# Emotion Analysis from Bangla Text Data

Arindam Dey
ID: 18-38458-2
Section: E
Group: 12
*Department of Computer Science and Engineering*
*American International University- Bangladesh*
Bagerhat, Bangladesh
deyarindam515@gmail.com

Umma Khadiza
ID: 18-38311-2
Section: E
Group: 12
*Department of Computer Science and Engineering*
*American International University- Bangladesh*
Kushtia, Bangladesh
ummakhadiza786@gmail.com

Tahmeed Mahbub
ID: 18-38743-3
Section: E
Group: 12
*Department of Computer Science and Engineering*
*American International University- Bangladesh*
Dhaka, Bangladesh
tahmeedmahbub@gmail.com

***Abstract*—Emotions are a vital element of human connection in everyday life. Written language, vocal communication, and facial expressions can all be used to convey emotions. The habit of expressing emotion on social media or blogs has exploded in popularity in recent years. People express their views and ideas on a variety of political and global topics through writing. All of these social activities have necessitated the collection and analysis of human emotion from the text. Although the topic of emotion detection has been extensively researched for the English language, research into this domain for the Bangla language is still in its early stages. The authors offer a Sentiment Analyzer that detects the sentiment or opinion of Bangla speakers on a given topic from Bangla text. They create certain phrase patterns and determine the sentiment orientation of those patterns. They create the phrase pattern for positive and negative mood by adding tags to words in the Bangla text.**

***Index Terms*—affection, emotion, definitions, history, passion, semantics**

## I. Introduction

Sentiment analysis is a method of detecting emotional biases concealed within a string of words to better comprehend an online mention's views, beliefs, and emotions. An overall context polarity or emotional response to a material is determined via sentiment analysis [1]. Sentiments are by definition subjective. Due to the volume of daily postings on social media, collecting people's views is challenging [2]. Various individuals may perceive a text's attitude differently. Bangla has the second-most speakers and is the sixth most commonly spoken language [3]. The main language of over 200 million people, 160 million of whom are Bangladeshis [4]. The usage of social media has increased the number of people expressing their views on many issues through websites such as Facebook, Twitter, and others [5]. 90 percent of today's data was collected in the past two years, and analyzing it is not easy [6]. Because these companies have little or no human interaction, it is difficult to monitor and analyze market trends, especially when tracking customer responses to their goods or services. [7]. In natural language processing (NLP), affect analysis recognizes the emotional component of text. The same material may be delivered in many emotional ways [8]. This sentiment analysis technique's primary aim is to identify whether a written document in Bangla is happy/sad/angry/tender/excited/frightened. We choose these emotion classes because they may represent a wide range of emotions. Words like joy, grin, optimism, laugh, and pleasure are examples [9]. Sentiment analysis has several levels of scope.

Identify applicable funding agency here. If none, delete this.

## II. Importance of Research

Sentiment analysis is a highly significant field in the processing of natural languages. Generally speaking, sentiment categorization implies the analysis to identify whether a speaker's views on a certain topic are positive or negative. Stimulus analysis may significantly affect everyone in their actual lives with the fast development of e-Commerce. For example, product reviews on the web have become an essential information source when consumers decide to purchase any goods. As consumers frequently get too many reviews, how to categorize and detect their feelings automatically has become an essential research issue. In this proposal we propose a feelings analyzer that identifies the feeling of Bangladesh or the viewpoint of a Bangladeshi topic. We create certain sentence patterns and determine the direction of their feelings. Social networks are the key means of collecting information on people's opinions and feelings about various subjects while spending hours on social media every day and sharing their opinions. In this technical article we demonstrate how sentimental analysis is used and how Twitter is connected and how sentimental questions are analyzed. We conduct tests on many topics from politics to mankind and present the findings. We noticed that the tweet neutrality is considerably high, which clearly indicates the limits of the efforts now underway. Research is more important for:

1) Presentation of a Bangladesh emotional corpus, which includes the variety of emotional expressions of fine grain in the social media language.
2) To use traditional machine-learning methods, usually good in identifying the six emotional categories described above.
3) To evaluate the performance of the machine-learning classification with a baseline to find the optimal model for the identification of emotion.
4) Pre-processing data in a manner that researchers can easily use.
5) Application of deep recurrent models on text corpus Bangla and Romanized Bangla.

6) Pre-train data set for another label to show its utility (vice versa).

### A. Research Question

- What feeling analysis can be explored?
- Who is involved to analyze emotions?
- How can we determine consumer feelings from their views (good or negative)?
- How precisely can we grasp this feeling?

## III. Objective

As we all know, our project concept is centered on emotional analysis from text data depending of Bangladesh. Emotion Analysis is used instead of dividing the corpus solely on the basis of positive and negative emotions. We determine the feeling of a Banglade phrase or paragraph utilizing the validity of a word in this suggestion. We tried to think of more precise descriptions of emotions such as sadness, joy, disdain, surprise, fear and anger, all based on human sensation. There are thus six basic kinds of emotion in total. We employ a model, formulas and algorithms to determine the value of a word in the past. Six types of values correspond with the meaning of every word: sadness, pleasure, disgust, surprise, fear and anger. The senses vary according to the speech components in the linked sentence.

We may also use Bengal terms from other social media. Social media may be news, Facebook, Twitter, etc. Our primary emphasis is,

- We gathered and presented Bangla datasets for our project, which we made available.
- We conducted linguistic statistical analysis on datasets.
- We utilized state-of-the-art machine learning methods and obtained acceptable accuracies for the data sets collected.

## IV. Scope

This proposal offers a sentiment analysis method of text data produced in English. This method may automate the study of the user response to a particular emotion such as movies or TV shows in any social media. With more and more individuals expressing their views freely on social networking sites, analysis of the feeling of remark on a certain text shows how they feel. The project proposal has a lot of scope.

- We provide the detection of six fundamental emotions in this project, including anger, contempt, fear, happiness, sadness and surprise for the Bangladesh text.
- Supervised and uncontrolled learning technique for emotional analysis or opinion mining in Banglade text for a particular domain of text may be readily used. The material is partially covered by the Support Vector Machine method, various models and techniques.

## V. Design of Case Study

Some of our work in these areas is motivated by previous work and others are designed to broaden our knowledge. Our idea was mainly driven by many papers attempting to classify different groupings, such as aggressiveness and emotions. And this is one of the major works we followed. They try to extract emotions from text messages. Sensitivity analysis or text detection is a well-studied research topic for highly resourced languages such as English, Arabic and other European languages[10]. Some of them used TF-IDF to increase classification accuracy and categorize the vector support machine [11]. They are quite similar to the technique we employed in our first study[12]. However, they used the Vector Space Paradigm (VSM) as a paradigm for document representation. We introduced a large quantity of digital data (mainly opinionated texts such as statuses, comments, debates etc.) like we never had until social media appeared on internet, e.g. Facebook, Twitter, discussion forums, reviews and their fast development. Since the early 2000s, sentiment analysis has been one of the most active areas of NLP study[13]. As far as we know, some study was carried out using an examination of valuability to identify sentiments or views from Bangladesh. Some work has been done on the validity analysis of Bangladesh verbs[14][15]. Nevertheless, there are many studies dealing with emotion detection in Bangladesh. [16] suggested an automated Bangla text sentiment identification method utilizing a machine learning model. [17] utilized the Santi Word Net[18] and the WordNet Affect[19] to suggest an emotion tracking system based on a subject or event via the application of meaning affect algorithms for the grading of annoted news and blog companies in Bangladesh. They achieved a maximum emotional accuracy of 94.5% from their data set. The latest research [19] for Bengali use social communication data to detect the polarity of a bengali text whether it is emotional or not. They have utilized the Bangladesh Post-Tagger package and the Support Vector Machine and Maximum Entropy algorithms to compare the performance of these two approaches by evaluating the various sets of features. A similar collection was made in [18], which mechanically gathered 1400 Bangla tweets. However, these datasets are not accessible to the public and are very small in size. The use of recurring neural networks or RNNs have all been used for speech recognition, handwriting recognition, natural language processing and other applications. [10]. Taher et al. discovered that in their tests, SVM had the highest accuracy and 0.86 recall. N-gram Basic Bangladesh Text Mining Using Vector Support Machine [13] Negativity The separation of a word's negative postfix and its focus on the negative effect of the whole phrase were new strategies in this work. Another major investigation[17] utilized opinion mining and mood extraction in which the text's polarity was assessed as positive, negative or neutral. [11] carried out a sentiment analysis and opinion mining survey by analyzing the text categorization. In addition,[14] carried out a survey of text data for sentiment analysis. [13] looked examined and monitored the emotions conveyed in the English and Bengali texts. Through this literature, we investigated Bengali language processing.

## VI. Methodology

Methodology may assist create a single blueprint for the solution presented. In this research, emotional analysis of

Bangla text data was suggested. We want to assess our work on the basis of computing speed and accuracy. A precision and recall factor will be used to evaluate accuracy. In this research, six different methods were utilized.

### A. Methods

1) The ratio of speech components.
2) Cosine similarity TF-IDF.
3) Calculate cosine similarity using a custom TF-IDF.
4) Uni-gram and Nave Bayes stammer model
5) a Bi-gram Bayes model and a normalizer are employed.

## VII. Experiments of the Methods

### A. Parts of Speech Ratio

In this approach, the favourable and unfavourable data sets are utilized as classifiers. The pleasant and unpleasant data sets are tagged with a POS tagger first. Then the nouns, adjectives, verbs, pronouns and conjunctions of both classifiers are tallied. The POS ratio is computed for both beneficial and harmful classifiers. The POS is tagged and the POS ratio computed when a query is received. Then the hamming distance is calculated between the desirable and undesirable classifier and the query. The classifier is defined according to the minimal distance.

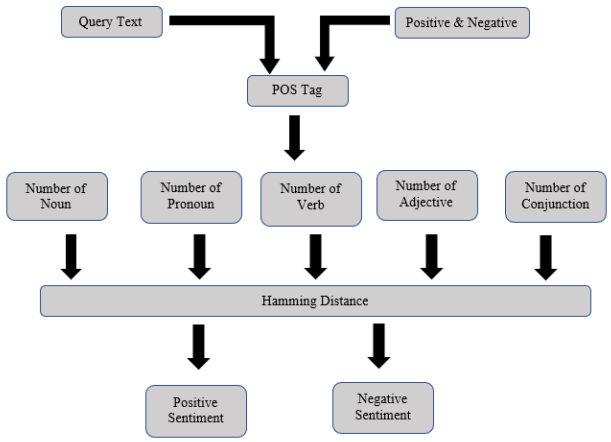


Fig. 1. Parts of speech ratio model.

### B. Cosine Similarity Using TF-IDF

The cosine similarity of documents is calculated here. The positive data set is D0C1, the negative data set is D0C2, and the query dataset is D0C3. The primary aim is to find the greatest measure of correlation between D0C1 and D0C3 and D0C2 and D0C3. Three vectors have been generated to detect cosine similitude: D0C1 as PV, D0C2 as NV and D0C3 as NV as a QV (Query Vector). Term Frequency (TF): Every single word in the document is listed and its frequencies are tallied. The number of documents containing the single phrase (document frequency, or DF). The vectors are computed when the TF, DF and IDF are acquired for both documents and queries.

### C. Cosine similarity using custom TF-IDF

Because of the many parts of speech in Bangla, the same words might have multiple meanings. As an example, ‘’(star) can be used as Noun. Again, ‘’(they) can be used as Pronoun. Here, the Bangla word ‘’ has different meanings when it’s used in different sentences. Like, “ ” “ ” A custom POS tagger is employed to solve this problem, which recognizes the portions of speech in a text. The words “” (star) (Noun) and “” (they) (Pronoun) are treated differently while looking for the TF and DF. The technique is then repeated as in the preceding section.

TABLE I
Cost Estimate

| Sl. No. | Tasks | Quantity | Unit Price | Total Price |
|---|---|---|---|---|
| 1. | Data Collection | N/A | N/A | 2000tk |
| 2. | Experiments | 5 | 1000tk | 5000tk |
| 3. | Implementation | NA | 1000tk | 1000tk |
| 4. | Analysis | 5 | 500tk | 2500tk |
| 5. | Final Execution | NA | 2000tk | 2000tk |

## VIII. Work Schedule

| Task Name | 20/5 | 30/5 | 19/6 | 2/7 | 16/7 | 4/8 | 19/8 | 25/8 |
|---|---|---|---|---|---|---|---|---|
| Proposal Submission | | | | | | | | |
| Project objective | | | | | | | | |
| Deliverable and due dates | | | | | | | | |
| Rules and responsibilities | | | | | | | | |
| Tracking and Execution tools | | | | | | | | |
| Simulation | | | | | | | | |
| Coding | | | | | | | | |
| Progress Report | | | | | | | | |
| Implement and analysis | | | | | | | | |
| Writing and presentation | | | | | | | | |
| Final Report | | | | | | | | |

## IX. Conclusion

The limits of the study are those elements of the design or technique that have affected or impacted the analysis of the findings. Like all other perspectives, feeling is basically subjective and may even be irrational. When trying to quantify feelings, a large — and appropriate — sample of data is essential to utilize. There is no data point that is not important. It’s the whole thing that matters. One or more indirect factors may influence a person’s sentiment towards a brand or product, such as someone having an awful day and tweeting a negative remark about doing something they normally saw as very neutral. With a sufficiently enough sample, the outlets are diluted. Also, because a person’s mood, global events and so on are highly likely to alter the feeling over time, it is generally essential to examine the data from the time point.” A quasi-experiment was conducted to verify the emotion recognition applicability in an e-learning environment. It was based on

a typical online course for using a software program with monitoring of user emotion recognition channels. Today's examination of feelings is exact. However, nuances such as irony, humor, or sarcasm with a basic examination of feelings are difficult to discern.

## REFERENCES


[1] M. Rahman, "Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique," International Journal of Computer Applications , vol. 176 , no. 0975 – 8887, pp. 13-17, 17, April 2020.

[2] "Survey on Text-Based Sentiment Analysis of Bengali Language," 1st International Conference on Advances in Science, Engineering and Robotics Technology , vol. 978, no. 7281, pp. 1-6, 2019.

[3] M. S. H. Rumman Rashid Chowdhury , "Analyzing Sentiment of Movie Reviews in Bangla by Applying Machine Learning Techniques," International Conference on Bangla Speech and Language Processing, vol. 7281, no. 978, pp. 22-27, 2019.

[4] M. R. A. Asif Hassan, "Sentiment Analysis on Bangla and Romanized Bangla Text (BRBT) using Deep Recurrent models," International Conference, vol. 567, no. 192, pp. 12-17, 2014.

[5] M. H. R. Chowdhury, ""Bangla handwritten character recognition using convolutional neural network with data augmentation," International Conference on Bangla Speech and Language Processing, vol. 376, no. 1998, p. 04, 2019. [6] "Sentiment Detection from Bangla Text using Contextual Valency Analysis," K. M. Azharul Hasan, vol. 668, no. 12, pp. 2324, 2014.

[6] Rahaman, M. A., Jasim, M., Ali, M. H., Hasanuzzaman, M. (2020). Bangla language modeling algorithm for automatic recognition of hand-sign-spelled bangla sign language. Frontiers of Computer Science, 14(3), 1-20.

[7] S. C. a. P. Bhattacharyya, "Valency Analyzer of Verb Arguments for Bangla" proceeding," International Conference on Bangla Speech and Language Processing, vol. 766, no. 345, pp. 13-17, 2011.

[8] C. Strapparava, "The affective weight of the lexicon," International Conference on Language Resources, vol. 4481, no. 474, pp. 34-38, 2006.

[9] k. khan, "Data analysis," International Conference on Computer and Communication Systems, vol. 998, no. 123, pp. 7982, 2019.

[10] H. V. L. T. P. Le, "Aspect analysis for opinion mining of vietnamese text," international conference on advanced computing and applications (ACOMP), vol. 1, no. 1109, pp. 118-123, 2015.

[11] J. D. S. a. P. S. Haddela, "A term weighting method for identifying emotions from text content," 4th International Conference on Electrical Information and Communication Technology (EICT), vol. 978, no. 7281, p. 381–386, 2019.

[12] M. K. B. G. S. R. A. R. Mehra, "Sentimental analysis using fuzzy and naive bayes.," Computing Methodologies and Communication (ICCMC), 2017 International Conference , vol. 168, no. 991, p. 945–950, 2017.

[13] B. L. L. a. S. V. Pang, "sentiment classification using machine learning techniques.," the ACL-02 conference on Empirical methods in natural, vol. 10, no. 654, pp. 217-220, 2002.

[14] S. C. a. P. Bhattacharyya, "Valency Analyzer of Verb Arguments for Bangla," International Conference on Bangla Speech and Language Processing(ICBSLP, vol. 78, no. 4799, pp. 278-282, 2011.

[15] B. a. L. L. Pang, "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval," International Journal of Information Engineering and Electronic Business(IJIEEB), Vols. 2(1-2), no. 135, p. 1–135, 2008.

[16] M. S. I. K. M. Azharul Hasan, "Sentiment Recognition from Bangla Text," 17th Int'l Conf. on Computer and Information Technology, vol. 3, no. 6, pp. 316-327, 2013.

[17] A. . B. S. Das, "Phrase level polarity identification for Bengali," International Journal of Computational Linguistics and Applications, vol. 1, no. 2, p. 169–181, 2010.

[18] A. a. S. Esuli, "SENTIWORDNET: A publicly available lexical resource for opinion mining," Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), vol. 06, no. 5, pp. 12-25, 2006.

[19] C. a. V. A. Strapparava, "Wordnetaffect: An affective extension of WordNet," International Conference on Bangla Speech and Language Processing(ICBSLP), vol. 2, no. 4, pp. 1083-1086, 2004.