Beyond Likes and Loves: Uncovering Complex Sentiments in Bengali Food Reviews on Facebook

Ataf Fazledin Ahamed
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1705066@ugrad.cse.buet.ac.bd

Sadia Sharmin
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
sadia@teacher.cse.buet.ac.bd

***Abstract*—As a result of the technological advancements of the internet, Bangladeshi users are increasingly active on social networks. In this sense, social media influencers are becoming more well-known and attracting a growing number of users. Bangladeshi food review influencers are becoming more and more well-known every day. The most sophisticated Bengali sequence classification model was used in this study's analysis of social network interaction data. Through an extensive exploration of the social media landscape, we delve into the realm of food reviews. We used the sequence classification model to classify the comments collected from social media for our study. Our findings reveal that the majority of viewers hold a positive perception of Bengali food reviews on social media, while a small number of outliers may express contrasting opinions. Notably, our classifier, BanglaBERT, achieves an impressive prediction accuracy of 83.76%, emphasizing the reliability and effectiveness of our approach.**

I. Introduction

With the technological advancement in the 21st century, more and more people are getting internet access. As of January 2023, around 5.16 billion people worldwide have internet access. In Bangladesh, around 38.9% of the population has internet access [1]. According to a 2023 report by DataReportal [1], around 43.25 million Bangladeshis are Facebook users, making it the most popular.

Today, Facebook is not just a platform for connecting with friends and family. It has transformed into a multidimensional tool, for example, an online marketplace, advertisement medium, and a digital workplace. Facebook as a medium of entertainment is another feature that is taking the world by storm. Every day, new content is being created and uploaded to the platform, mostly in video format. Stories, reels, and videos are the main types of content that have a higher engagement rate. Even the YouTubers are not limited to YouTube; they are constantly sharing their contents on Facebook and generating huge revenue from Facebook. On the other hand, people from all walks of life are getting easy access to content from their phones. Different types of content, such as travel videos, cooking videos, food review videos, DIY (Do It Yourself) videos, real-life hacks, etc., are now considered major sources of entertainment.

In recent years, the number of Bangladeshi social media influencers has increased. Businesses collaborate with influencers to generate traffic. On the other hand, people generally pay attention to influencers when making decisions for themselves. Bangladeshi food influencers are increasing in popularity daily, relying on social media sites like Facebook and Instagram.

The organization of the rest of the paper is as follows: the following section describes a few relevant works. Section 3 demonstrates our methodology. Sections 4 and 5 deal with Dataset collection and preprocessing. Feature selection and corresponding model selection are listed in Sections 6 and 7. Finally, we present the results in Section 8, followed by the conclusion.

II. Related Works

There have been several works in the social media domain on sentiment analysis. Since social media mainly depends on user interactions, the scope of sentiment analysis and natural language processing has always been present.

1) **"Sentiment Analysis on Bengali Facebook Comments to Predict Fan's Emotions Toward a Celebrity" [2]:** In this study, the authors have performed sentiment analysis on the comments posted by different fans on a celebrity's Facebook page. In this work, they have applied various machine learning algorithms to predict the sentiments of the comments. The algorithms used by the authors are *Random Forest Classifier, Support Vector Machine, Neural Network, Naive Bayes Classifier*. In this study, the authors categorized the emotions into classes such as: 'Surprised', 'Abusive', 'Angry', 'Religious', 'Happy', 'Excited', 'Sad'.
2) **"Hateful Speech Detection in Public Facebook Pages for the Bengali Language" [3]:** The study describes an approach to hate speech detection using machine learning techniques in the Bengali language. In this paper, some basic classifiers such as *Stochastic Logistic Gradient Classifier, logistic Regression along with Principle Component Analysis* were used along with *Gated Recurrent Unit, Long Short Term Memory, and Word2Vec* models. The paper shows high performance for neural network-based models with a macro F1 score of 0.69 and an accuracy of 70%.

There has been notable progress in sentiment analysis in the food review/recommendation domain. For example, the works

mentioned below give an overview of the approaches usually taken by the authors-

1) **"Sentiment Analysis Techniques in Food Reviews Using Machine Learning" [4]:** The study employs machine learning techniques such as the *Naive Bayes Classifier, Support Vector Machine Classifier, Decision Tree, K-Nearest Neighbors*. The article focuses on the food review collected from Zomato and Amazon Foods. Based on the reviews and ratings submitted by customers, sentiment analysis performed on them gives an idea about the food recommendation ecosystem.
2) **"Sentiment Analysis in Food Review using Machine Learning Approach" [5]:** The article uses the Yelp review dataset in their study. Sentiment analysis using machine learning algorithms for English text was used for the research.
3) **"Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naive Bayes" [6]:** The paper follows the same approach as the previous paper. However, in this study, the Yelp review dataset was translated into Bengali. The paper applies *Multinomial Naive Bayes* algorithm to perform sentiment analysis on the translated Bengali reviews.

Based on a thorough study and a literature review of existing papers and research, we can conclude that no work has yet been done in the domain of food review in the context of the Bengali language. As a result, our proposed study justifies itself as an unexplored topic.

III. Methodology

To classify the video content of the influencers collected from Facebook, we employ two approaches. The first approach depends on the interactions between followers and viewers of the page. In contrast, the second one relies on the comments posted on the videos by the viewers stating their opinion.

A. Interaction Based Approach

In this approach, we prioritize different interaction metrics, such as the reactions to the videos. Facebook has several reactions: 'Like', 'Love', 'Care', 'Wow', 'Haha', 'Angry', and 'Sad'. Each of these reactions conveys different feelings. Since an individual can only give one type of reaction, we classify these reactions into two main categories: i) positive and ii) negative.

The approach follows the method employed in this [7] paper. In this paper, the authors performed sentiment analysis of Facebook posts through special reactions. The same approach can also be found discussed in this [8] paper. In this article, the authors conducted a study on understanding Facebook reactions to different articles. This method uses a score-based system to classify Facebook content into two categories. In this approach, the reactions are categorized into two classes, and a metric called polarity is calculated based on them. The formula is as follows:

$$polarity = \frac{n_{positive} - n_{negative}}{n_{positive} + n_{negative}}$$

$n_{positive}$ = Sum of all positive reactions
$n_{negative}$ = Sum of all negative reactions

Based on the polarity value, we classify the videos into four categories. The categories are shown in Table I

TABLE I
DIFFERENT CLASSES OF POLARITY BASED ON VALUE

Value	Category
$-1 \leq polarity < -0.5$	Class 1
$-0.5 \leq polarity < 0$	Class 2
$0 \leq polarity < 0.5$	Class 3
$0.5 \leq polarity \leq 1$	Class 4

B. Sequence Classification Based Approach

To analyze the sentiment of the viewers' comments, we propose using a neural network-based machine learning model. Since we are working on the domain of Bengali food review, most of the comments are expected to be in the Bengali language. As a result, we decided to use the BanglaBERT model [9]. BanglaBERT achieves state-of-the-art performance when it comes to Bengali natural language processing tasks.

In this method, we analyze the sentiment of each comment collected from food review videos. For each video, we determine the positive and negative sentimental comments with the help of the BanglaBERT model. To classify the text sequence, we must train the pre-trained version of the BanglaBERT model. We used three existing datasets to fine-tune the model. We used the "SentNoB" dataset [10] containing Bengali comments from more than 13 domains. We also used the "ABSA" dataset [11] containing different labeled comments on different aspects. In addition to these two, we used the translated Yelp restaurant review dataset mentioned in the article [6].

We merged the data sets and only used positive and negative labeled comments for training. We divided the dataset into train, test, and validation subsets for fine-tuning the pre-trained BanglaBERT model. After training the model for further classification tasks, we predicted the comments in each food review-related video. Based on the class having the highest number of comments, we label the post as belonging to that class.

IV. Dataset Creation

We chose five leading food review-related influencers from Bangladesh for our research. We chose video content over other types of content due to its higher interaction rate. In figure 1, the comparison between different content types can be seen.

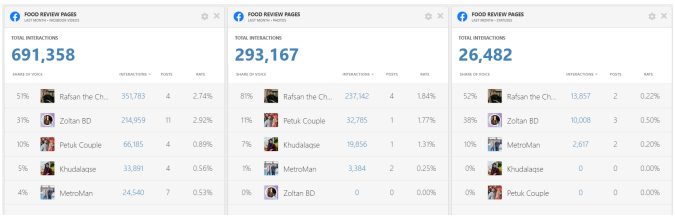

Fig. 1. Difference between different content types

We used "CrowdTangle"[1], a social media analytics tool. For our study, we only selected videos published between 2020 and 2022 and downloaded related data. After downloading, we filtered the data based on a few conditions:

- Post type must be video
- The content must be posted from their Facebook page
- The downloaded data must belong to food review contents

To fulfill the last condition, we manually visited each URL and checked whether it was a food review video. If it was not one, we discarded the data related to it. Some of the notable data collected from CrowdTangle contained the following information:

TABLE II
AVAILABLE INFORMATION FROM CROWDTANGLE

Post Created Date	Post Created Time
Post Type	Total Interactions
Total Comments	Shares
Likes	Love
Wow	Haha
Sad	Angry
Care	Haha
Sad	Total Views
Video Length	Message
Link	Score

Since the sequence classification-based approach requires Bengali text to analyze the sentiments of the food review videos, we had to download the comments related to the videos mentioned in the previous section. For the filtered data from the previous section, we manually downloaded the comments for each post since no public API exists for Facebook. For our study, we developed a plugin for Chromium-based browsers based on an existing one. The plugin helped us to extract the comments from Facebook easily. For each video, we extracted and saved the comments in an individual CSV file and linked the file to the CSV previously downloaded from CrowdTangle. Following the steps above, we downloaded around 26,004 comments from 905 Facebook food review-related videos.

V. PREPROCESSING

The data downloaded from CrowdTangle and Facebook needs to be preprocessed before we apply it to our proposed methods. The preprocessing steps mainly involve the data collected from the comment section of the food review videos. The comments collected in our research are mainly in Bengali, English, and Banglish.

To run the BanglaBERT model on the data, we must translate them into Bengali. We follow three basic steps for preprocessing the data. At first, we pass the comment text through the Avro phonetic parser [2]. The parser output is then passed through Google Translate API [3]. Even after these two steps, some comments were not translated into Bengali. As a result, we had to translate the text by hand manually. The human-level text filtering steps ensured the data quality for our research.

VI. FEATURE SELECTION

Among all the information received from CrowdTangle, the 'Likes' information related to a post, which denotes the total number of likes, was discarded. It was done due to several reasons-

1) The Facebook News Feed ranking algorithm gives less importance to it. [12]
2) The like count of a post bears no significance. It is treated as a neural reaction since it is thought to belong to a lower sentimental reaction.
3) The total like count of our collected data amounted to more than the sum of other reactions. As a result, taking it into calculation would nullify the importance of other reactions by a large margin. In figure 3, it is quite evident that the total number of likes outperforms all other reactions, even when combined.

Figure 4 portrays the different correlation coefficients of different features of our dataset. Among them, the value of coefficients between 'Love and Care' and 'Haha and Angry' are higher than others. It goes on to show a strong positive correlation between these two pairs of reactions. As a strong positive correlation exists between 'Love' and 'Care', we group them into the positive reaction class. And since 'Haha' and 'Angry' reactions have a positive correlation between them, we group them into the negative reaction class. Since the reactions in social media are disjoint and have no effects on each other, we cannot find any negative correlation between the features.

As we had previously discarded 'Like' as a significant reaction, only 'Wow' and 'Sad' remains to be grouped into classes. The reaction 'Wow' is considered to be a positive sentiment. As a result, we put it with 'Love' and 'Care' in the positive class. A 'Sad' reaction is considered a negative sentiment, so it is put in the negative class. Figure 5 shows our classification of the Facebook reactions. The paper *Pratama, et al.* [7] employed the same classification in their study of special Facebook reactions.

A. Features for interaction Based Method

Since the interaction-based approach uses many positive and negative reactions, we define two new features from our

[1]Website: https://crowdtangle.com

[2]GitHub: https://github.com/hitblast/avro.py

[3]Google Translate API: https://cloud.google.com/translate

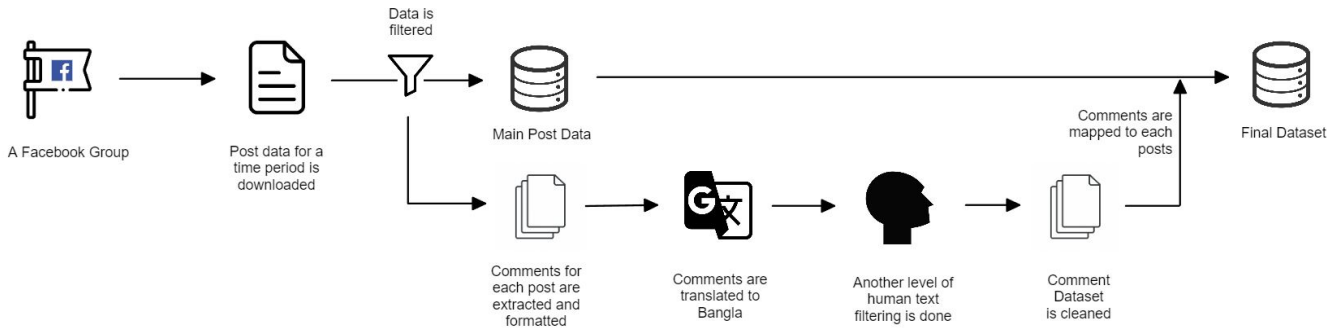

Fig. 2. Data Collection and Preprocessing Steps

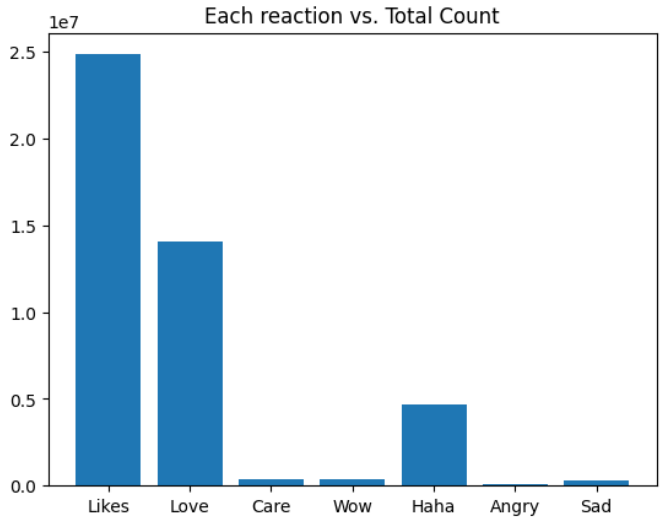

Fig. 3. Distribution of Different Reactions of Facebook Food Review Videos

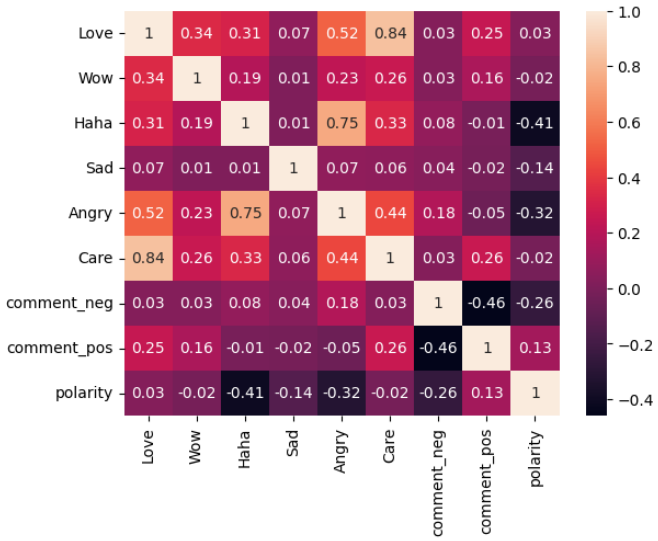

Fig. 4. Correlation Coefficients of Different Features

existing ones. The feature $n_{positive}$ denotes the total number of reactions belonging to the positive class, and the feature $n_{negative}$ denotes the total number of reactions belonging to the negative class. As a result, we write down the formula for 'polarity' below:

$$polarity = \frac{n_{positive} - n_{negative}}{n_{positive} + n_{negative}}$$

B. Features for Sequence Classification Based Method

Similarly, the same approach is followed for selecting the features for the sequence classification-based method. In this method, the number of comments belonging to positive and negative classes is computed. It leads to two new features called $comment_{positive}$ and $comment_{negative}$, which denotes the total number of positive and negative comments predicted by the fine-tuned BanglaBERT model.

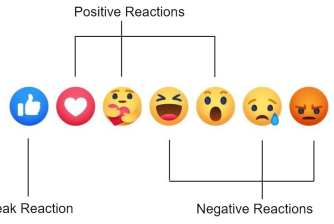

Fig. 5. Classification of Different Facebook Reactions

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT BERT MODELS FOR DOWNSTREAM TASKS

Models	Params	SC	NLI	NER	BLUB Score
mBERT	180M	67.59	75.13	68.97	70.29
BanglishBERT	110M	70.61	80.95	76.28	75.73
BanglaBERT	110M	72.89	82.80	77.78	77.09

TABLE IV
PERFORMANCE RESULTS OF BANGLABERT

Metrics	Value
Evaluation Accuracy	0.744
Prediction Accuracy	0.8376
Evaluation Macro F1	0.699
Prediction Macro F1	0.8375
Train Loss	0.7038
Evaluation Loss	0.64

In summary, we can list the features used by our methods as follows:

1) **Features for interaction-based approach**
   a) Total number of positive reactions
   b) Total number of negative reactions
2) **Features for sequence classification based approach**
   a) Total number of positive comments
   b) Total number of negative comments

VII. MODEL SELECTION

Since our study involves working with interaction-based sentiment analysis, we do not use any machine learning model for this purpose. However, for the sequence classification-based approach, we use the Bengali text-based variant of BanglaBERT due to its higher BLUB Score. From the table III collected from BanglaBERT paper [9], we can see that the performance of the BanglaBERT model exceeds that of the BanglishBERT one.

BanglaBERT offers three pre-trained models for the downstream task of the Bengali language. Due to the computation hardware and resources limitation, we chose to work with the 'BanglaBERT Base' model. The pre-trained weights of the model were downloaded from HuggingFace model hub [4]. For fine-tuning the model, we used the script provided with the model's code base.

VIII. RESULTS

The fine-tuning task of the BanglaBERT model is critical for our research as it was determined to be the ground truth for our sequence classification-based approach to sentiment analysis. Our model yields a better performance on the training dataset. During the training, we used cross-validation to validate the trained model against a completely unknown dataset. The performance achieved by our model is described in the tables below:

[4]https://huggingface.co/csebuetnlp

A. Sequence Classification Based Approach

The bar graph and the histogram below denote the total number of positive and negative comments received from the viewers. Figure 6 shows the total number of comments from each class collected from our dataset. It is quite evident that the positive comments are way more in number than the negative ones. The histogram shown in figure 7 denotes the distribution of the comments in terms of the polarity value.

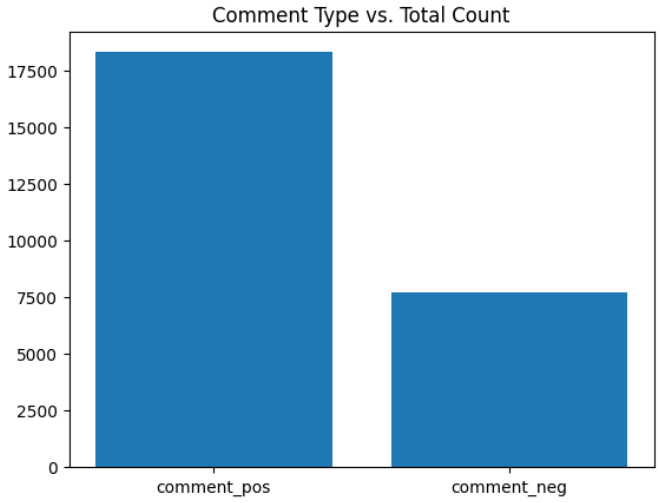

Fig. 6. Distribution of positive and negative comments

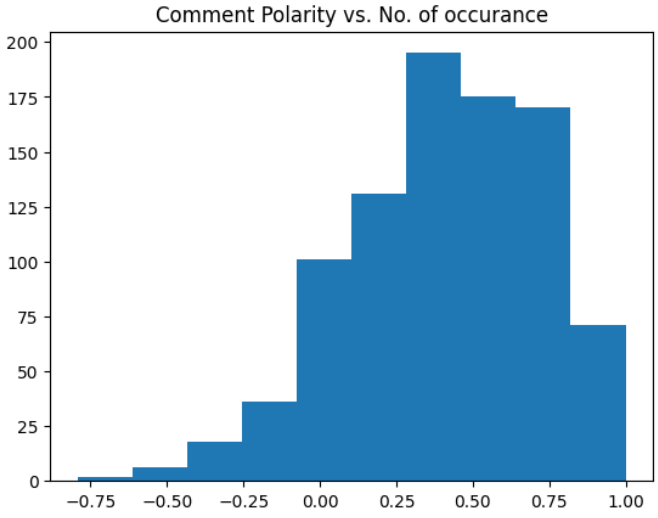

Fig. 7. Distribution of comments in terms of polarity

B. Interaction Based Approach

The interaction-based approach proposed in this study provides high-level insights into the overall food review situation in social media. The bar graph shown in figure 8 shows the

total number of videos belonging to each polarity class. The histogram shown in figure 9 denotes the polarity distribution of the food review contents.

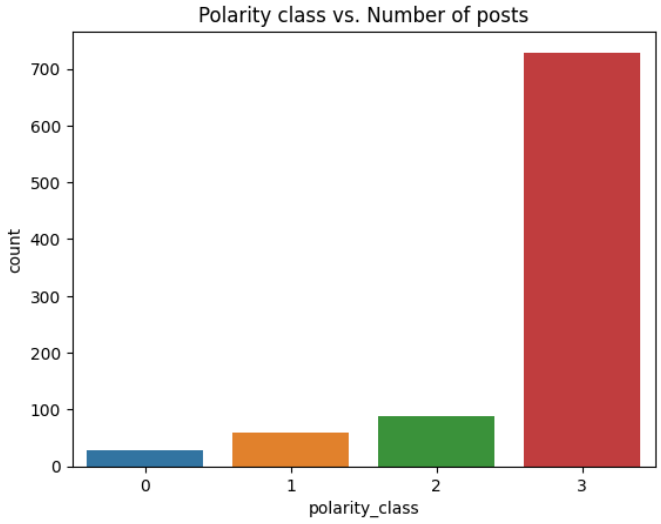

Fig. 8. Distribution of food review videos in terms of polarity class

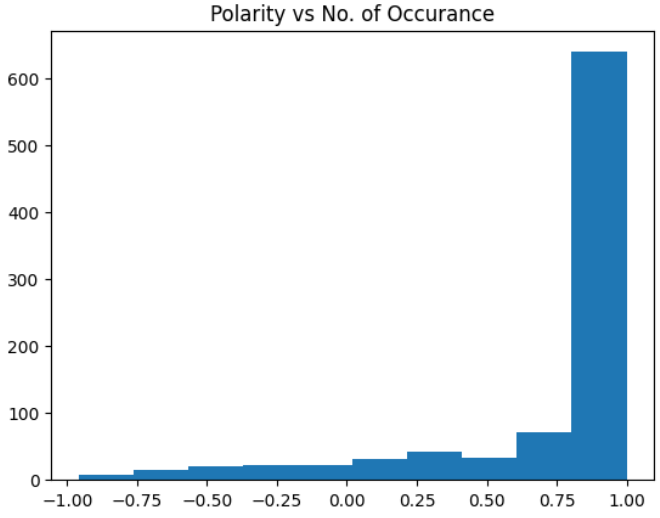

Fig. 9. Distribution of food review videos in terms of polarity score

From both approaches, we find significant insight into the sentiment related to food reviews in the social media domain in Bangladesh. From the interaction-based approach, we can see that more than 90% of the food review videos contain positive reactions, indicating positive feelings. The number of videos decreases along with a lower polarity score, which means that the minimum number of videos shows negative polarity.

On the other hand, the sequence classification-based approach shows a bit different insights. The maximum number of videos contains a polarity score between 0.25 and 0.5. There exist lower high-polarity videos than those of medium polarity. Also, the number of videos decreases along with the polarity level.

Both approaches proposed and implemented in this paper indicate a positive sentiment of the food review ecosystem of social networks in Bangladesh. The interaction and comments posted by the viewers indicate that there is a positive sentiment about the food review-related content shared on Bangladesh's social media space.

IX. Conclusion

The study conducted throughout this research dives into social network interaction, other analytical data, and a state-of-the-art sequence classification model. The findings show us that the food reviews shared on social media are taken in positive sentiment by most of the viewers, whereas some outliers denote a negative sentiment towards the reviews.

Since our study relies heavily on more and more data points, we intended to expand our dataset to get clearer insights. In the future, we hope to expand our work into different food review-related groups where we might find more insights regarding this industry that is increasing daily. Since the overall process shown throughout our study is data-dependent and lengthy, we plan to incorporate an automated system through our developed plugin. We believe both the users and the research community will benefit from the data collected through the plugin.

The interaction-based sentiment analysis introduced by Freeman et al. (2019) [8] and Pratama (2022) [7] can be applied to research areas working with social media. The special reactions classified into multiple categories, and a radar chart-based metric might help us look inside the data and provide rich insights. Being a cost-effective and resource-efficient approach, this method can lead to an initial hypothesis carried out at the initial research phase.

References

[1] S. Kemp, "Digital 2023: Bangladesh — DataReportal – Global Digital Insights — datareportal.com," https://datareportal.com/reports/digital-2023-bangladesh, [Accessed 20-May-2023].

[2] M. S. S. Khan, S. R. Rafa, A. E. H. Abir, and A. K. Das, "Sentiment analysis on bengali facebook comments to predict fan's emotions towards a celebrity," *Journal of Engineering Advancements*, vol. 2, no. 03, p. 118–124, Jul. 2021. [Online]. Available: https://scienpg.com/jea/index.php/jea/article/view/109

[3] A. M. Ishmam and S. Sharmin, "Hateful speech detection in public facebook pages for the bengali language," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 555–560.

[4] S. Gite, A. Udanshiv, R. Date, K. Jaisinghani, A. Singh, and P. Chetwani, "Sentiment analysis techniques on food reviews using machine learning," in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, P. K. Singh, S. T. Wierzchoń, S. Tanwar, M. Ganzha, and J. J. P. C. Rodrigues, Eds. Singapore: Springer Singapore, 2021, pp. 511–528.

[5] N. Islam, N. Akter, and A. Sattar, "Sentiment analysis on food review using machine learning approach," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, pp. 157–164.

[6] O. Sharif, M. M. Hoque, and E. Hossain, "Sentiment analysis of bengali texts on online restaurant reviews using multinomial naïve bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–6.

[7] A. Pratama, "Sentiment analysis of facebook posts through special reactions: The case of learning from home in indonesia amid covid-19," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 8, pp. 83–92, 03 2022.

[8] C. Freeman, M. K. Roy, M. Fattoruso, and H. Alhoori, "Shared feelings: Understanding facebook reactions to scholarly articles," in *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2019, pp. 301–304.

[9] A. Bhattacharjee, T. Hasan, W. Ahmad, K. S. Mubasshir, M. S. Islam, A. Iqbal, M. S. Rahman, and R. Shahriyar, "BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla," in *Findings of the Association for Computational Linguistics: NAACL 2022*. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 1318–1327. [Online]. Available: https://aclanthology.org/2022.findings-naacl.98

[10] K. I. Islam, S. Kar, M. S. Islam, and M. R. Amin, "SentNoB: A dataset for analysing sentiment on noisy Bangla texts," in *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 3265–3271. [Online]. Available: https://aclanthology.org/2021.findings-emnlp.278

[11] M. A. Rahman and E. Kumar Dey, "Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation," *Data*, vol. 3, no. 2, 2018. [Online]. Available: https://www.mdpi.com/2306-5729/3/2/15

[12] A. Lada, "How Does News Feed Predict What You Want to See? — Meta — about.fb.com," https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/, [Accessed 20-May-2023].