

Depression prognosis using natural language processing and machine learning from social media status

Md. Tazmim Hossain, Md. Arafat Rahman Talukder, Nusrat Jahan

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Jul 1, 2021

Revised Jan 5, 2022

Accepted Jan 23, 2022

Keywords:

Depression

Logistic regression

Machine learning

Multinomial NB

NLP

Social media

ABSTRACT

Depression is an acute problem throughout the world. Due to worst and prolong depression many people dies in every year. The problem is that most of the people are not concern of the fact that they are suffering from depression. In this research, our aim was to find out whether an individual is depressed or not by analyzing social media status. Therefore, we focused on real data. Our dataset consists of 2,000 sentences, which was collected from different social media platforms Facebook, Twitter, and Instagram. Then, we have performed five data pre-processing approaches for natural language processing (NLP) such as tokenization, removal of stop words, removing empty string, removing punctuations, stemming and lemmatization. For our selected model, we considered that processed data as an input. Finally, we applied six machine learning (ML) classifiers multinomial Naive Bayes (NB), logistic regression, liner support vector classifier, random forest, K-nearest neighbour, and decision tree to achieve better accuracy over our dataset. Among six algorithms, multinomial NB and logistic regression performed well on our dataset and obtained 98% accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nusrat Jahan

Department of Computer Science and Engineering, Daffodil International University

102, Sukrabad Mirpur Rd, Dhaka 1207, Bangladesh

Email: nusratjahan.cse@diu.edu.bd

1. INTRODUCTION

Depression is a usual and significant medical disease which negatively affects a person including their feeling and thinking approach. It is the key reason of inability and for suicidal attempt worldwide. Anxiety and other psychological disorders are connected with depression. Depression affected our social life and control our feelings and behavior. Around 80,000 people died for this acute disorder [1]. In addition, Asian people are more sufferer by this single disease. It also a concern that many country do not consider depression as a disease therefore it is still under-diagnosed [2]. People are losing their ability to do their daily work due to depression and this is pushing them towards more severe depression. The biggest problem is that the people who are depressed do not know that they are suffering from it. As a result, they gradually suffer from several mental and physical damages. Research also proposed, if the following symptoms appear only once a day for 2 weeks then diagnosis is essential [3], [4]: i) feeling of incompetence and worthlessness, ii) decrease of diurnal activity, iii) changes the pattern of sleep, iv) suicidal perception, v) reckless behavior or excessive anger and irritability, and vi) changes in appetite.

People are now expressing their thoughts and feelings through social media where their mental feelings are expressed through text, pictures and videos. Text data has become an extensively used and popular medium of communication. On the other hand, machine learning classifiers are useful for getting information from data. Machine learning algorithms are mainly separated among three sections—supervised,

unsupervised and reinforcement learning. To identify depression from text data we have applied different supervised learning techniques. In a study, support vector regression (SVR) and multinomial Naive Bayes (NB) was applied and compared their accuracy with other authors. SVR performed better than multinomial NB and accuracy was 79.7% [5].

In another study, where dataset was created manually and 7,500 sentences were collected for the corpus. Here, NB and topical approach classifiers were applied to predict depression. Topical approach performed better than NB and its accuracy was 90% [6]. Twitter application programming interface (API) is one of the major sources for data collection. On the other hand, recurrent neural network (RNN) performed better than most of the renowned classifiers. In another study, RNN, NB, and SVM was applied and RNN's performance (90.3% accuracy) was better than others [7]. Almouzini *et al.* [8] also worked with twitter data. Here, they worked with western Asian countries and collected 7,000 tweets from 97 twitter users. LIBLINEAR classifier worked better for their dataset and among four classifiers 87.5% accuracy was obtained.

In recent past, a study proposed six classifiers to identify saint and common form in Bengali text. Though they have used six classifiers, among all of those accuracy of NB was much efficient than other classifier's. Furthermore, count-vectorizer, tokenizing words, removal of stop words, part-of-speech (POS) tagging were the major steps for data preprocessing [9]–[11]. Different libraries and tools such as natural language toolkit (NLTK), TextBlob, Waikato environment for knowledge analysis (WEKA), and Beautiful Soup had been used for data preprocessing [12], [13].

People with critical depression may have a negative thinking on life and try to take a dangerous attempt like suicide. In the paper, Coppersmith *et al.* [14] proposed a natural language processing and deep learning based suicidal risk prediction model and data was collected from social media. In this case, total 418 users were observed who attempted suicide. In the paper, Caicedo *et al.* [15] reported a result on suicidal attempt. In this research they applied supervised classifier to predict a suicidal case from messages. They worked on 3,472 data. They applied several classifiers and initially they considered four categories. Depression prediction from speech is also a growing research area. In the paper, He and Cao [16] proposed an approach to predict depression using deep convolutional neural networks (DCNN) model. Their dataset was audio-visual emotion recognition challenge 2013 (AVEC2013) and AVEC2014 and got 91% accuracy. One study done by Priya *et al.* [17] predicted anxiety, depression and stress. For this purpose, they collected data through Google form and total 348 participants were involved. They chose five classifiers where NB performed better than other and obtained 85% accuracy.

In the paper, Zaghouani [18] considered 3,200 Twitter data for predicting youth depression. They focused on Middle East and North Africa (MENA) region to collect data and to analyze depression. Here, natural language processing (NLP) tools along with machine learning algorithms were used to predict depression [18]. In the paper, Sharma *et al.* [19] used machine learning with natural language processing to predict suicidal issues among youth. They considered twitter data from kaggle.com and obtained 88% accuracy from term frequency and inverse document frequency (TF-IDF) approach. The performance of the classifier varies for different datasets. In the paper, Nigam *et al.* [20] proposed a machine learning based approach on sentimental analysis where sentiment 140 (Stanford University) was used as a dataset. Different machine learning algorithms were applied but logistic regression performed better and accuracy was 82.59%. For few cases, SVM performed much better than other classifiers. In the paper, Laoh *et al.* [21] worked on hotel review to find out a review whether it is positive or negative. SVM had performed better with the accuracy of 94%. For the same dataset he had applied recursive neural tensor network (RNTN) classifier but the accuracy was 85%. In the paper, Hassan *et al.* [22] analyzed sentiment from sentence to measure depression. A comparison among SVM, NB, and maximum entropy had made by them. Data was collected from Twitter and 20 news group. In this study, SVM, NB and ME performed and their following accuracy was 91%, 83% and 80% [23]. Twitter data is also used to predict personality according to Garg and Garg [24]. In this study authors used six machine learning classifiers. They worked on 17,464 user's data. Here, four personality attitudes were considered to predict individual's personality.

In the paper, Patidar and Umre [25] proposed an approach to predict depression and analysis reason of the depression. For this study they collected total 1,695 data where stressed data were 535 and non-stressed were data 1,160. They applied naive Bayes algorithm to classify depression level. In the paper, Tadesse *et al.* [26] studied on Reddit social media forum to predict depression. They focused natural language processing along with machine learning algorithms to complete this work. Here, multilayer perceptron (MLP) classifier worked 91% accurately. They considered total 1841 posts for predicting depression where 1293 posts for depression related and 548 were standard posts. In the paper, Rustam *et al.* [27] considered 7528 tweets. They measured sentiment and applied different machine learning algorithms random forest (RF), XGBoost, support vector classifier (SVC), extra tree classifier (ETC) and decision tree. Among these ETC provided the best accuracy 92%.

To complete our study and to get our desire outcome which is showed in Figure 1, we have applied supervised machine learning techniques. The application process of this technique is easy to build and provide better results. Our dataset is labeled into two tags named as depressed and non-depressed. In recent days, depression is a major problem and young generation is in more vulnerable condition. Finally, our aim is to propose a model to predict depression and the objectives: i) the creation of a potential dataset for natural language processing, ii) an examination of a standard algorithm to predict depression and can be used for further research, and iii) evaluate the model performances based on proposed model and previous study. We organized our paper, in section 2 we discussed about the proposed method where data collection and data pre-processing procedure are the main part. In section 3, we compare and examine the results also present the better machine learning technique to predict depression. We discussed about our future work in section 4 including conclusion and limitation of our study.

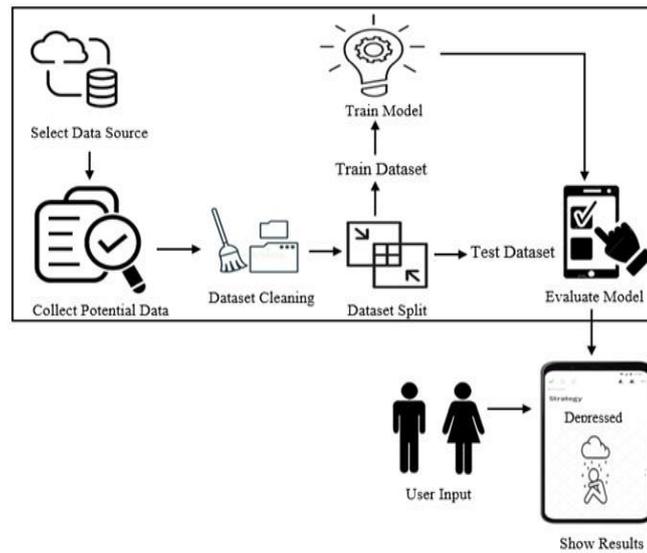


Figure 1. Structural design of our proposed model

2. PROPOSED RESEARCH METHOD

To find out our desire outcome, we have applied supervised learning method to predict depression form the text data. The working procedure is illustrating step by step in the following sections. Figure 2 presented the work flow of our proposed work.

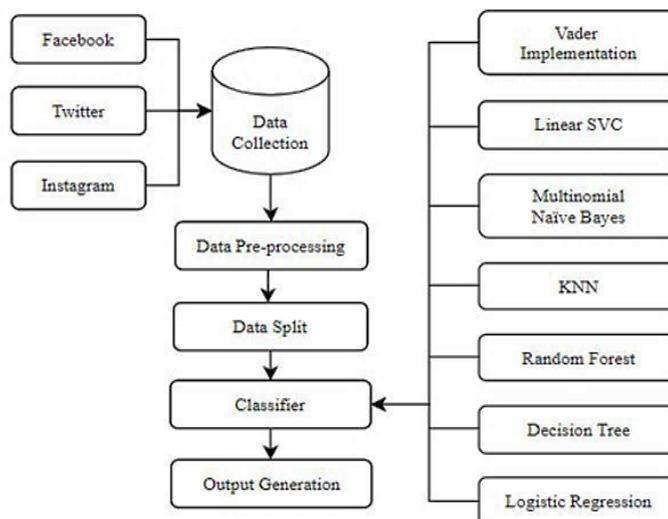


Figure 2. Flow chart of our proposed model

2.1. Data collection and analysis

Data collection was one of our major challenges. We collected data from different social sites manually and insert into an excel file. Our dataset consists of total 2,000 sentences where 1,025 tagged as non-depressed sentences and 975 tagged as depressed sentences. Here, Table 1 shows the distribution of data sources and the number of data we collected. We have used count plot to visualize data which is given in Figure 3. We have collected all sample data through three social media sites which is mentioned in Table 1. We considered user's comment and social status to create a potential dataset. After studying depression detection text and other related works, we find out the desire data. Different text has different meaning. Table 2 is created to present a sample dataset.

Table 1. Description of data sources

Social media sites	Number of user	Number of sentences
Facebook	100	1050
Instagram	70	450
Twitter	40	500
Total	210	2000



Figure 3. Data visualization

Table 2. Sample data

Sentence	Label
My heart keeps breaking.	Depressed
Life is awesome, overlook your difficulties, be cheerful, live your life, Feel so happy, Nice to see you, It is surprising	Non-Depressed
It is my bad luck, Do not upset, Everything will be fine, Life is just a drama, Need a tour, anyone want to join with me.	Depressed
Life is short, be happy, live your life, Enjoy your day, Think about yourself,	Non-Depressed

2.2. Data pre-processing

We applied five data pre-processing techniques on our dataset. Text pre-processing helps us to increase the quality of the dataset as well as the accuracy of the classifiers. We performed following pre-processing techniques to increase the accuracy of a classifier, which is showed in Figure 4.

- Data normalization: after importing our dataset we checked whether there was any null input or not. We found out two null inputs on our dataset then we had dropped those null inputs. All types of punctuations have been removed from our dataset as they do not put any impact in our research work. List of punctuations that we removed from our dataset are given in Table 3.
- Data tokenization: tokenization is the method of splitting or tokenizing a string [9]. Words are the token of a sentence and the sentences are the token of a paragraph. We applied sentence to word tokenization on our dataset. Table 4 is an example of tokenization approach.
- Removal of stop words: the most used words are known as stop words. Stop-word do not have any significant to identify depression level from a sentence [10]. Therefore, we removed those words from our dataset to reduce the unnecessary load. We built our own stop words considering the fact that they do not change the meaning of a sentence. Some of the examples are given in Table 5.

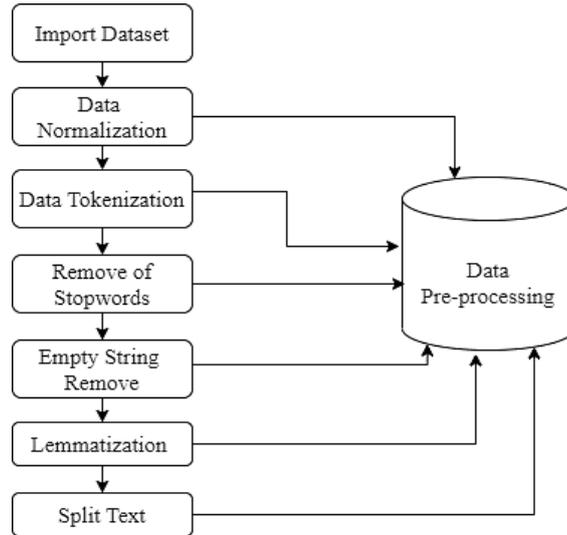


Figure 4. Data pre-processing approaches

Table 3. Punctuation list

List of punctuations					
'	!	“	#	\$	%
“	()	*	+	-
:	:	<	=	>	?
[]	^	_	{	}
~	`	/	\	@	

Table 4. Example of tokenization

Before tokenization	After tokenization
My heart keeps breaking.	[my, heart, keeps, breaking]
Life is short, be happy, feel so happy	[lifes, short, be, happy, feel, so, happy]
Everything will be fine	[Everything, will, be, fine]
That's how I feel now.	[That's, how, I, feel, now.]

Table 5. List of stop words

Our customized list of Stop words
'a', 'about', 'above', 'across', 'after', 'again', 'all', 'almost', 'along', 'already', 'also', 'although', 'always', 'am', 'among', 'amongst', 'amongst', 'amount', 'an', 'and', 'another', 'any', 'anyhow', 'anyone', 'anything', 'anyway', 'anywhere', 'are', 'around', 'as', 'at', 'back', 'be', 'became', 'because', 'become', 'becomes', 'becoming', 'been', 'before', 'being', 'below', 'beside', 'besides', 'between', 'bill', 'both', 'but', 'by', 'call', 'can', 'co', 'con', 'de', 'describe', 'detail', 'each', 'eg', 'eight', 'either', 'eleven', 'else', 'elsewhere', 'etc', 'even', 'ever', 'every', 'everyone', 'everything', 'everywhere', 'he', 'hence', 'her', 'here', 'hereafter', 'hereby', 'herein', 'hereupon', 'hers', 'herself', 'him',

- Empty string remove: empty string can be very sensitive while implementing the classifiers. It kills our memory space and we may get lower accuracy rate for a classifier. Therefore, we removed all empty string from our dataset. Table 6 to present an example of empty string remove.
- Lemmatization: lemmatization is the method of removing inflectional endings from a particular word and it returns the base or the dictionary form of that word. It is also help us to increase the accuracy of an algorithm. Here, Table 7 to present the results of Lemmatization technique.
- Data split: after applied all mentioned pre-processing techniques finally, we created a dataset for prediction depression. For getting more appropriate results, data split is also a major part of machine learning. Here, we considered 80% data for training purpose and 20% data to test the model's performance.

Table 6. Removing empty string

Before removing empty string	After removing empty string
[“You”, “”, “are”, “ ”, “looking”, “”, “gorgeous”]	[“You”, “are”, “looking”, “gorgeous”]

Table 7. Example of lemmatization

Before	After
Drinking	Drink
Is	Be
Done	Do
Goes	Go

2.3. Proposed model

We proposed a model using machine learning techniques to analyze human depression through text analysis. Valence aware dictionary for sentiment reasoning (VADER) is a model for analyzing text for sentiment. We implemented VADER to analyze the polarity of each sentence but the accuracy was 85%. After that, we applied following algorithms to get the desired output multinomial naive Bayes (MNB), linear SVC, KNN, RF, decision tree, and logistic regression (LR).

3. RESULTS AND DISCUSSION

To achieve our desire goal, we have completed all experiments in python environment. Python has many necessary libraries to implement a complete system. The following requirements are also necessary to develop our proposed system-OS (Windows/Linux/Mac), RAM (minimum requirements of 4 GB), HDD (minimum requirements of 128 GB).

We done all experiments based on six algorithms which we mentioned in previous section. Among six algorithms, MNB and LR provided the best result. MNB classifier is the collection of different algorithms. It is based on Bayes theorem. In NB classifiers all the algorithms share their common principle [24]. Our classification report successfully predicted the output with 98% accuracy.

Performance measurement unit can express an algorithm more precisely. F1-score is for controlling the balance between recall and precision. Precision is mentioning the ratio of total true positive value predicted and the total true positive class values predicted. Recall is also crucial to define a performance of an algorithm. It presents the number of positive value divided by the total number of true positives and false negatives [28]. All mentioned performance measurement units are defined:

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

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

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

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

For NB algorithm we calculated all those mentioned measurement units. We provided Table 8 for illustrating the best performance results of MNB classifier.

By implementing VADER, we tried to find out the initial accuracy. However, the accuracy was 85%. After that, we applied six above mentioned classifiers on our dataset to get more accurate result. Table 9 to represent the performance of our selected algorithms using four measurement unit.

Table 8. Performance of MNB on our dataset

	Precision	Recall	F1-score	Support
Depressed	0.96	0.99	0.98	195
Non Depressed	0.99	0.97	0.98	205
accuracy			0.98	400
macro avg	0.98	0.98	0.98	400
weighted avg	0.98	0.98	0.98	400

Table 9. Compared different proposed classifiers

Classifier	Precision	Recall	F1-score	Accuracy
Multinomial NB	0.98	0.98	0.98	98%
LR	0.97	0.97	0.97	97%
Linear SVC	0.96	0.97	0.96	96%
RF	0.93	0.92	0.93	93%
KNN	0.95	0.84	0.89	91%
DT	0.92	0.86	0.89	90%

To illustrate the performance of a model we used receiver operating characteristic (ROC) curve. ROC curve is a graph for presenting accuracy based on a threshold value and the plot created with two parameters-true positive rate and false positive rate. We can also compare algorithms through this graph [29]. Figure 5 shows the ROC curve on our proposed algorithms.

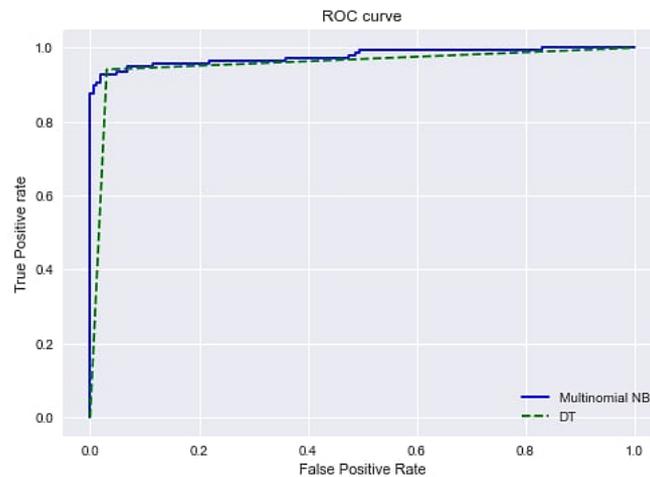


Figure 5. ROC curve for four selected algorithms

4. CONCLUSION

Our aim was to predict depression by using different social media posts. Especially, people in poor and least developed countries are suffering from this disease. Therefore, they are suffering from depression for a long time which is reducing their ability to survive and it slowly pushing them towards death. A proper solution is needed to overcome from the level of depression along with its impact. Using our model it will be easy to identify depression. Our proposed model performed on 2,000 sample text data, and got 98% accuracy for predicting depression. Here, MNB and LR provided the best result. However, to solve and propose an acceptable result the amount of data is crucial.

Still there are many ways to enhance our study. Bring out this research work in a mobile application will be a great further plan. That application will be helpful to check depression level at any time and people will be able to take necessary steps. A notification system can be used for generating awareness through social media. Moreover, we have a plan to identify how a patient can recover from depression and also want to work on large dataset to increase the accuracy rate as much as possible. Therefore, need to create a more appropriate dataset.

ACKNOWLEDGEMENTS

The authors would like to thank Daffodil International University, Dhaka, Bangladesh to complete this research work.

REFERENCES

- [1] W. H. Organization, "Depression," *World Health Organization*. <https://www.who.int/news-room/fact-sheets/detail/depression> (accessed Mar. 03, 2021).
- [2] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter, and A. K. Das, "An automated system of sentiment analysis from Bangla text using supervised learning techniques," in *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*, Feb. 2019, pp. 360–364, doi: 10.1109/CCOMS.2019.8821658.
- [3] "Recurrent major depressive disorder with a seasonal pattern," *Elsevier*. <https://elsevier.health/en-US/preview/recurrent-major-depressive-disorder-seasonal> (accessed Apr. 24, 2020).
- [4] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017*, Dec. 2018, pp. 858–862, doi: 10.1109/ISS1.2017.8389299.
- [5] P. Arora and P. Arora, "Mining twitter data for depression detection," in *2019 International Conference on Signal Processing and Communication, ICSC 2019*, Mar. 2019, pp. 186–189, doi: 10.1109/ICSC45622.2019.8938353.
- [6] V. Goel, A. K. Gupta, and N. Kumar, "Sentiment analysis of multilingual twitter data using natural language processing," in *Proceedings - 2018 8th International Conference on Communication Systems and Network Technologies, CSNT 2018*, Nov. 2018, pp. 208–212, doi: 10.1109/CSNT.2018.8820254.
- [7] D. Mahajan and D. Kumar Chaudhary, "Sentiment analysis using RNN and google translator," in *Proceedings of the 8th*

- International Conference Confluence 2018 on Cloud Computing, Data Science and Engineering, Confluence 2018*, Jan. 2018, pp. 798–802, doi: 10.1109/CONFLUENCE.2018.8442924.
- [8] S. Almouzini, M. Khemakhem, and A. Alageel, “Detecting Arabic depressed users from Twitter data,” *Procedia Computer Science*, vol. 163, pp. 257–265, 2019, doi: 10.1016/j.procs.2019.12.107.
- [9] N. J. Ria, S. A. Khushbu, M. A. Yousuf, A. K. M. Masum, S. Abujar, and S. A. Hossain, “Toward an enhanced Bengali text classification using saint and common form,” *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*. IEEE, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225358.
- [10] N. Al Asad, M. A. Mahmud Pranto, S. Afreen, and M. M. Islam, “Depression detection by analyzing social media posts of user,” in *2019 IEEE International Conference on Signal Processing, Information, Communication and Systems, SPICSCON 2019*, Nov. 2019, pp. 13–17, doi: 10.1109/SPICSCON48833.2019.9065101.
- [11] M. R. Hasan, M. Maliha, and M. Arifuzzaman, “Sentiment analysis with NLP on Twitter data,” 2019, doi: 10.1109/IC4ME247184.2019.9036670.
- [12] S. A. S. A. Kulasinghe, A. Jayasinghe, R. M. A. Rathnayaka, P. B. M. M. D. Karunarathne, P. S. Silva, and J. A. Jayakodi, “AI based depression and suicide prevention system,” in *Int. Conf. on Advancements in Computing, (ICAC)*, Dec. 2019, pp. 73–78, doi: 10.1109/ICAC49085.2019.9103411.
- [13] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun, “Facebook social media for depression detection in the thai community,” *Proceeding of 2018 15th International Joint Conference on Computer Science and Software Engineering, JCSSE 2018*, Sep. 2018, doi: 10.1109/JCSSE.2018.8457362.
- [14] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, “Natural language processing of social media as screening for suicide risk,” *Biomedical Informatics Insights*, vol. 10, Jan. 2018, doi: 10.1177/1178222618792860.
- [15] R. W. Acuña Caicedo, J. M. Gómez Soriano, and H. A. Melgar Sasieta, “Assessment of supervised classifiers for the task of detecting messages with suicidal ideation,” *Heliyon*, vol. 6, no. 8, Aug. 2020, doi: 10.1016/j.heliyon.2020.e04412.
- [16] L. He and C. Cao, “Automated depression analysis using convolutional neural networks from speech,” *Journal of Biomedical Informatics*, vol. 83, pp. 103–111, Jul. 2018, doi: 10.1016/j.jbi.2018.05.007.
- [17] A. Priya, S. Garg, and N. P. Tigga, “Predicting anxiety, depression and stress in modern life using machine learning algorithms,” *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020, doi: 10.1016/j.procs.2020.03.442.
- [18] W. Zaghouni, “A large-scale social media corpus for the detection of youth depression (project note),” *Procedia Computer Science*, vol. 142, pp. 347–351, 2018, doi: 10.1016/j.procs.2018.10.483.
- [19] M. Sharma, B. Pant, V. Singh, and S. Kumar, “STP:Suicidal tendency prediction among the youth using social network data,” in *Advances in Intelligent Systems and Computing*, vol. 1162, Springer Singapore, 2021, pp. 161–169.
- [20] S. Nigam, A. K. Das, and R. Chandra, “Machine learning based approach to sentiment analysis,” in *Proceedings-IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018*, Oct. 2018, pp. 157–161, doi: 10.1109/ICACCCN.2018.8748848.
- [21] E. Laoh, I. Surjandari, and N. I. Prabaningtyas, “Enhancing hospitality sentiment reviews analysis performance using SVM N-grams method,” in *2019 16th International Conference on Service Systems and Service Management (ICSSSM)*, Jul. 2019, pp. 1–5, doi: 10.1109/ICSSSM.2019.8887662.
- [22] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, “Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression,” in *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017*, Oct. 2017, vol. 2017-December, pp. 138–140, doi: 10.1109/ICTC.2017.8190959.
- [23] D. Tanna, M. Dudhane, A. Sardar, K. Deshpande, and N. Deshmukh, “Sentiment analysis on social media for emotion classification,” in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, May 2020, pp. 911–915, doi: 10.1109/ICICCS48265.2020.9121057.
- [24] S. Garg and A. Garg, “Comparison of machine learning algorithms for content based personality resolution of tweets,” *Social Sciences & Humanities Open*, vol. 4, no. 1, 2021, doi: 10.1016/j.ssaho.2021.100178.
- [25] H. Patidar and J. Umre, “Predicting depression level using social media posts,” *IEEE 13th Int. Symposium on Autonomous Decentralized System (ISADS)*, vol. 8, no. 12, pp. 234–237, 2021, doi: 10.29121/granthaalayah.v8.i12.2020.1972.
- [26] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, “Detection of depression-related posts in reddit social media forum,” *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.
- [27] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, “A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis,” *PLoS ONE*, vol. 16, no. 2, 2021, doi: 10.1371/journal.pone.0245909.
- [28] T. Basu and C. A. Murthy, “A feature selection method for improved document classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Berlin Heidelberg, 2012, pp. 296–305.
- [29] X. H. Wang, P. Shu, L. Cao, and Y. Wang, “A ROC curve method for performance evaluation of support vector machine with optimization strategy,” in *IFCSTA 2009 Proceedings - 2009 International Forum on Computer Science-Technology and Applications*, 2009, vol. 2, pp. 117–120, doi: 10.1109/IFCSTA.2009.356.

BIOGRAPHIES OF AUTHORS



Md. Tazmim Hossain     received the B.Sc. Degree in Computer Science and Engineering (CSE) from Daffodil International University (DIU), Dhaka, Bangladesh. He has experience on Machine Learning and Web development. His current research interests are in Natural Language Processing, Deep Learning, Image Processing, Artificial intelligence, Cloud Computing. He has expertise knowledge on C, C++, Python, PHP, and Javascript. He is very enthusiastic in research work and programming. He can be contacted at email: tazmim15-10390@diu.edu.bd.



Md. Arafat Rahman Talukder    received the B.Sc. Degree in Computer Science and Engineering (CSE) from Daffodil International University (DIU), Dhaka, Bangladesh. He has experience on Web development and Artificial Intelligence. He had worked as an intern on a Software Development Company. His current research interests are in Natural Language Processing System, Deep Learning, Image Processing, Artificial intelligence, Cloud Computing. He can be contacted at email: arafat15-10404@diu.edu.bd.



Nusrat Jahan    received the bachelor's degree from the Department of Computer Science and Engineering, Patuakhali Science and Technology University, Bangladesh and Master in Information Technology from Institute of Information Technology (IIT), University of Dhaka. She has 5 years of teaching experience. She is currently a Senior Lecturer in Computer Science and Engineering Department at Daffodil International University, Bangladesh. Her research interest focuses on machine learning, data science, bioinformatics and NLP. Her can be contacted at email: nusratjaha.cse@diu.edu.bd.