iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37359

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Detection and Analysis of Depression Level Using Social Media

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

[1, 2, 3, 4, 5, 6]*Department Of information technology, Karmaveer Dadasaheb Kannamwar College of Engineering, Nagpur*

Abstract: Detection of depression through messages sent by a user on social media are often a fancy task because of the recognition and trends in them. In recent years, messages and social media has over up being a really shut illustration of a person's life and his status. This is often an enormous stockpile of information a couple of person's behaviour and might be used for detection of varied mental sicknesses (depression in our case) victimisation tongue process. This project is regarding constructing a model victimisation NLP to predict such mental disorders. Short-term memory networks square measure well-suited to classifying, process and creating predictions supported statistic knowledge, since there are often lags of unknown length between necessary events during a statistic.

Keywords: Depression, social media, mental illness, NLTK

I. INTRODUCTION

Depression as a typical psychological state disorder has long been denned as one illness with a group of diagnostic criteria. It usually co-occurs with anxiety or alternative psychological and physical disorders; and has a bearing on feelings and behaviour of the affected people. In keeping with the World Health Organization study, there are 322 million folks calculable to suffer from depression, cherish four.4% of the worldwide population. In today's world, communication through social media is rising as an enormous deal. They're willing to share their thoughts, stories and their personal feelings, mental states, needs on social network sites, blogging platforms etc... Receivers use the manuscripts from emails and alternative varieties of social media comments to make correct reasoning and to correct the mistakes. Once folks write digitally on social media, their texts are processed mechanically. Linguistic communication process techniques are accustomed infer people's mental behaviour.

According to World Health Organization, depression could be a common worldwide folio that affects a colossal quantity of people no matter their age. There are multiple factors that interfere period of time detection and treatment like lack of skilled specialists, social shaming, improper identification so on. The ever-lasting depression disorder could lead on to suicide if the depressed people don't seem to be equipped correct consultance, instant facilitate and might additionally suffer from anxiety. This work is targeted on the detection of depression and anxiety from tweets. The experiment conducted throughout this work needs the text knowledge that the chosen knowledge supply is Twitter wherever folks tweet regarding their feelings, hopes, desires, thoughts, stories and mental states. The goals of our analysis are: collect the in public on the market media messages of healthy and self-diagnosed people that contains mixed emotions therefore measure the extracted Twitter knowledge and apply NLTK and text blob to predict depressive and anxiety tweets. We are able to hunt for an answer to a performance increase through a correct options choice and their multiple feature mixtures. First, we decide the foremost helpful linguistic options applied for depression identification to characterize the content of the posts. Second, we have a tendency to analyze the correlation significance, hidden topics and word frequency extracted from the text. We have a tendency to compare the performance results supported 3 single feature sets and their multiple feature mixtures. In our experiment, we have a tendency to use knowledge collected from the Reddit social media platform.

II. AIMS AND OBJECTIVE

A. Identify the foremost effective deep neural spec among some of elect architectures that were with success utilized in tongue process tasks.

B. 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.

C. To investigate the result of depression detection, we have a tendency to proposed technique as associate economical and ascendable technique.

D. 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.

E. This study are often extended within the future by considering additional deciliter models that live} extremely unlikely to over-fit the used knowledge and notice a additional dependable thanks to measure the features' impact.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

III. LITERATURE SURVEY

Michael M. Tadesse, Hongfei Lin, Bo Xu, and Liang Yang: we can significantly improve performance accuracy. The best single feature is bigram with the Support Vector Machine (SVM) classifier to detect depression with 80% accuracy and 0.80 F1 scores. The strength hand effectiveness of the combined features (LIWC+LDA+bigram) are most successfully demonstrated with the Multilayer Perceptron (MLP) classifier resulting in the top performance for depression detection reaching 91% accuracy and 0.93 F1 scores.

P.V. Rajaraman, AsimNath, Akshaya.P.R, ChaturBhuja.G: Messages and social media has ended up being a very close representation of a person's life and his mental state. This is a huge stockpile of data about a person's behaviour and can be used for detection of various mental illnesses (depression in our case) using Natural Language Processing.

Akshi Kumara, Aditi Sharmab, Anshika Arorac: - This mixed anxiety-depressive disorder is a predominantly associated with erratic thought process, restlessness and sleeplessness. Based on the linguistic cues and user posting patterns, the feature set is defined using a 5-tuple vector <word, timing, frequency, sentiment, contrast>. An anxiety-related lexicon is built to detect the presence of anxiety indicators. Time and frequency of tweet is analyzed for irregularities and opinion polarity analytics is done to find inconsistencies in posting behaviour. The model is trained using three classifiers (multinomial naïve bayes, gradient boosting, and random forest) and majority voting using an ensemble voting classifier is done.

Michael M. Tadesse , Hongfei Lin , Bo Xu , Liang Yang: We can significantly improve performance accuracy. The best single feature is bigram with the Support Vector Machine (SVM) classifier to detect depression with 80% accuracy and 0.80 F1 scores. The strength and effectiveness of the combined features (LIWC+LDA+bigram) are most successfully demonstrated with the Multilayer Perceptron (MLP) classifier resulting in the top performance for depression detection reaching 91% accuracy and 0.93 F1 scores. According to our study, better performance improvement can be achieved by proper feature selections and their multiple feature combinations.

HoyunSong, Jinseon You, and Jin-Woo Chung Jong C. Park: we propose Feature Attention Network (FAN), inspired by the process of diagnosing depression by an expert who has background knowledge about depression. We evaluate the performance of our model on a large scale general forum (Reddit Self-reported Depression Diagnosis) dataset. Experimental results demonstrate that FAN shows good performance with high interpretability despite a smaller number of posts in training data. We investigate different aspects of posts by depressed users through four feature networks built upon psychological studies, which will help researchers to investigate social media posts to find useful evidence for depressive symptoms.

Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir, Ahmed Javed Ferzund, Basit Shahzad, Alexander Gelbukh : A sample of their recent tweets collected ranges from (200 to 3200) tweets per person. From their tweets, we selected 100 most frequently used words using Term Frequency-Inverse Document Frequency (TF-IDF). Later, we used the 14 psychological attributes in Linguistic Inquiry and Word Count (LIWC) to classify these words into emotions. Moreover, weights were assigned to each word from happy to unhappy after classification by LIWC and trained machine learning classifiers to classify the users into three classes of depression High, Medium, and Low. According to our study, better features selections and their combination will help to improve performance and accuracy of classifiers.

Kali Cornn: a dataset of scraped Reddit comments, this project aims to classify depression in comments. Focusing on the setting of social media, this project explores methods of machine learning and neural network architectures for identifying depression in digitally shared text entries. This project developed machine learning (logistic regression, support vector machines), a BERT-based model, and neural networks with and without word embeddings (CNN) for this classification task.

IV. DESIGN AND IMPLEMENTATION

A. Proposed System

Mental illness detection in social media will be thought of a posh task, primarily because of the difficult nature of mental disorders. In recent years, this analysis space has began to evolve with the continual increase in quality of social media platforms that became associate degree integral a part of people's life. This shut relationship between social media platforms and their users has created these platforms to replicate the users' personal life with completely different limitations. In such associate degree surroundings, researchers are conferred with a wealth of knowledge relating to one's life. Additionally to the amount of quality in distinguishing mental sicknesses through social media platforms, adopting supervised machine learning approaches like deep neural networks haven't been wide accepted because of the difficulties in getting decent amounts of annotated coaching information. Because of these reasons, we tend to attempt to establish the foremost effective deep neural spec among a number of of hand-picked architectures that were with success utilized in tongue process tasks.

We tend to gift a brand new model NLTK with textblob. The chosen architectures ar accustomed observe users with signs of mental sicknesses (depression in our case) given restricted unstructured text information extracted from the Twitter social media platform.

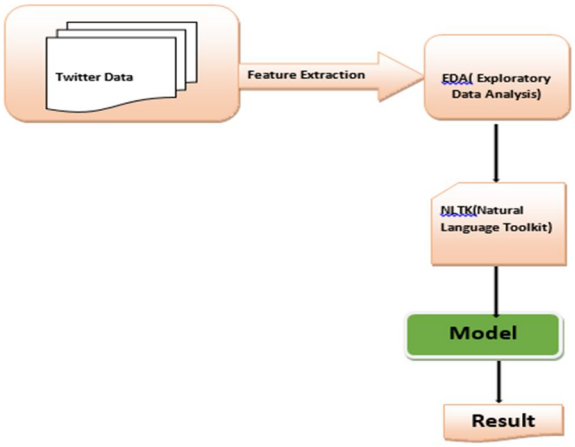


Fig. 1 System flow Diagram

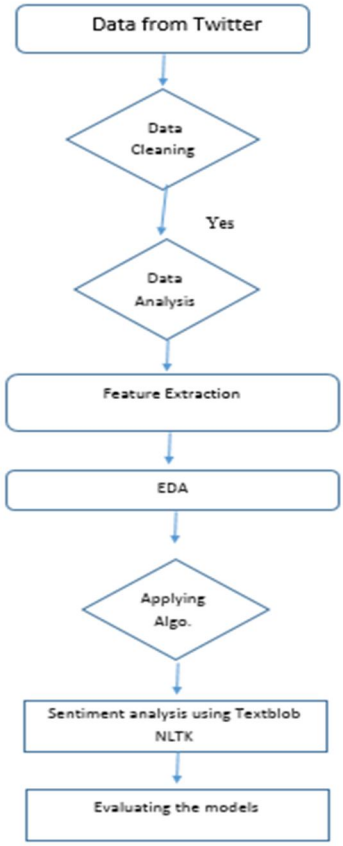


Fig. 2 Data Flow Diagram

Data for the model are extracted from the Twitter web site. Information are in unstructured format and it'll have uncountable useless information. Thus we'll initial do information improvement. Once we have a tendency to clean the info we'll apply EDA for locating the polarity of the info like positive, neutral, or negative mechanically or several complicated sentiments like happiness, sadness, anger, joy, etc. Once EDA is finished can pass the info to the NLTK model that we can choose the model giving the simplest accuracy.

B. Modules

1) *Information Selection*: A coaching set is employed to coach the machine learning method to know the potential relationship between the informative variables and target variable.
2) *Information Cleaning*: information improvement suggests that filtering and modifying your information specified it's easier to explore, understand, and model. Filtering out the elements you do not wish or would like in order that you ought not to consider or method them.
3) *Information imputation*: Machine learning algorithms need numeric input values, and a worth to be gift for every row and column in a very dataset. As such, it's common to spot missing prices in a very dataset and replace them with a numeric value.
4) *Information Analysis*: a way of knowledge analysis that automates analytical model building. It's a branch of computer science supported the concept that systems will learn from information, determine patterns and create selections with least human intervention.
5) *Information image*: information visualization is that the illustration of {information} or information in a very graph, chart, or alternative visual format. Machine learning makes it easier to conduct analyses like prophetic analysis, which might then function useful visualizations to gift.
6) *Coaching:* Training information is that the information you employ to coach associate degree algorithmic rule or machine learning model to predict the result you style your model to predict.
7) *Testing*: A take a look at dataset may be a dataset that's freelance of the coaching dataset, however that follows identical likelihood distribution because the coaching dataset.
8) *Algorithmic* rule selection: Machine learning algorithms ar the engines of machine learning, which means it's the algorithms that flip an information set into a model.

C. Planned Methodology

1) *Linguistic communication Toolkit (NLTK):* The linguistic communication Toolkit, or a lot of usually NLTK, may be 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 and text blob 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 a 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. NLTK and text blob may be a powerful Python package that gives a collection of numerous natural languages algorithms. It is free, open supply, straightforward to use, massive community, and well documented. NLTK and text blob consists of the foremost common algorithms like tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. NLTK helps the pc to analysis, pre-process, and perceive the written communication.

V. RESULT AND DISCUSSION

```
,Tweets,User,User_statuses_count,user_followers,User_location,User_verified,fav_count,rt_count,tweet_date
0,@urstrulyMahesh @pluspluSMovies @sashiTikka @vivekkrishnani @majorthefilm @prakashraaj,Salman
Khan,42942,42411934,MUMBAI,True,14885,3255,2021-04-12 10:51:45
1,Isse kehte hain dhamakedaar teaser! Really happy and proud to launch this. Congratulations to the team.. and salute…
https://t.co/tw3KaPBKri,Salman Khan,42942,42411934,MUMBAI,True,28723,6128,2021-04-12 10:51:45
2,"Happy  birthday brother, wish u all the health, happiness, success.. stay blessed stay safe . @ajaydevgn https://t.co/4x8ty23h8b",Salman
Khan,42942,42411934,MUMBAI,True,79460,5631,2021-04-02 17:49:49
3,"Karo apne aap ko aur family ko protect with the FRSH germ protection spray!

#RahoFrshRahoSafe

Click now-… https://t.co/hx1WCR6ff9",Salman Khan,42942,42411934,MUMBAI,True,31661,3706,2021-03-31 05:30:07
4,Happy bday brother .... Wish you all the success and happiness. @AlwaysRamCharan,Salman Khan,42942,42411934,MUMBAI,True,93549,10008,2021-
03-27 17:46:23
5,"Arre waah! @arbaazSkhan
This looks truly amazing just like the Mangoes at @aamwalla

Place your order today:
Visit : https://t.co/eltkqlBl6b",Salman Khan,42942,42411934,MUMBAI,True,21359,2791,2021-03-27 05:30:23
6,"Double good news hai aapke liye. Being Human Clothing ki Spring Summer'21 Collection is here. Aur, Chandigarh &amp;amp; Sri…
https://t.co/EIAHRm0Jdt",Salman Khan,42942,42411934,MUMBAI,True,38596,4404,2021-03-26 10:30:03
7,"#DareToLeap with more pixels!
Aayie le chale aapko better clarity ki duniya me with the Camera Flagship… https://t.co/EHjGBvWUxs",Salman
Khan,42942,42411934,MUMBAI,True,20109,3423,2021-03-24 15:31:18
8,Took my first dose of vaccine today....,Salman Khan,42942,42411934,MUMBAI,True,113840,8228,2021-03-24 12:36:29
9,"Eid ka commitment tha, Eid par hi aayenge kyun ki ek baar jo maine.....

#RadheOn13thMay #2MonthsToRadhe… https://t.co/VGeaEcV0EH",Salman Khan,42942,42411934,MUMBAI,True,79586,11440,2021-03-13 06:30:03
10,"Congratz farhan teaser looks good ...

#Toofaan @FarOutAkhtar @excelmovies @210Kassim @ritesh_sid https://t.co/rKEkSbqRVW",Salman Khan,42942,42411934,MUMBAI,True,29205,3403,2021-
03-12 11:19:10
11,"PEPSI AB AUR BHI ZYADA REFRESHING
#WhattayGarmi
#WhattayRefreshingPepsi
#HarGhoontHeinSwag
```

Fig 3.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, emoji's, stock words, etc.

Fig 4.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("senti")

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

Out[134]:

	Tweets	User	User_statuses_count	user_followers	User_location	User_verified	fav_count	rt_count	tweet_date	clean_tweets	senti
1	Isse kehte hain dhamakedaar teaser! Really hap...	Salman Khan	42942	42411934	MUMBAI	True	28723	6128	2021-04-12 10:51:45	Isse kehte hain dhamakedaar teaser Really happ...	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 succe...	Salman Khan	42942	42411934	MUMBAI	True	93549	10008	2021-03-27 17:46:23	Happy bday brother Wish you all the success a...	positive
5	Arre waah! @arbaazSkhan\nThis looks truly amaz...	Salman Khan	42942	42411934	MUMBAI	True	21359	2791	2021-03-27 05:30:23	Arre waah arbaazSkhan\nThis looks truly amazin...	positive
6	Double good news hai aapke liye. Being Human C...	Salman Khan	42942	42411934	MUMBAI	True	38596	4404	2021-03-26 10:30:03	Double good news hai aapke liye Being Human Cl...	positive
...	...	...	...	...	...	...	...	...	...	...	...
186	RT @SKFilmsOfficial: This International Men's ...	Salman Khan	42942	42411934	MUMBAI	True	0	2094	2019-11-19 12:03:05	RT SKFilmsOfficial This International Mens Day...	positive
187	RT @saieemmanjrekar: Khushi badi khush with he...	Salman Khan	42942	42411934	MUMBAI	True	0	1664	2019-11-18 20:09:18	RT saieemmanjrekar Khushi badi khush with her ...	positive
188	Happy children's day today and every day ... h...	Salman Khan	42942	42411934	MUMBAI	True	41529	3761	2019-11-14 15:13:35	Happy children's day today and every day htt...	positive
196	Congratz daisy on a good clean child friendly ...	Salman Khan	42942	42411934	MUMBAI	True	15094	1700	2019-11-10 05:30:08	Congratz daisy on a good clean child friendly ...	positive
197	Best wishes to not just a superstar. The only...	Salman Khan	42942	42411934	MUMBAI	True	35361	7044	2019-11-07 12:15:50	Best wishes to not just a superstar The only s...	positive

82 rows × 11 columns

Fig 5.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_statuses_count	user_followers	User_location	User_verified	fav_count	rt_count	tweet_date	clean_tweets	senti
ing Human Clothing is Flat 50% today.\nVh...	Salman Khan	42942	42411934	MUMBAI	True	47179	5945	2020-12-27 05:30:00	Being Human Clothing is at Flat today\nVihat a...	negative
Bhaiyo, beheno aur on\nIn difficult times ...	Salman Khan	42942	42411934	MUMBAI	True	66227	7445	2020-10-08 07:51:10	Bhaiyo beheno aur mtron\nIn difficult times m...	negative
Thank u Balaji Rao of ky's Chicken for the ...	Salman Khan	42942	42411934	MUMBAI	True	49696	4204	2020-04-25 20:42:10	Thank u Balaji Rao of Venky's Chicken for the ...	negative
ye...@realmemobiles mere saath humaar...	Salman Khan	42942	42411934	MUMBAI	True	40012	4448	2020-03-05 09:39:04	Aayiye realmemobiles aur mere saath humaari is...	negative
Is Valentine's Day par, ingle mat bolo, bolo ...	Salman Khan	42942	42411934	MUMBAI	True	68694	8235	2020-02-10 07:42:10	Is Valentines Day par single mat bolo bolo swa...	negative
Very very sorry n addened to hear of the demi...	Salman Khan	42942	42411934	MUMBAI	True	55015	5158	2020-01-25 12:34:56	Very very sorry n saddened to hear of the demi...	negative
(2/2). n their family go rough this extreme ...	Salman Khan	42942	42411934	MUMBAI	True	31332	7056	2019-11-30 12:24:44	n their family go through this extreme agony n...	negative
liceForPriyankaReddy These r the worst kin...	Salman Khan	42942	42411934	MUMBAI	True	57333	11851	2019-11-30 12:24:44	JusticeForPriyankaReddy These r the worst kind...	negative

Fig 6. Grouping of negative tweets

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

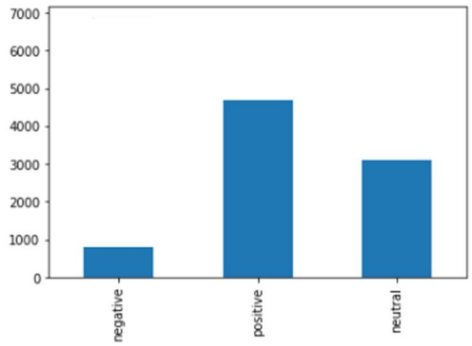


Fig 5.chart

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

Fig 6.Result

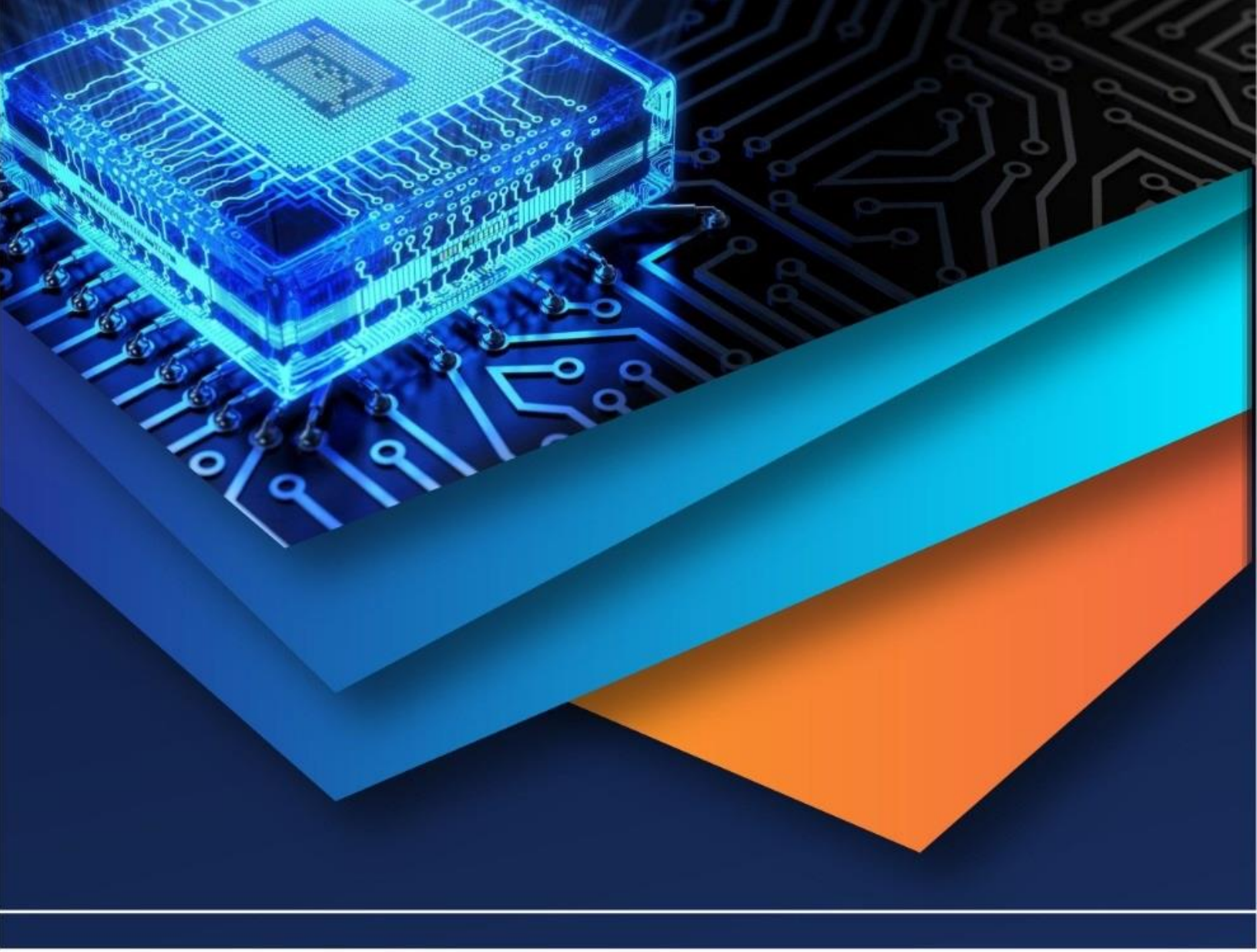
Our module detect the level of mental health depending on the word polarities that is positive, negative and neutral which has higher amount.

VI. CONCLUSION

We have exhibited the capability of using twitter as a tool for measuring and detecting major depression among its users. To give a clear understanding of our work, numbers of research challenges were stated. The algorithms are designed to analyze the tweet for emotion detection as well as for detection of suicidal thoughts among people on social media. The mechanism does analysis of the tweets for prediction of depression without checking the validity of tweets. Social media is an open platform where many people refrain from telling their true emotions that might relate to depression they are facing, and so the model analysis here are mostly based on the prediction from posts using various machine learning algorithms. The main requirement of model is to be perfectly able to predict the result as there are a number of implementations that require verification of data before predicting the thoughts or posts of the person as suicidal or non-suicidal we are developed new model with combination NLTK & text blob for better Performance result.

REFERENCES

[1] P.V. Rajaraman , AsimNath ,Akshaya.P.R , ChaturBhuja.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, AditiSharmab, AnshikaArorac, 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 MesfinTadesse ,ongfei Lin ,Bo Xu, and Liang Yang ,Detection of Suicide Ideation in Social Media Forums Using ,State Key Laboratory of Cognitive Intelligence, Algorithms 2020, 13(1), 7,24 December 2019



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY