Depression Detection Using Sentiment Analysis of Tweets

Chaitanya Bhargava[a], S. Poornima[b], Shivam Mahur[c], M. Pushpalatha[d]

[a]Dept. of CSE, SRMIST, Kattangulathur, Tamil Nadu, India. E-mail: cd4640@srmist.edu.in
[b]Assistant Professor, Dept. of CSE, SRMIST, Kattangulathur, Tamil Nadu, India.
E-mail: poornims@srmist.edu.in
[c]Dept. of CSE, SRMIST, Kattangulathur, Tamil Nadu, India. E-mail: sm7759@srmist.edu.in
[d]Professor, Dept. of CSE, SRMIST, Kattangulathur, Tamil Nadu, India. E-mail: pushpalm@srmist.edu.in

Abstract: Depression is among the most common mental issue and a major reason for suicides. The rise in the use of social media platforms has given an opportunity for early diagnosis of depression using the everyday language of the users. The motive of this study is to classify tweets having depressive traits among the tweets scraped from Twitter. The present systems fail to capture all the language features used to train the classifiers. This model attempts to maximize the utilization of all the available linguistic features present in the tweets and makes use of proper cleaning and pre-processing techniques for more accurate diagnosis of depression. Rule based Vader Analyzer and a hybrid model of CNN-LSTM is used. The results are calculated and shown using the performance metrics such as recall, precision, F1-score and accuracy.

Keywords: Depression, Machine Learning, Mental Health, Sentiment Analysis, Social media, Twitter.

1. Introduction

Depression, commonly known as "major depressive disorder" is among the most widely recognized psychological wellness problems all around the world and more often than not leads to disability. According to WHO, 322 million people of all ages worldwide are affected by depression alone. WHO has also stated that 7.5 % of the total Indian population is suffering from some kind of mental disorder World Health Organization has predicted that soon approximately 20 per cent of India may be suffering from different kinds of mental illness (swachhindia.ndtv.com, n.d.). Depression currently records for about 4.3% of the global diseases, and it is the main cause of disease burden in wealthy nations by 2030 (WHO, 2011). The symptoms of major depression include: Fatigue, Irritability, Difficulty concentrating, Feeling hopeless and self-hate, Social isolation, sudden weight loss/gain, Talks of threat or suicide.

The best form of treatment of mental diseases like depression are early detection and professional intervention. Delayed diagnosis and subsequently delayed treatment can often lead to severe case of the condition. This leads to suicidal thoughts and they reach the conclusion that ending it all is a much better alternative, as opposed to seeking treatment, etc.

Recent progress in Machine Learning, NLP & Big data Analytics can provide a solution to this problem. Another factor is the growth of multiple social media platforms which has proved a boon for mental health detection studies. Studies have demonstrated that people with stigmatized illnesses are bound to use the web to obtain information, interface with others and share experiences over various social media platforms. Social media sites such as Twitter, Facebook, Reddit forums contain a substantial amount of data having insight into the feelings and sentiments of users. This throws light on the fact how important it is to help those individuals who have depressive symptoms and how it can be done easily through the social media content of these individuals. The outcomes cannot only be utilized as a primer type of conclusion for proficient intercession but also can likewise fill in as an extra information for the professionals.

Extensive research has been performed in this area and it has shown that social media is one of the most used platforms by an individual and hence can provide a lot data and information regarding possibility of depression or presence of depressive traits in a person. Our paper uses data mining techniques in psychology, especially the field depression, finding those who are depressed in social networking services (SNS). This paper constructs a model which takes tweets as an input for training and predicts depression in user by classifying tweets as depressive or not as output. Also, this paper uses sentiment analysis technique which itself is a hot topic for research which helps in identifying the nature of text i.e. Depressed or non-depressed. We can find that there is huge amount of data for analysis now a days on social media platforms like Facebook posts, images but it rises to privacy concerns so we are using twitter data which we scrapped with the help of Twint tool [15]. Many numbers of machine learning algorithms are available for segmenting tweets, some are Naïve Bayes, KNN, SVM, Random forests, LSTM algorithms etc. But studies show the best results of the LSTM, Convolution Neural Networks, the study is limited

in these algorithms. This paper deals with mainly the tweets i.e., short messages tied to a limit of 280 characters. In this format only, users generally display their own feelings & emotions about what is going on in their lives and the world around them.

2.   Related Work

Mandark Deshpande and Vignesh Rao[5] used NLP to analyze emotion that focuses on depression on twitter. Further, tweets are categorized as positive or negative, based on a list of words to detect how much user is tend to show depressive tendency. They generated dataset using 10,000 tweets collected through the twitter API and used Naïve Bayes classifier for predicting depression with 83% accuracy. Jini Jojo Stephen & Prabu (2019)[14] attempted detecting not only depression but also the measure of depression in the text with the help of sentiment analysis. Dataset was collected from twitter which was of 52 users containing around two thousand tweets each with the help of python library named Tweepy. For base emotion calculation, they used Syuzhet package, which contains its own dictionary and accordingly categorizes the words into different emotions. Emotions include disgust, anticipation, anger, surprise, joy, trust and fear. Also basic sentiment was analysed whether it is positive or negative. Hemanthkumar M, Latha A[9] used Naïve Bayes, SVM techniques for depression detection in user tweets. Dataset used was of 43000 tweets which was collected from Kaggle website. The dataset was partitioned into training and testing sets in a fraction of 70 by 30. For feature generation, they used Bag of Words method to count the frequency of every word in dataset. From the feature generated from the Bag of Words model, then model is trained using Naïve Bayes and SVM. Salma Almouzini, Maher Khemakhem & Asem Alageel[1] used Random Forest, Naïve Bayes, Liblinear for detection of depression. For scrapping tweets, they used a scrapper tool named Twitter4j. For depressed tweets, they performed a search for users who declared themselves in their tweets that they have suffered or are suffering from depression, so for this they got 35 twitter users who were depressed and 62 users who were non-depressed. Chempaka Seri Abdul Razak and Muhammad Ameer Zulkarnain[12] used Vader Sentiment Analysis and a couple of machine learning approaches, Naïve Bayes and CNN for detection of depressed users on twitter platform. They presented a system development by the name of Tweep, which helps a user to assess depression level using Machine Learning based on the user's tweets. Nafiz Al Asad and Md. Appel Mahmud Pranto[3] proposed a data analytic model for depression detection in any person. For this they collected data from user social media post like Facebook and Twitter. They used Natural Language processing and Naïve Bayes algorithm to detect depression in an easy and efficient manner compared to others. Julius Jacobson[8] employed various machine learning classifiers to predict whether tweet was posted by a depressed user or non-depressed user. Up to 3200 tweets were scraped for each verified user. He basically used four different classifiers and they were Naïve Bayes, Random Forest, SVM and Logistic Regression. This study was only applicable to Korean data. Xiaohui Tao and Xujuan Zhou[16] used sentiment analysis approach for early detection of depressed users on social media. They basically used a knowledge base storing depressive sentiment vocabulary for calculating sentiment scores of tweets of users and if found negative for a long time, the system will alert the user for diagnosis. Zhicho Peng, Qinghua Hu and Jianwu Dang[11] used a multi kernel support vector machine approach for early detection of depressed users. Three different types of features, user behaviours, user social text and user profile, were obtained from their respective social media to show from what condition and situation user is going through. They also build a special emotion dictionary for extracting features for calculating word counts statistics. Data was taken from Chinese Social Media Site for training and finally achieved an accuracy of 83.46%. Abu Raihan M. Kama and Naznin Sultana[7] used Facebook data for detection of depressed on social media. So, for this they basically used K-nearest neighbour for detecting depressed emotions. They also used sentiment analysis for detecting different level of depression in Facebook data. Dataset used was Facebook user's comments on which sentiment analysis technique was applied. Finally, they achieved an accuracy of 81.82%.

On analyzing the existing research and techniques, we found out that some of them failed to train their model on an unbiased dataset due to forming their dataset based on only one or two linguistic markers (for example, only "depression").

3.   Proposed Work

The above section demonstrated a scrutiny of the prevailing methodologies. In this section, the proposed model is explained. Data collection is the first step after which pre-processing, feature selection is performed as prerequisites for model training by a hybrid CNN-LSTM model.

A.   Dataset

The dataset consisted of two different kind of tweets, some random tweets with no indication of depressive traits and tweets which contain features indicating depression. A dataset of random tweets can be taken from the

Sentiment140 dataset available on Kaggle. Sentiment 140 contains about 1.6 million tweets which consist of 6 different attributes. For the tweets with depressive traits, a twitter scraping tool named TWINT was utilized due to the absence of such kind of publicly available datasets. Keywords such as depression, hopeless, lonely, antidepressants, etc. were utilized for the same. Manual cleaning was done to ensure the quality of the dataset by only keeping tweets indicating emotional depression.
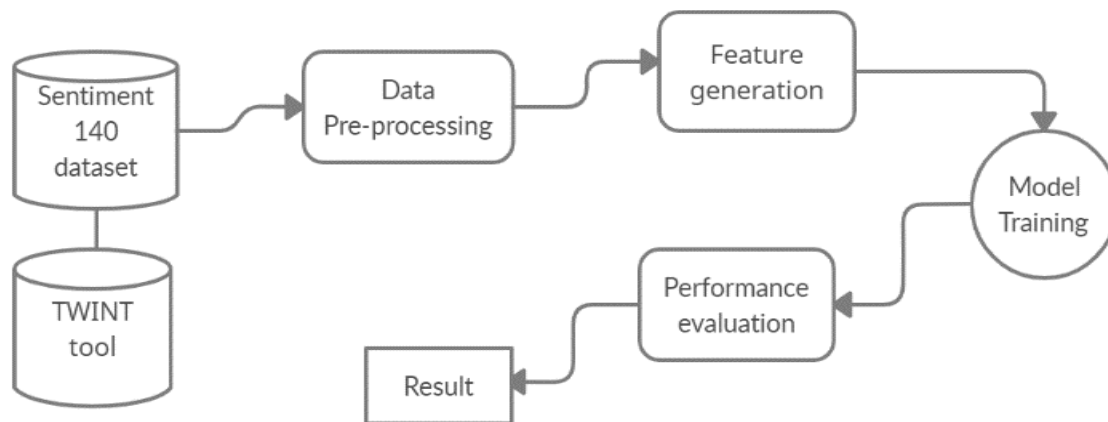

Fig. 1. Architecture Diagram

B. Preprocessing

Fig 1 shows the architecture diagram of our proposed model. The substructure of our design begins with pre-processing of the dataset. The purpose behind pre-processing is to remove words and punctuations that only decrease the efficiency of the model and has no beneficence in sentiment analysis. This step is one of the most important step as it provides foundation to the project and removes all un-necessary words that are not needed. Hence, the pre-processing is a primitive step.

Most of tweets consist of three different sections like first section consist of the person for whom tweet is meant or who is being referenced. Second part consist of the actual message the user wants to tweet or send. Third section consist of hashtag denoted by "#". This basically tells for what the tweet is related to. Hence the tweets undergo pre-processing to filter and remove the first and last section as they have no significance in analysis. And, afterward the message is further pre-processed to obtain keywords which are significant in detecting depression.

- **Text case change:** First, all of the words in each tweet are changed to lowercase.
- **Slang substitution:** Slang play a vital role in depression detection and can also increase the model efficiency by substituting the full forms of abbreviations like xD, LOL etc.
- **Hyperlink Removal:** Hyperlinks are the noisy data in tweets so removal of these make the model more efficient.
- **Digits Removal:** Also, digits don't hold much significance, hence need to be filtered out.
- **Spelling Checking:** Mostly tweets contain a lot of spelling errors which can be corrected in two ways - Shortening and Correction.
- **Lemmatization:** It is a phenomena of mapping word into their lemma like, their root word according to the meaning. Basically, It is better than stemming process as stemming directly eliminates the suffix (last part of the word) of the original word leading to change in meaning also. Ex. The word better can be substituted by good with the help of lemmatization.
- **Removal of Stop words:** Stop words are those words that are the mostly used words having no significance that can be ignored since their presence can affect the model efficiency in adverse manner. Ex. "are", "I", "is" etc.

Extra blank spaces, extra punctuations are also removed as they do not contribute towards feature generation.

For data visualization purpose, word cloud analysis of the pre-processed tweets in both datasets was done. Fig. 2 shows the presence of words found in the dataset formed from random tweets which do not contain any depression indicating traits. Fig. 3 clearly shows the presence of certain words which depict negative emotion and presence of such words in the language of a person is an indicator of depressive tendency.

Fig. 2. Random Tweets

Fig. 3. Depressive Tweets

C.   Feature Generation

1.   **Tokenization -** It is a process of separating words in a text in a way that a machine can comprehend. It essentially breaks down text into smaller pieces called tokens which can either be words, characters or sub-words. It is the initial step in modelling text information. Tokenization can be used to generate features for model training by considering texts as tokens which allows having counts of tokens.

2.   **Sentiment Extractions:** Sentiment extraction is basically extracting the polarity to get the general emotion of each tweet. Valence Aware Dictionary and sEntiment Reasoner, known as VADER (Hutto & Gilbert, 2014) was used for a comprehensive sentiment analysis of Tweets. It is a rule-based and lexicon analysing tool that is particularly adjusted for emotions expressed in social media for general sentiment analysis that is precisely attuned to sentiment in microblog-like contexts[6]. It takes into consideration not only the sentiment classification but also the related sentiment intensity scores. VADER doesn't need preparing information yet is developed from a human-curated, valence-based, generalizable conclusion vocabulary which is adequately quick to be utilized with streaming information. While VADER cannot detect depression in text, it does provide a base for perceiving the overall emotion of the data.

3.   **Word2Vec -** The word2vec tool gives output in form of word vectors on accepting input in the form of text corpus. The framework of Word2Vec is pretty straightforward. This tool can be used to provide a systematic execution of the continuous bag-of-words and skip-gram frameworks utilized in determining words as vector representations. It is a feed-forward neural network containing only one hidden layer. Pretrained word embeddings are perhaps one of the strongest way of portraying text because they are likely to capture the semantic and syntactic definition of a word. It first develops a lexicon from the training text information and then grasps a vector portrayal of words. Many machine learning and NLP applications can make use of the accomplished word vector file as a feature. In our model, pre-trained vectors have

been used which were prepared as explained in (Google's word2vec, n.d.). Fig. 4 shows the embedding layer as shown in (Google Developers, n.d.)The phrases were obtained using a straightforward data-driven methodology portrayed in (Mikolov, Sutskever, Chen, Corrado, & Dean). The archive can be installed from here:
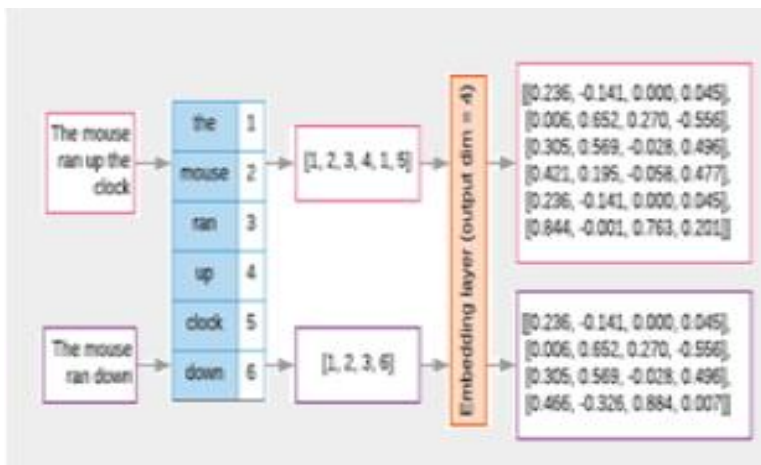

Fig. 4. Word2Vec

D.   Model Training

Long-Short Term Memory Neural Networks: LSTM networks are a kind of RNN architecture that is intended to "recall" recently read values for some random timeframe. LSTMs normally have three gates that control the flow to and from their memories. The "input gate" controls the contribution of new data to the memory. The "forget gate" has control over how long particular values are held in memory. At last, the output gate manages the amount of the value contained in memory which affects the output activation of the block. These functionalities are shown in Fig. 5 as shown in (Sosa, 2017). Intuitively, the benefit of utilizing LSTMs while doing any kind of text analysis is that the network will recollect what it has read before, and thus it has a better understanding of the data. In this particular situation, it very well may have the option to deal with sentences with changing opinion, for example, "I hated reading books, until I read Asimov" (Sosa, 2017).
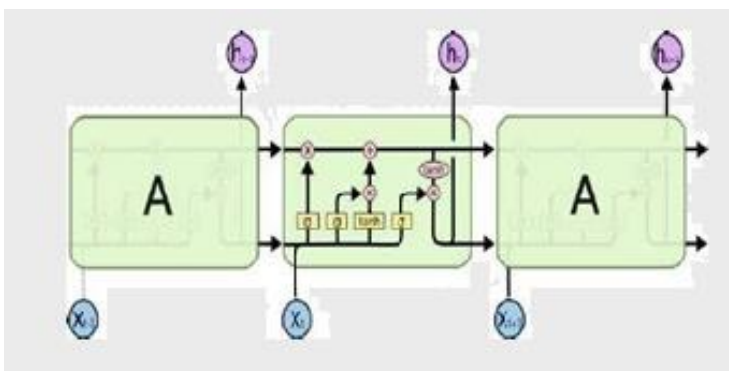
Fig. 5. LSTM Model

Fig. 6 as suggested in (Sosa, 2017) shows the model implemented here where input is fed and afterwards it produces a number as output addressing the likelihood that the tweet contains depressive traits. 60% of the dataset is used as the training set while remaining part is divided equally between validation set and training set. Each tweet is input into the model, substituted with its embedding's, then a convolution layer is added to the model which makes use of that embedding vector. CNNs are great at taking in spatial input from data, which is then used by the convolution layer to learn some structure from the successive data then pass into a conventional LSTM layer. To wrap things up, the yield from the LSTM layer is delivered to a standard dense model. This can now be utilized for prediction. The model contains a convolution layer after an embedding layer and then a Dense layer. It makes use of a max pooling layer and the dropout is set to 50%. Loss function used is binary cross-entropy and a nadam optimizer is utilized along with a Rectified Linear Unit activation function in the primary layer, and in the dense layer a sigmoid activation function is used.
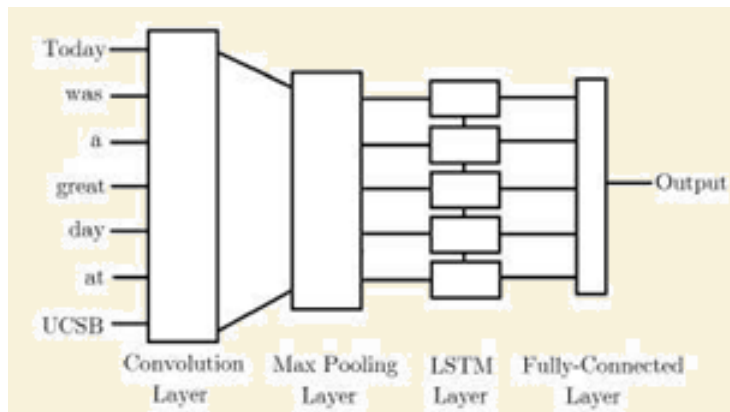
Fig. 6. CNN-LSTM Model

4. Results

Accuracy refers to the correctly predicted data points out of all the available data points. It encourages us to quantify how frequently the algorithm groups the data point accurately.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision is the portion of true positives (significant instances) out of all recovered instances.

$$Precision = \frac{TP}{TP + FP}$$

Recall or sensitivity is the portion of true positive among all applicable occurrences i.e. true positive and false negative.

$$Recall = \frac{TP}{TP + FN}$$

Here, TP means True Positive,
FP means False Positive,
FN means False Negative and
TN means True Negative.

F1 score refers to the harmonic mean of the recall and the precision which suggests that utilizing F1 score in a way assists us in consolidating both recall and precision. The higher is the F1 score, the closer are the precision and recall.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The LSTM model and the hybrid CNN-LSTM model are run and the results obtained are presented in Table I.

Table I. Performance of proposed classifier

PERFORMANCE PARAMETERS				
Model	Recall	Precision	F1-score	Accuracy
LSTM	0.91	0.91	0.91	90.33%
CNN-LSTM	0.91	0.92	0.91	91.35%

Fig. 7 shows a graph representing the accuracy of the model on training data and validation data throughout the epochs.
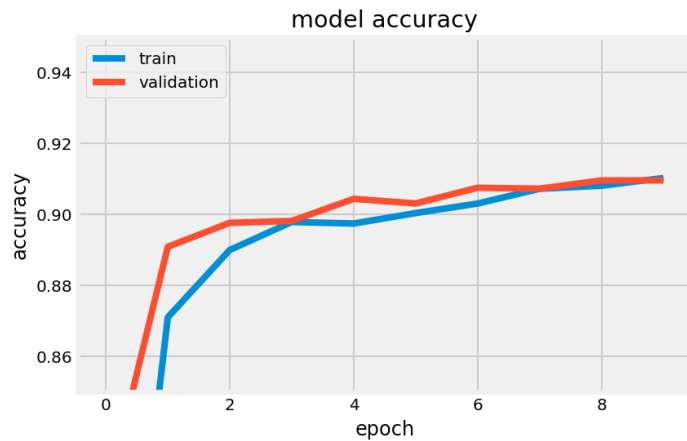
Fig. 7. Model Accuracy

Fig. 8 shows a graph representing the model loss on the training data and the validation data throughout the epochs.
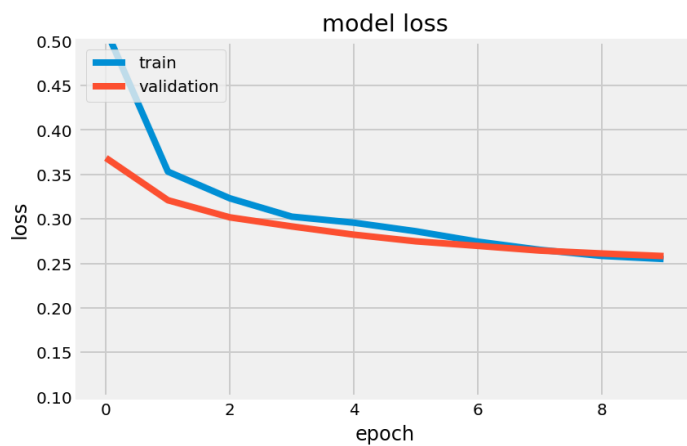

Fig. 8. Model Loss

On comparison with existing models, it can be said our model performs slightly better in all performance metrics. Table II presents a comparison with the previous studies.

Table II. Comparison with existing work

Classifiers	(Deshpande & Rao, 2017)	(Islam, et al., 2018)	(Peng, Hu, & Dang, 2017)
Accuracy	83%	81.82%	83.46%

5. Conclusion

A model for detecting depressed tweets with the use of sentiment analysis has been proposed in this paper and after observing the results and performance metrics, we can say that both LSTM and CNN-LSTM has performed really well in depression detection as can be seen in Table I. CNN-LSTM model came to be the better one among the two in accuracy as well as precision. Many pre-processing steps are performed including dataset preparation, filtering, feature extraction etc. This study can also help the professionals getting the initial data for analysing the depressed persons. This study can be expanded in the future by considering how social network of a user and his interactions can affect his depressive state and find a more reliable way to measure the impact of factors.

Acknowledgement

References

1. Almouzinia, S., khemakhema, M., & Alageel, A. (2019). Detecting Arabic Depressed Users from Twitter Data. 16th International Learning & Technology Conference 2019 (p. 9). *Procedia Computer Science* 163.
2. Arora, P., & Arora, P. (2019). Mining Twitter Data for Depression Detection. 2019 *International Conference on Signal Processing and Communication (ICSC).* Noida.
3. Asad, N., Mahmud Pranto, M., Afreen, S., & M.M., I. (2019). Depression Detection by Analyzing Social Media Posts of User. 2019 *IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON),* (pp. 13-17). Dhaka, Bangladesh.
4. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Choudhury, M.D., Gamon, M., Counts, S., & Horvitz, E. (2013). *Association for the Advancement of Artificial Intelligence,* 10.
5. Deshpande, M., & Rao, V. (2017). Depression detection using emotion artificial intelligence. *Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017)* (p. 5). IEEE.
6. Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media 216* (p. 10). AAAI.
7. Islam, M.R., Kamal, A., Sultana, N., Islam, R., Moni, M., & ulhaq, A. (2018). Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique. 2018 *International Conference on Computer (p. 4). Communication, Chemical, Material and Electronic Engineering (IC4ME2).*
8. Jacobson, J. (2018, 08). Detecting Language from Depressed Users with Korean Twitter Data. Seoul National University. Retrieved from s-space.snu.ac.kr: https://s-space.snu.ac.kr/handle/10371/143754
9. M, Hemnathkumar., & A, Latha. (2019). Depression Detection with Sentiment Analysis of Tweets. *International Research Journal of Engineering and Technology (IRJET),* 5.
10. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (n.d.). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems* 26 (NIPS 2013), 9.
11. Peng, Z., Hu, Q., & Dang, J. (2017). Multi-kernel SVM based depression recognition using social media data. *International Journal of Machine Learning and Cybernetics* volume, 15.
12. Seri, C., Zulkarnain, M., Ab Hamid, S. h., Anuar, N., Jali, M., & Meon, H. (2020). *Tweep: A System Development to Detect Depression in Twitter Posts.* doi:10.1007/978-981-15-0058-9_52
13. Sosa, P.M. (2017). Twitter Sentiment Analysis using combined LSTM-CNN Models. Eprint Arxiv, 1-9.
14. Stephen, J.J., & P,P. (2019). Detecting the magnitude of depression in Twitter users using sentiment analysis. *International Journal of Electrical and Computer Engineering,* 9.
15. TWINT tool. (n.d.). Retrieved from https://github.com/twintproject/twint
16. Xiaohui, T., Xujuan, Zhang, Ji, & Yong. (2016). Sentiment analysis for depression detection on social networks. 12th *International Conference on Advanced Data Mining and Applications* (ADMA 2016). Australia.
17. WHO (2011) https://apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf
18. Google News Vectors. (n.d.). GitHub - mmihaltz/word2vec-GoogleNews-vectors: word2vec Google News model.