DEPRESSIVE AND SUICIDAL TEXT-BASED SENTIMENT ANALYSIS IN BANGLA USING DEEP LEARNING MODELS

Rafiul Hoque*[1], Md. Samiul Islam[2], Sagor Sarkar[2], Sultana Umme Habiba[3],

Md. Moshiur Rahman[3], Md. Rajibul Palas[3], Muhammad Hoque[4]

[1]Green University of Bangladesh,Dhaka,Bangladesh,rafiulhoque21@gmail.com
[2]Green University of Bangladesh,Dhaka,Bangladesh,mdsamiulislam2172@gmail.com,
[2]Green University of Bangladesh,Dhaka,Bangladesh,sagorsarkar2396@gmail.com,
[3]Khulna University of Engineering & Technology, Dhaka, Bangladesh, habiba.kuet@gmail.com,
[3]Bangabandhu Sheikh Mujibur Rahman Digital University, Dhaka, Bangladesh, moshiur0001@bdu.ac.bd,
[3]Green University of Bangladesh, Dhaka, Bangladesh, rajibul@cse.green.edu.bd
[4]Sefako Makgatho Health Sciences University, Pretoria, South Africa, muhammad.hoque@smu.ac.za*

Abstract

This research aims to create and apply efficient sentiment analysis methods for the Bengali language. It also aims to investigate how people in Bangladesh communicate their feelings and mental health issues on social media platforms with a particular emphasis on depression and suicidal thoughts. The process of applying deep learning models to sentiment analysis of suicidal and depressing writing in Bangla entails a few thorough stages. First a dataset of 1076 data points is created by carefully classifying data from a variety of sources including news articles, Facebook, YouTube, and any other online resources into three categories: depressive, non-depressive, and suicidal. Tokenization, stop word removal, and stemming are important preprocessing techniques that help to improve the text. The dataset is split into training and testing sets to train various algorithms. Confusion metrics are used for evaluation and LSTM has the best accuracy (92.01%). This study advances the understanding of sentiment analysis in Bengali by exploring various methodologies and addressing specific challenges in this area. The usefulness of LSTM models is notably highlighted, and it shows that deep learning may be used to achieve accurate sentiment classification. The study compares the simplicity of use of machine learning with the superior performance of deep learning in managing contextual information. The goals are to employ sentiment analysis more widely in interdisciplinary fields and to improve existing methods.

Keywords: Algorithm, deep learning models, depression and suicidal thoughts, sentiment analysis

JEL Classification: M 100

DOI: https://doi.org/10.14311/bit.2024.02.13

Introduction

Depression and suicidal behavior present formidable global health challenges with far-reaching societal and individual implications. According to the World Health Organization (WHO), depression affected over 264 million people in 2020, emerging as the leading cause of disability worldwide [1]. Despite increased awareness of mental health, a significant number of individuals silently endure suffering, with timely intervention proving challenging, particularly in regions where mental health services are scarce. Bangladesh as a densely populated South Asian country serves as an example of such a setting where mental health issues are often underreported and stigmatized [2].

Suicide being the predominant global cause of death, often receives inadequate attention from researchers, healthcare professionals, policymakers, and physicians. Reports indicate an alarming increase in daily suicides, reaching nearly 32 per day in 2019 compared to 29 and 30 in 2015 and 2017, respectively. Bangladesh, with a suicide mortality rate of 39.6 per 100,000, exhibits a unique trend of higher female suicides compared to most Asian nations [3].

Various risk factors contribute to suicidal behavior, encompassing aspects like young age, low education, student status, nuclear families, family history of suicide, substance use, workplace issues, financial hardship, affairs, domestic abuse, divorce, and physical illness. Social media platforms such as Facebook and Twitter have become outlets for discussing suicide risk factors, particularly among teenagers and young adults who may not disclose such concerns to healthcare professionals [4].

Sentiment analysis in Bangla is vital for understanding human emotions, monitoring mental health, preventing suicide, tailoring healthcare, and promoting emotional content regulation. It also aids in public health research, educational assessment, and language learning [5, 6]. Therefore, the objective of this study was to develop effective Bengali language profiling techniques and identify suitable sentiment analysis methods for practical applications in social media monitoring brand perception analysis and market research.

Literature Review

The analysis of Bangla text on depression and suicidal tendencies is crucial for global mental health, but challenges like limited datasets and unique linguistic characteristics hinder its processing. The analysis of Bengali depressive and suicidal expressions using methods like lexicon-based collections, deep learning models, and machine learning algorithms is crucial for public health monitoring and intervention efforts [7]. For example, a lexicon-based collection of 61,582 Bangla words that can be used for sentiment analysis, emotion identification, and sentiment mining in this language. To address the special difficulties the Bangla language poses in digital situations, they represent the union of linguistic expertise with computational techniques. To enable researchers to use it with the same coding structure, the dataset is formally designated as the English Sent WordNet dataset [8].

"Datasets for Aspect-Based Sentiment Analysis in Bangla and Its Baseline Evaluation" addresses the scarcity of benchmark datasets for Aspect-Based Sentiment Analysis (ABSA) in the Bangla language. It contributes two datasets focusing on cricket and restaurant domains facilitating aspect category extraction and polarity identification. The paper's strengths include meticulous dataset creation involving participant collaboration and diverse categories providing a valuable resource for sentiment analysis. However, challenges arise due to the subjective nature of sentiment expression leading to reduced precision and recall in the baseline results. Nevertheless, this work lays a crucial foundation for further research in Bangla ABSA [9].

Sentiment Analysis of Bangla Language Using Deep Learning Approaches analyzes the method that is using a hybrid CNN-LSTM model to improve the accuracy. It discusses the significance of sentiment analysis, reviews related work and employs various word embedding techniques. The CNN-LSTM hybrid, particularly with Word2Vec, enhances accuracy and F1 scores but acknowledges limitations like an imbalanced dataset and a narrow emotion focus. This study presents a valuable Bangla sentiment analysis framework with potential for enhancement in dataset balance and emotion scope [10].

The study utilizes deep learning techniques to analyze Bangla sentiment in Bengali sports news comments. It employs CNN, LSTM, and DNN for sentiment classification with the CNN model showing strength in both 5-class and 2-class scenarios. The evaluation extends to Hindi datasets, where the CNN model performs well. Despite promising results and insights, the paper acknowledges limitations due to the small dataset size [11].

The research uses LSTM-based RNNs for sentiment analysis of Bangla cricket-related text for achieving 95% accuracy. Notable strengths include meticulous dataset preparation, word embedding for vectorization and effective LSTM implementation for capturing long-term dependencies. However, potential areas for improvement such as advanced preprocessing and expanding target classes for higher accuracy are noted [12].

Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language emphasizing the growing importance of e-commerce in Bangladesh. It utilizes a substantial dataset of 1020 Bangla product reviews and deploys various machine learning algorithms, with SVM achieving an impressive 88.81% accuracy. However, the paper lacks a detailed explanation of the NLP techniques and preprocessing methods used and provides limited insight into the practical implementation of its proposed future work [13].

Methodology details

To conduct sentiment analysis on depressive and suicidal text in Bangla using Deep Learning Models and employ a comprehensive methodological approach. Initially we perform data processing by assembling a dataset that includes both depressive, suicidal, and non-depressive Bangla text. This dataset undergoes preprocessing tasks, including stemming, stop word removal, tokenization, and feature extraction [8, 9]. Subsequently we eliminate common non-informative words through stop-word removal. For feature extraction, we transform the text data into numerical features using techniques such as TF-IDF or word embeddings, enabling the model to comprehend the textual content [10]. The dataset is then partitioned into training and following this training the model undergoes evaluation on the test set with confusion metrics employed to measure the ratio of true positives to total positive predictions. The overall process of our proposed methodology is shown in Figure I.

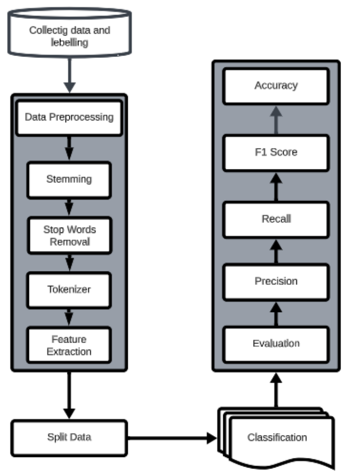

Figure I: Proposed Methodology

A. Data Collection and labeling:

We have gathered data from diverse sources, including Facebook, YouTube, and news articles. We meticulously categorized this data into three classes: depressive, non-depressive, and suicidal. The dataset encompasses 454 data points labeled as depressive, 501 as non-depressive, and 121 as suicidal, resulting in a total of 1076 data points. To assign labels to the dataset, we conducted a survey among our team members gathering feedback and votes on whether each data point should be labeled as depressive, non-depressive or suicidal. The labels were then assigned based on the majority votes obtained from the survey.

B. Data Prepossessing:

Bangla language prepossessing involves a series of essential steps to refine and prepare textual data for analysis. The prepossessing steps are detailed in the following sections. Table 1 illustrates the preprocessing steps for the Bengali language accompanied by an example.

1.Bangla tokenizer:

Bangla tokenizer is an NLP tool designed to break Bengali (Bangla) text into words or tokens, a fundamental step in NLP tasks like text analysis and language modeling. In Bengali, where words lack spaces, tokenization identifies word boundaries, separating words, punctuation, and other elements. This process is vital for segmenting text into meaningful units [13].

Here's an example of Bangla text and its tokenization:

Original Text (Bengali):bn "আমি বাংলায় গান গাই।"

Tokenized Text: ["আমি", "বাংলায়", "গান", "গাই", "।"]

In this example, the Bangla tokenizer has separated the text into individual words and also recognized the punctuation at the end.

2.Stop Removal Words:

In the Bengali (Bangla) language, stop words are common words that are often filtered out or removed from text data during natural language processing (NLP) and text analysis tasks. These words are of little value in text analysis because they appear frequently in most texts and do not carry significant meaning on their own. Removing Bengali stop words can help reduce the dimensionality of text data and improve the efficiency and focus of NLP algorithms.

3.Bangla stemmer:

A Bengali (Bangla) stemmer is a natural language processing (NLP) tool designed to reduce Bengali words to their root or base form by removing suffixes. Stemming is a common preprocessing step in text analysis and information retrieval tasks. It helps in reducing the dimensionality of text data and improving the efficiency of NLP algorithms by converting different inflected forms of a word into a common base form. The Bengali language has a rich system of inflections and suffixes that can change the form of words. A Bengali stemmer considers these linguistic variations and simplifies words to their root form.

4.Feature extraction:

Feature extraction is a crucial machine learning step where relevant information is selected or transformed from raw data to create input features for the algorithm. Effective feature extraction enhances model performance by reducing dimensionality, capturing patterns, and improving generalization. In this work we utilize the Tfidf Vectorizer tool, a popular text preprocessing and feature extraction tool in Python's scikit-learn library. Tfidf Vectorizer converts raw text documents into a matrix of TF-IDF features a numerical statistic reflecting word importance within a document relative to a corpus.

C. Split Data:

Supervised machine learning is about creating models that precisely map the given inputs (independent variables, or predictors) to the given outputs (dependent variables, or responses). The training set is applied to train, or fit, your model. For example, you use the training set to find the optimal weights, or coefficients, for linear regression, logistic regression, or neural networks. The test set is needed for an unbiased evaluation of the final model. You shouldn't use it for fitting or validation.

D. Train with Model:

So far, we have trained our data set in various models, including Random Forest, Decision Tree, Naive Bayes, Linear SVM Model, LSTM, MLP Classifier, Logistic Regression, Bagging Classifier, KNeighbors Classifier, Gradient Boosting Classifier, SGD Classifier, Passive Aggressive Classifier. Among these models and classifiers, we got the most accuracy in LSMT (Accuracy: 92.01%). LSTM-based sentiment analysis models require a substantial amount of labeled data for training and a careful selection of hyperparameters for optimal performance.

Tale 1: Data preprocess steps with raw data example:

Method	Data	Comment
Text	প্রতি নি:শ্বাসে মৃত্যুর দিন গুনছি।	Raw Data
Tokenizer	["প্রতি", "নি", ":", "শ্বাসে","মৃত্যুর", "দিন", "গুনছি"]	Tokenized Text
Stemming	["প্রতি", "নি" ":" "শ্বাস","মৃত্যু", "দিন", "গুনছি"]	Stemming Word
Feature Extraction	[[0. 0. 0. ... 0. 0. 0.].. ..[0. 0. 0. ... 0. 0. 0.]	Feature Extraction using TF_DF

Experimental Result

In our experiment, we sought to analyze depressive and suicidal texts, which often contain subtle cues and context-dependent sentiment. We employed a range of machine learning and deep learning models to address this challenging task, including Random Forest, Decision Tree, Linear SVM, LSTM, MLP Classifier, Logistic Regression, Bagging Classifier, KNeighbors Classifier, Gradient Boosting Classifier, SGD Classifier, Passive Aggressive Classifier, and KNN. Each of these models was tested and evaluated for its performance in sentiment analysis [8-10]. The choice of models aimed to cover a broad spectrum of machine learning techniques to determine which one could effectively capture the nuances in depressive and suicidal texts. We considered metrics such as accuracy, precision, recall, and F1-score to assess the models' abilities to identify and understand the complex sentiment within these texts. The experiment aimed to identify the most effective model for analyzing subtle cues and context-dependent sentiment in depressive and suicidal texts that is aiming to improve mental health support. The experimental results on confusion metrics using machine learning algorithms are presented here:

1.Random forest:

Random forest is a supervised Machine Learning algorithm. This algorithm creates a set of decision trees from a few randomly selected subsets of the training set and picks predictions from each tree. It's a suitable choice because it can handle both numerical and categorical features and can capture complex relationships in the data for sentiment analysis [14].

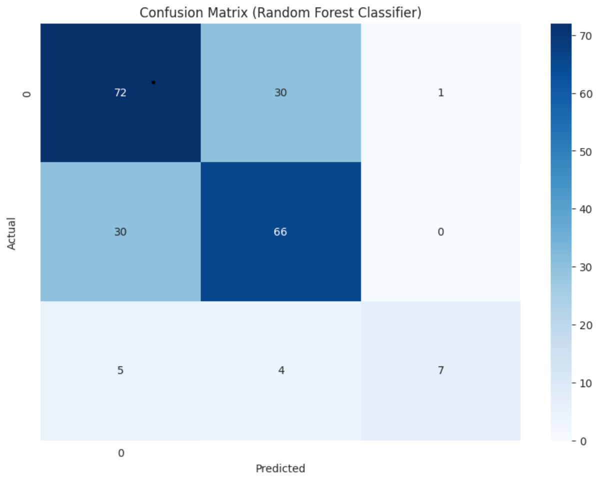

Figure 2:Random forest

2.L STMs (Long Short-Term Memory):

LSTM is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem in traditional RNNs. It is particularly well-suited for sequential data tasks, such as time series forecasting, natural language processing, text based and speech recognition. In sentiment analysis, LSTM networks are employed to capture the temporal dependencies in language, allowing them to recognize sentiment shifts, negations, and context within text data. By processing text sequentially and retaining information over extended sequences, LSTMs can effectively discern the sentiment expressed in longer sentences or documents.

LSTMs, or Long Short-Term Memory networks, are powerful for analyzing depressive and suicidal sentiment in Bangla text. They excel at capturing sequential patterns in data, which is crucial for dealing with the complex structures and dependencies in languages like Bangla. LSTMs can handle variable text lengths, making them versatile for short social media posts and lengthy narratives. This flexibility enables them to capture both short-term fluctuations and long-term sentiment trends in depressive and suicidal texts, which often contain subtle cues. Their ability to maintain context through hidden states helps understand the emotional state of the text. This is beneficial when dealing with languages like Bangla, where linguistic features may be less standardized or documented [15].

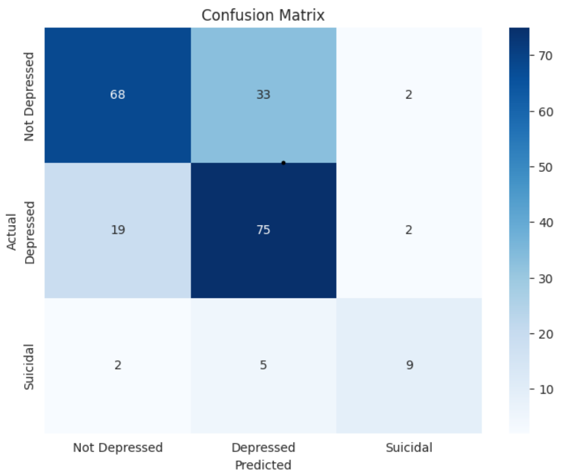

Figure 3: LSTMs (Long Short-Term Memory)

3.Support Vector Machine (SVM):

Support Vector Machine is a powerful and versatile supervised machine learning algorithm used for classification and regression tasks. SVM works by finding a hyperplane in a high-dimensional space that best separates data points into different classes while maximizing the margin between these classes. 16].

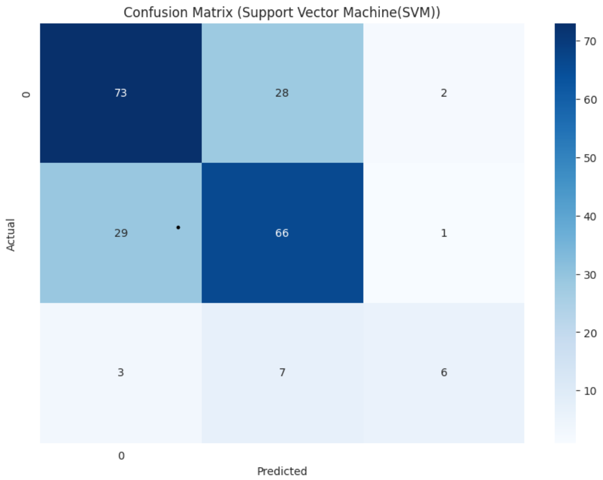

Figure 4: Support Vector Machine (SVM)

4.Naïve Bayes:

The Naïve Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification. In sentiment analysis, Naive Bayes is used to classify text sentiment. The approach assumes features (words) are independent given the sentiment. It calculates the probability of a text belonging to each sentiment class based on word frequencies. Then it assigns the class with the highest probability [17].

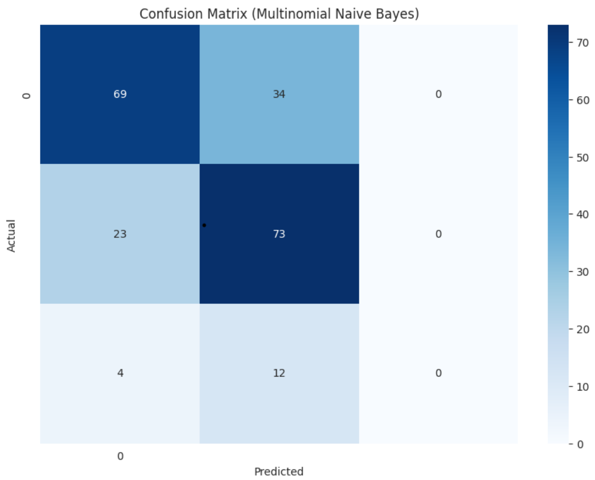

Figure 5: Naïve Bayes

5.K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a machine learning algorithm that can be applied to sentiment analysis tasks particularly in scenarios where you want to classify text data based on its similarity to other labeled examples. In sentiment analysis, the KNN Classifier assigns sentiment labels to a given text document by considering the sentiments of its nearest neighbors in a feature space. In this study, setting k=5 in KNN means that when you want to make a prediction for a new data point, consider the five closest neighbors in the training dataset to that data point to determine its class or value [18].

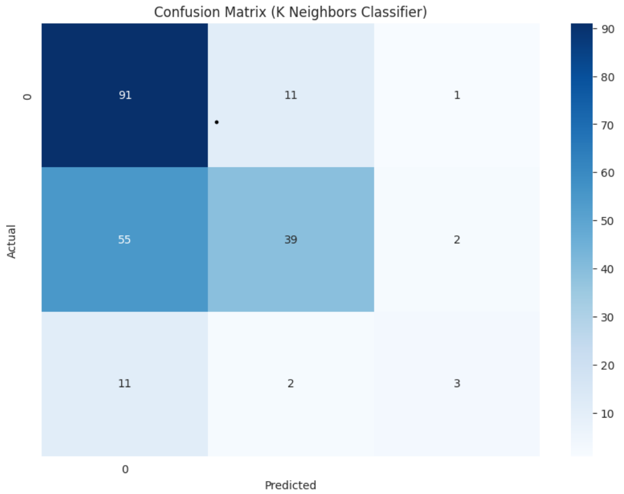

Figure 6:K-Nearest Neighbors (KNN)

6.Decision Tree:

Decision Trees are versatile machine learning models used in various tasks, including sentiment analysis. They can be applied effectively to classify text data based on sentiment. In sentiment analysis, a Decision Tree represents a sequence of binary decisions made based on the features of a text document. These features are typically derived from the text, such as word frequencies or presence/absence of specific words. The Decision Tree recursively splits the data into subsets based on these features which aim to separate classes of sentiments [19].

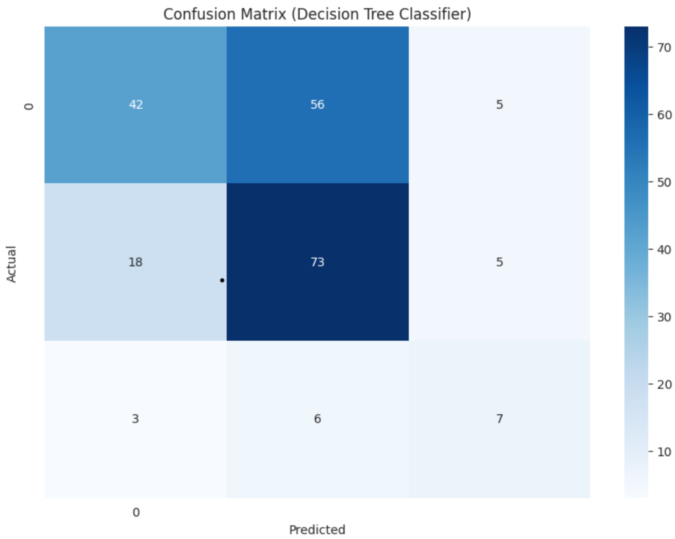

Figure 7: Decision Tree

7.MLP (Multi-Layer Perceptron):

The MLP (Multi-Layer Perceptron) Classifier is a type of neural network used extensively in sentiment analysis and various other machine learning tasks. It belongs to a class of feed-forward neural networks having various layers of perceptions. MLPs also have connected input and output layers, and their number is the same. MLP Classifiers are powerful tools for sentiment analysis, particularly when handling complex textual data with nuanced sentiment expressions. They are widely used in modern natural language processing applications, often achieving state-of-the-art results in sentiment classification tasks when combined with appropriate data preprocessing and regularization techniques [20].

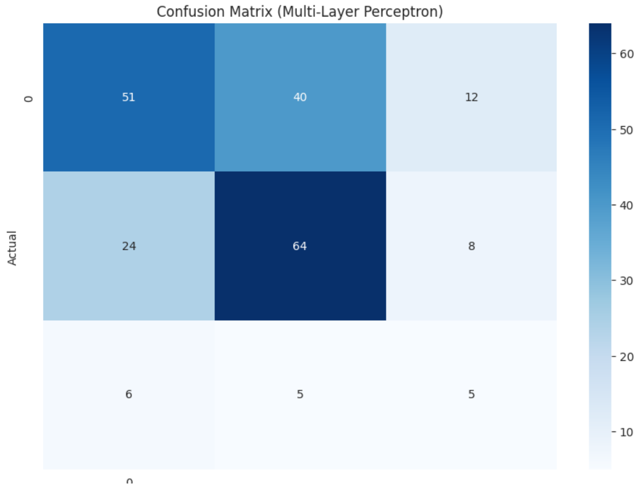

Figure 8: MLP (Multi-Layer Perceptron)

8.Logistic Regression:

Logistic Regression is a fundamental and widely used machine learning algorithm that can be effectively applied to sentiment analysis tasks, especially binary sentiment classification. In sentiment analysis, Logistic Regression models predict the probability that a given text document belongs to a particular sentiment class. Logistic Regression is a valuable choice for binary sentiment classification tasks in sentiment analysis, particularly when interpretability and computational efficiency are essential. For more complex tasks, deep learning models or ensemble methods may be preferred to capture subtle nuances in language and context [21].

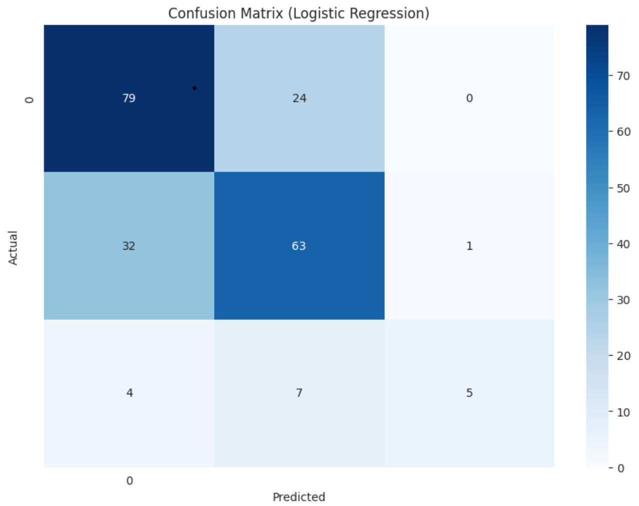

Figure 9: Logistic Regression

9.Bagging:

Bagging, which stands for Bootstrap Aggregating, is an ensemble learning technique, and the Bagging Classifier is a common implementation of this method. Bagging is used to improve the accuracy and robustness of machine learning models and it can be effectively applied in sentiment analysis tasks. The Bagging Classifier can enhance the performance of base classifiers in sentiment analysis by combining their predictions [22].

Figure 10: Bagging

10.Gradient Boosting Classifier:

Gradient Boosting Classifier is a powerful ensemble machine learning algorithm commonly used in sentiment analysis to improve the accuracy of sentiment predictions especially in tasks involving binary or multi-class sentiment classification. The Gradient Boosting Classifier is a sentiment analysis tool that employs a sequential combination of multiple weak learners typically decision trees to create a robust predictive model [23].

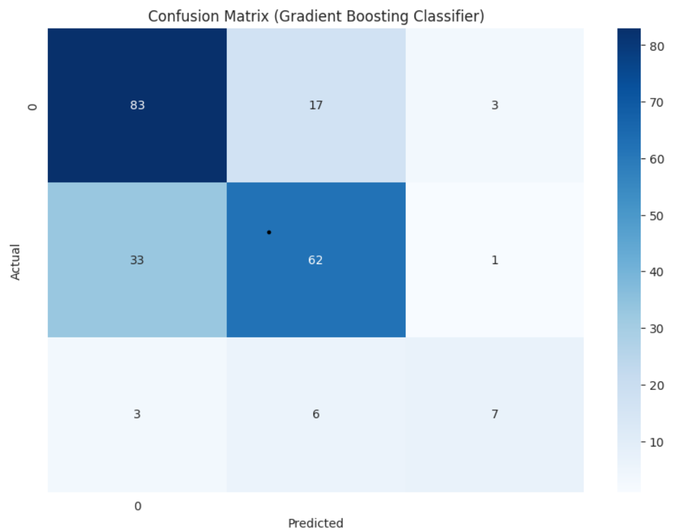

Figure 11: Gradient Boosting Classifier

11.The Stochastic Gradient Descent (SGD):

The Stochastic Gradient Descent (SGD) Classifier is a machine learning algorithm that can be applied to sentiment analysis tasks especially when efficiency and scalability are essential considerations. In sentiment analysis, the SGD Classifier is used to classify text data into sentiment categories such as depressive, non-depressive or suicidal [24].

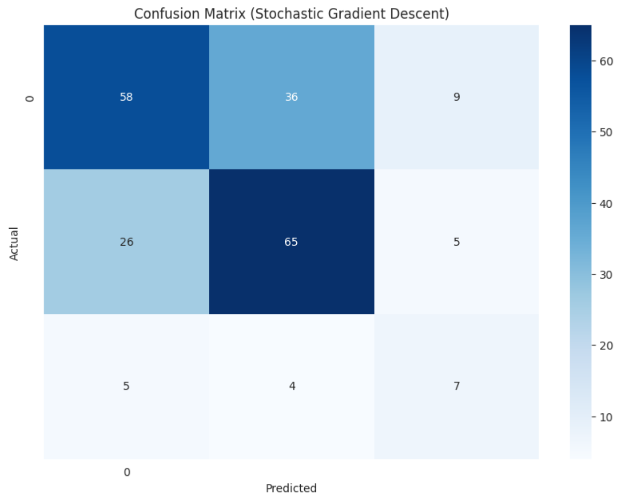

Figure 12: The Stochastic Gradient Descent (SGD)

12.Passive Aggressive (PA) Classifier:

The Passive Aggressive (PA) Classifier is a machine learning algorithm that can be effectively applied to sentiment analysis tasks particularly when dealing with data streams or situations where the model needs to adapt to concept drift (changes in data patterns over time). In sentiment analysis, the Passive Aggressive Classifier is used to classify text data into sentiment categories such as depressive, non-depressive and suicidal [25].

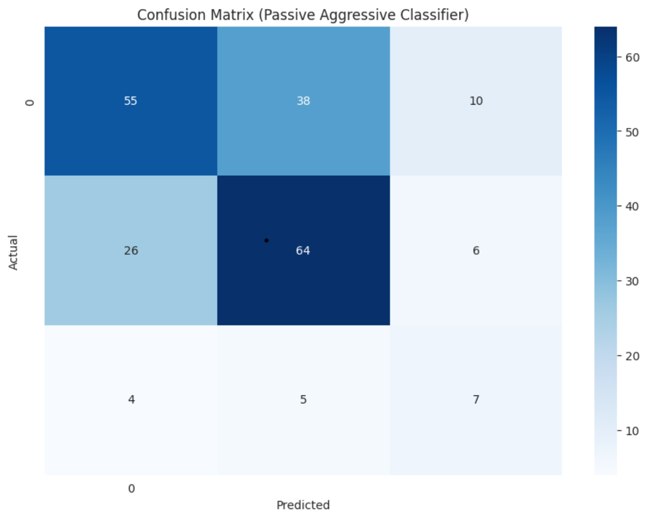

Figure 13: Passive Aggressive (PA) Classifier

Accuracy:

Accuracy is a metric used to evaluate the performance of a classification model. It measures the proportion of correctly classified instances out of the total instances. After implementing all those machine learning algorithms, we got more accuracy at the LSTM model. Here is the graph of accuracy:

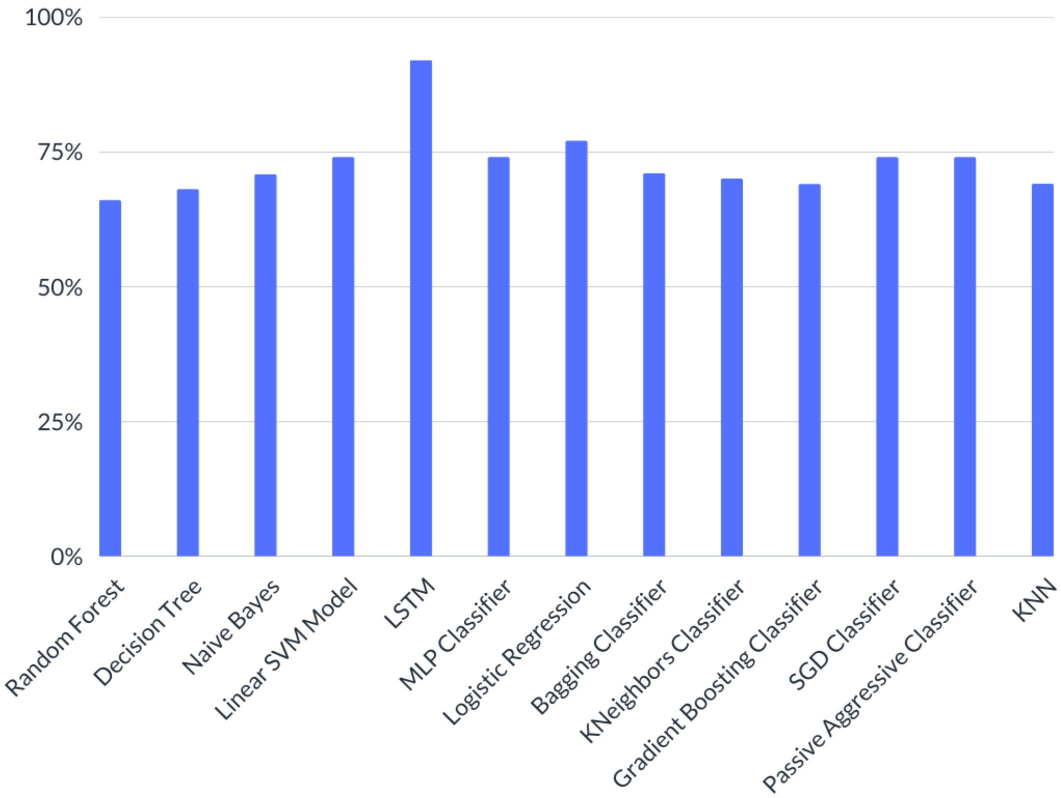

Figure 14: Performances of different models on our collected data.

Discussion and Conclusion

The research on depressive and suicidal text-based sentiment analysis in Bangla using deep learning models marks a significant advancement over existing literature. Previous studies largely relied on lexicon-based methods [8], such as the adapted Bangla SentiWord dataset while foundational were constrained by the static nature and comprehensiveness of word lists. This study, however, employs a hybrid CNN-LSTM model [10] demonstrating its effectiveness despite limitations like an imbalanced dataset and a narrow focus on emotions. Additionally, LSTM-based RNNs [12] have shown high accuracy in analyzing Bangla cricket-related text effectively capturing long-term dependencies but only distinguishing between depressive and non-depressive levels. Aspect-Based Sentiment Analysis and sentiment analysis in domains like Bengali sports news and product reviews have similarly utilized binary classifications of depressive and non-depressive data [9]. In contrast this research introduces a three-level dataset (depressive, non-depressive, suicidal) achieving an impressive 92.01% accuracy with the LSTM model. This represents a substantial improvement over previous studies which often struggled with dataset imbalances and binary classifications. In addition, we integrated advanced preprocessing techniques tailored to the unique characteristics of the Bangla language. This includes the development of effective Bangla language preprocessing methods which are critical for the accurate analysis of sentiment in Bangla text. These preprocessing steps ensure that the proposed model can effectively manage the exactness and complexities inherent in Bangla leading to more appropriate sentiment detection. The comprehensive evaluation of various models including Random Forest, Decision Tree, Linear SVM, MLP Classifier, Logistic Regression, Bagging Classifier, KNeighbors Classifier, Gradient Boosting Classifier, SGD Classifier, Passive Aggressive Classifier, and KNN that are underscores the robustness of the research highlighting the strengths and weaknesses of each model in sentiment analysis.

In conclusion, this study significantly contributes to understanding sentiment analysis in the Bengali language through a thorough examination of various methodologies. The findings underscore the specific challenges in Bengali sentiment analysis and propose effective solutions. The research demonstrates the feasibility of accurately classifying sentiment in texts using deep learning emphasizing the enhancement achieved by incorporating contextual information. The rapidly growing field of sentiment analysis is driven by various algorithms which plays a crucial role in meeting user expectations and facilitating necessary adjustments based on user reactions. The study compares the simplicity and ease of incorporating machine learning techniques with the complexity and superior performance of deep learning and a combination of both. Deep learning methods particularly those using LSTM that prove the effectiveness in sentiment analysis. LSTM is a deep learning method for sentiment analysis, but it lacks the ability to consider multiple objects with different sentiment scores in an image and its contextual information. The study highlights areas for improvement and suggests future enhancements, while also suggesting collaboration with psychology to develop more accurate sentiment analysis models. The ultimate objective is to enhance the efficiency of existing methods and improve the integration of sentiment analysis techniques for a broader range of interdisciplinary applications.

References

[1] Abdel-Bakky, M. S., Amin, E., Faris, T. M., & Abdellatif, A. A. H. (2021). Mental depression: Relation to different disease status, newer treatments and its association with COVID-19 pandemic. *Molecular Medicine Reports, 24*(6), 1-17. https://doi.org/10.3892/mmr.2021.12430

[2] Gary, F. (2006). Stigma: Barrier to mental health care among ethnic minorities. *Issues in Mental Health Nursing, 26*(10), 979-999. https://doi.org/10.1080/01612840500280638

[3] Ferdous, M. Z., & Alam, A. S. M. M. (2021). Present situation of suicide in Bangladesh: A review. *MedRxiv, 2021-02*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/2021.02.23.21252279

[4] Arafat, S. M. Y., Mali, B., & Akter, H. (2018). Demography and risk factors of suicidal behavior in Bangladesh: A retrospective online news content analysis. *Asian Journal of Psychiatry, 36*, 96-99. https://doi.org/10.1016/j.ajp.2018.07.005

[5] Rajesh Kumar, E., Rama Rao, K. V. S. N., Nayak, S. R., & Chandra, R. (2020). Suicidal ideation prediction in Twitter data using machine learning techniques. *Journal of Interdisciplinary Mathematics, 23*(1), 117-125. https://doi.org/10.1080/09720502.2020.1802097

[6] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering, 23*(5), 649-685. https://doi.org/10.1017/S1351324917000157

[7] Al-mashhadani, M., Hussein, K., Khudair, E., & Ilyas, M. (2022). Sentiment analysis using optimized feature sets in different Facebook/Twitter dataset domains using big data. *Iraqi Journal for Computer Science and Mathematics, 3*, 64-70. https://doi.org/10.52866/ijcsm.2022.01.01.007

[8] Ali, H., Hossain, M. F., Shuvo, S., & Marouf, A. (2020). BanglaSenti: A Dataset of Bangla Words for Sentiment Analysis. In IEEE (Eds.), *Proceedings of the International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-17). https://doi.org/10.1109/ICCCNT49239.2020.9225565

[9] Rahman, M. A., & Kumar Dey, E. (2018). Datasets for aspect-based sentiment analysis in Bangla and its baseline evaluation. *Data, 3*(2), 15. https://doi.org/10.3390/data3020015

[10] Hoq, M., Haque, P., & Uddin, M. N. (2021). Sentiment analysis of Bangla language using deep learning approaches. In *International Conference on Computing Science, Communication and Security* (pp. 140-151). Springer. https://doi.org/10.1007/978-3-030-70904-7_13

[11] Rahman, M., Haque, S., & Saurav, Z. R. (2020). Identifying and categorizing opinions expressed in Bangla sentences using deep learning technique. *International Journal of Computer Applications, 176*(17), 13-17. https://doi.org/10.5120/ijca2020920128

[12] Wahid, M. F., Hasan, M. J., & Alom, M. S. (2019). Cricket sentiment analysis from Bangla text using recurrent neural network with long short-term memory model. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICBSLP47725.2019.201377

[13] Shafin, M. A., Hasan, M. M., Alam, M. R., Mithu, M. A., Nur, A. U., & Faruk, M. O. (2020). Product review sentiment analysis by using NLP and machine learning in Bangla language. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCIT51783.2020.9392716

[14] Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data, 10*(1), 5. https://doi.org/10.1186/s40537-023-00632-9

[15] Fauzi, M. A. (2018). Random forest approach for sentiment analysis in Indonesian. *Indonesian Journal of Electrical Engineering and Computer Science, 12*, 46-50. https://doi.org/10.11591/ijeecs.v12.i1.pp46-50

[16] GeeksforGeeks. (n.d.). Deep learning - Introduction to long short term memory. Available online at: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/ (accessed 25 July 2023)

[17] GeeksforGeeks. (n.d.). Support vector machine algorithm. Available online at: https://www.geeksforgeeks.org/support-vector-machine-algorithm/ (accessed August 12, 2023)

[18] GeeksforGeeks. (n.d.). Naive Bayes classifiers. Available online at: https://www.geeksforgeeks.org/naive-bayes-classifiers/ (accessed 25 August 2023)

[19] Javatpoint. (n.d.). K-Nearest Neighbor (KNN) algorithm for machine learning. Available online at: https://www.javatpoint.com/machine-learning-algorithms (accessed 30 September 2023)

[20] Guia, M., Silva, R., & Bernardino, J. (2019). Comparison of Naïve Bayes, support vector machine, decision trees and random forest on sentiment analysis. In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 1:* pp. 525-531). https://doi.org/10.5220/0008364105250531

[21] Qiu, N., Shen, Z., Hu, X., & Wang, P. (2019). A novel sentiment classification model based on online learning. *Journal of Algorithms & Computational Technology, 13*. https://doi.org/10.1177/1748302619845764

[22] Tyagi, A., & Sharma, N. (2018). Sentiment analysis using logistic regression and effective word score heuristic. *International Journal of Engineering and Technology (UAE), 7*(2.24), 20-23. https://doi.org/10.14419/ijet.v7i2.24.11991

[23] Sawarn, A., Ankit, & Gupta, M. (2020). Comparative analysis of bagging and boosting algorithms for sentiment analysis. *Procedia Computer Science, 173*, 210-215. https://doi.org/10.1016/j.procs.2020.06.025

[24] Athanasiou, V., & Maragoudakis, M. (2017). A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: A case study for modern Greek. *Algorithms, 10*(1), 34. https://doi.org/10.3390/a10010034

[25] Vidyashree, K. P., & Rajendra, A. (2023). An improvised sentiment analysis model on Twitter data using stochastic gradient descent (SGD) optimization algorithm in stochastic gate neural network (SGNN). *SN Computer Science, 4*. https://doi.org/10.1007/s42979-022-01607-x

[26] Gulati, K., Kumar, S., Boddu, R., Sarvakar, K., Sharma, D., & Nomani, M. (2022). Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic. *Materials Today: Proceedings, 51*, 38-41. https://doi.org/10.1016/j.matpr.2021.04.364