Opinion Analysis and Machine Learning Modeling for Depression Detection

Adebisi A. Baale[1], Olawumi R. Olasunkanmi[2], Felicia E. Adelodun[3] and Adeniyi A. Adigun[4]
[1,2]Faculty of Computing and Informatics, Ladoke Akintola University of Technology, Ogbomoso.
[3]Computer Science Department, The Polytechnic, Ibadan.
[4]Centre for Excellence in Agricultural Dev. and Sustainable Environment, Federal University of Agriculture, Abeokuta.
Email: roolasunkanmi20@lautech.edu.ng

***Abstract*—** Many people express opinions on social media sites when they suffer from mental disorders like depression, anxiety, and tension due to pressures, external environment, and other reasons. Such posts shared via Twitter, Facebook, and Instagram are used to identify a person's state of mind. The situation ideation, which is rarely noticed on time until after a tragic consequence, are often earlier expressed overtly or covertly in social media posts. As a result, this research aimed at implementing four (4) classifiers- Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Decision Tree (DT) on two text-feature extraction techniques- Term Frequency- Inverse Document Frequency (TF-IDF) and Bag of Words (BOW). We split the Sentiment140 downloaded dataset from Kaggle into 75%, 25% training, and testing data to predict mental health depression in the tweet's dataset. TF-IDF models produced the highest accuracy with DT (99%) and RF (99%), while the BOW extends the same performance with LR (99%). However, to mitigate the challenges of erroneous classification of depressive individuals as neutral, Receiver Operating Characteristic / Area Under Curve (ROC_AUC) scores of classifiers used was obtained. At the same time, the RF and DT produced 99%, the highest ROC_AUC score. Overall performance of models revealed that tree-based models performed better on the test data used in this research to classify and predict mental health depression in the tweet's dataset.

***Keywords*—** Bag of Words, Depression, Social Media, Machine Learning.

I. INTRODUCTION

Past research confirmed that it is possible to study persons' psychological state by analysing their everyday language. The recognisable psychological features are depression, anxiety, tension, all of which occur due to pressures, external environment, and other reasons [1]. These psychological factors may influence people's life severely and sometimes lead to suicide. Unfortunately, people secretly live with these emotional well-being issues without consulting a therapist for psychotherapy. These resultantly narrow the depression detections to societal assumptions, intuitions, or outcomes of self-reported questionnaires. These, in recent times, has led to suicides. Nonetheless, a quantitative survey with questionnaires has been the most adopted instrument for understanding people's susceptibility to depression. In contrast, research showed that not all individuals suffering from depression would willingly consent to participate in such a survey. Even when they participate, 80% are sceptical about sharing these burdens in a research survey [1, 2]. Instead, they take to their social media page to covertly share their burden. Thus present an avenue for the domain experts, followers, and friends to detect the situation ideation on time in online posts before any tragic consequence.

It is now an undeniable fact that social media is used in diverse applications ranging from contents creation, information sharing, social networking, opinion expression, sentiments mining, and many others. Merriam Webster Dictionary [3] defined web-based media as a type of electronic correspondence for long-range interpersonal communication, enabling clients to make online networks, share data, thoughts, individual messages, recordings, and other substances. Past studies showed that people leverage electronic communication sites such as Twitter, Facebook, Instagram, and many more, to post opinions useful in identifying a person's state of mind [2].

Meanwhile, early recognition has proven to offer the best treatment for many Americans influenced by psychological sickness identified in social media posts. However, the earlier, precise, and non-intuitive identification of psychological illness dependent on organic bio-markers has demonstrated slippery [4]. Whereas recent advances in machine learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) may proffer solution tools to shortcomings [5]. These tools enhance the extraction of pointers of psychological sickness from hidden parts of individuals' everyday discourse as reflected in their specific word decisions, syntactic developments, and expression manners. These semantic markers may consider the aggregation of an advanced phenotype and computationally determined attributes of a person investigating psychological instability [6, 7].

All rights are reserved by **UIJRT.COM.**

II. LITERATURE BACKGROUND

Social media are intelligent PCs with advances that encourage creating and sharing data, thoughts, vocation interests, and different articulation types through virtual networks and organisations [8]. These platforms exhibit some connections with Depression, Narcissism, Anxiety, and other related challenges. For instance, Shaw and Gant [9] gave evidence for an inverse association between Internet use and depression. They suggested that more social forms of Internet use like chatting and gaming reduce the risk of depression [10]. Researchers thereby leverage the available data in social media space for mental well-being analysis. In a bid to build an Online Health Monitoring System, Wang et al. [11] calculated the depression inclination of various microblogs and identified ten psychological attributes associated with depression.

Moreover, [12] developed a tweet crawler that mines both consumers' tweets and consumers' followers' opinions and subjects their thoughts to analyse via a user-friendly system called "Tweep." The system relied on inferential models with "Vader," a rule-based sentiment analysis framework, Naïve Bayes, and Convolutional Neural Networks. Hemanth and Latha [13] identified tweets polarity using NLP on word lists curated from Twitter feeds. They conducted emotion analysis and finally implemented Naive-Bayes and support vector machine algorithms to classify the contents aggregated from the Twitter feed. Though the research employ evaluation criteria like accuracy, precision, recall, and confusion matrix, Naïve Bayes produced the highest accuracy and precision score of 73% and 75%, respectively.

Researchers have attempted to predict the psychological wellness of web users. A number of them utilised Machine Learning Algorithms without a comprehensive comparison of classifiers to best predict the mental illness in Twitter data at high accuracy.

Therefore, this research focuses on mining users' contents and interactions on social networking sites (Twitter) to analyse their association with depression symptoms. Four classifiers- Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Decision Tree (DT) on two text-feature extraction techniques-Term Frequency- Inverse Document Frequency (TF-IDF) and Bag of Words (BOW).

III. MATERIALS AND METHODS

A. The Dataset

The Sentiment 140 dataset downloaded from Kaggle has 1,600,000 tweets, annotated as 0 for and 1 for positive. The dataset has three fields; the label, representing the tweet's polarity; ids, the serial id of the tweets, and the text, which are the social media users' tweets as Depicted in Fig. I

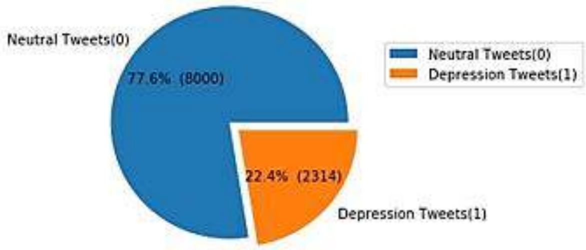


Fig.I Polarity of Tweets in the Dataset

There are 8000 (77.6%) instances of non-depressive tweet class against the 2314 (22.4%) entries in the depression tweet; therefore, not balanced. As a result, a class imbalance problem may occur.

In such cases, standard classifiers tend to be overwhelmed by the large classes and ignore the small ones. Consequently, this research will explore the performance of benchmark models for imbalanced learning.

B. Pre-processing and Analysis of the Dataset

85% of a particular tweet is punctuation marks, which implies that the machine learning model will learn unrelated words if trained with unprocessed data.

These will alter the classifier's effectiveness. The pre-processing steps include Tokenisation, StopWords removal, which we implemented using standard python library for regular expressions.

C. Pre-processing and Analysis of the Dataset

The tweets' descriptive analysis with the overtly used depressive and non-depressive words in the tweet dataset is depicted in Fig II.

II (a) Neutral II (b) Depressive Words

Fig II Word of Neutral and Depressive Tweets

As depicted in Fig. III, out of the over 70 grammatically identified words in the depressive tweets, the most frequent comments in which the word "Depression" appears about 160 times.

All rights are reserved by UIJRT.COM.

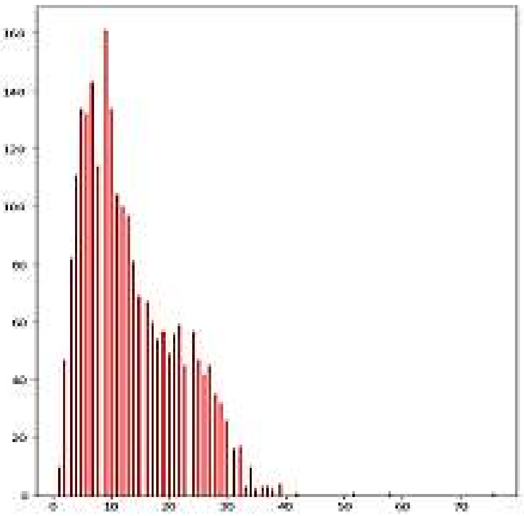
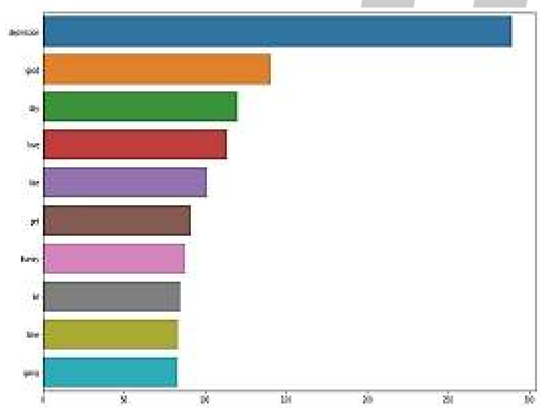
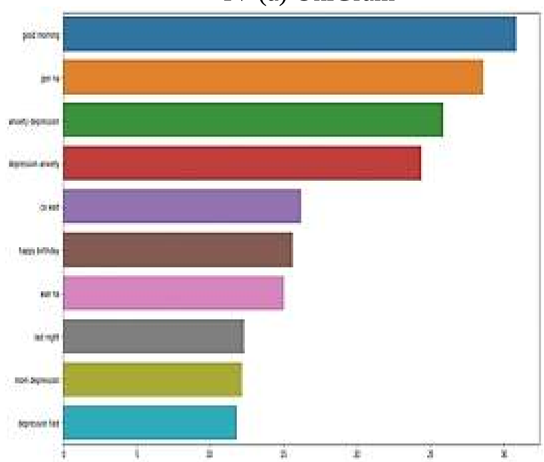


Fig. III Most Frequent Comments with Keyword "Depressive"

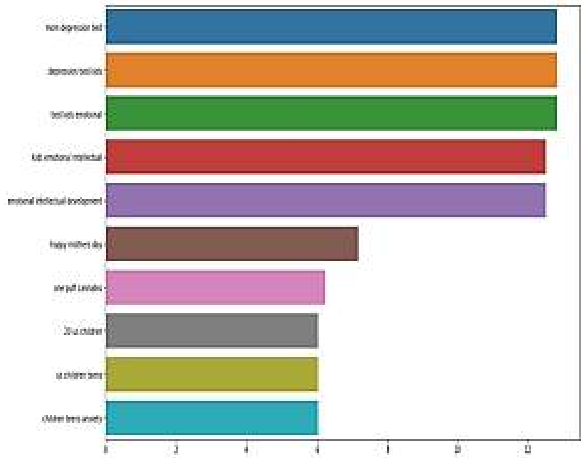
Furthermore, we compute the Gram analysis of the top ten terms in UniGram, BiGram, and Trigram shown in Fig. IV a, b, and c. These ensure an understanding of the proportion of words and how they are related to one another. The UniGram considers each expression's frequency in the entire tweet; the BiGram Analyses each pair of words present in the tweets. At the same time, Trigram depicts the count of three-word notations in the tweets



IV (a) UniGram



IV (b) BiGram



IV (c) TriGram

Fig. IV Gram Analysis of Words in the Tweets Dataset

D. Feature Extraction

We encoded the pre-processed words as integers or floating-point values for input to a machine learning algorithm; this is feature extraction (Vectorisation). This research examines two approaches to extract features from the tweet words in the dataset.

Bag-of-Words Model (CountVectorizer)

A simple and effective model for thinking about text documents in machine learning is called the Bag-of-Words (BOW) Model. A BOW model encodes a record by assigning each word a unique number. It generates an encoded vector with the length of the entire vocabulary and integer count for each text's number of times appeared in the document.

Word Counts with TF-IDF Vectorizer (Term Frequency – Inverse Document Frequency)

An alternative is to calculate word frequencies, which are the components of the resulting scores assigned to each word. Frequency scores identify more represented words, in a statement and across statements [14]. A simple ranking function is computed by summing the TF-IDF for each query term. The TF-IDF model is sometimes used for stop-words filtering in various subject fields such as text summarisation and classification. Also. TF-IDF possesses weight, which helps in information retrieval and text mining. Many more sophisticated ranking functions are variants of this model.

Train Test Split

After encoding the 10314 instances of the tweets' text data, the feature extractors generated 2308 numerical features that we split into 7735 (70%), 2579 (30%) for training and testing, respectively.

All rights are reserved by UIJRT.COM.

D. Model Implementation

To determine whether a given tweet is depressive or not. The general overview of the stages involved in the sentiment analysis is shown in Fig. V.

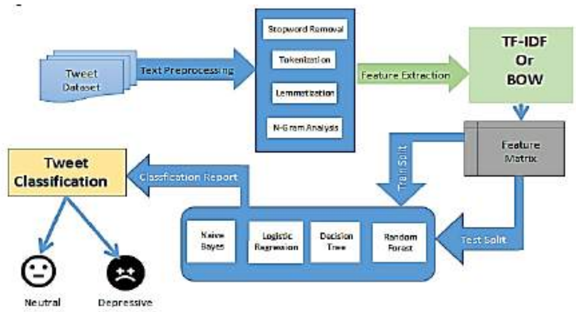


Fig. V Sentiment Analysis Work-flow

IV. RESULTS AND ANALYSIS

In this research, we performed a descriptive analysis shown in Fig. I-IV, to understand in context, the tweets captured in the dataset. With the model-ready feature matrix obtained from the two feature selection techniques adopted, Logistic Regression, Decision Trees, Random Forest and Naïve Bayes were trained to predict tweet polarity as either depressive or neutral in the test dataset fed into them, and their performances are analysed as follows:

A. Classification Performance

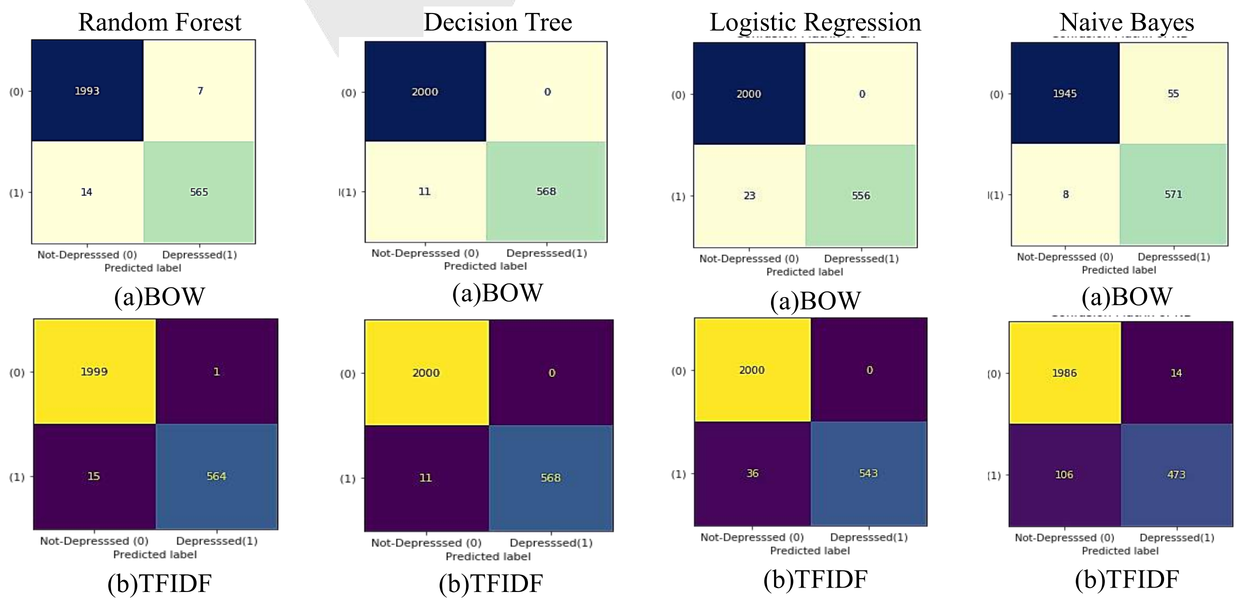
Confusion Matrices in Fig. VI show that the LR, DT and RF models had moderate overall performances on the held-out test data. Still, DT performed better with both TF-IDF and BOW feature matrix in classifying tweets into their various labels. The detailed report was also considered as depicted in Table I to enhance the hybrid evaluation metric for the imbalanced classes.

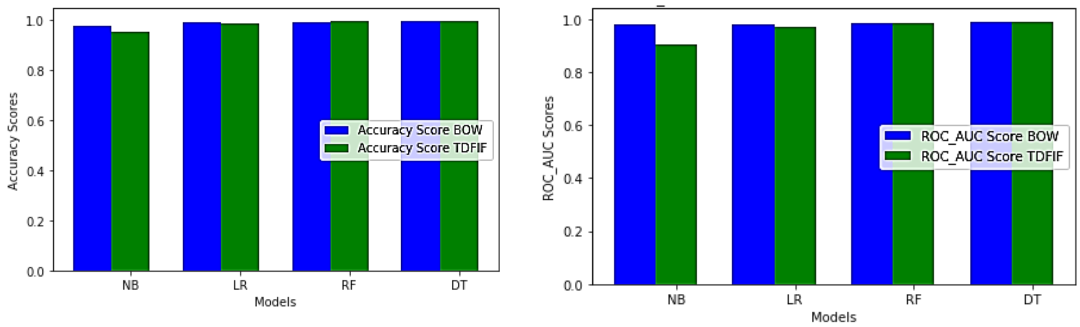
Table I. Classification Reports for the 4 Models with BOW and TF-IDF on the Test Set

Models	**BOW**					**TF-IDF**				
	Accuracy	**rocauc**	**Precision**	**Recal**	**F1**	**Accuracy**	**rocauc**	**Precision**	**Recal**	**F1**
NB	.98	.98	.98	.98	.98	.95	.90	.95	.95	.95
LR	.99	.98	.99	.99	.99	.98	.97	.98	.98	.98
RF	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
DT	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99

From Table I, TFIDF and BOW produced the highest accuracy (99%) with DT and RF while the BOW extends the same LR performance. However, to mitigate the challenges of erroneous classification of depressive individuals as neutral, the RF and DT consistently produced 99% ROC_AUC score with both TFIDF and BOW feature matrix. These establish the results depicted in the confusion matrices in Fig VI and Fig VII, where the tree-based models had the least False Positive and False Negatives for TFIDF and BOW models.

Fig. VI Confusion Matrices





(a) Accuracy Score

(b) ROC_AUC

Fig. VII Accuracy and ROC_AUC Score for the four models with BOW and TF-IDF

In this research, the tree-based algorithms like DT, and RF generalises well on the imbalanced datasets of both BOW and TFIDF features extracts. The splitting rules look at the class variable used in creating the trees, addressing both classes. With three depth of representation in Fig. VIII, the Attribute Selection Measures(ASM) of the Decision tree shows that the best attributes constitute words like “depression”, “anxietydepression”, “anxiety”, “depressionanxiety” etc.

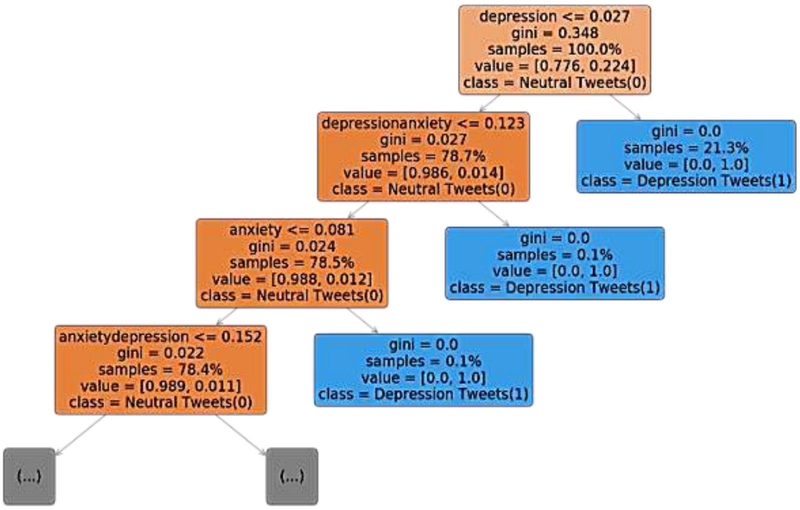


Fig. VIII Attribute Selection Measures(ASM) of the Decision Tree Classifier

The initial polarity of tweets in the original dataset and the predicted polarity that explains the DT model’s performance with BOW and TFIDF is given in Fig. IX below.

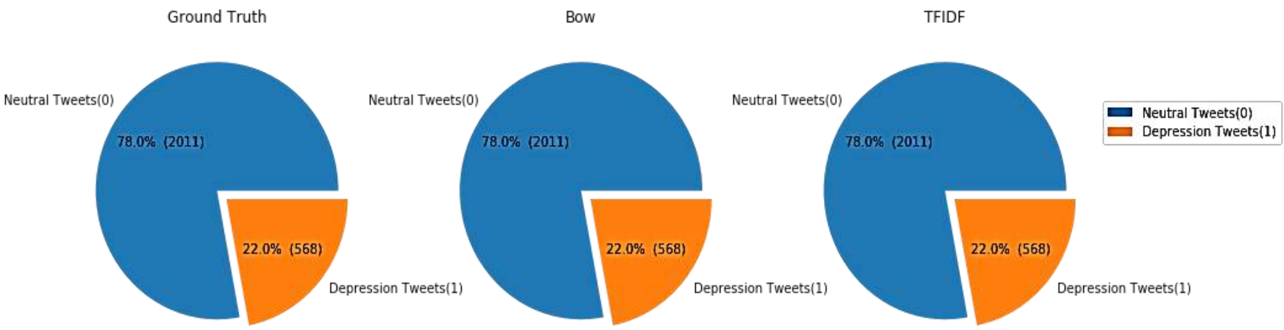


Fig. IX Polarity of Tweets in the Original Dataset, Decision Tree Classifier with BOW and TFIDF Model

The depression tweets have the least number of data entries with 22%, and the neutral tweets have the most data with 2011 (78%) entries.

All rights are reserved by UIJRT.COM.

V. CONCLUSION

Social media has revolutionised the way individuals interact and express their opinions in global space. This growth trend has enhanced the use of social media sites such as Twitter and Facebook Instagram to express deep thoughts, feelings, and ideas that are usable in identifying a person's state of mind. This state of mind, if not appropriately detected, may result in depression and unforeseen tragic consequences. In this research, we analysed 10,314 tweets downloaded from Kaggle and proposed Supervised learning models to identify depression disorder in tweet texts. Furthermore, we obtained feature matrices from the BOW and TF-IDF feature selection technique and trained the four supervised models. The trained models predicted tweet polarity as either depressive or neutral in the test dataset fed into them. Eventually, we compared the model's performances. The overall classification performance revealed that, given that the tweet dataset is imbalanced, tree-based algorithms like DT, and RF generalises well on the imbalanced datasets with the highest rocauc score (99%) in the two feature selection categories adopted in this research.

VI. RECOMMENDATION

Future works could explore Neuro-fuzzy and deep learning models for superlative prediction performance. Furthermore, the model could be tested on different user's base: geographic, age and profession. Though the developed model predicts users' mental health status, with few modifications, it could be used by psychiatrists, psychologists, and hospitals to mine posts of a single individual and predict mental health status individual.

REFERENCES

[1] T. Insel, "Digital phenotyping: technology for a new science of behaviour," Journal of the American Medical Association, Vol. 318, Issue 13, 1215-1216, 2017.

[2] J. S. Jini, and P. Prabu, "Detecting the magnitude of depression in Twitter users using sentiment analysis," International Journal of Electrical and Computer Engineering (IJECE) Vol.9, No.4, 3247-3255, 2019.

[3] https://www.merriam-webster.com/dictionary/social%20media#:~:text=Log%20In-Definition%20of%20social%20mediaother%20content%20(such%20as%20videos)

[4] S. Kapur, A. Phillips, and T. Insel, "Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it?" Molecular Psychiatry, Vol. 17, No.12, 1174, 2012.

[5] R. Thorstad, and P. Wolff, Behav Res, Vol.51: 1586–1600, 2019.

[6] S. Jain, B. Powers, J. Hawkins and J. Brownstein, "The digital phenotype," Nature Biotechnology, Vol.33, No.5, 462-463, 2015.

[7] B. Elvevag, A. Cohen, M. Wolters, H. Whalley, V. Gountouna, K. Kuznetsova, and K. Nicodemus, "An examination of the language constructs in NIMH's research domain criteria: time for reconceptualisation," American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, Vol.171, No.6, 904-919, 2016.

[8] J. H. Kietzmann, and H. Kristopher, "Social media? Get serious! Understanding the functional building blocks of social media," Business Horizons, Vol.54, No.3, 241–251, 2011.

[9] L. H. Shaw, and L. M. Gant, "In defence of the Internet: The relationship between Internet communication and depression, loneliness, self-esteem, and perceived social support," CyberPsychology and Behavior, Vol.2, 157–171, 2002.

[10] B. A. Mesfin, "Systematic review on the wide-spreading social media use and consequences of social media: mental health perspective, " International Journal of Development Research, Vol.09, No.10, 30712-30714, 2019.

[11] X. Wang, C. Zhang, Ji Y., L. Sun, L. Wu, Z. Bao A, "Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network," In Li J. et al. (eds) Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2013, Lecture Notes in Computer Science, Vol.7867. Springer, Berlin, Heidelberg, 2013.

[12] CSA Razak., M.A. Zulkarnain, S.H.A. Hamid, N. B. Anuar, M. Z. Jali, H. Meon, "Tweep: A system development to detect depression in Twitter posts," In Alfred R., Lim Y., Haviluddin H., On C. (eds), Computational Science and Technology. Lecture Notes in Electrical Engineering, Vol.603, 2020.

[13] M. Hemanthkumar A. Latha, "Depression detection with sentiment analysis of tweets," International Research Journal of Engineering and Technology (IRJET), Vol.06, No.05, 1197, 2019.

[14] Jason. "Machine Learning Mastery," ACN: 626 223 336. https://machinelearningmastery.com/gentle-introduction-bag-words-model/