Machine Learning Techniques for Depression Analysis on Social Media- Case Study on Bengali Community

Debasish Bhattacharjee Victor, Jamil Kawsher, Md Shad Labib, Ms. Subhenur Latif
Computer Science and Engineering Discipline, Daffodil International University, Dhaka, Bangladesh

victorbhattacharjee97@gmail.com, jamilkawsher@gmail.com, shadlabibofficial75@gmail.com,
snlatif@daffodilvarsity.edu.bd

Abstract - **Depression is a prevalent illness in todays society. It changes and influences our entire method of thought and our emotional, cognitive, and everyday behavioral behaviors. It affected over 264 million people, and the proportion increases every day. Mainly when it lasts for a prolonged time, it becomes a severe issue or health topic. It leads the trustworthy person to also malfunction, and that person commits suicide in his final position. There are several causes for depression, though social networking like Facebook, Twitter, and other networking plays a critical role in getting us more depressed. Most people in Asia use Facebook, Twitter, and various chat applications, and there they express their emotions. That is why our research initiative picks social media. Some work has been done on depression but depression detection on the Bengali community is done very rarely. So it has become a strong demand for today. The social media has intialted a study based on depression, tweets, and numerous chat app responses, and gathered Bengali data and projected depression posts and commentaries. Diverse approaches of machine learning have been used to evaluate these data and forecast depression and for algorithm purpose Support vector machine, Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes (Multinomial Naive Bayes), Logistic Regression has been used. The desired results can be obtained by adding those algorithms. Moreover, different algorithms send us different results as trends were common, but ultimately the precision was the same for all algorithms applied to our dataset.**

Keywords - *Machine Learning, Super-Vised Learning,* **Support Vector Machine, Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes**

I. INTRODUCTION

Every year, numerous individuals undergo mental disease. Measuring how prevalent mental disorder is significant. Depression steals our confidence, optimism, and motivation, making it hard to take action to make us feel better. According to [14] Dhaka Tribune approximately 17 percent of Bangladesh's adults suffer from the condition where 16.8 percent is the guy and 17 percent are women, and the remainder of the proportion may not take any diagnostic measures. That first survey was held in 2005. Among the total population of Bangladesh, there is a percentage of 16.1% mental disorder in adult in Bangladesh where 15.3% are household members, and it also shows that 6.7% people have a depressive disorder, 4.5% people have anxiety, and 2.1% people have somatic syndrome disorder and where 0.9% people have sleeping wake mental problems. Besides, 0.7% suffer from OCD that means obsessive-compulsive disorder, and 0.3% have neurodevelopment disorder were apart from this, there are 0.3 % of people suffer from neurocognitive disease. Moreover, some people have an addictive disorder where the ratio among different depressive people is 0.2%, and many others suffer from impulse control problems and sexual dysfunction. The situation, though, is very different in rural areas and is less than 16.5% versus a metropolitan region ratio of 18.7%. In comparison to the study, 14 % of children aged 7 to 17 suffer from mental health disorders and 94.5% do not call for psychiatric assistance. Besides, 56% of the people seeking care are left, 33% will treat private psychiatrists, while 23% treat Ayurvedic psychiatrists, homeopathists, and neuropathic psychiatrists. For example, depression may put women at risk of chronic diseases. [16]According to a study, women who experience depression symptoms even without a clinical diagnosis are at an increased risk of developing multiple chronic diseases. According to the IANS news agency, the study, published in the journal American Psychological Association Health Psychology, examined 7,407 middle-aged women from 45 to 50 for more than 20 years. According to IANS news agency. 43.2 percent of women experienced elevated depression symptoms during the study period and just under half of the cohort

were diagnosed with or received depression treatment. Of the total, multiple chronic diseases developed by 2,035 or 63.6 percent. These days many people suffer from multiple chronic diseases such as diabetes, heart disease, stroke, and cancer. Turning to the current epidemic, there is a pandemic on the planet, dubbed Covid19. Mental well-being is not an important problem for most people, but the life is to live with this for this moment. COVID 19 unexpectedly, which cannot be readily observed or visualized, impacts our mental health. Indeed, the unprecedented era of natural disasters of all sorts raises social tension and tears down our state of mind.[15] During this pandemic time, most people are self-confined and can't go anywhere and maintain a social distance. Their children are very young; the case is the same here because they can't go to work, kindergarten, colleges, visit friends or express their feelings, and don't communicate with anybody and too many issues. As a result, countless workers are isolated and the economy rapidly dropped tremendously.[1] According to the Centers for Disease Control and Prevention called CDC, this pandemic triggers more anxiety, concerns about their health, jobs, changes in eating behavior, and sleep schedules, and decreases the attention that affects our minds in general. Often it sounds like isolation. It also causes frustration for this dilemma. To alleviate their loneliness, those involved with multiple social media are often more interested in overcoming their boringness and expressing their feelings with others to receive support and empathy from others. Instead, though, they become more depressed because he or she from whom he or she wishes to obtain this emotion is also in a similar state. [2] Dhaka Tribune estimates that about 11,000 people in Bangladesh die last year. On average, 172 people commit suicide in each district. Many of those who commit suicide are college and school students. For numerous causes, our youth suffer from rage and depression. Such requirements are protested or rejected by suicides. Among the reasons for suicide were several cases due to depression in social media. [17] A real-time example that the Bengali community uses social media at a huge rate in Bangladesh, particularly as Dhaka city. Dhaka has 22 million active Facebook users at the moment. According to a 2017 report jointly conducted by the Social and Hootsuite, this is 1.10 percent of the overall monthly active social networking site users worldwide. Bangladesh ranked third in terms of the 'increasing number of Facebook users' in the same research conducted in 2019. In Bangladesh, about 92 million people have access to the Internet. Of this number, 34 million people are users of social media. In 2019, this is about 20 percent of the total population. As such, from 13 %

of the total population, the number of Facebook users has risen by 7.0%. Psychologists have found out that 'Fear of Losing Out' or FOMO, a phenomenon where consumers of social media are afraid of being removed from the flashy and trending materials on these sites, is induced by social media. This, through social media engagements, creates anxiety among them. The more you use social media, the more likely you are to be influenced by it. Some studies have been conducted on English dataset depression, while a handful of studies have focused on the Bangla dataset. On these papers, authors experimented with data using deep learning and other methodologies and working on over 30 thousand data here with greater precision. This paper has aimed to create a model that can predict depressed or not depressed according to social media dissection. So it becomes a novel work to minimize the ratio of depression rate by reducing the dependency on social media. The machine learning models is applied by which can justify that statistical information by mathematical term. Each person has their own style of communication.Different groups of Bengalis, have our etymological theme. From this point on, the dataset is generated according to these highlights. Custom AI calculations were used for the best choice in conjunction with various highlights. The SVM, Random Forest, Logistic regression, Naïve Bayes, KNN, Decision Tree, and retrieve our data via the grabber script and various APIs.

II. Related Works

Many works were undertaken to cover the problems and to create a research report. This article focuses on the Bengali community depression dependent on Bengali datasets. In [3] authors implemented Naive-Bayes and took the help of Support Vector Machines classifier to identify Messages, showing depression symptoms and mental ill-health. It uses Twitter API to create datasets and obtained 80:20 ratios for data sampling. This data was gathered to train and validate datasets. A total of 10,000 tweets were gathered to produce word-list training and testing results. NLP facilitated the identification of sentimental research. The research word-list is composed of a compiled collection of terms indicating depressive tendencies such as depressed, hopeless, suicide Tweets were selected at random for the research dataset, both supportive and negative elements. Results were measured based on F1 score and precision. Multinomial Naive Bayes precision was 83 percent and was 79 percent in SVM. Text-based emotion AI effectively used Twitter data to diagnose depression. In [4] authors derived data from Twitter streaming, which consists of all recent tweets and classifies

tweets composed of phrases like Stress, Anxiety, and Mental Disorder. Multinomial naïve Bayes and support vector regression approaches were used to identify and separate wellbeing tweets for stress, anxiety from mixed tweets. By deleting the emoticons, the words pause & erase the nostalgic terms the job processed. This full dataset has 3754 posts at an accuracy of Naïve Bayes 76%, K-means clustering 77.17% & SVM 78.8%. In [5] authors suggested a model capture messages, Facebook posts are obtained manually with certain user's permission. Then data sets were interpreted and processed and feed into the machine learning model was utilizing natural language processing. NLP is then used to uncluttered the data which has been obtained. SVM algorithm was applied to look at posts vigorously, and the Naïve algorithm was used to assess posts status over their positivity and negativity or neutral states. From an optimistic and pessimistic post, they considered depression accuracy 74%. In [6] authors placed particular efforts on perfecting handling the tweets. They made an average accuracy on various emotions and from the Unigram model, they got 74% accuracy, 78%, 92% for positive, sad & surprise. Using Recurrent-based methodology, they tried to create a single model in a given language. By utilizing Machine Learning, Deep Learning & Statistical approaches centered on thoughts & emotions, their study concentrated on this. They used Long Short-Term Memory Strong Persistent Network to examine Bengali social media depression. The highest accuracy was 86.3 percent. In [8] A model for Twitter distress identification was developed with emotion analysis by the authors. First, they suggested a form of sentiment analysis using linguistic laws, vocabulary. They then developed a model based on the proposed approach and employed ten features taken from troubled peoples in psychologist observation. They used three forms of classification in their work, and their accuracy is 80 percent. Their model explored user experiences in microblogs, they used emoticons and individual and plural first-person pronouns for the identification of depression. But only 180 Twitter user's data was collected. And their concept works in Chinese. In[9] authors suggested a method for predicting depression via social media. They collect 100 last year's Twitter status for this job, diagnosed with clinical depression. They also collect tweets from 476 Twitter users for training purposes. They noticed behavioral characteristics linked to social interaction, emotion, language, and linguistic types, ego networks by studying depressed people. After that, they construct the classifier (SVM) to predict the probability of being depressed. Their recognized features gave their classifier 70% accuracy. Only English tweets can predict the

process. In[10] authors took an approach to predict suicidal ideas in Chinese micro-blogs using a psychological dictionary lexicon. They examined language characteristics first and then developed a classification based on those characteristics of the algorithms such as Naive Bayes, Logistic Regression, J48, Random Tree, SMO, and SVM. They had 79% precision for suicide diagnosis and 94% full precision. 53 reported suicide users with 30,000 posts and 600,000 posts from 1,000,000 random non-suicidal users. However, 40 percent of suicidal places were misidentified. And their analysis refers only to English entries. In this research [11] authors collected 849 cases from the hospital. Those scores were then used to evaluate four classifiers, including Boosting, Linear Discriminate Analysis, Fisher Linear Discriminant Analysis, and Support Vector Machine (SVM). They were manually scored by psychological standards or questionnaires. They found that it was possible to use the scores they gave the individuals manually to classify suicidal individuals using classifiers.

III. Methodology

In the Bengali dataset, it was comparatively hard to flow as English info and to work hard to build our own dataset, and after making the dataset, labeling our dataset was still really necessary and perform that by a professional psychologist. Various process like pre-processing data, eliminating duplication, post-processing data which are very important part of our entire task.

A. Data Creation:
As our project mission is linked to the Bengali culture, various approaches like Facebook, Twitter and other media are used for data collection. Different Facebook status, messages and tweets from various social sites and tag messages from other television media on other channels are collected to broadcast in Bangladesh for other subjects. To collect data the Facebook Graph API is used to collect texts from Facebook and Twitter API for Twitter and used grabber script. Our main dataset contains 35,000 messages. From analyzing those texts, discovered that most of the data contain pure Bangla along with different emoji characters and uncommon texts only contain some mixed Bangla and English sentences. The data are stored in excel on xlsx type.

*B. Data Tagging***:**
A well-qualified psychologist labeled the dataset using [12] BDI1, [13] BDI2 books along with standards for human social actions and emotions, and theoretical expertise. BDI books include several depression-related questions and each question has a dedicated score from 0 to 3 where the scores are

related to options and she measured each text by considering cumulative questions and analyses the text as depressed or not depressed along with the other indications.

a) Non-depressive post:

অনেকেই অটো লাইক এর কথা শুনেছেন কিন্তু শিখতে পারেন নাই। আজকে আমি একটা লিঙ্ক দিলাম যেখানে লগইন করে প্রোফাইল পিকচার এবং স্ট্যাটসে অটো লাইক দিতে পারবেন তাহলে আর দেরি কেন এখনই নিচের দেয়া লিঙ্কে যান আর প্রোফাইল পিকচার এবং স্ট্যাটসে অটো লাইক দিতে শুরু করেন। '

b) Depressive post:

তদন্ত করে লাভ কী ? কোন দিন কোন তদন্তের পর কিছু হয়েছে বলে কী আপনার মন পরে ? রাজনৈতিক শক্তির সামনে তদন্ত কিছুই করতে পারবে না ! আর তাই আমাদের সামান্য মাথাব্যাথা, পেট ব্যথা হলে বা কলে পরিস্কার পানি না পেলে, সঠিক বিচার না পেলে আমরা আমাদের মাননীয় প্রধামন্ত্রীর দৃস্টি আর্কষন করে থাকি ! কিন্তু সকল দায়িত্ব কী তার একা ? তিনিই বা কেন লাথি দিয়ে এদের বহিস্কার করতে পারে না ?'

C. Data Preprocessing:

After preprocessing, recovered 30508 data and 3178 depressive texts and the majority of them were non-depressive. There were around 5000 redundant data and then extract the unorganized state with so many unique characters, emoji, tab symbol, punctuations. There is no well-decorated Bangla library for stop words, so using Spacy as the default library and making our own Bangla stop words with 354 Bengali stop words and connect to the default Nlp pipelines. There were also icons & pictographs, travel & chart icons. CountVectorizer and TfidfVectorizer are used for the word to an array of vector conversion which is a feature extraction technique.

bn_stop=["টি", "যথেষ্ট", "মোট", "চেয়ে", "চায়", "চান",
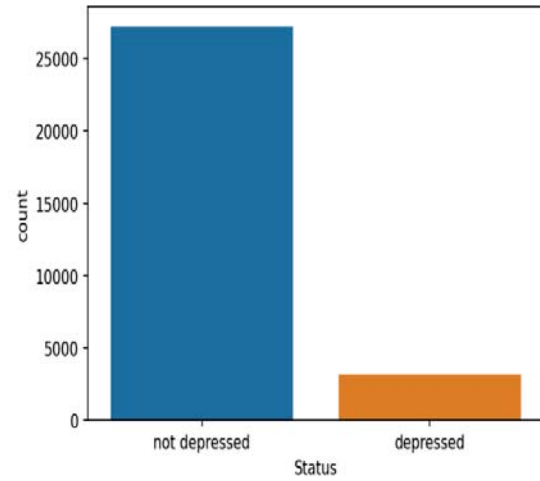bn_stop

Fig 1: Status of Non-depressive and Depressive

D. Data Postprocessing:

The preprocessing steps are applied to the dataset and received accurate data. There were no special characters, single character, emoji punctuation marks, and no Bengali stop words and without English words.
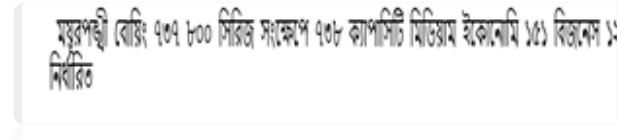
a) **Before processing:**

এগিয়ে যাচ্ছে বাংলাদেশ 👏👏

বরই লজ্জার বিষয় !!! এত বড় চুর দুর্নীতিবাজ কর্মকর্তা একটা দেশের
****

ময়ূরপঙ্খী Brand new sky bird of Bangladesh Biman বোঃ ৭০৭-৮০০ সিরিজ (
বিজনেস ১২) কৃট- শর্ট ইন্টারন্যাশনাল কৃট (গতকাল আসছে মাত্র_ কৃট এবং নির্বাচিত হয়নি)

b) **After Preprocessing:**

এগিয়ে বাংলাদেশ

বরই লজ্জার বিষয় বড় চুর দুর্নীতিবাজ কর্মকর্তা একটা দেশের প্রধান বিমান বন্দরের নিরাপত্তা প্রধান চি হি

ময়ূরপঙ্খী বোয়িং ৭৮৭ ৮০০ সিরিজ সংক্ষেপ ৭৮৮ ক্যাপাসিটি মিডিয়াম ইকোনমি ১৪১ বিজনেস ১৯ নিবারিত

As it is an imbalanced dataset to balance it by the unexpectedly accuracy got reduced by 4 percentages.

E. Working Lifecycle

A structural model approach is used to level the depressed and non-depressed people. This model is trained and tested for our system by splitting the data and apply with six algorithms one by one to ensure the best accuracy after labeling, preprocessing, and cleaning, and stop words removal .Our system features include Decision Tree, Random forest, SVM, Naïve Bayes, K Nearest Neighbors, Logistic Regression.
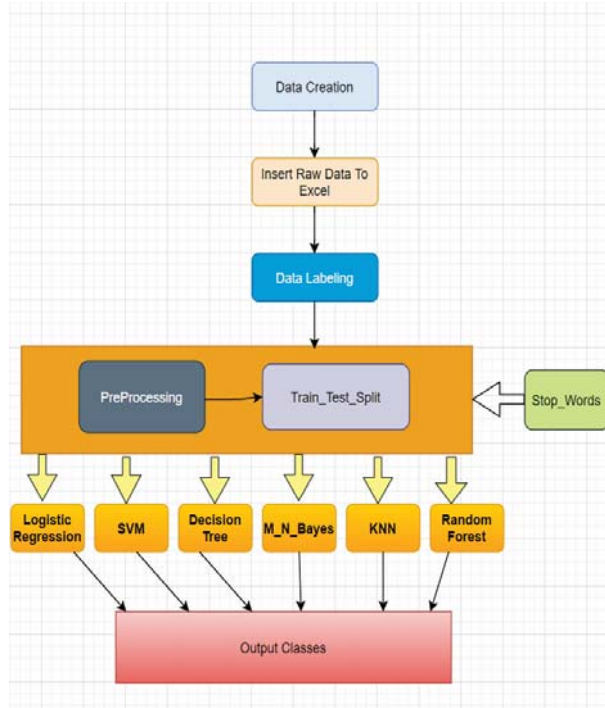

Fig 2: Working Structure

IV. Result and Evaluation

The different algorithms are applied such as SVM, Logistic Regression, Random Forest, K-Nearest Neighbors, Random Forest, Naïve Bayes. To deal with the text data, it is required to reformat the Bangla dataset. The data are analyzed to obtain data that has both English and Bengali parts. So processed that in data preprocessing and apply Tfidf and CountVectorizer for feature extraction technique which converts text to vectors. First, the

CountVectorizer is used as it gives a simple way for both tokenizations of a collection of text documents and builds a vocabulary of known words alongside it encode new documents by using that vocabulary. Further, Tfidf is used as it is also a scoring measurement technique is widely used in summarization and reflect how relevant a term is in that text data. Actually, this is done by multiplying two metrics, to know how many times a word appears in a document and the inverse document frequency of the word across a set of documents.

   a) **True Positive (tp):** It means the number of texts from the test set that is correctly labeled by the classifier as belonging to a particular class or label.

   b) **True Negative (tn**): It denotes the number of text from the test set, correctly labeled by the classifier as not belonging to a particular class or label.

   c) **False Positive (fp):** It signifies the number of text from the test set, incorrectly labeled by the classifier as belonging to a particular class or label.

   d) **False Negative (fn):** It simplifies the number of texts from the test set, that are not labeled by the classifier as belonging to a particular class or label but should have been.

$$tf.idf(Q, D) = \sum_{q_i \in Q} wtf(q_i, D) \cdot idf(q_i)$$

$$wtf(q, D) = \frac{tf(q, D)}{tf(q, D) + 0.5 + 1.5\frac{l(D)}{al}}$$

$$idf(q) = \frac{\log \frac{N}{n_q}}{N + 1}$$

N = number of documents in the corpus

n_q = number of documents in the corpus containing q

$tf(q, D)$ = number of times q appears in D

$l(D)$ = length of D in words

al = avg length in words of a D in the corpus

For performance measurement metrics::

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

$$Precision\ (P) = \frac{tp}{tp + fp}$$

$$Recall\ (R) = \frac{tp}{tp + fn}$$

$$F - measure = \frac{2 * P * R}{P + R}$$

a) *Decision Tree*

Decision tree models can handle categorical variables without one-hot encoding and for such cases, one hot encoding seriously decreases model performance so A decision tree is used whereas a parameter maximum depth as 10 which means it expands until minimum splitting criteria is lesser. And for criteria, the Gini Index is used as it is famous and easier than entropy and use Tfidf and CountVectorizer for feature extraction technique and keeps rest features as default.

Precision= .611
Recall =.51
$F1_{score}$ = .50

b) *Random Forest*

As the random forest is the extension of decision tree and One advantage of decision tree-based methods like random forests is their ability to natively handle categorical predictors without having to first transform them so the random forest is used. The proposed model trained with the hyperparameter using a grid search CV. From setting values as a list of the max depth of 2 to 8, with bootstrap and max features of auto and none and Gini's criterion, entropy passed then to CV and got our best parameter as Gini, max depth:8, and with the allowance of bootstrap. Before fitting, check Out Of Bag Error (OOB) and reset the parameter with n_estimator as 700 as the stable line of error. Then fit the model and get our result as expected.

Precision= .62
Recall =.52
$F1_{score}$ = .51

c) *Support Vector Machine*

In SVM classification, the hyper-surface is needed to be emphasized and as it directly inducts the surface of support vectors, and along with the hyper-surface reflecting with the relation between categorical data and result data is also necessary and as it promotes or increases the effect of performance used this classifier and use C=1.0 and kernel of 'linear' and gamma as 'scale' and keep rest parameters as it is.

Precision= .667

Recall =.5260
$F1_{score}$ = .5269

d) *Multinomial Naïve Bayes*

The data is categorized in nature so Naive Bayes classifier is suitable for classification with specific features that are categorically separated. The categories of each text are drawn from a certain categorical distribution. Here each array gives a certain number of samples that are found for each class and category of the particular feature. In our algorithm, the Alpha dataset as 2 and other values such as fit prior, etc. as default and fit the model after creating a bag of words.

Precision= .658
Recall =.51
$F1_{score}$ = .4964

e) *K-Nearest Neighbors*

KNN is used when matching a point with its most close k neighbors in the multi-dimensional space. It can be used for data that are continuous, discrete, ordinal, and categorical which makes it particularly useful for dealing with all kinds of missing data which are good reasons to use that algorithm. As there is no fixed parameter value for any dataset, the hit and try different nearest neighbors and other parameters and set parameters neighbors as 6 and leaf_size of 40.

Precision= .594
Recall =.5024
$F1_{score}$ = .4807

f) *Logistic Regression*

As it is as similar as linear regression it can manipulate and accommodate continuous and categorical data for investigating the potential effects for different data and for increasing the model performance along with different scoring methods applied it by setting the values as same as default values such as verbose 0, and the random state as 42 and got our result.

Precision= .6944
Recall =.5476

$$F1_{score} = .56203$$

g) *Our Result*

As this dataset is an imbalanced dataset that means the depressive class is lower than the non-depressive class. Most machine learning algorithms work best when the number of samples in each class is about equal. This is because most algorithms are designed to maximize accuracy and reduce errors. However, as the data set is in the imbalanced form than in such cases obtain a pretty high accuracy because by predicting the majority class not at the majority and minority at the same time. For that kind of data precision, recall and f1 score are very good measuring tool so used those terms as they don't compute score on a whole as accuracy score and according to these three techniques, the Logistic Regression gave us the best performance and with that accuracy metric perspective, this classifier is also best.

h) Confusion Matrix, Comparison Graph

This corresponding matrix for LR shows that it correctly predicted non-depressive as 2712 and wrongly predicted 264 and correctly predicted 32 as depressive post and wrongly predicted 35.
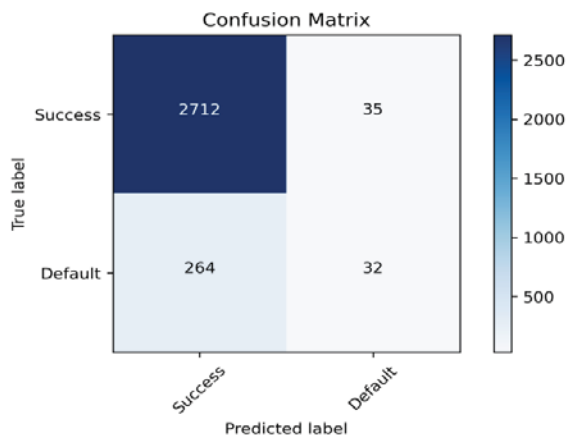

Fig 3: Confusion Matrix

This graph shows the accuracy of different algorithms for CountVectorizer and Tfidf. Here each plot is divided into two subplots for two different feature extraction techniques. On average, all these classifiers provide almost the same accuracy and that was 90 percent.
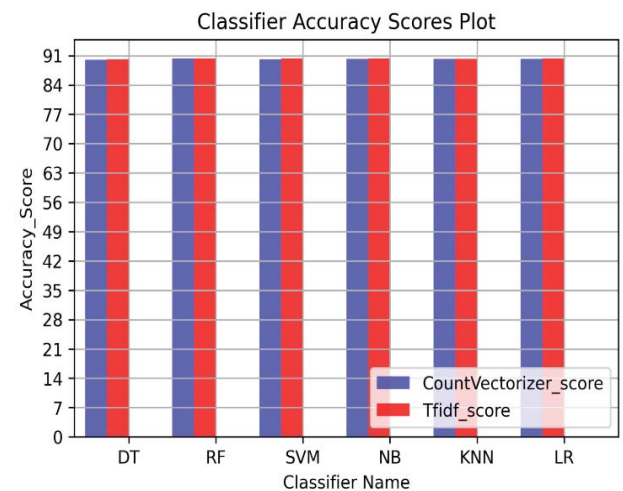

Fig 4: Accuracy Plot

i) Comparison Summary

For Decision tree(DT) our accuracy was 90% for CountVectorizer and 90.1% for Tfidf and Random Forest(RF) accuracy was 90.3% for Count Vectorizer and Tfidf and for Support Vector Machine(SVM) the accuracy was 90.1% and for Tfidf it was 90.3% and Naïve Bayes(NB) and Logistic Regression(LR) the accuracy were same and they were 90.2%,90.3% for CountVectorizer and Tfidf and for K-Nearest-Neighbors(KNN) it was 90.2% both for CountVectorizer and Tfidf.

Table 1:Accuracy Chart

Classifier	Accuracy With CV(%)	Accuracy With Tfidf(%)
DT	90	90.1
RF	90.3	90.3
SVM	90.1	90.3
LR	90.2	90.3
NB	90.2	90.3
KNN	90.2	90.2

V. Conclusion and Future Scope

In this research work, depression on two categories were classified and applied among total of 6 algorithms and every algorithm gives the best accuracy and that was on an average of 90% of test data and 91% for train data. The model is developed and tested with one single feature and that was the Bangla text. A single data based on our mode and model are predicted as depressive as 1 means depressive.

আমরা গরিবরা প্রতিদিন যুদ্ধ করে যাচ্ছি।
আমাদের মতো এ‌তিম অসহায় গরিব মানুষের
কাছে জীবন একটা যুদ্ধ কেন যেখানে প্রতিদিন
যুদ্ধ করে খাবার যোগাতে হয়। যেখানে প্রতিদিন
অত্যাচার জুলুম নির্যাতন সহ্য করতে হয়।
যেখানে প্রতিদিন পুষ্টির অভাবে মারা যায়
কোটি কোটি মানুষ। এখন একটাইই প্রশ্ন দেশে
আর কত লাশ দেখতে হবে? আর কত অত্যাচার
সহ্য করতে হবে? মানুষ কি কখনো ভালো হবে না?
আল্লাহ পাক মানুষকে বিবেক দিয়েছেন তাপড়ও কেন
কুকুরদের মত আচরন করে মানুষ? আজ টেকনোলোজির
যুগ আসছে..আছে শিক্ষিত সব ব্যাক্তি..তারপড়ও কেন
মানুষের বিবেক নেই? নেই কোন ধর্ম..মানবতা কেবল
মানুষের মধ্যাইই থাকে তাহলে কোথায় সেই মানবতা?
আমি নিশ্চিত দেশে দিনে দিনে উন্নতি হবে। হবে ডিজিটাল
বাংলাদেশ। কিন্তু গরিবদের কিছু উন্নতি হবে না।
তারা অতিতে যেমন ছিলো বর্তমানে ঠিক তেমন
থাকবে।মাননিয় প্রধান মন্ত্রি আপনি যতদইই দেশের
উন্নতি করেন কোনদিন এই গরিবদের অভাব পূর্ন করতে
পারবেন না।এবং তাদের অভিসাব থেকে মুক্ত হতে পারবেন না।

```
]:
rra([‘আমরা গরিবরা প্রতিদিন যুদ্ধ করে যাচ্ছি। আমাদের মতো এতিম অসহায় গরিব মানুষের কাছে জিবন
tta)
 ‹
]: array([1], dtype=int64)
```

The data was randomly collected from the Bengali community and can further classified into the class for depression as a major, mild, or minor because of time management. In the future, the proposed work can be extended by collecting more depressive text along with certain gender and work on that. Then it will be easy enough to separate all the three categories of depression is to find a certain group of people as well as which gender is suffering from which type of depression and can also find the intensity of depression. By the extension of this novel work, it is possible to detect the most common symptoms or features of depression disease and can create an alarm for this mental disorder and from the common syndrome of depression. The root causes for that is to find and take early precautionary steps for not spreading out of this disease. And with that extension, it can be applied for any free online counseling service such as chatbot or apps which can tell any gender which depression label he or she is belonging. Furthermore, this work will be helpful for others in depression analysis researches using Deep Learning models or other models with this large dataset and encourage more future works on Bangla social media dataset.

References
[1] Centers for Disease Control and Prevention. 2020. Coronavirus Disease 2019 (COVID-19). [online]Available at: <https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html>

[2] Dhaka Tribune. 2020. Psychologist: 11,000 Suicides Every Year In Bangladesh. [online] Available at: https://www.dhakatribune.com/bangladesh/nation/2019/01/20/psychologist-11-000-suicides-every-year-in-bangladesh

[3] Ieeexplore.ieee.org. 2020. Depression Detection Using Emotion Artificial Intelligence - IEEE Conference Publication. [online] Available at: <https://ieeexplore.ieee.org/document/8389299>

[4] P. Arora and P. Arora, "Mining Twitter Data for Depression Detection," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 186-189

[5] N. A. Asad, M. A. Mahmud Pranto, S. Afreen and M. M. Islam, "Depression Detection by Analyzing Social Media Posts of User," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 13-17

[6] A. H. Uddin, D. Bapery and A. S. M. Arif, "Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2019, pp. 1-4, doi: 10.1109/IC4ME247184.2019.9036528.

[7] Who.int. 2020. Depression. [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/depression>

[8] Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L. and Bao, Z., 2013, April. A depression detection model based on sentiment analysis in micro-blog social network. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 201-213). Springer, Berlin, Heidelberg.

[9] De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E., 2013, June. Predicting depression via social media. In Seventh international AAAI conference on weblogs and social media

[10] Huang, X., Zhang, L., Chiu, D., Liu, T., Li, X. and Zhu, T., 2014, December. Detecting suicidal ideation in Chinese microblogs with psychological lexicons. In 2014 IEEE 11th Intl

Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (pp. 844-849). IEEE.

[11] Delgado-Gomez, D., Blasco-Fontecilla, H., Alegria, A.A., Legido-Gil, T., Artes-Rodriguez, A. and Baca-Garcia, E., 2011. Improving the accuracy of suicide attempter classification. Artificial intelligence in medicine, 52(3), pp.165-168.

[12] Ismanet.org. 2020. [online] Available at: <https://www.ismanet.org/doctoryourspirit/pdfs/Beck-Depression-Inventory-BDI.pdf>

[13] Kdheks.gov. 2020. [online] Available at: <https://www.kdheks.gov/cf/integration_toolkits/Beck_Depression_Inventory_II.pdf>

[14] Dhaka Tribune. 2020. 16.8% Bangladeshi Adults Suffer From Mental Health Issues. [online] Available at: <https://www.dhakatribune.com/bangladesh/dhaka/2019/11/07/survey-nearly-17-of-bangladeshi-adults-suffer-from-mental-health-issues

[15] Longdom.org. 2020. Mental Health And Depression During COVID-19. [online] Available at: <https://www.longdom.org/special-issue/mental-health-and-depression-during-covid19-991.html>

[16] Bangla Tribune. 2020. Depression May Put Women At Risk Of Chronic Diseases | Banglatribune.Com. [online] Available at: <https://en.banglatribune.com/lifestyle/news/50825/Depression-may-put-women-at-risk-of-chronic>

[17] Express, T., 2020. Social Media Or Social Isolation?. [online] The Financial Express. Available at: <https://thefinancialexpress.com.bd/views/social-media-or-social-isolation-1562602219>