Mental Health Status Prediction Using ML Classifiers with NLP-Based Approaches

Israt Jahan Dristy
Department of Computer Science & Engineering
Brac University
Mohakhali, Dhaka-1212, Bangladesh
israt.jahan.dristy@g.bracu.ac.bd

Abdullah Matin Saad
Deparment of Computer Science & Engineering
Brac University
Mohakhali, Dhaka-1212, Bnagladesh
mabdullah@bracu.ac.bd

Annajiat Alim Rasel
Deparment of Computer Science & Engineering
Brac University
Mohakhali, Dhaka-1212, Bnagladesh
annajiat@gmail.com

2022 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET) | 979-8-3503-3270-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRPSET57982.2022.10188544

***Abstract*—Nowadays, people use different online platforms for their day-to-day communication. It is possible to use Natural language processing approaches to infer people's mental states from what they share on these platforms via text or speech. Predicting mental health problems by SMS is a proactive step toward improved treatment. NLP is changing the way mental health experts evaluate the free expression designed to check and diagnose mental diseases in patients. Using machine learning techniques can provide new ways to study human attitudes and behaviors as well as identifying signs and symptoms of mental illness. In this paper, we used NLP for detecting the status of mental health through a text message and our research gives a detailed investigation of several supervised classifier algorithms. Peoples suffer from different types of mental illnesses, but bipolar, panic disorder, depression, stress, anxiety, and PTSD are popular. For this analysis, We classified data using Logistic Regression, BernoulliNB, K-Nearest Neighbors, Random Forest, and Decision Trees. In our proposed method, Logistic Regression shows the best performance compared with the other four classifiers. The experimental outcome supports the proposed methodology's ability to classify patient data more accurately. The proposed model was shown to be effective with an accuracy rate of 93 percent.**

Keywords— machine learning, NLP, Supervised algorithm, LR, KNN, BernoulliNB.

I. Introduction

In this era, people used to communicate with others using various types of social media and chat applications such as WhatsApp, Twitter, phone SMS, etc. So, it's too hard to recognize by seeing someone's message to know about her/his mental state. Negative messages from anyone can detect as a mentally unstable state.When people feel insecure or when they are not able to stay in their normal state, they cannot communicate with anyone in a good mood. For this reason, the negative message may know as mental illness.

When people enjoy themselves or they feel the love they send a lovely message to their close ones. So positive message may denote a normal status of mental health. In our research, we considered the Emotions dataset based on people sending SMS to their friends and family. To predict their mental status, we train our model based on these data. Then we analyze the comparison of those supervised algorithms. Logistics Regression, Random forest, BernoulliNB, Decision Tree and K-nearest Neighbor used here.

II. Previous Works

To improve sentiment-based classification, Raj et al. used natural language processing (NLP). Feature of Sentiment To improve the validity and reliability of data for these objectives, a vector and the commencement of a word have been used. In this work, the several machine learning methodologies such as naive Bayes, SVM, and greatest entropy are usually used, as also the Disambiguation of Words. To enhance classification accuracy, NLP and WSD are implemented.[1]

This study employs a system that first collects data from the Twitter social media platform, which is then cleaned using several techniques. Feature vectors are taken from the data used for training, and then To categorize the data, a various machine learning classifiers are utilized.. This study also makes use of Python 2.7.3 and the NLTK module for NLP techniques.[1]

According to the findings of this study, employing sunsets can result in a 3% to 6% gain in performance. Future studies for this type of research could focus on reducing costs and improving approaches at the paragraph level, according to the author.[1]

Calvo et al. hse use NLP to create a common language that combines NLP, HCI, and mental health to provide psychological support. NLP may be utilized to construct marketing apps based on customer reviews and feelings, as well as to generate AI applications and design mental health applications, according to the author. Models based on data, labels, and interventions have been proposed in this study. They analyzed data and interventions using NLP. The author has looked at how various types of data can be used to create mental health applications using NLP.[2]

Katchapakirin et al. examined posts by Thai People on social media to detect depression. Natural language processing has been used in research to examine algorithms for identifying depression. Micro-blogging websites and the TMHQ psychological health survey have been used to carry out this research. SVM with Weak and random forest techniques with rapid miners have been used to optimize classifications. Positive and negative polarity data are split into two categories. Users who express terrible perspectives in private with only me are much more depressed than people that share thoughts with everybody and from dawn to noon, as per this survey. [3]

979-8-3503-3270-4/22/$31.00 ©2022 IEEE

Nigam et al. demonstrated the taxonomy of many sentiment analysis approaches using NLP and demonstrated that logistic regression has a higher accuracy when compared to other techniques. This study uses Twitter data and a supervised machine-learning method. The proposed system cleans data by removing HTML decoding, URLs, and symbols, among other things. Then explore terms across both favorable and unfavorable circumstances. It visualizes phrases in both negative and good situations, but this time without the stop words. The various machine learning approaches are used to classify and analyze data and results when features are extracted.[4]

Mental health issues are now of significant concern in Malaysia. In Malaysia, one in five persons experiences depression, two in five experience anxiety, and one in ten experience stress, according to the National Health and Morbidity Survey 2017. Students in higher education run risk of becoming a member of the impacted community. With the aid of machine learning algorithms, this study will group pupils according to their levels of stress, sadness, and anxiety. Students at a higher education facility in Kuala Terengganu provided the data. [5]

To determine if and where teenagers with depression sought out mental health therapies, this study evaluated both individual depression symptoms and sociodemographic factors. 53.38 percent of teenagers with significant depressive symptoms sought treatment of some type.[6]

Women without a history of depression underwent sub-analyses. Prenatal sadness and anxiety, as well as traits associated with toughness & psychology, were the factors that placed women at the highest risk for PPD. Future clinical models that might be put into action right away after birth might think about incorporating these characteristics to help identify women who are at high risk for postpartum depression and to enable cost-effective and personalized follow-up.[7]

Research on the application of machine learning in mental health difficulties has just been spurred by the increase in psychological problems and the urge for high-quality medical care. By exploring dependable databases, we gather studies and publications on a machine learning methods for forecasting mental health issues. Then, we aggregate the research articles we've acquired into categories based on mental health issues like schizophrenia, bipolar illness, anxiety & depression, posttraumatic stress disorder, and issues with children's mental health.[8]

The interpretation of the various models used to analyze the data from the mental health survey led us to the conclusion that the healthcare role a frontline worker holds, along with recent sleep patterns, the amount of COVID-19-related news they typically consume daily, their age, and their use of alcohol and cannabis, are the most significant predictors of mental health decline.[9]

One of the conditions for which there hasn't been a complete cure is a mental disease. The biggest problem is determining whether a person has a mental disorder in the first place. Somebody may be going through a particular scenario for several reasons, including society, Workplace stress, family, etc. Our exploration into this dilemma will be restricted to forecasting like as illness in person's bodies and determining what the patient is experiencing using the pre-classified dataset. Furthermore, the test outcome from this application can serve as a real-world illustration of IoT in healthcare.[10]

III. Methodology

Fig. 1 displays the sequence diagram of the method we've suggested. In the beginning, we are going to take our dataset and check for the null value. If there is a null value, we clean those data. In our dataset, if there is any duplicated data, then drop data and reindex the data. After that, we labeled our dataset. Just because our dataset is textual data we label it between 0 and 1. We labeled normal health status as a 0 and unstable stage(depression, bipolar,a stress, panic disorder, anxiety) as a 1. Then we make a sentiment analysis on positive and negative on the dataset. We preprocessed our textual dataset which is shown in fig 2. In preprocessing step, at first, we use the function to expand the contracted phrase into normal words. Then clean the texts and remove the stop words using lemmatization. By using stemming we extract the actual meaning of words and also used tokenization. So now we are going to split the dataset. 30 percent set as a test set and the rest part set as a training part. Every one of the supervised classification techniques will be trained using these training data. Following the training of the model parameters, we will utilize them to forecast where we will provide prediction output for test data and comparison among those classifiers. Then for the final evaluation, we will use the highest accuracy classifier on test data.

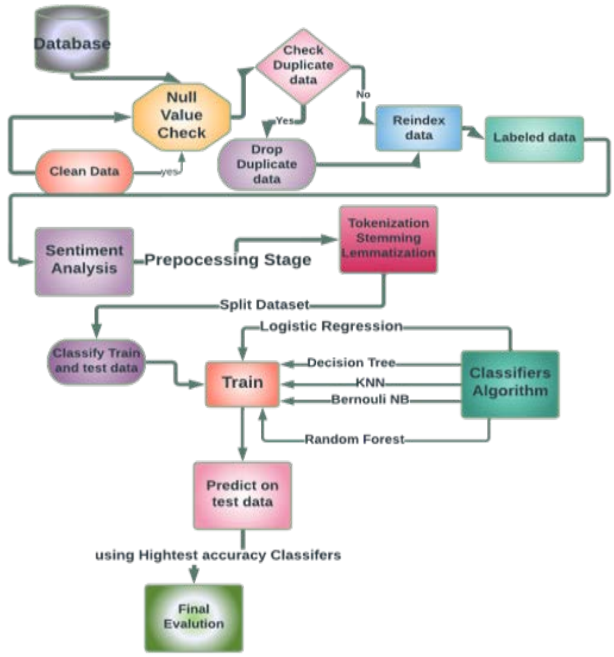


Fig. 1. Methodology.

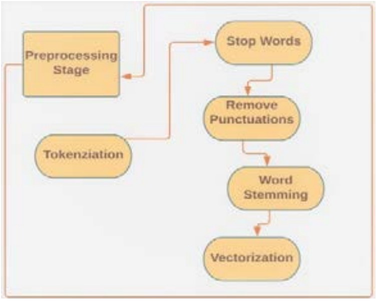


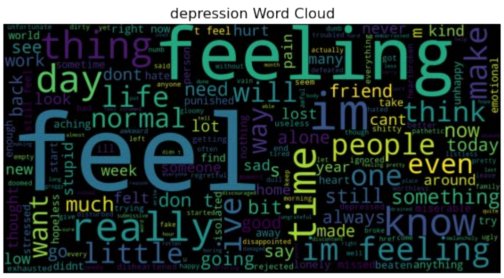
Fig. 2. Preprocessing step.

A. Obtaining Data

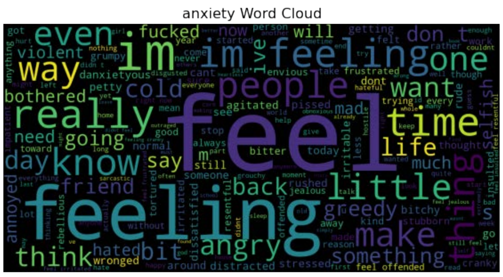
Mental health encompasses a wide range of our intellectual, cognitive, and societal well-being. It influences our ideas, feelings, and actions, as well as our mental health, social relationships, and judgment. At all ages, mental wellness is essential, from birth to adolescence to maturity. So, for the data set, we used a data set from Kaggle which was based on emotion [11]. After researching this dataset we remodeled our dataset for binary classification where we declared the anxiety, depression, panic disorder, stress, and bipolar as unstable states indicating 1 and the normal state as 0. Our dataset contains 16000 texts and after dropping duplicated data, it consists of 15893 features.

B. Data Analysis

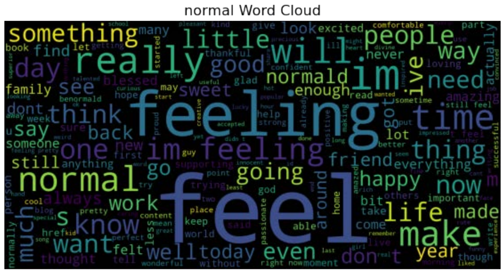
Data visualization is a powerful tool for both exploring complex concepts and communicating knowledge. Many various sorts of approaches, such as graphs, histograms, and matrices, are available for visualizing Sentiment Analysis data. Interactive Maps, Word clouds, and other methods are among the most popular. Here Fig. 3 shows the word cloud from the dataset which is a glimpse of the whole data. Here it respectively represents depression, anxiety, and normal state. Fig. 4 represents the distribution of the mental health status of the whole dataset. Here in our work anxiety, depression, and panic disorder all are depicted as unstable states which are class 1.



(a)



(b)



(c)

Fig. 3. (a)Depression (b)Anxiety (c)Normal Word Cloud.

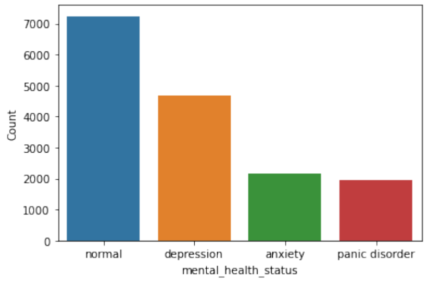


Fig. 4. Distribution of mental health states.

IV. Model Specification

We used five classifier algorithms in our proposed methodology: Logistic Regression, Bernoulli Naive Bayes, K-Nearest Neighbor classifier, Random Forest, & Decision Tree.

A. Natural Language Processing (NLP)

Text mining findings are inputs for NLP, which is one of the most sophisticated approaches in artificial intelligence. The ability of NLP is that humans can communicate words. It is a technique for transforming natural language output (spoken or written) into usable results. NLP is an intriguing task to accomplish since it necessitates both computer and human interaction. NLP is a branch of computer science that examines and comprehends the relationship between computers and human language. Developers can use these tools to create useful tech applications. In NLP, several areas of interest have emerged. The most important tasks in the core areas therefore involve mining named individuals, retrieving expertise through documents, interpreting text messages among language, synthesizing manuscripts, employing advancement in technology to infer answers, and sorting & aggregating papers.Theoretical concepts are frequently discussed in an academic setting. Statistics and dynamic mathematical computations are used in NLP, which is a subfield of data science.[12]

B. Supervised Algorithms

The goal of supervised machine learning algorithms is to anticipate an outcome (dependent variable) based on a set of characteristics (independent variables). They discover how attributes and outcomes are related using labeled datasets linked to anticipate the outcomes of fresh, unlabeled experiments datasets. They do this by modifying incrementally.The performance was excellent. After that, the algorithm is frequently put to the test with a second dataset in which it tries to forecast known outcomes from the qualities that go with them. In general, supervised machine learning algorithms are divided into two categories. There are two types of algorithms: regression algorithms and classification algorithms. Classification algorithms (or classifiers) anticipate categorical outcomes, whereas regression algorithms (or classifiers) tend to predict continuous outcomes (for example, blood pressure or mortality risk) (e.g., positive or negative, benign or malignant). It's worth mentioning that in machine learning, the term "regression" is utilized more distinctly than in classical statistics, where regression can also be used to track outcomes.[13]

C. *Logistic Regression*

Within the field of supervised learning, logistic regression (LR) is one of the machine learning algorithms that is utilized the most frequently. The classified explanatory variables are predicted by utilizing a predetermined set of independent factors. The value of this logistic function is between zero and one.[14]

D. *Bernoulli Navie Bayes*

Bernoulli One variation of the Naive Bayes algorithm used in machine learning is called Naive Bayes. When the dataset has a binary distribution and the output label is either present or absent, it can be used very effectively. This algorithm's key benefit is that it only recognizes features as binary values, such as: False or True: Ham or Spam No or Yes, and 0 or 1.[15]

E. *K-Nearest Neighbour*

The k-nearest neighbors (KNN) method employs the data points that are closest to a given data point in order to ascertain the likelihood that a data point falls into one of two categories. Using k-nearest neighbor, classification and regression issues can be resolved.However, the majority of the time, it is applied to problems with categorization. [16]

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (1)$$

F. *Random Forest*

The supervised studying paradigm is hired by the famous class set of rules Random Forest (RF). It can be applied to classification and regression problems in machine learning. It is premised on supervised methods, a strategy for aggregating different classifiers to tackle a challenging problem and improve the performance of the model.[17]

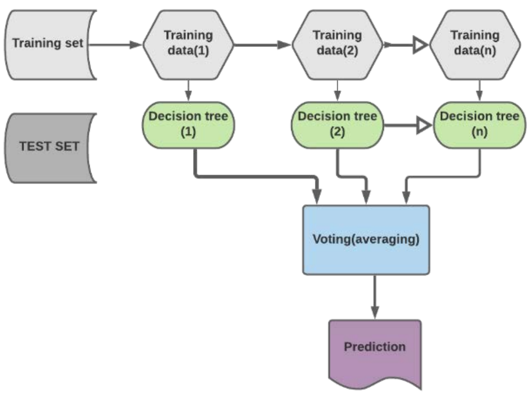


Fig. 5. Block Diagram of Random Forest.

G. *Decision Tree*

Whilst it has the potential to address machine learning problems, the decision tree (DT) is most frequently utilized to address classification challenges. It is a tree-structured extractor, with core nodes describing dataset characteristics, paths denoting prediction model, and node of the tree reflecting the outcome.[18]

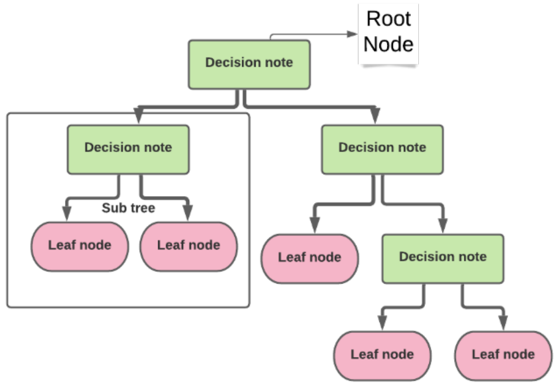


Fig. 6. Block Diagram of Decision Tree.

V. Result Analysis

K-fold (i.e. K=5) cross-validation is used in each experiment to determine the efficacy of the proposed methodology Accuracy, Reliability, and Efficiency are used as performance measures. Each experiment employs precision, recall, and the F-measure [19].

A. *Evaluations*

True Positives (TP) are positive values that are accurately predicted and indicate that the actual and forecast class values are both yes. For instance, if the definite value indicates that he or she has a mental condition, and the expected class indicates that he or she is likewise mentally not good. True Negatives (TN) are negative values that have been correctly predicted, implying that the true class is no, as is the projected class value. As an example, if the genuine result illustrates that his or her mental health status is not okay, we can assume that the predicted class is also the same. Whenever the expected class is yea but the real class is no, it's a False-positive [FP]. For example, the person is mentally not good, but the predicted value indicates that he or she is not. False Negatives occur when the actual class emerges but not the predicted one (FN). For instance, if the true class value is evident, he or she is in good health, whereas the anticipated class value implies that he or she is mentally unstable. After we understand those four characteristics, we could compute Accuracy, Precision, Recall, and F1 score. [19].

For assessing an algorithm's success in classification tasks, accuracy is the metric that is most frequently employed and, in many circumstances, the first option. The right data item to witness proportion is most often referred to as the proportion of successfully predicted data items to the overall number of observations. [19]

$$Accuracy = \frac{TP + TN}{TP+FP+FN+TN} \quad (2)$$

Precision is defined as "the quantity of significant data bits from among a group of data items." To put it another way, what percentages of an algorithm's significantly predicted observations are positive? [19]

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

In recall, the number of significant data items that have been identified is displayed. [19]

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

When calculating an algorithm's performance, the f-score incorporates both precision and recall[19]

$$F1\ Score = \frac{2*(Recall*Precision)}{Recall+Precision} \quad (5)$$

B. Estimation of errors

The mean squared error is accustomed to evaluating how closest a set of information is to the linear regression. To do this, the lengths in between series of points as well as the regression model are squared. The "imperfections" are these lengths. The final squaring is required to remove any undesirable signals and provide more emphasis on significant differences. It is referred to as the mean squared error as you are looking at the aggregate of a series of errors.[19]

It's best if that MSE score is as low as possible. This might be tough to achieve a very small mean squared error.The MSE is commonly referred to as the next phase (from about the premise) of the cost function since it takes into account for both estimator's dispersion and bias. The range of an unbiased estimator is defined as MSE. MSE utilizes the same measurement system since even the squares from the under parameter, which also is equivalent to variance prediction.[19]

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (6)$$

The Root Mean Squared Error conveys how precisely the data fit the given, i.e. the difference between the observed and anticipated data. According to the response variable, this tends to be in the same unit. The smaller the RMSE, the better the projected data matches the actual data. [19]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{obs,i} - x_{model,i})^2}{n}} \quad (7)$$

R2 is a figure that measures how unique variations in one variable might be at several in the other one. R-squared is a simple scale that lies in the range of 1, with 0 signifying that the proposed approach does not strengthen forecasting over through the mean theory and 1 suggesting this has perfect accuracy. The regression model's R-squared is increasing proportionally, showing that it is increasing.[19]

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2)]}} \quad (8)$$

C. Performance Analysis by Classifiers

The classification results for the True Positive (TP), False Positive (FP), False Negative (FN), and True Negative elements of the confusion matrix are shown in Table 1. (TN).Regarding the data in Table II, we computed precision, recall, and f1-score. From Table II, as may be seen the precision value is higher in the LR in a mentally stable state but in Class 1 which means the mentally ill state, RF shows the best precision value. Then for the mentally stable state BernoulliNB (NB), RF, DT, and KNN have the highest value. In (class=1) mentally ill state after RF and LR, DT, NB, and KNN show a high precision value. So for that dilemma, we take a look at the F1 score. Since the F1 score exhibits a balance between recall and precision. Low false positives and false negatives imply a good f1-core, indicating correct detection. We can see in Table 2 that in the f1 score In both states, LR is the standout performer, after NB, RF, DT, and KNN is the lowest in a mentally stable state, and mentally ill state NB, RF, DT, and KNN. But the F1 Score of LR is high in both states. LR has a great F1 score than other sorts of classifiers which is 0.94. Table 2 shows that compared to the other four classifiers, Logistic Regression has a greater accuracy. So here Logistic regression is the best performer.

TABLE I. TRUE POSITIVE (TP), FALSE POSITIVE (FP), FALSE NEGATIVE (FN), AND TRUE NEGATIVE (TN) FROM THE CONFUSION MATRIX

Classifiers	TP	FP	FN	TN
LR	2474	160	130	2031
RF	2098	93	452	2098
NB	2440	253	158	1938
KNN	2149	573	455	1618
DT	2097	164	507	2027

TABLE II. METRICS FOR PERFORMANCE ASSESSMENT (PRECISION, RECALL, F1-SCORE, ACCURACY)

Classifier	**precision**	**recall**	**f1-score**	**Accuracy**
LR	0.93	0.95	0.94	0.93
RF	0.95	0.82	0.88	0.88
NB	0.90	0.93	0.92	0.91
KNN	0.78	0.82	0.80	0.78
DT	0.92	0.80	0.86	0.86

The error analysis, which includes meaning squared error (MSE), root mean squared error (RMSE), and r-squared error, has been shown (r2-score). For the MSE and RMSE, the lower the values, the better the classifier.Table-2 shows that compared to the other four classifiers, Logistic Regression has a greater accuracy. So here Logistic regression is the standout performer. So that to categorize, we used logistic regression as the training model shown in fig 7.

TABLE III. ERROR EVALUATION METRICS (MSE, RMSE, AND R2-SCORE) FOR BOTH TRAINING AND TEST DATA

Classifier	**Train Data**			**Test Data**		
	MSE	***RMSE***	***R2***	***MSE***	***RMSE***	***R2***
LR	0.025116	0.15848	0.8985	0.060479	0.24593	0.75627
RF	0.000983	0.031356	0.99603	0.116579	0.34144	0.5302
NB	0.050947	0.22572	0.79412	0.085714	0.29277	0.65458
KNN	0.116821	0.34179	0.52791	0.214389	0.46302	0.13603
DT	0.000983	0.031356	0.99603	0.146611	0.3829	0.40917

```
text=['life is so unfair' ]
test_result = lr_clf.predict(vectorizer.transform(text))
print(test_result)
```

```
[1]
```

```
text=['I am good' ]
test_result = lr_clf.predict(vectorizer.transform(text))
print(test_result)
```

```
[0]
```

Fig. 7. An evaluation example representation of the best training model.

VI. CONCLUSION

We give a detailed comparison of the five Classifier Algorithms discussed above in our paper. According to the findings of the experiments, Logistic Regression has the highest performance, whereas others specialize in certain scenarios. In the future, we will work with other social site data and other classifier algorithms can also be considered.

REFERENCES

[1] M. Kanakaraj, and R M. R Guddeti, "NLP based sentiment analysis on Twitter data using ensemble classifiers".*In 2015 3Rd international conference on signal processing, communication, and networking (ICSCN)* (pp. 1-5). IEEE 2015, March

[2] N. Glozier, 2017. Effectiveness of eHealth interventions for reducing mental health conditions in employees: A systematic review and metanalysis. *The Technical Writer's Handbook. Mill Valley, CA: University Science*, PloS one, 12(12), p.e0189904M. Young, 1989.

[3] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun," Facebook social media for depression detection in the Thai community". *In 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-6). 2018, July IEEE.

[4] S. Nigam, A. K. Das, and R. Chandra, "Machine Learning Based Approach to Sentiment Analysis."*Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking*, ICACCCN 2018, 157–161, 2018.

[5] S. Mutalib, "Mental health prediction models using machine learning in higher education institution". *Turkish Journal of Computer and Mathematics Education* , *12*(5), pp.1782-1792,2021

[6] M. L. Dobias, M. B. Sugarman, M C. Mullarkey, and J. L. Schleider, "Predicting mental health treatment access among adolescents with elevated depressive symptoms: Machine learning approaches". *Administration and Policy in Mental Health and Mental Health Services Research*, *49*(1), pp.88-103,2022.

[7] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods." *Scientific reports*, *11*(1), pp.1-15,2021.

[8] J.Chung, and J.Teo," Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges."*Applied Computational Intelligence and Soft Computing*, *2022*.

[9] M. Rezapour, and L. Hansen," A machine learning analysis of COVID-19 mental health data". *Scientific reports*, *12*(1), pp.1-16,2022

[10] P. Kumar, R. Chauhan, T. Stephan, A. Shankar, and S. Thakur, "Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection." In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 568-574). IEEE,2021

[11] Emotion_Classification | Kaggle."[online]. Avaliable: https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp? [Accessed: 04-16-2020].

[12] D. H. Maulud, S. R. Zeebaree, K. Jacksi, M. A M. Sadeeq, and K. H. Sharif, "State of art for semantic analysis of natural language processing". *Qubahan Academic Journal*, *1*(2), pp.21-28,2021.

[13] C. J. Harrison, and C. J. Sidey-Gibbons," Machine learning in medicine: a practical introduction to natural language processing". *BMC Medical Research Methodology*, *21*(1), pp.1-11,2021.

[14] J. Han, M. Kamber, and D. Mining, "Concepts and techniques". *Morgan Kaufmann*, *340*, pp.94104-3205,2006.

[15] A. Kharwal. "Bernoulli Naive Bayes in Machine Learning | Aman Kharwal." Bernoulli Naive Bayes in Machine Learning. Jul. 27, 2017. https://thecleverprogrammer.com/2021/07/27/bernoulli-naive-bayes-in-machine-learning/ (accessed: Nov. 14, 2022).

[16] H. Elmunsyah, R Mu'awanah, T. Widiyaningtyas, I. A. Zaeni, and F.A. Dwiyanto," Classification of employee mental health disorder treatment with k-nearest neighbor algorithm". In *2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE)* (Vol. 6, pp. 211-215). IEEE, 2019, October

[17] M. Z. Islam, J. Liu, J. Li, LL. Liu, and W. Kang, "A semantics aware random forest for text classification". In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1061-1070), 2019.

[18] B. Charbuty, and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning". *Journal of Applied Science and Technology Trends*, *2*(01), pp.20-28, 2021.

[19] M. A. Haque, I. J. Dristy, and M. G. R. Alam, "Symptomatic & Non-Symptomatic Hepatocellular Carcinoma Prediction using Machine Learning." In *2021 International Conference on Electrical, Computer, Communications, & Mechatronics Engineering (ICECCME)* (pp. 1-6). IEEE, 2021.