A Combination Study Of Multiple Sentiment Analysis Using Artificial Intelligence

[a]Ismot Jahan Sarna, [b]Projna Samajdar, [c]Syeda Lamima Farhat

ismot15-10418@diu.edu.bd, projna15-10406@diu.edu.bd, syeda15-10452@diu.edu.bd
[a]Daffodil international university, Dhaka-1212, Bangladesh
[b]Daffodil international university, Dhaka-1212, Bangladesh
[c]Daffodil international university, Dhaka-1212, Bangladesh

Abstract

Sentiment analysis is a form of NLP that determines whether data is positive, neutral or negative. Emotion extraction, Opinion mining and trend prediction all benefit from sentiment analysis. We read some papers that sentiment analysis is used in different ways like aspect based. A small amount of work has been done in sentiment analysis in Bangla. In this survey paper represents the comparison of emotion in sentiment analysis. There are very little works in sentiment analysis in Bengali language. Recently Covid -19 problem as people are in lockdown, they cannot shop in physically so online is the better solution for this problem. People are highly attracted online platform at that moment. People benefit from online product service providers, but it also raises concerns about product quality and service. It is another sentiment analysis platform. Sentiment analysis is a process that the user opinions and views against any service or a product. YouTube is one of the most popular Video sharing platforms obtaining millions of views. So YouTube is also another sentiment analysis platform. There has been a lot of work in the sentiment analysis sector with deep learning, NLP, Machine learning.

Keyword: Sentiment analysis, Bangla language, Machine learning, NLP, Natural Language Processing, Bangla Sentiment analysis.

1. INTRODUCTION

We are surrounded by billions or trillions of data in this age of the internet, and sentiment analysis is one of the most common methods in order to comprehend it. Sentiment analysis detects polarity (positive, strongly positive, negative, strongly negative, neutral) with in a sentence or a word. English, Bangla, Hindi, Urdu, Romanized, Chinese and many others language used in same and different analysis. This sentiment analysis works to emotion detection, news categorization, authorship attribution, reviews in product, movie,restructure, tweets in tweeter. This sentiment analysis is required for emotional detection, news categorization, author specialization, product review, movie review, restructure review, Tweet in Twitter related work. Such a level of research is the most in-depth edition, capable of detecting nuanced viewpoints in feedback. In ABSA (Aspect Based Sentiment Analysis), there are two main tasks to complete. A first step is really to isolate the particular areas or factors listed in the opinion analysis, followed by determining the polarization (positive, negative, or neutral) of every factor. One of the first is to identify the particular fields or topics identified in the analysis and forecast, and the next is to determine the polarization (positive, negative, or neutral) to every aspect. This paper uses two categories of datasets which are "cricket" and "restaurant"[3].Text categorization is a well-known topic in NLP today, with a variety of real-world implementations. It refers to the process of categorizing textual units such as sentences, requests, chapters, and papers. Text classification is used in a

variety of applications, including emotion analysis, news classification, client purpose identification, controlling, and so on. Web pages, social media, internet news portals, newsletters, online retailers, user feedback, and questions and answers from customer support are used to gather data for text classification. Naive Bayes, Support Vector Machine, Hidden Markov Models, and random forests have all been included in this analysis article [4]. Natural language processing (NLP) and machine learning techniques are used in a text sentiment analysis application to attribute weighted sentiment scores to people, topics, patterns, and groups within a sentence or phrase. Machine learning techniques are used in sentiment analysis. Two methods are used here unsupervised and supervised. Unsupervised models are more efficient than supervised model [2]. Sentiment Analysis is a branch of NLP that develops programs that attempt to extract viewpoints from text in natural language comprehension. The annual SemEval (Semantic Evaluation, a well-known seminar in the field of NLP) competition was added to the ABAS task list in 2014. At first there have used English dataset for the ABSA and then later they used multilingual (Arabic, Czech, and French dataset) [1]. In this paper, we overview and analyzing some paper we got some idea about sentiment analysis. Using those idea, we compare those model and algorithm and find best output for sentiment analysis and best accuracy for sentiment analysis result. And we proposed a method which is standard and latest method for sentiment analysis.

2. METHODOLOGY

Various studies have been carried on the analysis of sentiment. Different types of technology have been used for sentiment analysis. This paper's approach is divided into different categories: data acquisition, data pre-processing, data interpretation, and visualizing the results. This paper using in different techniques. The method of writing the methodology of each review paper is different. To build our proposal dataset, we took several systematic steps to gather, pick, format, and translate Bengali language terms [5]. Sentiment analysis includes work related to deep learning algorithms. The authors use LSTM and CNN to identify feelings and emotions from YouTube comments.A comparison of traditional - SVM - and deep learning - also CNN and LSTM - algorithms for critical identification of Bengali News comments was given in the report. The researchers used multi - channel Convolutional-LSTM (MConv-LSTM) systems to capture embeddings in order to model different forms of hateful speech, text analysis, and emotion observation in Bengal. Because of the prevalence of Romanized Bengali text on media platforms, researchers are now using LSTM to develop and test a sentiment analysis model. [5]. In ABSA using data collection, processing, Feature extraction, fitting algorithm to train. Also using mechanic learning algorithm such as SVM (Support Vector Model), RL (Random Forest), LR (Linear Regression), KNN (N-Nearest Neighbor) in ABSA analysis.

• Data collection:

We will create the dataset first. We extract comments or emotion from different platform. We utilize Google Translate to determine whether each comment is written in Bangla or English. Romanized Bengali refers to sentences that haven't been assigned to a language. We use a threeor five-class sentiment detection issue to annotate each sentence in the database. It can be linked to both emotional experiences. Surprisingly, the number of basic human emotions is quite large [9].

• Processing:

Text analysis requires pre - processing in order for the model to grasp the data. Because text data contains a lot of noise, cleaning the messages is difficult. Removing special characters is the first step. They can be perplexing at times. Second, we eliminate punctuation marks. The removal of punctuations is a common and widely used preprocessing technique. Collected data samples are manually annotated into one of three categories: positive (1), negative (0) and Neutral (A). Each text sample was independently manually annotated by two different native Bangla speaking individuals for total two validations. Every annotator independently

verified the data without being aware of the decisions made by the others. This guarantees that the validations are both objective and personal [6].

• Feature:

In the case of ABSA, they have divided the texts into some features to facilitate their work. this are: BOW, TFIDF, TF, IDF [1].

• Fitting algorithm to train:

Finding a high-performing machine learning technique for a specific dataset might be difficult. We used a "trial - and - error" method to come up with a sufficient list of algorithms that work with those datasets. They've chosen the most commonly used classification algorithms. RF, NB, LR, KNN, LSTM and SVM are examples.

Table: Summary of results

Papers	Techniques	Data set	Approaches	Accuracy
Aspect Based	SVM,KNN RF, ,LR, NB	Cricket	RF	37%
	SVM,RF,KNN,LR, NB	Restaurant	LR	43%
Evaluation	SVM, RF, KNN	Cricket	RF	37%
	SVM, RF, KNN	Restaurant	KNN	42%
Bangla Text Classification	LSTM, CNN, NB, SVM	YouTube 3 Class	LSTM	63.5%
	SVM, CNN, LSTM	Comment 3 Class	LSTM	79.29%
Sentiment analysis of E-commerce product	RF, LR, SVM, KNN, XGBoot	Daraz	KNN	96.25%
Product review sentiment analysis	KNN,Decision tree, SVM, RF, LR	chaldal.com, Daraz	SVM	88.81%

Sentiment and Emotions from Bangla YouTube Comments	LSTM, CNN, NB, SVM	YouTube 5 Class	LSTM	54.24%

3. PROPOSE METHOD

Sentiment analysis is being used for spam review / comment detection as well as fraud app identification. As a result, expanding research to include more non-English languages and aspects can lead to a more precise understanding of users and their reviews, which can help businesses make better decisions and increase cyber security. A system that will assess client feedback from an internet platform and provide a ratio of good and negative feedback from past customers written in Bangla.

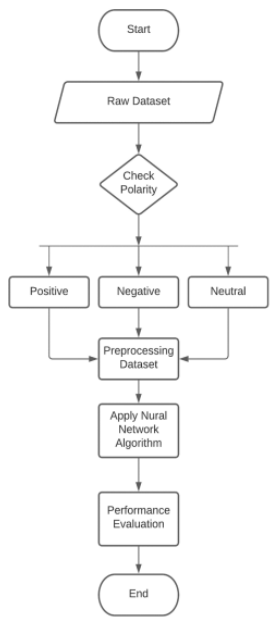


Fig 1: Proposed Methodology Diagram

In figure-1, We want to apply our proposed method in Bangla sentiment analysis. After Data collection we will preprocess our dataset for our propose model. We will annotate each sentence in the database according to three (positive, negative and neutral) or five (positive, negative, neutral, highly positive and highly negative) Classes of sentiment identification problems. Then we will apply the neural network algorithms on the dataset. Because we know that, neural network algorithm is very powerful for this scenario. In future, we aim to include multiple aspects and topic information in sentiment and emotion detection. next time, we want to do language independent through sentiment analysis using multiple languages like (Bangla & English) together and find out its polarity. The development of a multilingual approach to sentiment analysis is therefore a major challenge.

4. DISCUSSION AND CONCLUSION

In this paper, we have reviewed some papers those are related in sentiment analysis. This paper we have high priority of Bangla Language. All of the reviewed paper is using to different strategy and also using multiply algorithm. Machine learning has used several algorithms to get good efficiency in all the papers and has got good results using the algorithms.After applying the neural network algorithms, we will see which algorithm's efficiency is better. The better the performance of the algorithm, the higher its accuracy. We get a better result from sentiment analysis of E-commerce product. Here, they use some old classification and analyzing algorithm. They get 96.25% accuracy. After study all the paper, we do not get any paper solutions or best fit for future technology. Although, we finish study more and more about sentiment analysis technique, but we have some limitations. First limitation, we cannot find our valuable dataset for our model. So, we cannot implement our propose method properly. As things currently stand now, we collect data for making a proper dataset for our propose method. Every study intended to increase the efficiency of the plagiarism detection system by using a Natural Language Processing approach. In future, we want to work with Bangla sentiment analysis tool and multiple sentiment detection.

Acknowledgements

We would like to express our gratitude to our PetaAI Solutions for providing all research resources. Our supervisor deserves special thanks for his patience and support in helping us overcome the lot of obstacles we encountered during our research.

References

Clark, T., Woodley, R., De Halas, D., 1962. Gas-Graphite Systems, in *"Nuclear Graphite"* R. Nightingale, Editor. Academic Press, New York, p. 387.

Deal, B., Grove, A., 1965. General Relationship for the Thermal Oxidation of Silicon, Journal of Applied Physics 36, p. 3770.

Deep-Burn Project: Annual Report for 2009, Idaho National Laboratory, Sept. 2009.

Fachinger, J., den Exter, M., Grambow, B., Holgerson, S., Landesmann, C., Titov, M., Podruhzina, T., 2004."Behavior of spent HTR fuel elements in aquatic phases of repository host rock formations," 2nd International Topical Meeting on High Temperature Reactor Technology. Beijing, China, paper #B08.

Fachinger, J., 2006. Behavior of HTR Fuel Elements in Aquatic Phases of Repository Host Rock Formations. Nuclear Engineering & Design 236, p. 54.

Haque, S., Asif khan Shakir, A. k., & Shakir, A. k. (n.d.). ASPECT BASED SENTIMENT ANALYSIS IN BANGLA DATASET BASED ON ASPECT TERM EXTRACTION.

Bhagat, A., Sharma, A., & Chettri, S. K. (6, June 2020). Machine Learning Based Sentiment Analysis for Text. IJCAT – International Journal of Computing and Technology, Volume 7.

Rahman, M. A., & Dey, E. K. (2018). Datasets for AspectBased Sentiment Analysis in Bangla and Its Baseline Evaluation. www.mdpi.com.

Alam, T., Khan, A., & Alam, F. (n.d.). Bangla Text Classification using Transformers.

Ali, H., & Hossain, M. F. (2020). BanglaSenti: A Dataset of Bangla Words for Sentiment Analysis.

Hassan,, A., Amin, M. R., Abul Kalam Al Azad, A. K., & Mohammed, N. (n.d.). Sentiment Analysis on Bangla and Romanized BanglaText (BRBT) using Deep Recurrent models.

Akter, M. T., Begum, M., & Mustafa, R. (2021). Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors. IEEE.

Bhowmik, N. R., Arifuzzaman, M., M. Rubaiyat Hossain Mondal, M. H., & Islam, M. (2021). Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary. Vol. 1(3-4).

Shafin, M. A., Hasan, M. M., Alam, M. R., Mithu, M. A., Nur, A. U., & Faruk, M. O. (2021). Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language. IEEE.

Tripto, N. I., & Ali, M. E. (2018). Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments. ICBSLP.

Ahmed, A., & Yousuf, M. A. (2020). Sentiment Analysis on Bangla Text Using Long Short-Term Memory (LSTM) Recurrent Neural Network. Springer, Singapore.

Dev , N. R., & Hossen, M. S. (2021). An Improved Lexicon Based Model for Efficient Sentiment Analysis on Movie Review Data.
Mukherjee, Sudipta. "Sentiment analysis." ML. NET Revealed. Apress, Berkeley, CA, 2021. 113-127.
Basiri, Mohammad Ehsan, et al. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." Future Generation Computer Systems 115 (2021): 279-294.
Farha, Ibrahim Abu, and Walid Magdy. "A comparative study of effective approaches for arabic sentiment analysis." Information Processing & Management 58.2 (2021): 102438.
de Oliveira Carosia, Arthur Emanuel, Guilherme Palermo Coelho, and Ana Estela Antunes da Silva. "Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning." Expert Systems with Applications 184 (2021): 115470.



Ismot Jahan Sarna was born Dhaka, Bangladesh on january 5, 2000. Now she is studying Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh. she is currently working on research related. Gaining knowledge about data mining, machine learning.



Projna Samajdar was born Dhaka, Bangladesh on April 5, 2000. Now she is studying Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh. she is currently working on research related. Gaining knowledge about data mining, machine learning.



Syeda Lamima Farhat, She has recently completed BSC in CSE from Daffodil International University. She is currently working on research field and getting prepared to study MS abroad. She has a keen interest on Artificial Intelligence related field. She has a good capability of adapting knowledge easily.