Domain-Specific Aspect-Based Sentiment Analysis in Bangla Using Modern NLP Techniques

1st Akash Ahmed
dept. Computer Science and Engineering (of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
Student ID: 2021-3-60-242

2nd Amin Ocin
dept. Computer Science and Engineering (of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
Student ID: 2021-3-60-135

3rd Tasnova Haque Mazumder
dept. Computer Science and Engineering (of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
Student ID: 2021-3-60-235

4th Dilruba akter
dept. Computer Science and Engineering (of Aff.)
East West University (of Aff.)
Dhaka, Bangladesh
Student ID: 2021-3-60-077

***Abstract*—In our study, we evaluated several models for aspect-based sentiment analysis on a Bengali dataset. Among the models tested, the Bangla BERT model by Sagor Sarker demonstrated the best performance, while the Llama model performed the worst. Due to the dataset's language-specific characteristics, certain feature extractions, such as part-of-speech tagging, were not feasible. However, language-specific tokenization techniques proved effective. Each sentence in the dataset was analyzed for both aspect and sentiment. The results highlight the importance of selecting language-appropriate models and techniques for ABSA tasks. Furthermore, the study contributes to advancing sentiment analysis for Bengali, a low-resource language.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Aspect-based sentiment analysis (ABSA) identifies the polarity of sentiment for applications such as product reviews and social media analysis. Although advanced in English, it has scarce resources for Bangla. To perform this analysis we use the BANGLA ABSA dataset (Mendeley) in four subsets to assess the performance and adaptability of Bangla BERT, mBERT, DistilBERT, and LLAMMA 3.2 over Bangla ABSA Tasks. Objectives of this study are: Compare model performance across domains. Address challenges in Bangla sentiment analysis, including linguistic nuances and multi-faceted sentences.

II. Related Work

A. Islam and Mahmud (2021)

Islam and Mahmud [**?**] proposed a BERT-based model for Aspect-Based Sentiment Analysis (ABSA) on Bengali text. Their approach utilized pre-trained BERT models fine-tuned for the task of aspect-based sentiment classification. The model achieved significant improvement over traditional machine learning techniques, demonstrating its effectiveness in handling complex linguistic structures in Bengali. However, the study was limited by the use of only two models for comparison, and the dataset they employed did not cover a wide range of domain-specific texts, which might affect its generalizability.

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

B. Bhattacharya and Ghosh (2019)

Bhattacharya and Ghosh [**?**] used a hybrid LSTM-CNN model for sentiment analysis in Bengali, focusing on both syntactic and semantic features. They combined a Long Short-Term Memory (LSTM) network with a Convolutional Neural Network (CNN) to capture both the sequential nature of the text and local features. While the model performed well on standard sentiment analysis tasks, it had limitations such as a small number of models tested and the absence of transformer-based architectures, which are now popular for ABSA tasks. Although the reported accuracy was competitive, it could be further improved by including state-of-the-art models like BERT.

III. Methodology

A. Dataset Overview

The dataset used for this study contains a total of 3,725 sentences across four different categories:

- **Car**: 1,149 sentences
- **Mobile Phone**: 975 sentences
- **Movie**: 800 sentences
- **Restaurant**: 801 sentences

Each sentence in the dataset is annotated with two labels:

- **Aspect**: The specific entity or feature related to the sentiment.
- **Sentiment Polarity**: The sentiment expressed (positive, negative, or neutral).

The dataset poses specific challenges due to the nature of the Bengali language, particularly in tokenization and feature extraction. Despite these challenges, the dataset provides a rich resource for evaluating Aspect-Based Sentiment Analysis (ABSA) models for Bengali text.

B. Preprocessing of Data

- **Sentence Splitting:** Compound and complex sentences were split into simpler sentences. This splitting was performed at conjunctions to isolate individual aspects and sentiments, ensuring each sentence carried only one aspect and one sentiment polarity.
- **Stop Word Removal:** A list of common Bangla stop words was used to remove non-informative words from the dataset. This process reduced noise and improved the signal-to-noise ratio, making the features more relevant for sentiment analysis.
- **Handling Missing Values:** Rows with missing values were either deleted or filled with a ”neutral” sentiment polarity to preserve the dataset’s integrity. This step ensured that incomplete or undefined data did not interfere with model training.
- **Label Correction:** Corrected any misspelled or inconsistent sentiment labels in the dataset to ensure accuracy, which in turn improved model reliability.
- **Data Tokenization and Encoding:** The dataset was broken down into individual words or subword units, and numerical encoding was used to convert the text data into a format that machines can understand. This tokenization process made it possible to use machine learning models, particularly those based on BERT and similar architectures.

C. Model Selection and Architecture

BERT: (Bidirectional Encoder Representations from Transformers) is an open source machine learning framework designed for natural language processing (NLP). It was developed by Google AI language researchers in 2018. It leverages a transformer-based neural network to understand and generate human-like language. BERT wmploys an encoder only architecture. In the original transformer architecture there are both encoder and decoder modules. The reason behind using encoder only is to focus on the input sequence rather than the output sequence. We used 3 variants of BERT model to make a comparative decision.

- **Bangla-Bert:** We used sagorsarker/bangla-bert-base as our bangla bert. It used BNPL package for training bengali sentences piece model with vocab size of 102025. It follows the bert-base-uncased model architecture (12-layer, 768-hidden, 12-heads, 110M parameters) with 10 Million training steps trained using Google Cloud GPU.
- **multiLingualBert:** google-bert/bert-base-multilingual-cased is use as the multiLingualBert for our analysis. this model is trained in a self supervised fashion. It means the model were trained on raw text only without any human intervension. [] It was pre trained with 2 objectives Masked language modeling (MLM), Next sentence prediction (NSP).
- **DisTilBert:** [] DisTilBert is a distilled version of BERT. It has 40% less parameters than google-bert/bert-base-uncased, runs 60% faster while preserving over 95% of BERT’s performances as measured on the GLUE language understanding benchmark. This model was contributed by victorsanh. This model jax version was contributed by kamalkraj.

Llamma: Llama is one of the leading state of the art open source large language model released by Meta in 2023. []This model network is based on transformer architecure. The main difference of this model from the actual transformer is the it uses Pre-normalization [GPT3]. To improve the training stability, SwiGLU activation function [PaLM], Rotary Embeddings [GPTNeo]. We used Llama 3.2 1B version for our analysis.

- **Llama-3.2-1B:** Llama 3.2 is an auto-regressive language model that uses an optimized transformer architecture. The tuned versions use supervised fine-tuning (SFT) and reinforcement learning with human feedback (RLHF) to align with human preferences for helpfulness and safety. It has a mix of publicly available online data with multilingual text. There are 1.23B parameters. It has a 128k context length with shared embeddings.

D. Fine Tuning

We used above mentioned model on our dataset to fine tune the model to capture the context specific knowledge. We used AdamW optimizer with lr=5e-5 and CrossEntropyLoss as the loss function. Every model mentioned here was trained on the dataset for 10 epochs. Since we processed our dataset properly and perform proper EDA. We find out the max lenght of our attention mask is 19. So, we used 19 as the MAX LENGTH.

IV. XAI WITH NLP MODELS

A. shap

SHAP (SHapley Additive exPlanations) was utilized to interpret the feature importance of the Bangla BERT model, providing insights into how specific tokens contributed to aspect and sentiment predictions.

Fig. 1. Shap to expalin the effect of feature in outcome

B. lime

LIME (Local Interpretable Model-Agnostic Explanations) was employed to generate localized explanations, highlighting the role of key phrases and words in influencing classification outcomes.

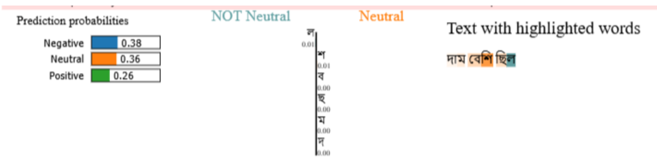

Fig. 2. Lime to expalin the effect of feature in outcome

V. Result and Discussion

This dataset targets aspect specific sentiment analysis (ABSA) in Bengali within the four domains of cars, mobile phones, cinemas, and restaurants, with a collection of 3,725 samples. Every entry captures aspect categories, including Bengali user comments, safety, performance, price, and sentiment polarization, being positive, negative, or neutral. These datasets have been merged into a single corpus for blended sentiment analysis for fine-grained analysis. We evaluated four models—**mBERT**, **BERT**, **DistilBERT**, and **LLaMA**. Below is a summary of their performance:

A. Bangla-Bert:

BERT generates an outstanding performance, much more on the Bangla text; this is due to its adaptability to specific languages and datasets. It strikes a balance between precision and recall, hence being effective in both the aspect and sentiment analyses. Here goes the table for relevant overall metrics:

TABLE I
Performance Overview of BERT

Metric	Aspect Classification	Sentiment Classification
Accuracy	0.78	0.82
Macro Avg Precision	0.70	0.74
Macro Avg Recall	0.70	0.77
Macro Avg F1-Score	0.69	0.76

B. multiLingualBert:

In our models, mBERT is the best overall performer, scoring highest on all metrics. The model's support for numerous languages and contextual settings makes it one of the most flexible models available, especially for sentiment and aspect classification tasks.Here's the table for the relevant overall metrics:

TABLE II
Performance Overview of mBERT

Metric	Aspect Classification	Sentiment Classification
Accuracy	0.70	0.75
Macro Avg Precision	0.65	0.50
Macro Avg Recall	0.61	0.51
Macro Avg F1-Score	0.62	0.51

C. DisTilBert

DistilBERT is a lightweight version of BERT with greater speed and resource efficiency than BERT. While being somewhat behind BERT in terms of performance, it gives a good trade-off between accuracy and computational efficiency. Best for large-scale deployment or resource-constrained systems. Here is the table for relevant overall metrics:

TABLE III
Performance Overview of DisTilBERT

Metric	Aspect Classification	Sentiment Classification
Accuracy	0.70	0.75
Macro Avg Precision	0.65	0.50
Macro Avg Recall	0.61	0.51
Macro Avg F1-Score	0.62	0.51

D. Llamma:

It demonstrates the best performance for broad and context-heavy tasks that require creative reasoning or are general-purpose text processing; hence, our dataset, which involves very specialized NLP tasks such as sentiment analysis, fine-tuned models that do not give us that much efficiency. Here is a summary table based on the reports provided:

TABLE IV
Performance Metrics for Aspect and Sentiment Classification

Metric	Aspect Classification	Sentiment Classification
Accuracy	0.28	0.56
Macro Avg Precision	0.22	0.43
Macro Avg Recall	0.17	0.36
Macro Avg F1-Score	0.16	0.30

E. comparison of result

Given the metrics, it can be concluded that BERT had the best overall performance in both aspect and sentiment classification. It reached the highest accuracy, precision, recall, and F1-score values across tasks with balanced macro-averages, while Lamma is the weakest model when taking into consideration classification performance.

TABLE V
Comparison of Models for Aspect and Sentiment Classification

Model	Aspect Accuracy	Sentiment Accuracy
Bangla BERT (Sagor)	0.78	0.82
mBERT	0.70	0.75
DistilBERT	0.70	0.75
LLAMMA	0.28	0.56

VI. Conclusion

The evaluation of four models—Bangla BERT, mBERT, Gemma, and LLAMMA—revealed significant differences in their performance on Bangla aspect-based sentiment analysis (ABSA) tasks. The outstanding results obtained by Bangla BERT run across all metrics of as well sentiment classification, and it explained the importance of language-specific models in capturing specifics of language nuance. DisTilbert also did show a decent performance, but not as good as the Bangla Bert. Mbert held a decent performance, but the fact that it did not tune to the language was a slight limitation compared to BERT.. Then, LLAMMA had the smallest score, mostly because of extremely low performance, inadequate adaptation to Bangla, and sparse representation of the vocabulary and bare embeddings when trained. The research highlights the need to select specific models for their languages while addressing other resource constraints, as they will deliver optimal performance.

References

[1] Liu, B. (2017). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

[2] Zhang, Y., Qian, T., Liu, B. (2018). Aspect-based Sentiment Analysis for Social Media Reviews. Journal of Artificial Intelligence Research.

[3] Zhou, W., Zhang, Y. (2017). Neural Approaches to Aspect-Based Sentiment Analysis. Proce