e-palli

AMERICAN JOURNAL OF AGRICULTURAL SCIENCE, ENGINEERING, AND TECHNOLOGY (AJASET)

ISSN: 2158-8104 (ONLINE), 2164-0920 (PRINT)

VOLUME 6 ISSUE 3 (2022)

PUBLISHED BY: E-PALLI, DELAWARE, USA

American Journal of Agricultural Science, Engineering, and Technology (AJASET)

Volume 6 Issue 3, Year 2022
ISSN: 2158-8104 (Online), 2164-0920 (Print)
DOI: https://doi.org/10.54536/ajaset.v6i3.990
https://journals.e-palli.com/home/index.php/ajaset

A Systematic Review of Sentiment Analysis from Bengali Text using NLP

Suma Hira[1], Atish Kumar Dipongkor[1], Saumik Chowdhury[1], Mostafijur Rahman Akhond[1*], Syed Md.Galib[1]

Article Information

Received: November 10, 2022

Accepted: December 04, 2022

Published: December 06, 2022

Keywords

Bangla, NLP, Review, Sentiment, Survey

ABSTRACT

Sentiment Analysis (SA) is the method of studying a person's comments and statements through computational means. It is a sub-domain of Natural Language Processing. A sentiment analysis (SA) system is created by training a significant number of positive, negative or neutral sentence datasets. Although there are many research papers on this subject in English, the work of sentiment analysis in Bengali has not become very popular due to the complexity of the Bangla language and the insufficient presence of the Bangla language online. But now the use of Bengali language has increased in online news, social media, and blogs. At the same time, the number of researches on Bengali NLP is also increasing. But what is the current state of sentiment analysis, its limitations and is there still room for improvement in some places, are not being properly reviewed and research is lagging behind. We have conducted a review on sentiment analysis so that future researchers of sentiment analysis can easily find out about the current state of sentiment analysis. In this research, we have tried to survey the current context of sentiment analysis (SA) and at the same time, we have created a sequence of comparatively better research from the existing ones. To create this sequence we followed a method called is TOPSIS. We have also discussed the challenges to overcome for improving the Sentiment analyzer.

INTRODUCTION

A Systematic Literature Review (SLR) is the technique of recognizing, appraising, and interpreting all available investigations relevant to the specific study topic. SLR can be used to review the available data on a certain research issue, identify any research gaps, provide recommendations for further research, and collaborate to generate new ideas (Mandal & M Mainul Hossain, 2017). SLR necessitates a greater amount of effort than ordinary literature evaluations. SLR aims to find as much relevant material about a specific study domain as feasible. Conducting a thorough literature study on a specific issue can be quite helpful and valuable.

There have done several works in Sentiment Analysis and we found very few surveys in Bangla SA but none of them created rankings based on the dataset, performance, etc. So, we realized the need for a systematic literature review of the Bangla language in SA. Moreover, it is clear from the definition of SLR that the researcher will be able to know the current situation in a specific research area. So in our Systematic review of Bengali Sentiment Analysis, we highlight all the details of the present stage of Bengali Sentiment analysis. We think that it will be helpful for those researchers who are willing to research the topic of Bengali Sentiment Analysis.

Tapasi (Chavan *et al.*, 2014) highlights some of the basics of a few Bengali SA research. They have divided the proposed methods into different categories and also mentioned what kind of dataset has been used there. The dataset is further analyzed with the corresponding sentiments (Sharif *et al.*, 2019). Banik (Prasad *et al.*, 2017) has reviewed several aspects of Bangla sentiment analysis. They discuss the existing pre-processing steps and research methods. Moreover, they have reviewed the results of various methods recently and provided guidelines for future researchers. Zubair (Tabassum & Khan, 2019) has reviewed the various methods of sentiment analysis divided by category. In that systematic review, the researchers focused on which topics in the Sentiment Analysis Approach focused more on and which topics they focused on in the lesson. In our systematic review, we discussed the existing proposed method, and pre-processing method as well as discussed the limitations. Moreover, we gave an opinion on how to further improve the Bengali sentiment analysis system. Finally, we explained comparatively better research from the existing ones by applying the TOPSIS method.

Our contributions to this research paper are as follows:

- First, we discussed several research papers' existing methods.
- In the next step, we showed how to improve the Sentiment Analysis system. At this stage, we discussed some of the complex sentences or articles for example in the Bengali language where the special rules should be applied to improve Bengali SA.
- And lastly, we created a sequence of ranking among the existing ones by applying a mathematical model called the TOPSIS method.

The paper is organized as follows: In section 2, we deliberated the methodology of the work. Section 3 presents the result analysis of this SLR. Finally, the conclusions of the work are outlined in Section 4.

[1] Jashore University of Science and Technology, Bangladesh
[*] Corresponding author's e-mail: mr.akhond@just.edu.bd

MATERIALS AND METHODS

Our systematic literature review has been performed in compliance with the systematic review guidelines (Alam *et al*, 2018) and consists of a few steps.

Research Procedural Questions

Research questions are important to complete successful research work. Only specific questions can help a researcher to set the right goals. In the case of a literature review, research questions can arrange the review paper more specifically so that the current state of research that has already been conducted on this subject can be easily discovered by researchers. We found some research questions at the time of conducting this study.

RQ#1. What are the current methods used for developing Bengali sentiment analysis?

RQ#2. What are the limitations of the existing research method?

RQ#3. How can the Sentiment Analysis Method be further improved?

RQ#4. Which research work is comparatively better?

Search Strategy

We used some keywords to search for the relevant paper for our study. For each search term, we used alternative spelling and synonyms. The search string in our searches are given below:

- (Bengali OR Bangla) AND (sentiment OR emotion) AND (analysis OR detection)
- (Bengali OR Bangla) AND sentence AND polarity
- (Bengali OR Bangla) AND (emotion OR sentiment) AND mining

Study Selection Criteria

The research questions are the basis of our study selection criteria. We included some papers on Sentiment Analysis in the Bengali language and the work must be issued in either a Journal or a Conference proceeding. We included only those papers if the full texts of the potentially relevant publications are obtained. However, some research has been done on Bengali sentiment analysis but it is comparatively fewer than the number of other most spoken languages of the world. We included a lot of paper and extracted information by answering our research question.

Study selection process

We have chosen specific sources for our process of searching. A manual search procedure was followed to obtain the answers to our research question from the following sources-

- IEEEXplore
- ACM Digital Library
- Google Scholar
- Research Gate

Most of the informative journal and conference papers are published here that's why we chose those sources. After collecting the paper we selected papers according to our research question.

Data Extraction

To record the information, collected from the primary studies, we have used a data extraction form, shown in Table 1. It is representing the entity of Study Id, Paper Name, Author, Training Data, Proposed Method, Accuracy, Limitations, etc.

Table 1: Summary of the research

Author name	Method	Year	E-library
Sharif *et al.*	NB	2019	IEEE
Prasad *et al.*	NB, DT	2016	IEEE
Tabassum & Khan	RF	2019	IEEE
Tripto & Ali	LSTM, CNN	2018	IEEE
Nabi *et al.*	Tf-.Idf	2016	IJCA
Chowdhury & Chowdhury	SVM, MaxEnt	2014	IEEE
Aziz *et al.*	RNN	2018	R. Gate
Tuhin *et al.*	NB	2019	IEEE
Sarkar & Bhowmick	NB	2017	IEEE
Shamsul *et al.*	SVM, Tf-.IDf	2018	IEEE
Alam *et al.*	CNN	2017	IEEE
Dey & Sarkar	Lexicon	2019	IEEE
Eftekhar *et al.*	NB		
Tapasy *et al.*	Backtracking	2019	IEEE
Islam *et al.*	NB	2016	R. Gate
Ismail *et al.*	MNB	2019	ACM
Khondoker *et al.*	BERT	2020	IEEE
Al-Amin *et al.*	CBOW; SG	2017	IEEE
Asif *et al.*	LSTM	2016	R. Gate
Sazzed & Jayarathna	RR,SVM,LR,RF	2019	IEEE

RESULTS AND DISCUSSION

RQ1: What are the methods used for developing Bengali sentiment analysis?

Sharif *et al.*(Sharif *et al.*, 2019) proposed a method that could classify the sentiment of customer reviews according to their feelings into either positive or negative. They applied multinomial naïve Bayes, decision trees (Chavan, 2014), and Random forests in their system. But, they didn't introduce the neutral class and POS tagger in their important system but they didn't apply this rule.

Prasad *et al.* (Prasad *et al.*, 2017) presented a method to extract feelings from Tamil and Bengali tweets. They extracted two types of features- One is the tf-idf score of Unigram and Bigram and another one is the tweet-specific feature where they considered Hashtags and Emoticons. Finally, they applied a naïve Bayes classifier and decision tree classifier to implement the system of sentiment mining for the Bengali language. However, they didn't use POS taggers and emoticons.

Tabassum & Khan (2019) have developed a system for classifying sentiments that quantify the total positive and negative effects of the documents. They implemented the system with a combination of POS tagging, unigram, negation handling, and Random Forest Classifier which was a better result. Then they set a rule to detect positive sentences the summation of negative or negation words must be less than the total number of positive words. Similarly, if the summation of the negative and negation word is greater or equal to the positive words the sentence was considered negative.

Tripto & Ali (2018) have built a deep learning-based model to identify sentiment and extract emotions from Bengali texts. They labelled the sentiment in three-class and five-class shows with 65.97% and 54.24% accuracy. They applied the Word2vec algorithm (Mikolov *et al.*, 2013)to the vector representation of each sentence and also implemented the Continuous Bag of Words (CBOW) and Skip Gram (SG) model. After completing all preparation, they applied two methods for both sentiment and emotion identification. They are LSTM with an embedding layer and CNN as its core layer. They utilized SVM and Naïve Bayes as their baseline method and also used tf-idf to generate the feature set for each sentence.

Asif *et al.* (Hassan *et al.*, 2017) applied a deep recurrent model, LSTM to analyze sentiment on Bengali along with Romanized Bengali text. Their 9337 textual data amount of Bengali and Romanized Bengali entries was 6698 and 2639 respectively. Additionally, they manually categorized those data in one of three categories: positive (1), negative (0), and ambiguous (A) by two different Bangla-speaking annotators. Then they applied a recurrent neural network, LSTM. Keras model-level library was used as it has all the required features. They also used the Embedding layer to implement word to a vector representation. They used their first validation data to pre-train the second validation data and used the second validation data to pre-train the first validation data.

Nabi *et al.*(Mahmudun *et al.*, 2016) proposed a method where they got TF-IDF is better to get the solution of sentiment analysis to extract the positive, negative or neutral features out of different features and have more accurate results. The most frequent letters were calculated as beginning and ending letters, with 750 positive and negative Bangladeshi words being collected. Then, they applied the tf-idf algorithm to find the most informative words and analyze the Bengali sentences (i.e.: Sentence length, Pattern analysis).

Chowdhury & Chowdhury(2014) proposed a semi-supervised bootstrapping approach to detect the polarity of text from Bengali microblog posts. In the case of POS tagging, they used NLTK's Brill tagger. Then, they built up their Bengali Sentiment Lexicon in a total of 737 single words and used the self-training bootstrap method, and set rule-based classifiers. After setting up the rule, they extracted different features (n-grams, stemming, emoticon, lexicon, POS-tagging, Negation, and combination of all features) from each tweet.

Hasan *et al.* (Hasan *et al.*, 2014) proposed a method to detect sentiment from Bengali text by using valence analysis. They translated the Bengali word into English words by sentiwordNet3.0 and found the different senses (positive, negative, and neutral) and their polarity for different POS. Then, a different sense is formed by WorldNet 3.0 in each case of the word. Finally, the system detects the sentiment of writing a Bengali by comparing the percentage of numbers. The major limitations of this paper are that they use WorldNet and sentiwordNet which are designed for the English language. So, it is important and necessary to develop WorldNet and sentiwordNet for the Bangla language.

Sazzed & Jayarathna (2019) proposed a method in which they classify the sentiment of the Bengali language along with its corresponding Machine Translated English corpus. They compared the sentiment classification accuracy of Bengali and Machine translated corpus. Then they steamed and tokenized both the Bengali and English corpora using Scikit-learn (Hao & Ho, 2019). For the deep learning model LSTM, they used Keras (Keras: The Python Deep Learning API, n.d.) built-in tokenizer. Finally, they applied six machine learning techniques and compared the Sentiment between Bengali and English corpora. However, they didn't use a POS tagger in their system.

Khan *et al.*(Khan *et al.*, 2020) proposed a model to detect the sentiment from the Bengali paragraph. Their proposed system can detect two types of sentiment categories that are, happy and sad. There were two columns in their dataset, where one for sadness and another for happiness. After that, they tokenize their text using Countvectorizer. Here, researchers have used six different classification models and found that MNB gives the best result to classify the sentiment polarity. There are some drawbacks such as the dataset is not so large, wrong spelling in Bangla language, and Writing English or other language words in Bangla letters.

Khondoker *et al.* (K. I. Islam *et al.*, 2020)manually created two datasets of Sentiment Analysis for 2-class and 3-class in Bengali. They applied an unsupervised language representation model named BERT to train this model of analyzing sentiment and also extended the model with three different end-to-end deep network layers: GRU(Cho *et al.*, 2014), LSTM(Hochreiter & Schmidhuber, 1997), and CNN(Fukushima, 1988). Finally, for dimension reduction, this vector is passed through a completely connected neural dense layer, and then Softmax is used to define emotions using the final reduced variable. They also implemented two other deep-learning architectures with pre-trained word embedding, Word2Vec, and fast Text.

Sarkar & Bhowmick (2018) proposed a method that detects the sentiment polarity of Bengali tweets using two machine learning approaches, SVM and Multinomial Naïve Bayes. The proposed method used the dataset of the SAIL contest 2015. Then, they used an n-gram tokenizer as the primary feature extractor, and SentiWordnet features are contained in a tweet. Then they applied their classifier model from Weka. They chose MNV and SVM algorithms from the Weka model. According to the author, the details of Weka and the other features are given in (Witten & Frank, 2005). They train each classifier using the SAIL 2015 contest's training data and then put it to the test using the contest's test data.

Alam *et al.* (Alam *et al.*, 2018)proposed a method to generate sentiment classification of Bengali comments using a Convolution Neural Network. They considered the Bengali comment written in English letters and convert them into Bengali. Then, they used Vocabulary creator and Matrix Generator. The model is divided into two parts. In the first part convolution operation and max-pooling techniques are performed. The second part is the fully connected layer. They also applied regularization techniques on neurons, Cross-Entropy to calculate loss, and Softmax-classifier to calculate probabilities. They used an Adam optimizer to optimize the gradient in their work. They obtained the highest accuracy of 99.87%. But they didn't introduce a neutral class and their comments on a dataset that is not so large.

Islam *et al.* (M. S. Islam *et al.*, 2017) proposed a model to extract the sentiment of Bengali Facebook status by applying the Naïve Bayes classification model. They classified two types of sentiment (Positive and Negative) in this research. They applied to stem by using nltr open-source software ('Snltr-Software' [Online]. Available: Http://Nltr.Org/Snltr-Software/ [Accessed: - Google Search, n.d.)and normalized the data by the rules of Bengali Grammars (বাংলা ব্যাকরণ – Bangla Library, n.d.). Then they counted the Unigram and Bigram of the data words as features in the Naïve Bayes classification model. Then they measured the Prior probability and conditional probability by the Naïve Bayes classifier. Then they applied Laplacian smoothing on NB and finally, they calculated the polarity by choosing the maximum score from a class between two classes.

Shamsul *et al.* (Arafin Mahtab *et al.*, 2018) proposed a method to extract sentiment that extracted sentiment of people's comments about Bangladesh Cricket on social sites and news portals. The system can identify the positive, negative, and neutral sentiments of a sentence. They also upgrade their work to detect criticism, praise, and sadness. They have collected a dataset called ABSA (GitHub - Atikdu/Bangla_ABSA_Datasets, n.d.) Initially which contains 2979 data with five columns. They also collected a cricket dataset from the Facebook group of Bangladesh Cricket and the "ProthomAlo" newspaper. They have extracted features of data by using TF-IDF methods. Finally, they applied an SVM classifier with a linear kernel to implement the system.

Aziz *et al.*(Aziz Sharfuddin *et al.*, 2018) proposed a model for Classify Sentiment of Bengali Text using RNN with Bidirectional LSTM. For this research, at first, they managed the dataset of (Md Al-Amin *et al.*, 2017) that contained about 15000 Facebook comments. As the dataset didn't meet their expectations, they collected their own collected dataset, consisting of 30,000 Facebook comments. Among them, they only select 10,000 comments (5000 positive and 5000 negatives). Then they removed the emoji, symbols, numbers, stickers, and all the English letters. That means their dataset contained only plain Bengali text. The comments that consist of more than fifty words are also neglected in this research. Then they trained their dataset using LSTM, which is an artificial RNN architecture.

Al-Amin *et al.* (Md Al-Amin *et al.*, 2017)presented a sentiment classification with word2vec of Bengali comments. They started their implementation with 2500 comments where they trained 90% of their collected data step by step. They calculate the negative and positive scores of a query by two syntactic rules. They trained their model in six steps, where each step they introduced 2500 new comments. So that the total dataset size is 15000. Then they calculate the modified positive and negative scores. If the positive score of a query is greater than or equal to the negative score then the comment is positive otherwise, it is negative.

Tuhin *et al.* (Tuhin *et al.*, 2019)extracted six types of emotion classes in both sentences and article-level using two machine learning techniques – Naïve Bayes and Topical approach from any Bengali text. The Naïve Bayes classifier is applied to extract the emotion of the sentence. In this case, they also calculated the probabilistic score of each emotion class for each sentence, like sentence level. Then the maximum value among six emotion classes was considered as the article's corresponding emotion. In the article level emotion detection, they also applied the Topical approach where the concept of tf-idf was used to calculate the emotional value.

Eftekhar *et al.* (Hossain *et al.*, 2021)introduced a technique to determine positive or negative sentiment from Bengali book reviews. The multinomial Naïve Bayes approach was applied in this sentiment analysis system. They developed their dataset containing 2000 textbook reviews by following the strategy of (Dash & Ramamoorthy, 2018) Then they distributed the whole dataset in three subsets; training set (TR), validation set (VD), and test set (TS). Then they extracted N-gram features and tf-idf feature values from the pre-processed dataset. Then they used LR, DT, RF, MNB, KNN, SVM, and stochastic gradient descent (SGD) algorithm to classify the sentiment of Bengali text.

Ismail *et al.*(Siddiqi Emon *et al.*, 2019) proposed a technique that extracts the sentiment of Bengali online reviews written in the English alphabet. They have applied the Multinomial Naïve Bayes method to implement the system. After that, they used the word Net database to tag the text with their corresponding parts of speech(Saranya & Jayanthy, 2018). Then by using Sentiwordnet(De Silva & Haddela, 2013), (Mohammad, 2016)database they assigned Numerical values to an adjective. Then finally they applied the multinomial Naïve Bayes classifier and calculated how positive or how negative the review was.

Dey & Sarker (2019) proposed a lexicon dictionary-based approach to detect Bengali text's positive, negative or neutral sentiment. At first, they collected 5200 sentences. They developed a lexicon dictionary with 5100 words where a sentiment score was assigned to each word. To calculate the polarity of sentiment words, they also considered boost words and negation words. Then they constructed the sentiment result by normalizing the total sentiment score. If the normalization score is between zero to one, then the text was considered positive if the normalization score is between zero to minus one, then the text was deemed negative.

Tapasy *et al.* (Rabeya *et al.*, 2019) proposed a model of Sentiment Analysis that analyze the acceptance rate of a star from his song review. For detecting the sentiment here they applied backtracking. In the backtracking approach, the main part is the sentiment lexicon. They omitted the words usually absent in song reviews. They also added positive and negative English words written in the Bengali language. Then they generated their expressions from the input sentences. They have denoted affirmative, negative, and নাবাচক as respectively 1, 2, and 3. After that, they applied a backtracking algorithm to the generated expression.

RQ2: What are the limitations of the existing research method?

In most cases, the limitations of our discussed methods are almost the same. But still, there are some differences between them. In our study, positive, negative, and neutral are three basic categories of sentiment. Neutral sentiment is as well as important as positive and negative sentiment. As it takes the result in the right direction(Qiu & Li, 2016), that's why it should be considered positive and negative sentiment. In the research of (Sharif, 2019), (Tabassum, 2019), (Chowdhury, 2014), (Khan, 2020), (Alam, 2017), (Islam K. I., 2020), (Islam M. S., 2016), (Hassan, 2016), (Abdullah Aziz Sharfuddin, 2018), (Tuhin *et al.*, 2019), (Hossain *et al.*, 2021) .They didn't introduce the neutral class in their proposed method.

There is a linear relation between accuracy and dataset size, if dataset size increases then another one also increases (Md Al-Amin *et al.*, 2017). For any NLP application large amount of dataset, the percentage of noisy data is comparatively less than the small data. Because of the large dataset, a model won't be affected by a few noisy data. That's why a large dataset is important in Sentiment analysis. But none of the research without (K. I. Islam *et al.*, 2020), (Hassan *et al.*, 2017), (Aziz Sharfuddin *et al.*, 2018) , (Tuhin *et al.*, 2019) that we have studied has worked with a large dataset.

When it occurs a negative word after a negative word then it is considered positive together. So this rule is important to express the sentence class more accurately. But in the research of (Prasad *et al.*, 2017) , (Tripto & Ali, 2018) ,(Mahmudun *et al.*, 2016), (Chowdhury & Chowdhury, 2014), (Hasan *et al.*, 2014), (Sazzed & Jayarathna, 2019), (Khan *et al.*, 2020), (K. I. Islam *et al.*, 2020), (Sarkar & Bhowmick, 2018), (Alam *et al.*, 2018), [26], (Hassan *et al.*, 2017), (Aziz Sharfuddin *et al.*, 2018), (Tuhin *et al.*, 2019), (Arafin Mahtab *et al.*, 2018), (Hossain *et al.*, 2021) they didn't apply this rule. As a result, if any sentence contains negation after negative that will not be considered positive together.

In recent years, the use of emoji's use on the Internet has increased rapidly. Ideogram and smiley allow users to communicate the text in e-mails and web pages more easily. People use them frequently when their expressions are hard to explain with words alone. For this reason, emoticons need to be considered. But none other than the research of (Prasad *et al.*, 2017), (Tripto & Ali, 2018), and (Chowdhury & Chowdhury, 2014) didn't consider emoticons in their proposed method.

A compound sentence consists of several clauses combined through conjunctions. It is easy to express an opinion in simple sentences as it is easy to analyze. But in the case of compound sentences, it is difficult to express opinions since they use various forms of conjunctions to attach clauses (Savanur & Sumathi, 2018). We found that all the research papers without (Alam *et al.*, 2018), (M. S. Islam *et al.*, 2017) that we have read on our research purpose didn't focus on establishing any rules to extract the sentiment of compound sentences.

The POS tag of the original text is the basic building block of many NLP pipelines, for example, disambiguation of word sense, question answering, sentiment analysis (Twitter NLP Example: How to Scale Part-of-Speech Tagging with MPP (Part 1), n.d.), etc. Adjectives, adverbs, interjections, and some verbal expressions of different senses (like positive, and negative). On the other hand, other parts of speech don't express any sense so these parts of speech are used as stop words and removed in the pre-processing steps. But in the research of (Prasad *et al.*, 2017) , (Tripto & Ali, 2018) , (Chowdhury & Chowdhury, 2014), (Hasan *et al.*, 2014), (Sazzed & Jayarathna, 2019), . (K. I. Islam *et al.*, 2020), (Sarkar & Bhowmick, 2018), (Alam *et al.*, 2018), (Hassan *et al.*, 2017), (Aziz Sharfuddin *et al*., 2018), (Tuhin *et al*., 2019), (Arafin Mahtab *et al*., 2018), (Hossain *et al*., 2021) they didn't use POS tagger in their system.

Elongated words and Punctuation marks also contain emotion (Tripto & Ali, 2018). So the rule should be considered for Sentiment Analysis in the case of Elongated words and Punctuation marks. Though (Tripto & Ali, 2018) considered both the elongated word and the punctuation mark and (Alam *et al*., 2018)considered the punctuation mark only, no other research considered a separate rule for them. Some Limitations also found in the Research Papers are:

• (Rabeya *et al*., 2019) Failed to detect sentiment from those positive sentences consisting of only negative words.

• (Alam *et al*., 2018) removed the slang in pre-processing steps but it may mean negative sentiment.

• None of the researchers introduced a spell checker as spelling may be incorrect when pressing another key unintentionally while typing.

Q3: How can the Sentiment Analysis Method be further improved?

Earlier we highlighted the various limitations of the conventional methods of sentiment analysis. If these limitations could be overcome, the performance of the sentiment analysis system would be further improved. Moreover, we have highlighted some more challenges here. We think that if we can implement them, Bangla Sentiment Analyzer will become much more. We have documented several Sentiment Analysis System challenges to find opportunities for improvement scope in this research domain. Most of the research papers we read did not meet those challenges. So, future researchers should meet those challenges to improve the system of Sentiment Analysis.

Challenge1

Let's look at two examples

"আমি একটি লাল রঙের ব্যাগ কিনেছি।" (I bought a red bag).

"লাল রঙ সুপ্রিয়ার খুব পছন্দের। সে একটি লাল রঙের ব্যাগ কিনেছে।" (Red is Supriya's favorite color. She bought a red bag). In the field of the first case, the sentiment should be neutral as there is no presence of a strongly positive or negative sense. Now in the second case, analyzing the two sentences here, it is understood that Supriya has been able to buy the thing of her choice, which indicates a positive sense. But individually, both of the sentences carry no sentiment. A human can quickly detect sentiment for this kind of sentence or paragraph, but it is not as easy as a human to detect sentiment in such cases by an automatic sentiment analysis system.

Challenge 2

Let's start with an example again "রেস্টুরেন্ট দেখতে অনেক সুন্দর ছিল।এখানে ওয়েটারদের ব্যবহার বেশ ভাল বলে শুনেছি। তাছাড়া এটি আমাদের বাড়ি থেকে বেশ কাছে অবস্থিত। এজন্য আমরা এখানে অনুষ্ঠান করার সিদ্ধান্ত নিই। কিন্তু বাস্তবে এখানকার খাবারের মান খুবই খারাপ।" (The restaurant looked very nice. I heard the waiters here are pretty good. Moreover it is located quite close from our house. That's why we decided to hold the event here. But in reality the quality of food here is very bad). This type of writing is mainly found in food review groups. Basically, in a restaurant, we go to eat. So first of all its food quality must be good. Here it is noticeable that the last sentence of the above paragraph has changed the whole paragraph's sentiment polarity. In a word, this sentence contains the sentiment of the entire section.

Challenge 3

Sometimes, a negative word sits with different words and expresses different sentiments. As example:

"তমার হাতের লিখা অসম্ভব সুন্দর।" ("Toma's handwriting is incredibly beautiful.")

"আমার পক্ষে এ কাজ করা অসম্ভব।" (It is impossible for me to do this).

Challenge 4

Sometimes, the purpose of the sentence may be to compare different famous/infamous people or objects. Although it is easy for people to determine the sentiment of such a sentence, it may be quite difficult for an automatic system like sentiment analysis.

"শান্তর স্বভাব হিটলারকে হার মানায়।" (Shantu's nature defeats Hitler)

"রমাকে আমরা সবাই বিশ্ববিদ্যালয়ের মাদার তেরেসা বলে জানি।" (We all know Rama as the Mother Teresa of the University).

Challenge 5

Sometimes, the sentence may have more than one entity and the gist of the sentence may be positive for some and negative for others. For example:

"ব্রাজিল এবং আর্জেন্টিনা শ্বাসরুদ্ধকর ম্যাচে আর্জেন্টিনা জয়।" ("Argentina won the breath-taking match between Brazil and Argentina.") Of course, this is good news for Argentina but at the same time, it carries a negative meaning for Brazil. Several challenges have been mentioned in the example above. It is very easy for people to understand the sentiment of this kind of sentence. But in the case of business or web portal sentiment analysis, people can't decide by analyzing the sentiment of all kinds of comments, so there is a need for an automatic sentiment analysis system. But in the case of the above examples, we find it difficult to do sentiment analysis automatically. We think that the Sentiment analyzer can be further improved by applying the rule for analyzing the sentiment of such sentences. Besides, there is also a lot of research done at the level of the sentences, but we have not been able to find such research that works with Article-Level Sentiment Analysis. We think that implementing article-level sentiment analysis will be much more effective in real life.

RQ4: Which model is comparatively better?

In the previous section, we have reviewed different papers and we propose to use Saaty (Saaty, 2002) introduced the

AHP weighted TOPSIS method to prioritize the papers. Create a decision matrix consisting of p alternatives and q attributes. Determine the attributes-Accuracy, Sentiment Category, Size, Sensitivity Noise where the Noise attribute convert into numerical values {very low= 1, low=2, medium=3, high=4, very high=5} and in terms of alternatives, various papers name Table 2 So, to employ the TOPSIS method, next steps should be followed:

Table 2: Decision Table

Alternatives	Accuracy	Category	Size	Noise
Sharif *et al.*	80.48%	2	1000	v. low
Prasad *et al.*	74.00%	3	999	low
Tabassum & Khan	87.00%	2	1050	low
Tripto & Ali	66.00%	5	8910	low
Nabi *et al.*	83.00%	3	1500	medium
Chowdhury & Chowdhury	93.00%	2	1300	v. high
Aziz *et al.*	85.67%	2	10,000	low
Tuhin *et al.*	90.00%	6	7500	v. low
Sarkar & Bhowmick	45.00%	3	1000	v. high
Shamsul *et al.*	73.49%	3	1601	v. high
Alam *et al.*	99.87%	2	850	low
Dey & Sarker	92.00%	3	5200	low
Eftekhar *et al.*	84.00%	2	2000	v. low
Tapasy *et al.*	71.23%	2	6368	low
Islam *et al.*	72.00%	2	2000	v. low
Ismail *et al.*	79.06%	2	10,000	v. low
Khondoker al.	71.00%	3	17,852	low
Al-Amin *et al.*	75.5%	2	15000	low
Asif *et al.*	78.00%	3	9337	low
Sazzed & Jayarathna	65.40%	3	3505	high

Normalization: The decision matrix M_{pq} is then normalized to form the matrix $M_{\sim\sim}$

$$M`_{pq} = \frac{M_{ij}}{\sqrt{\sum_{j=1}^{q} M_{ij}^2}} \quad (1)$$

Where, i= 1,2,3...p and j= 1,2,3...q. Then, we calculate the weighted normalized matrix using

$$C = M`_{pq} * W_i \quad (2)$$

Where, W_i= [0.059246, 0.4828, 0.31385, 0.1441] following the Saaty scale shown as Table 3.

Table 3: Weighted normalization matrix

Alternatives	Accuracy	Category	Size	Noise
Sharif *et al.*	0.0134	0.0734	0.0096	0.0119
Prasad *et al.*	0.0123	0.1101	0.0096	0.0239
Tabassum & Khan	0.0145	0.0734	0.0100	0.0239
Tripto & Ali	0.0110	0.1835	0.0856	0.0239
Nabi *et al.*	0.0138	0.1101	0.0144	0.0359
Chowdhury & Chowdhury	0.0155	0.0734	0.0124	0.0598
Aziz *et al.*	0.0143	0.0734	0.0961	0.0239
Tuhin *et al.*	0.0150	0.2202	0.0721	0.0119
Sarkar & Bhowmick	0.0075	0.1101	0.0096	0.0598
Shamsul *et al.*	0.0122	0.1101	0.0153	0.0598
Alam *et al.*	0.0167	0.0734	0.0081	0.0239
Dey & Sarker	0.0153	0.1101	0.0499	0.0239
Eftekhar *et al.*	0.0140	0.0734	0.0192	0.0119
Tapasy *et al.*	0.0119	0.0734	0.0612	0.0239
Islam *et al.*	0.0120	0.0734	0.0192	0.0119

Ismail *et al.*	0.0132	0.0734	0.0961	0.0119
Khondoker al.	0.0118	0.1101	0.1716	0.0239
Al-Amin *et al.*	0.0126	0.0734	0.1442	0.0239
Asif *et al.*	0.0130	0.1101	0.0897	0.0239
Sazzed & Jayarathna	0.0109	0.1101	0.0336	0.0478

Calculate the ideal best alternative and worst alternative solution

The best alternative ideal solution has the form L^+ and the worst ideal solution has the form L^- where accuracy, category, and size beneficial criteria are low-value desire, and noise non-beneficial criteria are high-value desire that is presented in Table 4.

Table 4: Ideal Solution

Alternatives	L+	L-
Accuracy	0.0167	0.0075
Category	0.2202	0.0734
Size	0.1716	0.0081
Noise	0.0119	0.0598

Rank the Alternatives:

Equation 3 explains the calculation of the TOPSIS, which is the determination of the closeness, S of alternatives. E is calculated from the Euclidian distance table.

$$S_p^+ = \frac{E_p^-}{(E_p^+ + E_p^-)} \quad (3)$$

The alternatives with the highest closeness score are considered the best-preferred alternative. In our case, the paper turns out to be the best-preferred paper among those measured in this work followed by papers.

Table 5: Euclidean distance and Ranking

Alternatives	E^{+p}	E^{-p}	S^{+p}
Sharif *et al.*	0.2186	0.0482	0.1808
Prasad *et al.*	0.1963	0.0506	0.2081
Tabassum & Khan	0.2186	0.0366	0.1435
Tripto & Ali	0.0944	0.1393	0.5962
Nabi *et al.*	0.1934	0.0447	0.1877
Chowdhury & Chowdhury	0.2217	0.0091	0.0395
Aziz *et al.*	0.1655	0.0952	0.3652
Tuhin *et al.*	0.0995	0.1673	0.6269
Sarkar & Bhowmick	0.2018	0.0367	0.1539
Shamsul *et al.*	0.1970	0.0377	0.1606
Alam *et al.*	0.2200	0.0370	0.1441
Dey & Sarkar	0.1645	0.0666	0.2884
Eftekhar *et al.*	0.2116	0.0495	0.1897
Tapasy *et al.*	0.1841	0.0642	0.2585
Islam *et al.*	0.2116	0.0493	0.1890
Ismail *et al.*	0.1651	0.1003	0.3778
Khondoker *et al.*	0.1108	0.1713	0.6071
Al-Amin *et al.*	0.1498	0.1407	0.4843
Asif *et al.*	0.1377	0.0965	0.4120
Sazzed & Jayarathna	0.1802	0.0464	0.2048

CONCLUSIONS

In our systematic review, we tried to bring out the procedure that exists along with its limitations and achievements. In our review paper, we not only demonstrated the existing method but also highlighted the pre-processing steps, features, improvement scope, performance evaluation, etc. We also tried a model named the TOPSIS model by which we have tried to specify the best model among the several models of Sentiment Analysis for Bengali. Due to the abundance of Bengali language data on the Internet at present, researchers are becoming interested in implementing this sentiment analysis model in Bengali. This SLR will pave a clear view of research in this domain.

In the future, we will study how to get high performance in this domain for the Bengali language. We are also determined to design a model for the implantation of Bengali Sentiment Analysis.

REFERENCES

Alam, M. H., Rahoman, M. M., & Azad, M. A. K. (2018). Sentiment analysis for Bangla sentences using convolutional neural network. 20th *International Conference of Computer and Information Technology, ICCIT* 2017, 1–6. https://doi.org/10.1109/ICCITECHN.2017.8281840

Arafin Mahtab, S., Islam, N., & Mahfuzur Rahaman, M. (2018). Sentiment Analysis on Bangladesh Cricket with Support Vector Machine. 2018 *International Conference on Bangla Speech and Language Processing, ICBSLP*, 1–4. https://doi.org/10.1109/ICBSLP.2018.8554585

Aziz Sharfuddin, A., Nafis Tihami, M., & Saiful Islam, M. (2018). A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification. 2018 *International Conference on Bangla Speech and Language Processing, ICBSLP* 2018, September. https://doi.org/10.1109/ICBSLP.2018.8554396

Chavan, G. S., Manjare, S., Hegde, P., & Sankhe, A. (2014). *A Survey of Various Machine Learning, 15*(6), 288–292.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014 *Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734. https://doi.org/10.3115/v1/d14-1179

Chowdhury, S., & Chowdhury, W. (2014). Performing sentiment analysis in Bangla microblog posts. 2014 *International Conference on Informatics, Electronics and*

Vision, ICIEV 2014. https://doi.org/10.1109/ICIEV.2014.6850712

Dash, N. S., & Ramamoorthy, L. (2018). Utility and application of language corpora. Utility and Application of Language Corpora, Sasaki 2003, 1–290. https://doi.org/10.1007/978-981-13-1801-6

De Silva, J., & Haddela, P. S. (2013). A term weighting method for identifying emotions from text content. 2013 *IEEE 8th International Conference on Industrial and Information Systems, ICIIS 2013-Conference Proceedings,* 381–386. https://doi.org/10.1109/ICIInfS.2013.6732014

Dey, R. C., & Sarker, O. (2019). Sentiment analysis on bengali text using lexicon based approach. 2019 22nd *International Conference on Computer and Information Technology, ICCIT 2019*, 1–5. https://doi.org/10.1109/ICCIT48885.2019.9038250

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks, 1*(2), 119–130. https://doi.org/10.1016/0893-6080(88)90014-7

GitHub-atikdu/Bangla_ABSA_Datasets.(n.d.). Retrieved May 29, 2022, from https://github.com/atikdu/Bangla_ABSA_Datasets

Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *Journal of Educational and Behavioral Statistics, 44*(3), 348–361. https://doi.org/10.3102/1076998619832248

Hasan, K. M. A., Rahman, M., & Badiuzzaman. (2014). Sentiment detection from Bangla text using contextual valency analysis. 2014 17th *International Conference on Computer and Information Technology, ICCIT 2014*, 292–295. https://doi.org/10.1109/ICCITechn.2014.7073151

Hassan, A., Amin, M. R., Azad, A. K. Al, & Mohammed, N. (2017). Sentiment analysis on bangla and romanized bangla text using deep recurrent models. IWCI 2016 - 2016 *International Workshop on Computational Intelligence,* 51–56. https://doi.org/10.1109/IWCI.2016.7860338

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hossain, E., Sharif, O., & Moshiul Hoque, M. (2021). Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naïve Bayes. *Advances in Intelligent Systems and Computing,* 281–292. https://doi.org/10.1007/978-981-33-4299-6_23

Islam, K. I., Islam, M. S., & Amin, M. R. (2020). Sentiment analysis in Bengali via transfer learning using multi-lingual BERT. ICCIT 2020 - 23rd *International Conference on Computer and Information Technology, Proceedings*, 19–21. https://doi.org/10.1109/ICCIT51783.2020.9392653

Islam, M. S., Islam, M. A., Hossain, M. A., & Dey, J. J. (2017). Supervised Approach of sentimentality extraction from Bengali facebook status. 19th *International Conference on Computer and Information Technology, ICCIT 2016,* March 2018, 383–387. https://doi.org/10.1109/ICCITECHN.2016.7860228

Keras: the Python deep learning API. (n.d.). Retrieved August 29, 2022, from https://keras.io/

Khan, M. R. H., Afroz, U. S., Masum, A. K. M., Abujar, S., & Hossain, S. A. (2020). Sentiment Analysis from Bengali Depression Dataset using Machine Learning. 2020 11th *International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*. https://doi.org/10.1109/ICCCNT49239.2020.9225511

Mahmudun, M., Tanzir, M., & Ismail, S. (2016). Detecting Sentiment from Bangla Text using Machine Learning Technique and Feature Analysis. *International Journal of Computer Applications, 153*(11), 28–34. https://doi.org/10.5120/ijca2016912230

Mandal, P., & M Mainul Hossain, B. (2017). A Systematic Literature Review on Spell Checkers for Bangla Language. *International Journal of Modern Education and Computer Science, 9*(6), 40–47. https://doi.org/10.5815/ijmecs.2017.06.06

Md Al-Amin, Islam, M. S., & Das Uzzal, S. (2017). Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words. ECCE 2017 - *International Conference on Electrical, Computer and Communication Engineering,* 186–190. https://doi.org/10.1109/ECACE.2017.7912903

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. 1st *International Conference on Learning Representations, ICLR 2013* - Workshop Track Proceedings, 1–12.

Mohammad, S. M. (2016). Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. *Emotion Measurement,* 201–237. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Prasad, S. S., Kumar, J., Prabhakar, D. K., & Tripathi, S. (2017). Sentiment mining: An approach for Bengali and Tamil tweets. 2016 9th *International Conference on Contemporary Computing, IC3 2016,* 1–4. https://doi.org/10.1109/IC3.2016.7880246

Qiu, R., & Li, D. (2016). *The Importance of Neutral Class in Sentiment Analysis of ArabicTweets, 8*(2), 17–31. https://doi.org/10.5121/ijcsit.2016.8202

Rabeya, T., Chakraborty, N. R., Ferdous, S., Dash, M., & Al Marouf, A. (2019). Sentiment Analysis of Bangla Song Review- A Lexicon Based Backtracking Approach. Proceedings of 2019 3rd *IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2019,* 1–7. https://doi.org/10.1109/ICECCT.2019.8869290

Saaty, T. L. (2002). Decision making with the Analytic Hierarchy Process. *Scientia Iranica, 9*(3), 215–229. https://doi.org/10.1504/ijssci.2008.017590

Saranya, K., & Jayanthy, S. (2018). Onto-based sentiment classification using machine learning techniques. Proceedings of 2017 *International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS* 2017, 2018-Janua, 1–5. https://doi.org/10.1109/ICIIECS.2017.8276047

Sarkar, K., & Bhowmick, M. (2018). Sentiment polarity detection in Bengali tweets using multinomial Naïve Bayes and support vector machines. 2017 *IEEE Calcutta Conference, CALCON 2017* - Proceedings, 2018-Janua, 31–35. https://doi.org/10.1109/CALCON.2017.8280690

Savanur, S. R., & Sumathi, R. (2018). Feature Based Sentiment Analysis of Compound Sentences. 2017 2nd *International Conference On Emerging Computation and Information Technologies, ICECIT* 2017, 1–6. https://doi.org/10.1109/ICECIT.2017.8453357

Sazzed, S., & Jayarathna, S. (2019). A sentiment classification in bengali and machine translated english corpus. Proceedings - 2019 IEEE 20th *International Conference on Information Reuse and Integration for Data Science, IRI 2019*, 107–114. https://doi.org/10.1109/IRI.2019.00029

Sharif, O., Hoque, M. M., & Hossain, E. (2019). Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes. 1st *International Conference on Advances in Science, Engineering and Robotics Technology 2019,* ICASERT 2019, 1–6. https://doi.org/10.1109/ICASERT.2019.8934655

Siddiqi Emon, M. I., Ahmed, S. S., Milu, S. A., & Mahtab, S. S. (2019). Sentiment Analysis of Bengali Online Reviews written with English Letter Using Machine Learning Approaches. *ACM International Conference Proceeding Series*, 109–115. https://doi.org/10.1145/3362966.3362977

'snltr-software' [online]. available: http://nltr.org/snltr-software/ [accessed: - Google Search. (n.d.). Retrieved May 29, 2022, from https://www.google.com/search?

Tabassum, N., & Khan, M. I. (2019). Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning. 2nd *International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, 1–5. https://doi.org/10.1109/ECACE.2019.8679347

Tripto, N. I., & Ali, M. E. (2018). Bangla YouTube Comments. 21–22.

Tuhin, R. A., Paul, B. K., Nawrine, F., Akter, M., & Das, A. K. (2019). An automated system of sentiment analysis from Bangla text using supervised learning techniques. 2019 IEEE 4th *International Conference on Computer and Communication Systems, ICCCS 2019*, September, 360–364. https://doi.org/10.1109/CCOMS.2019.8821658

Twitter NLP Example: How to Scale Part-of-Speech Tagging with MPP (Part 1). (n.d.). Retrieved August 30, 2022, from https://tanzu.vmware.com/content/blog/twitter-nlp-example-how-to-scale-part-of-speech-tagging-with-mpp-part-1

Witten, I. H., & Frank, E. (2005). Credibility: Evaluating What's been Learned. In Data Mining: Practical machine learning tools and techniques. shorturl.at/efgX1

বাংলা ব্যাকরণ – Bangla Library. (n.d.). Retrieved May 29, 2022, from https://www.ebanglalibrary.com/category/বাংলা-ব্যাকরণ