Sentiment Analysis from Bangla Book Reviews

Shaikh Ramisha Maliyat
ID: 20200204054
Group: B1 (01)
Department of Computer Science and Engineering

Nabila Rahman
ID: 20200204065
Group: B1 (01)
Department of Computer Science and Engineering

Arijit Paul
ID: 20200204067
Group: B1 (01)
Department of Computer Science and Engineering

***Abstract*—In this study, we perform sentiment analysis on Bangla book reviews to determine the overall sentiment expressed by readers. Utilizing natural language processing techniques, we analyze the text to classify reviews as positive, negative, or neutral. This research aims to provide insights into reader opinions and improve the understanding of sentiment trends in Bangla literature.**

I. Introduction

Sentiment analysis is a powerful tool in natural language processing that helps understand the emotional tone behind textual data. In the context of Bangla book reviews, sentiment analysis can provide valuable insights into readers' perceptions and experiences. By analyzing reviews written in Bangla, we aim to classify them into positive, negative, or neutral sentiments, thereby aiding authors, publishers, and potential readers in gauging the reception of books within the Bangla-speaking community.

II. Motivation

The rise of online platforms has allowed readers to share their opinions on books, significantly influencing the literary market and other readers' choices. While sentiment analysis is well-studied for major languages like English, there is a notable lack of research and tools for Bangla, the seventh most spoken language globally. Despite Bangla literature's rich history and vibrant contemporary scene, the sentiments in book reviews remain largely unexplored. Understanding these sentiments can provide authors and publishers with valuable feedback and help readers make more informed decisions, enhancing their overall experience.

Studies in other languages show that sentiment analysis can greatly impact marketing strategies and customer satisfaction. Extending these benefits to Bangla literature can create a more dynamic and responsive book market. Additionally, developing sentiment analysis tools for Bangla can advance natural language processing for underrepresented languages, promoting digital inclusivity. As the online Bangla-speaking population grows, tools catering to their linguistic and cultural context become increasingly essential. This research fills a critical gap and supports broader efforts to democratize access to digital tools across diverse linguistic communities.

III. Literature Review

There are some existing research on sentiment analysis for Bangla text that has focused on classifying emotions. Most of these studies have employed natural language processing (NLP) techniques to achieve this classification.

The paper [1] utilizes two main methods for sentiment analysis from Bangla text: Naïve Bayes Classification Algorithm and Topical Approach. The topical approach outperformed the Naïve Bayes classification algorithm for both sentence and article levels of scope. This paper detects sentiment from Bangla text documents by classifying text emotions into six basic categories: happy, sad, angry, tender, excited, and scared and the six emotion classes were mapped into two higher emotion classes, Positive and Negative. Above 90% accuracy for Topical approach has been achieved.

The article [2] proposed a hybrid framework for sentiment analysis on Bangla texts, combining machine learning and a rule-based approach. The combination of machine learning and rule-based methodologies in the hybrid framework contributed to achieving an impressive accuracy rate of 95.54% in the experimental analysis, surpassing the performance of previous works.

Combining word2vec co-occurrence scores with sentiment polarity enhances Bengali sentiment analysis by enriching word relationships and improving accuracy. This is shown in paper [3], with accuracy after 75.5% in a binary classification analysis.

The research [4] focuses on the sentiment analysis of Bengali book reviews using various machine learning algorithms. Among the five algorithms applied, the Random Forest algorithm stands out, achieving an impressive 98.39% accuracy in sentiment prediction. This high accuracy highlights the potential of the Random Forest algorithm for effective sentiment analysis. The results underscore the capability of machine learning techniques in handling the nuances of Bengali text for sentiment analysis.

The [5]th paper demonstrates the model's effectiveness in analyzing sentiment in Bengali text, overcoming challenges like the language's complex grammar and scarcity of labeled datasets. It uses multilingual BERT model with extensions of Gated Recurrent Unit (GRU), Long Short Term Memory

(LSTM), and Convolutional Neural Network (CNN) with accuracy: 71% for 2 class and 60% for 3 class using bert with extention GRU.

IV. Methodology

We used machine learning and deep learning techniques for automated emotion detection due to their high accuracy. Specifically, supervised learning was chosen for its ability to model and adapt to dynamic systems. Our study proposes four methods, Logistic Regression, Deep Neural Network, LSTM(Long Short-Term Memory) and CNN(Convolutional Neural Network), for emotion detection from Bangla Text Reviews across three distinct classes. Figure1 represent the simple working procedure for this study-

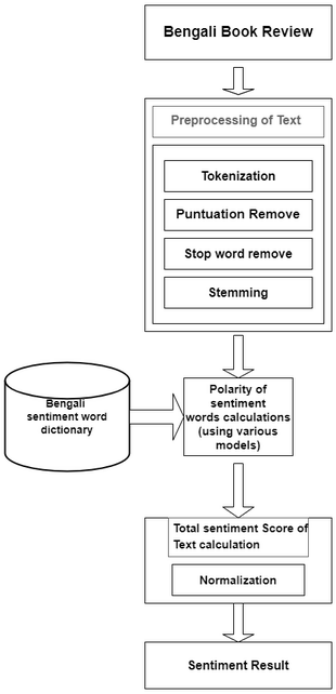

Fig. 1. Model of working mechanism

We used a sentiment analysis dataset [6] with 158,065 samples classified into positive, negative, neutral classes. The methodology for sentiment analysis of Bengali book reviews involves preprocessing the text through tokenization, punctuation removal, stop word removal, and stemming. Then, using a Bengali sentiment word dictionary and various models, the polarity of each word is calculated. These polarity scores are aggregated to determine the overall sentiment score of the text, which is then normalized. The final sentiment result is derived from this normalized score, providing an accurate assessment of the sentiment in the book review.

V. Result Analysis

Table 1 shows the performance of the various models through weighted average of three performance metrics- accuracy, precision, recall. The performance analysis reveals Logistic Regression as a reliable option for balanced performance, while Deep Neural Network shows potential false positive issues. LSTM and CNN models demonstrate superior accuracy, recall, and precision, particularly highlighting CNN's robustness in capturing complex data patterns, aligning well with its relevance in machine learning. This underscores the intricate dance between model complexity and dataset nuances, suggesting that while simpler models suffice for straightforward tasks, complex data necessitates sophisticated models for accurate representation and analysis.

TABLE I
Performance Metrics for Different Models

Model	Accuracy	Precision	Recall
LR	0.90	0.86	0.90
DNN	0.89	0.80	0.89
LSTM	0.91	0.86	0.91
CNN	0.91	0.88	0.91

References

[1] Rashedul Amin Tuhin, Bechitra Kumar Paul et al, 'An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques', 2019 IEEE 4th International Conference on Computer and Communication Systems

[2] Md. Motaleb Hossen Manik et al, 'A Hybrid Framework for Sentiment Analysis from Bangla Texts', 2022 25th International Conference on Computer and Information Technology (ICCIT)

[3] M Al-Amin et al, 'Sentiment Analysis of Bengali Comments With Word2Vec and Sentiment Information of Words', 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)

[4] Mst. Eshita Khatun et al, 'Machine Learning Approach for Sentiment Analysis of Book Reviews in Bangla Language', 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)

[5] Khondoker Ittehadul Islam et al, 'Sentiment analysis in Bengali via transfer learning using multi-lingual BERT', 2020 23rd International Conference on Computer and Information Technology (ICCIT), 19-21 December, 2020.

[6] Mohsinul Kabir et al, 'BanglaBook: A Large-scale Bangla Dataset for Sentiment Analysis from Book Reviews', arxiv.org/abs/2305.06595