family ko protect with the FRSH germ protection spray!
9 #RahofrshRahoSafe
10 Click now- https://t.co/hx1wCR6ff9",Salman Khan,42942,42411934,MUMBAI,True,31661,3706,2021-03-31 05:30:07
11 4,Happy bday brother .... Wish you all the success and happiness. @AlwaysRamCharan,Salman Khan,42942,42411934,MUMBAI,True,93549,10088,2021-
12 03-27 17:46:23
13 5,"Arre wahni @arbaazskhan
14 This looks truly amazing just like the Mangoes at @aamwala
15 Place your order today:
16 Visit : https://t.co/altkq1816b",Salman Khan,42942,42411934,MUMBAI,True,21359,2791,2021-03-27 05:30:23
17 6,"Double good news hai apke liye. Being Human Clothing ki Spring Summer'21 Collection is here. Aur, Chandigarh & Sri-
18 https://t.co/EIAHR00tdt",Salman Khan,42942,42411934,MUMBAI,True,38596,4404,2021-03-26 10:30:03
19 7,"@DareToLeap with more pixels!
20 Ayie le chale apko better clarity ki duniya me with the Camera Flagship.. https://t.co/EHjGBvWxs",Salman
21 Khan,42942,42411934,MUMBAI,True,20109,3423,2021-03-24 15:31:18
22 8,Took my first dose of vaccine today....,Salman Khan,42942,42411934,MUMBAI,True,113840,8228,2021-03-24 12:36:29
23 9,"Eid ka commitment tha, Eid par hi aayenge kyun ki ek baar jo maine.....
24 #RahdeOn13thMay #2MonthsToRahde.. https://t.co/VGeaEcV0EH",Salman Khan,42942,42411934,MUMBAI,True,79586,11440,2021-03-13 06:30:03
25 10,"Congrats farhan teaser looks good ...
26 #Toofaan @FarOutakhtar @excelmovies @310Kassin @ritesh_sid https://t.co/rKKSbqRWV",Salman Khan,42942,42411934,MUMBAI,True,29205,3403,2021-
27 03-12 11:19:10
28 11,"PEPSI AB AUR BHI ZYADA REFRESHING
29 #WhattayGarmi
30 #WhattayRefreshingPepsi
31 #HardhoonthleinSwag

```

Fig 4. Collection of data from twitter

Here, we firstly collect all the data from the tweets that user has been tweeted. This data includes name, time, data, emojis, stockwords, etc.



Fig 5. Data cleaning and create word cloud

We apply data cleaning on that collected data and create word cloud as shown in figure 4.

```
In [ ]: sentiment_df3.groupby("sentiment")
```

```
In [134]: sentiment.get_group("positive")
```

```
Out[134]:
```

	Tweets	User	User_status_count	user_followers	User_location	User_verified	fav_count	ret_count	tweet_date	clean_tweets	sent
1	Issa khathe baan @hananabakar lesser! Ready hai!	Salman Khan	42942	42411934	MUMBAI	True	28723	6126	2021-04-12 10:51:45	Issa khathe baan @hananabakar lesser Ready hai!	positive
2	Happy birthday brother, wish u all the health!	Salman Khan	42942	42411934	MUMBAI	True	79460	5631	2021-04-02 17:49:49	Happy birthday brother wish u all the health!	positive
4	Happy bday brother ... Wish you all the success!	Salman Khan	42942	42411934	MUMBAI	True	93549	10000	2021-03-27 17:46:23	Happy bday brother Wish you all the success!	positive
5	Are waah! @abac@ShahinThis looks huy amaz!	Salman Khan	42942	42411934	MUMBAI	True	21359	2794	2021-03-27 05:30:23	Are waah @abac@ShahinThis looks huy amaz!	positive
6	Double good news hai apple eye. Being Hemant C.	Salman Khan	42942	42411934	MUMBAI	True	30596	4404	2021-03-26 10:30:03	Double good news hai apple eye Being Hemant C.	positive
106	RT @SKFFansOfficial: The International Men's ...	Salman Khan	42942	42411934	MUMBAI	True	0	2094	2019-11-19 12:03:05	RT @SKFFansOfficial: The International Men's Day ...	positive
107	RT @saemmarajpar: Khushi hai khush with her.	Salman Khan	42942	42411934	MUMBAI	True	0	1664	2019-11-18 20:09:10	RT @saemmarajpar: Khushi hai khush with her.	positive
108	Happy children's day today and every day h.	Salman Khan	42942	42411934	MUMBAI	True	41529	3761	2019-11-14 19:13:35	Happy children's day today and every day h!	positive
195	Congrats daddy on a good clean child bendy!	Salman Khan	42942	42411934	MUMBAI	True	15094	1700	2019-11-10 05:30:00	Congrats daddy on a good clean child bendy!	positive
197	Best wishes to not just a superstar. The only.	Salman Khan	42942	42411934	MUMBAI	True	35361	7844	2019-11-07 12:15:50	Best wishes to not just a superstar. The only.	positive

82 rows x 11 columns

Fig 6. Grouping of positive tweets

After collecting all the data we do separation of positive tweets as shown in figure 4.

```
In [135]: sentiment.get_group("negative")
```

```
Out[135]:
```

	Tweets	User	User_status_count	user_followers	User_location	User_verified	fav_count	ret_count	tweet_date	clean_tweets	sent
1	ing Human Clothing is fat 50% today with ...	Salman Khan	42942	42411934	MUMBAI	True	47179	5945	2020-12-27 05:30:00	ing Human Clothing is fat 50% today with ...	negative
2	Bhaya, beheno aur mubrats difficult times ...	Salman Khan	42942	42411934	MUMBAI	True	66227	7445	2020-10-00 07:51:10	Bhaya, beheno aur mubrats difficult times ...	negative
3	Thank u Bhai! Rao of Ky's Chicken for the ...	Salman Khan	42942	42411934	MUMBAI	True	49696	4204	2020-04-25 20:42:10	Thank u Bhai! Rao of Ky's Chicken for the ...	negative
4	ye. @realmemobles sur more saah human is.	Salman Khan	42942	42411934	MUMBAI	True	40912	4440	2020-03-05 09:38:04	ye. @realmemobles sur more saah human is.	negative
5	In Valentine's Day par single mat bolo. bolo ...	Salman Khan	42942	42411934	MUMBAI	True	68694	8235	2020-02-10 07:42:10	In Valentine's Day par single mat bolo. bolo ...	negative
6	Very very sorry n addressed to her of the dem.	Salman Khan	42942	42411934	MUMBAI	True	55015	5150	2020-01-25 12:14:56	Very very sorry n addressed to her of the dem.	negative
7	(D) in their family go through this extreme agony n.	Salman Khan	42942	42411934	MUMBAI	True	31332	7056	2019-11-30 12:24:44	(D) in their family go through this extreme agony n.	negative
8	JustForPriyankaReddy These r the worst kin.	Salman Khan	42942	42411934	MUMBAI	True	57333	11851	2019-11-30 12:24:44	JustForPriyankaReddy These r the worst kin.	negative

Fig 7. Grouping of negative tweets

Similarly, we also do negative tweets separation as shown in figure 5.

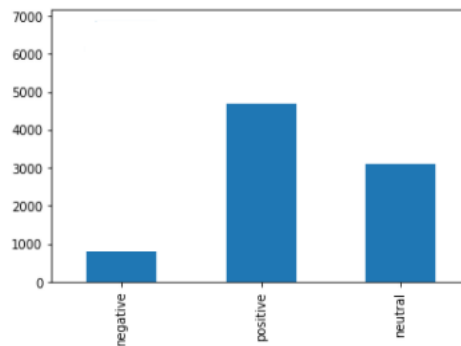


Fig 8. Chart

Depending on the number of negative, positive and neutral polarities we create chart.

```
In [136]: print("The neutral tweets are :",len(sentiment.get_group('neutral')))

print("The positive tweets are :",len(sentiment.get_group('positive')))

print("The negative tweets are :",len(sentiment.get_group('negative')))

The neutral tweets are : 110
The positive tweets are : 82
The negative tweets are : 8

In [137]: if len(sentiment.get_group('negative')) > len(sentiment.get_group('neutral')) and len(sentiment.get_group('negative'))>len(sentiment.get_group('neutral')):
else:
    print("person is mentally good")

person is mentally good
```

Fig 9. Result

VI. CONCLUSION

We collected the aptitude of victimisation Twitter as a tool for mensuration and a platform where share thoughts and investigation major depression among its users. provide a transparent understanding of our work, numbers of analysis challenges were explicit. The algorithms area unit designed to investigate the tweet for feeling detection in addition to the detection of thoughts among folks on social media. The mechanism will analysis of the tweets for the prediction of depression on faith the validity of tweets. Social media is AN open platform wherever telling our true emotions which may relate to their on going life so the model analysis here is usually supported by the prediction from posts victimisation varied machine learning algorithms. the most demand of the model is to be utterly ready to predict the result as there area unit variety of implementations that need verification of information before predicting the thoughts or posts of the person as self-destructive or non-suicidal we've developed a replacement model with a mix NLTK & LSTM for higher Performance result.

VII. REFERENCES

- [1] P.V. Rajaraman, Asim Nath, Akshaya. P.R , Chatur Bhuja.G Depression Detection of Tweets and A Comparative Test, International Journal of Engineering Research & Technology (IJERT) ,Vol. 9 Issue 03, March-2020
- [2] Akshi Kumara, Aditi Sharmab, Anshika Arorac, Anxious Depression Prediction in Real-time Social Data, ICAESMT19, International Conference on Advanced Engineering, Science, Management and Technology - 2019
- [3] Michael M. Tadesse , Hongfei Lin , Bo Xu , And Liang Yang, Detection of Depression-Related Posts in Reddit Social Media Forum, Digital Object Identifier 10.1109/ACCESS.2019.2909180, April 4, 2019
- [4] Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir Ahmed, Javed Ferzund, Basit, Shahzad, Alexander Gelbukh, A Multiclass Depression Detection in Social Media Based on Sentiment Analysis, 17th International Conference on Information Technology–New Generations (ITNG 2020) , 12 May 2020
- [5] Kali Cornn, Identifying Depression on Social Media, Department of Statistics Stanford University Stanford, CA 94305
- [6] Michael Mesfin Tadesse ,ongfei Lin ,Bo Xu, and Liang Yang ,Detection of Suicide Ideation in Social Media Forums Using Deep Learning, State Key Laboratory of Cognitive Intelligence, Algorithms 2020, 13(1), 7,24 December 2019