A Comprehensive study of Emotion Detection in Bangla Natural Language Processing

Rakin Bin Rabbani 1, Sumaiya Tanjil Khan 2, Prasenjit Das 3, Md Abu Ishak Mahy 4, Shafkat Hassan 5,
Md Humaion Kabir Mehedi 6, Md Motahar Mahtab 7
and Annajiat Alim Rasel 8
Department of Computer Science and Engineering
BRAC University
66 Mohakhali, Dhaka - 1212, Bangladesh
{*rakin.bin.rabbani, sumaiya.tanjil.khan, prasenjit.das, abu.ishak.mahy, shafkat.hassan,*
humaion.kabir.mehedi, md.motahar.mahtab}*@g.bracu.ac.bd*
annajiat@bracu.ac.bd

***Abstract*—Emotion detection or sentiment analysis is one of the major and challenging research subjects in the natural language processing field because of its vast range of practical applications like trends predictions in social media, extraction of emotions, opinion mining. In the English language domain Sentiment analysis has obtained a numerous attention in recent years, very less amount of study has been done in the context of Bangla which is one of the most extensively spoken languages around the world. Due to the popularity of this research topic, fresh researchers are looking for the best suitable approaches or best possible methods for better accuracy and improvement of their research. Moreover, since the Emotion Detection from the Bangla language using NLP research field is significantly new, it often becomes difficult to decide which methods should be worked with as reference availability is poor. As a result, a bulk amount of time is squandered experimenting the sentiment analysis approaches before coming to a deduction. This paper takes a detailed look at the current sentiment analysis approaches or algorithms used by researchers in the field of Emotion Detection from the Bangla language to help researchers identify appropriate algorithms for better results.**

***Index Terms*—NLP, Bangla Language, Emotion Detection, sentiment analysis, algorithms**

I. Introduction

Human emotion has always piqued my attention as a subject to research in psychology. Because they are Very important stuff to understand human nature. One of the most important uses of NLP and Artificial Intelligence is recognition of emotions from text. This is a crucial field of research in order to better the interaction of human and machine [1]. Recent technological advancements have allowed us to test human minds, particularly the emotional hardware involved in sensing the emotion, which refers to a more changed perspective of brain's capability and structure for emotion detection. Although this issue has gained huge amount of attention in the English language, it is relatively unexplored territory in the Bangla language till now. With over two hundred and twenty eight million native speakers, Bangla is the world's seventh most spoken language. Bangla is the mother tongue of Bangladeshi people. A large portion of the Indian people also speaks in this language . In social media, Bangla language is widely used by a huge number of individuals. Because of the growing development of Bangla language users, it is critical to concentrate on the research of emotion identification in Bangla. The majority of modern Bangla research relies on binary sentiment analysis. Many studies have been conducted in which good emotions are labeled as happy and negative emotions as sorrow. However, this is insufficient for analyzing a text. The majority of Bangla sentiment analysis research affixed on utilizing classic rule-based or machine learning-based approaches to detect positive or negative texts [2].

In this paper, we tried to analyze various methods of sentiment analysis on the Bangla language by studying various research papers. Through the study, we came across different algorithms related to NLP. These include Naïve Bayes classification algorithm, K-nearest neighbour, decision tree, support vector machine, K-means clustering, Multinomial Naïve Bayes, bidirectional long-short term memory, convolution neural network and many others. During the research, we found that some of the papers introduced a hybrid model by utilizing two or more algorithms while other papers experimented with just one specific algorithm. After studying all those algorithms, performance analysis was performed which is the novelty of this paper. Among all those algorithms, it is seen that accuracy degrades with an increasing number of classifications. It is preferable to use a deep learning model in such a situation.

The sections of this paper are organized as follows. Firstly, Literature Review in section II begins with an overview of the works, where multiple pieces of literature that are related to our topic have been reviewed to move forth with ideas being developed concurrently to the real-life application of the proposed works. Then in Methodology in section III we have discussed the different technologies used in the reviewed works. Moving on to the Performance Analysis in section IV, which explains the implementation and analysis of the different approaches and the Discussion in section V where we talk about the shortcomings of the different approaches. Finally in Conclusion in section VI, we concluded our paper.

II. Literature Review

In this section, we discussed the research we have conducted on different approaches of sentiment or emotion analysis on Bangla text. In paper [3] the authors categorized a given Bangla text with six individual emotion classes such as happy, sad, tender, excited, angry and scared using two machine learning techniques: Naïve Bayes Classification Algorithm and Topical approach. For both sentence and article-level of scope, the accuracy of the topical approach is much higher than the Naïve Bayes classification algorithm . Usually, positive and negative emotions are considered for Bengali sentiment analysis. Another paper [4] focused their research on fine-grained emotion analysis on Bengali text which identified 6 emotions like sadness,happiness,disgust, fear, anger, surprise using 5 basic machine learning techniques which are: Support Vector Machine (SVM), K-Means Clustering, Decision Tree,Na¨ıve Bayes and k-Nearest Neighbor (k-NN). Then the results of these machine learning techniques were compared.

Another paper [1] proposed Multinomial Naïve Bayes (NB) classifier along with various features such as parts-of-speech (POS) tagger, stemmer, term frequency-inverse document frequency (tf-idf), n-grams for Bangla text classify it into three emotion classes (sad, happy and angry) with an overall accuracy of 78.6%. [5] introduced the Bidirectional Gated Recurrent Unit called BiGRU along with the combination of the most classical deep learning algorithm, Convolutional Neural Network, and Bidirectional Long Short-Term Memory named CNN-BiLSTM to predict the affection of Bangla text by detecting six identical emotion concerned labels which are Happy, Sad, Love, Surprise, Fear, Angry.

With the growing rise of e-commerce, sentiment analysis can have a significant impact on everyone's daily lives. [6] introduced a machine learning-based model to predict a user's sentiment as in positive, negative, and neutral from a Bangla text review by applying five machine learning algorithms which are Support Vector Machine (SVM), Logistic Regression, Random Forest classifier, XGBoost algorithms and K-Nearest Neighbors (KNN) in the dataset which they collected from a Bangladeshi e-commerce site called "Daraz." KNN performs great among all these five algorithms in terms of all the performance measures. In one paper [2] the authors talked about their deep learning based model that can extract six basic emotions as in sadness, joy, fear, disgust, anger and surprise from Youtube Bangla comments. This [7] paper uses a Random Forest Classifier to classify sentiments and quantify total positive and negativity against a document or sentence.

In one paper [8] for a given Bangla document, the authors extracted a unique feature to generate the text review whether it is positive, negative or neutral by using the Tf.Idf (term frequency-inverse document frequency) model. [9] used character ngram features and a supervised machine learning technique named Multinomial Naïve Bayes to predict sentiment polarity in Bengali tweets. Their experimental results showed that character n-gram features are more effective than the traditional word n-gram features resulting in better performance than some existing sentiment polarity detection systems.

III. Methodology

A. Algorithms:

1) Naïve Bayes - NB: The Naïve Bayes method is one of the most effective classification algorithms to categorize text content. Based on likelihood, this technique assigns an emotional score to each word that corresponds to each emotional class. It attempts to determine emotion for the question text by utilizing this emotional score.

2) Decision Tree: The decision tree is the most powerful and extensively used classification and prediction tool. It is a flowchart-like tree structure consists of internal node, branch and leaf node. Each internal node denotes an attribute test, each branch denotes the test's conclusion, and each leaf node denotes a class label.

3) BiGRU: Bidirectional Gated Recurrent Unit also known as BiGRU is a RNN that can predict the present state using knowledge from previous and subsequent time steps. As input feature 64 characteristics were implemented in the author's model. The author also used gated unit for training those parameters.

4) CNN: Convolutional Neural Network is a deep learning model mainly used for extracting information from pictures and texts.

5) BiLSTM or Bidirectional Long Short Term Memory: BiLSTM is the process of constructing a neural network that can store sequence information in both directions (future to past) and forward (ahead to future) (past to future). Bidirectional LSTM differ from normal LSTM in that input flows in both directions. The standard LSTM only allows input to travel in one direction: backwards or forwards. Bi-directional input, on the other hand, flows in both directions to preserve both future and past data.

6) LR - Logistic Regression: The LR is a type of classification that is used to address the binary classification issue. In models with a twofold scenario, the result is commonly defined as 0 or 1.

7) XGBoost: The XGBoost is also called Extreme Gradient Boosting. It is a very efficient, versatile, and portable designed gradient boosting distributed toolkit . It utilizes the Gradient Boosting framework in order to develop Machine Learning algorithms. It helps to get rid of a wide range of data science issues quickly and correctly via parallel tree boosting.

8) RF - Random Forest: The first algorithm that comes to mind after hearing supervised machine learning method is Random Forest. It can be used to solve all kinds of classification and regression problems. The ensemble learning concept is used in this algorithm which entails combining many classifiers to tackle a difficult problem and improve the model's performance.

B. Implementation on BNLP:

Emotion detection is relatively scattered compared to other NLP tasks, especially for Bangla content. Tuhin et al. [3] and Azmin and Dhar [1] classified emotions from Bangla literature using the Nave Bayes algorithm. They combined NB with parts-of-speech (POS) tagger, stemmer, term frequency-inverse document frequency (tf-idf)and n-grams. POS Tagging is a typical Natural Language Processing method that refers to identifying words in a text according to a certain part of speech, depending on the term's definition and context. N-grams of texts are frequently applied in NLP tasks which are just a collection of co-occurring words inside a specific window. Tf-idf, is used to determine the relevance of a word in a corpus. Tuhin et al. [3] used a corpus dataset including 7500 hand written Bangla texts to conduct their investigation.

Rahman and Seddiqui [4], analyzed the outcomes of the 5 most familiar machine learning approaches. The 5 approaches are SVM, Naïve Bayes, k-NN, Decision Tree, and K-Means Clustering, with various feature combinations on their Textual Emotion Analysis task. Initially, they presented a labelled Bangla emotion corpus, which has a wide spectrum of quite well emotion expressions in social-media content. Following that, they implemented five algorithms and compared the outcomes. They have used outcomes from the K-NN algorithm with k=15 as their baseline model. Because the baseline k-NN model performed badly, they sought to modify parameters for the K-NN classifier in order to determine the ideal k-value for their data and obtain better results for k=5. The only unsupervised machine learning model they have used was K-Means Clustering. For the SVM classifier, they used both linear and non-linear SVM-kernel. According to their obtained experimented result SVM with a Non-Linear RBF-Kernel was the best among all. In their experiment, a Gamma value of 0.6 and C value of 2.0 were the optimal parameters on SVF.

Rayhan et al. [5] used BiGRU and a mix of two deep learning algorithms CNN and Bidirectional Long Short-Term Memory called CNN-BiLSTM to identify emotion from Bangla text. Moreover, since the BiLSTM model processes information in two ways, the model generates context for each letter in a particular text. In their work, they utilized both CNN and BiLSTM to classify emotions from Bangla text. By using this hybrid model, CNN captures whole text information and sends to BiLSTM, which accepts it as an input and looks for distinctive patterns from start to finish and end to start. The main purpose of merging CNN and BiLSTM was to extract as much information as possible from a text and better categorize emotion. The authors used a vocabulary size of 57,000, an embedding dimension of 64, and a maximum input length of 59 for both the BiGRU and CNN-BiLSTM models.

Using a dataset taken from the Bangladeshi e-commerce site "Daraz," Akter et al. [6] examined different classifiers like the Random Forest classifier, Logistic Regression, Support Vector Machine, K-Nearest Neighbors and XGBoost algorithms.

Tripto and Ali [2], used LSTM and CNN for multilevel sentiment and emotion classification from youtube comments. They used SVM and Naïve Bayes as their standard methods and compared output from LSTM and CNN with the outputs from two baseline methods. Tabassum and Khan [7], used Random Forest Classifier for sentiment analysis from Bangla sentences or documents.

Nabi et al. [8] used Tf.Idf to extract unique features from Bangla text and review emotions. Sarkar [9] presented Multinomial Naïve Bayes and character n-gram features for detecting the sentiment polarity on tweets that are written in bangla. Since Bengali tweets data is relatively noisy containing spelling errors, the author employed an N-gram tokenizer to extract features and tokenizes the original tweet data into character n-grams.

IV. Performance Analysis

In this section, we compared some of the models from current research on text classification in the context of Bangla Natural Language Processing. In terms of techniques and results, the TABLE I presents a comparison with relevant works. During the evaluation, we discovered that the dataset size, features, and classifier used for classification have a significant impact on the overall result of the job.

The best model by S. Azmin [1] uses a bigram-based tf-idf (frequency times inverse document frequency) with POS features, as well as a stopword and emoticon processor. They were able to achieve an overall accuracy score of 78.6 percent using this combo. They also used Support Vector Machine to classify the results for this same combination.

To assess the success of the learning algorithm, Mahmudun et al. [8] in their work used sets of characteristics to measure the positive and negative attitudes for each classifier, using the conventional precision and recall. The accuracy metric must then be used to compare the classifier's overall performance. A sentiment analysis task can be thought of as a classification task, with each classification label representing a different sentiment. As a result, the four metrics for each label (positive and negative) are defined and calculated in the same way as in a generic classification task.

The proposed best model by K. Sarkar [9] in their work outperforms the best system in the SAIL 2015 contest on sentiment analysis of Bengali tweets (IIT-TUDA). The winning system (IITTUDA) in the SAIL 2015 challenge uses a more advanced method that leverages distributional thesauri, support vector machine and sentence-level co-occurrence statistics to expand a (limited) Indian sentiment lexicon using an unsupervised approach. Because we employed character

TABLE I
PERFORMANCE OF DIFFERENT MODELS

Paper	Model	Accuracy
S. Azmin, 2019	Multinomial NB	78.6%
	SVM	71.6%
M. Mahmudun et al., 2016	Tf-Idf	83%
K. Sarkar, 2018	NB	48.5%
T. Rabeya, 2017	Lexicon Approach Backtracking Technique	77.16%
M.A. Rahman, 2019	SVM	52.98%
	NB	52.50%
	KNN	47.90%
	Decision tree	44.2%
	K-means clustering	49%
R. A. Tuhin et al., 2019	NB	90%
	SVM	93%
	Tf-Idf	83%
Rahman and Seddiqui, 2020	CNN-BiLSTM	66.62%
	BiGRU	64.96%
Tabassum and Khan,2019	RF	86%
	SVM	67%
	Maximum Entropy	68%
	Tf-Idf	83%
M. T. Akter et. al	RF	90.84%
	LR	90.33%
	SVM	94.35%
	KNN	96.25%
	XGBoost	90.56%
Tripto and Ali, 2018	LSTM	65.97%
	CNN	60.89%
	NB	60.79%
	SVM	49.81%

n-grams and SentiWordnet features to train the model that is Multinomial Naive Bayes classifier based system, their proposed approach is straightforward.

Following the experimental analysis by Rabeya et al. [10], in their hypothesis they were able to correctly conclude with an accuracy of 77.16 percent using a backtracking strategy in the form of a lexicon-based approach on 301 sentences gathered from various data sources. It is possible for a sentence to carry a sentiment or to be neutral. They identified the emotion (happy or sadness) connected with a sentence if it has sentiment. As a result, complete accuracy refers to the fact that while speaking or writing most individuals convey their feelings at the conclusion of a sentence.

According to comparison analysis the study by Rahman et al. [4], the "SVM with a Non-Linear RBF-Kernel" model was the best of all. Using this SVM model, they then experimented with alternative preprocessing and feature combinations. Having seen list of possible feature and preprocessing procedure combinations for the non-linear SVM model with an RBF-kernel, the most ideal parameter values in this experiment were a C-parameter value of 2.0 and a Gamma-value of 0.6. The most essential (tf-idf) word unigrams have the finest mix of attributes. As a result, the model's best accuracy score was 0.5298 (i.e., a 20.08 percent improvement over the baseline model) and an F1(macro) of 0.3324 (i.e., a 0.2174 improvement over the baseline model). They found that CNN-BiLSTM performs somewhat better than BiGRU (64.96 percent) with a 66.62 percent accuracy. This section's accuracy was calculated using the test set after the model had been trained with the training set. The performance difference between the two models was not significant. To analyze the output result, they contrasted their two models with certain Bangla texts that were not included in their dataset. The actual and projected emotions by the two proposed models are shown in Table V. During the evaluation, they discovered that both models perform similarly well on supplied Bangla texts, with CNN-BiLSTM marginally outperforming BiGRU.

In their last experiment, Tuhin et al. [3] compared their findings to two earlier research on sentiment analysis from Bangla text. These two articles were chosen based on their working methodology and level of performance. It is clear that no paper used more than three emotion classes to detect emotion. Apart from that, the training dataset was far more vibrant than the previous study articles, and that the dataset can be used for future research. Another crucial factor is that the writers of a compared paper employed SVM and gained greater accuracy as a result. However, the authors did not employ SVM. The fundamental reason for this is that SVM is quite effective when it comes to distinguishing two groups, but it becomes more difficult when it comes to classifying multiple classes with noisy input. Topical Approach, on the other hand, has demonstrated that it can work with several classes with reasonable accuracy. In light of the entire circumstances, it is reasonable to conclude that this research article outperforms others in terms of Bangla text in some respects.

Between data gathering and sentiment identification in the work by Akter et al. [6], they went through a lot of processes. For the dataset's sentiment identification, they deployed five machine learning techniques. They employed a split ratio of 80 percent and 20 percent for their algorithms' data in a regular train test. All of the methods have been subjected to tenfold cross-validation. It is demonstrated that KNN outperforms the others, with a 96.25 percent accuracy and a 96 percent f1-score.

Surprisingly, comments from review videos have a better accuracy in our method since they are more polarized (either praising or criticizing the product/video). Both of their methods are quite good at recognizing sentiment and emotion from comments in music, sports, and talk-show videos. The accuracy of news videos, on the other hand, decreases since the comments in this area cover a wide range of issues and contain ambiguous ideas.

V. DISCUSSION

The papers we have gone through, we found out Mahmudun et al. [8] model had a very good accuracy level which is

83%. They focused on features and detected negative and positive emotions or sentiments with a small data set. They did not work on detecting the 6 basic emotions like happy, sad, angry, fear, love and surprise. Rabeya et al. [10] also worked on finding out two basic emotions or sentiments from Bengali text and they are known as melancholy and happiness. They used backtracking techniques to determine these emotions with accuracy of 77.16%. This backtracking technique is so efficient in solving problems but it is so time consuming. Rahman et al. [4] combined some classic ML approaches. They used Naïve Bayes, K-mean clustering, SVM, KNN, Decision Tree to detect emotions from Bangla Language. They got an average accuracy of 52.50% using NV and 52.98% with SVM, where both of them are not as good as expected. Sarkar et al. [9] created a sentiment polarity detection system Using Multinomial Naïve Bayes. Because of their insufficiency of the training data wrongly labeled data they got only 48.5% of accuracy which is the lowest accuracy so far. Akter et al. [6] used a small dataset which was compiled and taken from 'Daraz'. In that study, the authors found the most outstanding result from the TF-IDF vectorizer and KNN algorithm which gained accuracy of 96.25%.Tripto et al. [2] got accuracy of 65.97% using LSTM. It was also average. Rahman et al. [4] worked on CNN-BiLSTM and BiGRU and got 66.62% and 64.96% of accuracy rates. Azmin et al. [1] used a combination of Multinomial Naïve Bayes, Bi-gram and Tf-Idf and got an accuracy of 78.6%. Though they used the dataset for 6 basic emotions but they work on 3 emotions so far. Their model works well and fast. They tested the same dataset with SVM then got 71.6% accuracy which was lower than their proposed model. Tabassum et. al [7] proposed a model which was combined with unigram, POS tagging, Random Forest classifier and negation handling. That model provided more accuracy in results. They worked on detecting sentiments(positive or negative) and got accuracy of 85% using (Unigram+POS) 87% accuracy using (Unigram+ Negation+ POS) on their proposed method . Raihan et al. [5] worked on two models and they were CNN-BiLSTM and BiGRU. Between these models, CNN-BiLSTM performed a little better comparing BiGRU (64.96%) with the accuracy of 66.62%. Their key challenge was the non availability of the Bangla Language's text dataset. They had to translate a English dataset into Bangla. It was a public dataset for six fundamental emotions which was prepared to a Bangla dataset. To translate the data set they used Google Translator. Tuhin et al. [3] detected six fundamental emotion classes using Topical Approach & Naive Bayes Algorithm. They need a noise-free dataset because a supervised method is a non-effective approach to dig information from a noisy dataset or a very big dataset. They got 90% accuracy-rate with the NV algorithm. This Model is the most accurate, fast & well-performed model so far.

VI. Conclusion

In this survey paper, we attempted to cover the sentiment or emotion detection area from the Bangla language using NLP algorithms. In every research paper, they came up with new ideas to extract a unique feature to create a better approach to understanding sentiment or emotion from Bangla text. They used different algorithms of NLP. Among them, Naive Bayes and Random Forest are commonly used. The process of detecting emotions from Bangla language text datasets through deep learning models is new to researchers till now. But the Deep Learning models also showed very good outcomes on emotion detecting and Multinomial Naive Bayes as well gave a good accuracy rate as expected. In this paper, we have gone through 10 papers on emotion detection from the Bangla language. In posterior time, we will study more research papers and survey more new models with various NLP algorithms to detect emotion more accurately from Bangla Language.

References

[1] Azmin, S., & Dhar, K. (2019, December). Emotion detection from Bangla text corpus using Naïve Bayes classifier. In 2019 4th International Conference on Electrical Information and Communication Technology (EICT) (pp. 1-5). IEEE.

[2] Tripto, N. I., & Ali, M. E. (2018, September). Detecting multilabel sentiment and emotions from bangla youtube comments. In 2018 International Conference on Bangla Speech and Language Processing (ICBSLP) (pp. 1-6). IEEE.

[3] Tuhin, R. A., Paul, B. K., Nawrine, F., Akter, M., & Das, A. K. (2019, February). An automated system of sentiment analysis from bangla text using supervised learning techniques. In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (pp. 360-364). IEEE.

[4] Rahman, M., & Seddiqui, M. (2019). Comparison of Classical Machine Learning Approaches on Bangla Textual Emotion Analysis. arXiv preprint arXiv:1907.07826.

[5] Rayhan, M. M., Al Musabe, T., & Islam, M. A. (2020, December). Multilabel Emotion Detection from Bangla Text Using BiGRU and CNN-BiLSTM. In 2020 23rd International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.

[6] Akter, M. T., Begum, M., & Mustafa, R. (2021, February). Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors. In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) (pp. 40-44). IEEE.

[7] Tabassum, N., & Khan, M. I. (2019, February). Design an empirical framework for sentiment analysis from Bangla text using machine learning. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-5). IEEE.

[8] Nabi, M. M., Altaf, M. T., & Ismail, S. (2016). Detecting sentiment from Bangla text using machine learning technique and feature analysis. International Journal of Computer Applications, 153(11), 28-34.

[9] Sarkar, K. (2018, January). Using character n-gram features and multinomial naïve bayes for sentiment polarity detection in Bengali tweets. In 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT) (pp. 1-4). IEEE.

[10] Rabeya, T., Ferdous, S., Ali, H. S., & Chakraborty, N. R. (2017, December). A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text. In 2017 20th international conference of computer and information technology (ICCIT) (pp. 1-7). IEEE.