2023 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 1-3, 2023

Sentiment Analysis of Restaurant Reviews from Bangladeshi Food Delivery Apps

Ehsanur Rahman Rhythm*, Rajvir Ahmed Shuvo, Md Sabbir Hossain, Md. Farhadul Islam and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{ehsanur.rahman.rhythm*, rajvir.ahmed.shuvo, md.sabbir.hossain1, md.farhadul.islam}@g.bracu.ac.bd, annajiat@gmail.com

***Abstract*—In this study, we conducted sentiment analysis on restaurant reviews from Bangladeshi food delivery apps using natural language processing techniques. Food delivery apps have become increasingly popular in Bangladesh, and understanding the sentiment of customer reviews can provide valuable insights for restaurant owners and food delivery app companies. In this research, we have created a dataset named "Bangladeshi Restaurant Reviews" by gathering customer reviews of restaurants available on Foodpanda and Hungrynaki, which are two popular food delivery apps in Bangladesh. We used Robustly Optimized BERT Pretraining Approach (RoBERTa), AFINN, and DistilBERT, a distilled version of Bidirectional Encoder Representations from Transformers (BERT) to perform the sentiment analysis. Overall, this research paper highlights the importance of sentiment analysis in the food delivery industry and demonstrates the effectiveness of different models in performing this task. It also provides insights for businesses looking to use sentiment analysis to improve their services and products. The accuracy of the models evaluated, RoBERTa, AFINN, and DistilBERT, were 74%, 73%, and 77% respectively.**

***Index Terms*—Sentiment analysis, Restaurant Reviews, Food Delivery, Text analysis, AFINN, RoBERTa, DistilBERT.**

I. Introduction

In recent years, food delivery apps have grown in popularity, offering customers a convenient way to order food from their favorite restaurants and have it delivered to their doorstep. Foodpanda [1], and Hungrynaki [2] are two major food delivery apps in Bangladesh that have gained widespread popularity. In contrast, sentiment analysis is a rapidly expanding subfield of natural language processing (NLP) with the aim of comprehending the sentiment or emotion underlying written or spoken language. It has numerous applications, including customer feedback analysis, brand reputation management, and market research. Sentiment analysis can be used in restaurant reviews to understand the overall satisfaction level of customers and identify areas for improvement.

The present research aims to conduct a sentiment analysis of restaurant reviews from Bangladeshi food delivery apps, specifically Foodpanda and Hungrynaki. A total of 20,000 reviews were collected and analysed using three different models: RoBERTa, AFINN, and DistilBERT. The results showed an accuracy of 74%, 73%, and 77%, respectively.

The motivation behind this research is to understand the sentiments expressed in the reviews of restaurants on these food delivery apps and to identify any patterns or trends in the data. This information can be helpful for both restaurants and food delivery apps to improve their services and customer satisfaction. Additionally, the results of this research can provide valuable insights for similar studies in other regions or countries.

Overall, the present research aims to contribute to the growing body of knowledge on sentiment analysis and its applications in the context of restaurant reviews. This study aims to provide valuable insights into the food industry and the field of natural language processing by conducting a thorough analysis of reviews from Bangladeshi food delivery apps.

II. Literature Review

There have been several studies that have applied sentiment analysis to restaurant reviews from various countries. For example, a survey by Laksono et al. used Naive Bayes to classify how satisfied customers were with restaurants in Surabaya [3]. The authors found that both approaches successfully gauge customer opinion, with the Naive Bayes method edging out TextBlob sentiment analysis by a margin of 2.9%. We also looked into sentiment analysis methods like [4], which provides a comprehensive overview of sentiment analysis, including techniques and applications. Furthermore, Hasan et al. conducted sentiment analysis for Twitter accounts using machine-learning techniques [5]. Rehman et al. developed a Hybrid CNN-LSTM Model (A model consists of Convolutional Neural Networks and Long Short-Term Memory Networks) to improve the precision of sentiment analysis for movie reviews, which outperforms traditional deep learning and machine learning techniques [6]. HarishRao et al. used VADER (Valence Aware Dictionary for sEntiment Reasoning) to classify the sentiment of unsupervised product review [7] while Pano et al. utilized the same model to analyse sentiment in Bitcoin (BTC) Tweets during the Era of COVID-19 [8]. In [9], Marcec and Likic used AFINN in tweets to analyse the sentiment toward SARS-CoV-2 vaccination. Moreover, Vijayarani et al. conducted a comparative analysis for open-sourced tokenization tools [10].

Despite the abundance of literature on sentiment analysis of restaurant reviews, no studies have specifically focused on reviews from Bangladeshi food delivery apps. Biswas et al. conducted the closest study to this, using the BERT machine learning technique to predict users' sentiments based on Facebook comments about online food delivery services [11]. However, there is currently no study that uses ratings and reviews made on food delivery apps as a primary data source. This gap in the literature is significant as it is essential to understand the sentiment of customers towards these apps in order to improve the quality of service provided. The proposed study aims to address this gap by using sentiment analysis to measure customer satisfaction with these apps.

III. Sentiment Analysis Approaches

A variety of approaches exist for conducting sentiment analysis. These approaches can be broadly categorized as non-machine-learning, or machine-learning-based [12].

One of the non-machine learning approaches used in this paper is the lexicon-based sentiment analysis tool. A lexicon-based approach to sentiment analysis involves using a pre-defined list of words (a lexicon) to identify the sentiment of a piece of text. This can be done by assigning a positive or negative score to each word in the lexicon and then summing the scores for all the text words to determine the overall sentiment.

On the other hand, approaches based on machine learning [13] rely on labeled datasets of text and their corresponding labels for training a model (e.g., positive, negative, neutral). After training, the model can predict the label for new, unseen text. Several types of machine learning algorithms can be used for sentiment analysis, including Support Vector Machine (SVM), Naive Bayes classifiers, and Deep Learning models such as Convolutional Neural Networks (CNNs) and transformer models.

One common approach to training a machine learning model for sentiment analysis is to represent the input text as a sequence of word embeddings, numerical representations of words that capture their meaning and context. The word embeddings are then passed through one or more layers of a neural network, which learns to classify the text based on its embeddings.

Another approach is to use pre-trained language models, such as DistilBERT, which have been trained on large datasets of text and can be fine-tuned for specific tasks, such as sentiment analysis. These models often achieve high accuracy on a wide range of natural language processing tasks, including sentiment analysis, due to their ability to capture complex relationships between words and their contexts.

In this paper, we will use AFINN, which is lexicon-based, whereas RoBERTa and DistilBERT are pre-trained machine-learning language models that we will use for sentiment analysis.

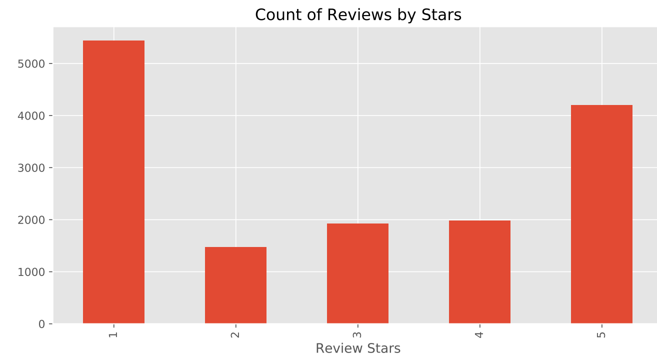

Fig. 1. Count of Reviews based on Rating from the Dataset

IV. Data Collection and Preprocessing

A. Data Collection

A total of 20,000 reviews were collected from two major Bangladeshi food delivery apps, Foodpanda and Hungrynaki. We used their public website API (Application Programming Interface) to get the reviews from each restaurant of the app. Only 2000 customer reviews were taken from Foodpanda, whereas we took 18000 from Hungrynaki. All of the data was collected from restaurants located in various places in the city of Dhaka. We collected all the data in CSV (Comma-separated Values) files separately and then merged them for later analysis. The CSV file contained the review date, the reviewer's name, the review text, and the rating value on a scale of 1 to 5 as shown in table I. Based on the dataset, figure 1 was created to give an overview of the distribution of reviews across different ratings.

TABLE I
Collected Dataset from Foodpanda and Hungrynaki

id	createdAt	text	reviewerName	ratingvalue
1	2021-09-27T13:05:59.993Z	everything is perfect...	Munia	4
2	2021-11-11T13:03:23.982Z	Need to improve food quality	Anonymous	2
3	2021-12-11T10:37:54.844Z	they have become the worst. the beef patty had...	Sazzat	1
4	2021-12-24T12:33:37.968Z	Awesome	Anonymous	5
5	2022-07-10T08:10:40Z	Food is very good	MD	5

B. Data Preprocessing

Before conducting our sentiment analysis, we performed several preprocessing steps on the collected data to ensure that it was in a suitable format for analysis.

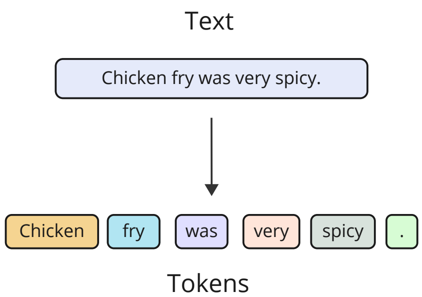

Fig. 2. Tokenization in NLP: Splitting a sample text into smaller units (tokens)

First, we checked for missing data or null values and removed any instances found with the Pandas library's help. We also checked for any errors in the data, such as incorrect ratings. Many customers just gave ratings without providing text reviews; we also removed those. We also dropped the unnecessary rows which came along the review from the API.

Next, we performed text normalization on the review text. This included converting all text to lowercase, removing punctuation, and removing the emojis from the reviews. A large number of rows contained Bangla reviews, as well as in Banglish. We translated the Bangla reviews and transliterated the Banglish reviews with the help of Google Cloud Translation API [14].

V. Methodology

To perform sentiment analysis on restaurant reviews from Bangladeshi food delivery applications, we employed three pre-trained models: AFINN, RoBERTa, and DistilBERT. These models were chosen based on their ability to analyze text in multiple languages and their suitability for the specific domain of restaurant reviews. RoBERTa and DistilBERT, both transformer-based language models, were trained on a large corpus of text and use a method called tokenization to break down text into smaller pieces before analyzing it as shown in figure 2. They also utilize a combination of linguistic and structural features, such as word embeddings, Masked Language Modeling (MLM), and Next Sentence Prediction (NSP), to understand the context of the text. In contrast, the AFINN model utilizes a pre-defined lexicon of words and their corresponding sentiment scores as features to classify text. One of the challenges in sentiment analysis is the limited availability of annotated data sets, particularly for specific domains, which can be observed in the AFINN model's approach. Sentiment analysis, in general, can be approached in two ways, non-machine-learning, or machine-learning-based, as previously discussed.

A brief description of the models used for sentiment analysis is given below:

A. AFINN

AFINN is a lexicon-based sentiment analysis tool that uses a list of words and their associated sentiment scores to classify the sentiment of a given text. The sentiment scores are based on the valence (i.e., positive or negative) of the words in the list, with higher scores indicating a more positive sentiment and lower scores indicating a more negative sentiment.

AFINN was developed by Finn Årup Nielsen in 2011 and has been widely used in sentiment analysis research [15]. It is a simple and fast model that can be a good choice for sentiment analysis tasks with relatively small and simple datasets.

An *Afinn* class is used to classify the sentiment of a given text. The *Afinn.score* method returns an overall sentiment score for the text, which ranges from -5 (most negative) to 5 (most positive). Alternatively, we can use the *Afinn.score* method to classify the sentiment of each word in the text and then use these scores to compute the overall sentiment of the text. However, we followed the first method.

TABLE II
Sentiment Analysis Using AFINN

Sentence	Score (Out of -5.0 to +5.0)
The Food is so great	3.0
The Food smells bad	-3.0
How ponderous this moonlight be, for methinks it weigh heavily on my resting mood.	0.0

AFINN is based on a manually-constructed list of words and their associated sentiment scores derived from human judgments of the valence of the words. The list includes more than 2,500 English words and effectively classifies the sentiment of a wide range of texts, including social media posts, news articles, and movie reviews.

Table II shows that "The Food is so great" has a positive score of 3.0. AFINN checks all the lexicons of the sentence, checks it with its store of lexicons, and gives out a score. The same goes for "The food smells bad," giving it a negative score of -3.0. Now, where AFINN falls short is in the third example. "How ponderous this moonlight be, for methinks it weighs heavily on my resting mood." is a big sentence. However, as AFINN is individually checking all the lexicons, it does not catch the human context of the sentence. Hence, it shows the sentence as a neutral sentence with a score of 0.0, even though it should have been a negative sentence. This is why the AFINN model is underwhelming when finding out the sentiment analysis of big reviews or paragraphs.

Overall, AFINN is a useful tool for sentiment analysis that can be easily applied to various tasks and datasets. It is particularly well-suited for tasks that require fast and reliable sentiment classification and can be a good choice when working with small and simple datasets. Nevertheless, as AFINN is a non-machine learning model, it lacks the ability to catch the human context of the sentence.

B. RoBERTa

To solve the problem of AFINN, which was not understanding the human context in a sentence, which resulted in wrong results in some cases, we decided to use RoBERTa. When it comes to processing input sequences and producing output sequences, RoBERTa, a transformer-based model, uses self-attention mechanisms to do the trick [16]. The model takes in a sequence of tokens (e.g., words or subwords as shown in 2) and produces a fixed-length representation for each token, known as an embedding. These embeddings are then passed through multiple transformer layers, which use self-attention to capture dependencies between the tokens in the input sequence.

A final linear layer receives the output from the transformer layers and uses it to predict the current task. For example, in the case of language translation, the model might predict the next word in a sentence given the previous words. In the case of text classification, the model might predict the class label (e.g., positive or negative sentiment) for a given piece of text.

RoBERTa is trained using a variant of the BERT training process known as "masked language modeling," in which the model is trained to predict masked tokens in a sequence based on the context of the remaining unmasked tokens. This process allows the model to learn the relationships between words and their meanings in a way that generalizes well to other tasks.

TABLE III
SENTIMENT ANALYSIS USING ROBERTA

Sentence	Neg	Neu	Pos
The Food is so great	0.0022	0.0119	0.9858
The Food smells bad	0.9505	0.0442	0.0053
How ponderous this moonlight be, for methinks it weigh heavily on my resting mood.	0.5537	0.4063	0.04001

Table III shows the same examples as AFINN but with much better results. The 1st two examples had the same result as AFINN in RoBERTa. However, in the case of the third one, it has a score of 0.5537 negative and 0.4063 neutral in contrast to 0.0 score, which is neutral of AFINN. AFINN had its weakness in the case of lengthy reviews, which RoBERTa solves as it still considers the human context in the whole sentence or paragraph.

C. *DistilBERT*

DistilBERT is a smaller, faster, and more efficient version of the BERT language model developed by Hugging Face [17]. It was trained on a version of the English Wikipedia and is designed to perform well on various natural language processing tasks, including sentiment analysis [18]. DistilBERT can be used for sentiment analysis by using its pre-trained weights as the starting point for a sentiment analysis model.

To use DistilBERT for sentiment analysis, we would typically start by fine-tuning the model on a labeled dataset of text annotated with sentiments (e.g., positive, negative, neutral). This can be done using a supervised learning approach, where we train a model to predict the sentiment of a given text based on its features.

Once the model has been fine-tuned, we can use it to classify the sentiment of a new text by inputting it into the model and using the output to predict the sentiment.

TABLE IV
SENTIMENT ANALYSIS USING DISTILBERT

Sentence	Label	Score
The Food is so great	POSITIVE	0.999885
The Food smells bad	NEGATIVE	0.999791
How ponderous this moonlight be, for methinks it weighs heavily on my resting mood.	NEGATIVE	0.992314

It is important to note that the performance of a sentiment analysis model can vary depending on the quality of the training data and the specific characteristics of the text being analyzed. To achieve good performance, it is usually necessary to fine-tune the model on a large and diverse dataset and carefully evaluate the model's performance on various test cases.

VI. RESULTS AND ANALYSIS

The sentiment analysis of the collected restaurant reviews from the Bangladeshi food delivery apps Foodpanda and Hungrynaki was conducted using three different models: RoBERTa, AFINN, and DistilBERT. A total of 20,000 reviews were analysed, and the results showed an accuracy of 74%, 73%, and 77%, respectively.

Table V compares the outcomes of the three models used to conduct the sentiment analysis. An improved balance between precision and recall is reflected in a higher f1 score. Precision measures how well the model can classify positive sentiment, while recall evaluates how well it can recognize all occurrences of positive sentiment.

The models were shown to have poor accuracy, probably due to the short amount of data used for the investigation. The inadequate data may have prevented the models from accurately classifying the reviews' sentiments. Additionally, the low accuracy could be attributable to the complexity of natural language and the subjectivity of sentiment analysis, as individuals' interpretations of sentiment can vary.

We used 1 non-machine learning and 2 machine learning models in this research. We wanted to see how effective these models are in cases of reviews that do not have the best grammar or reviewers who might sometimes give wrong review ratings. The models did have average success in sentiment analysis. All and all, we do think that a non-machine learning model specific to this case can be made and have a decent accuracy score.

The models used (RoBERTa, AFINN, and DistilBERT) were not explicitly trained for sentiment analysis of restaurant reviews from Bangladeshi food delivery apps. Instead, these models were pre-trained on large datasets and then fine-tuned or adapted for use in this specific context. RoBERTa and DistilBERT are pre-trained models developed by Meta and the

TABLE V
COMPARISON OF 3 DIFFERENT MODELS

Metric	AFINN	RoBERTa	DistilBERT
Accuracy	73%	74%	77%
F1 Score	0.81	0.69	0.73
Precision	0.77	0.76	0.79
Recall	0.86	0.63	0.68

Hugging Face team. On the other hand, the AFINN model is a non-machine learning model that uses a set of pre-defined rules and heuristics to identify and classify sentiment in text. Unlike machine learning models, which learn to identify patterns and relationships in data through training, non-machine learning models rely on a fixed set of rules that do not change or adapt over time. In addition to the small dataset size and limitations of the model, low accuracy in sentiment analysis may also be caused by noise in the data or mistranslated reviews. Suppose the data contains multiple layers of sentiment (e.g., sarcasm, irony) that are difficult for the model to interpret. In that case, it can be challenging for the model to classify the sentiment, resulting in lower accuracy accurately.

Despite the low accuracy, the results of this study provide insight into the sentiment of restaurant reviews on food delivery apps in Bangladesh. To further improve the accuracy of sentiment analysis models for restaurant reviews on food delivery apps in Bangladesh, obtaining a larger and more diverse dataset may be necessary. This could be achieved with the direct support of the companies operating the food delivery apps. In addition to a larger dataset, further research may consider additional preprocessing or feature engineering of the data and use different models or techniques for sentiment analysis.

VII. Conclusion

In conclusion, our study examined the use of DistilBERT for sentiment analysis on Bangladeshi food delivery app restaurant review data. DistilBERT surpassed two popular sentiment analysis tools, AFINN and RoBERTa, in accuracy. AFINN's sentiment analysis on ground rules overlooks the human context of reviews. This paradigm works well for short reviews, but this research focuses on how it may not work well for longer reviews. RoBERTa, however, solves this problem and is a fantastic review sentiment analysis model. DistilBERT's transformer design can record complicated word-context interactions, making it ideal for natural language processing. This, along with pre-trained language models, allows the transformer pipeline to achieve high accuracy on a wide range of NLP tasks, including sentiment analysis. Our findings imply that DistilBERT can help extract insights and opinions from Bangladeshi food delivery app restaurant review data. Other text data formats and sentiment analysis applications in business and consumer research might be examined using the DistilBERT model. Researchers and practitioners studying Bangladeshi food delivery consumer attitudes may find the DistilBERT model useful.

References

[1] "Foodpanda: Food & groceries delivery service in bangladesh." [Online]. Available: https://www.foodpanda.com.bd/

[2] "Hungrynaki — best online food delivery service in bangladesh." [Online]. Available: https://hungrynaki.com/

[3] R. A. Laksono, K. R. Sungkono, R. Sarno, and C. S. Wahyuni, "Sentiment analysis of restaurant customer reviews on tripadvisor using naive bayes," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 2019, pp. 49–54.

[4] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Boston, MA: Springer US, 2012, pp. 415–463.

[5] A. M. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," *Mathematical & Computational Applications*, vol. 23, p. 11, 2018.

[6] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A hybrid cnn-lstm model for improving accuracy of movie reviews sentiment analysis," *Multimedia Tools and Applications*, pp. 1–17, 2019.

[7] M. Harish Rao and D. Shashikumar, "Automatic product review sentiment analysis using vader and feature visulaization," in *International Journal of Computer Science Engineeringand Information Technology Research (IJCSEITR)*, vol. 7, 2017, pp. 53–66.

[8] T. Pano and R. S. Kashef, "A complete vader-based sentiment analysis of bitcoin (btc) tweets during the era of covid-19," *Big Data Cogn. Comput.*, vol. 4, p. 33, 2020.

[9] R. Marcec and R. Likić, "Using twitter for sentiment analysis towards astrazeneca/oxford, pfizer/biontech and moderna covid-19 vaccines," *Postgraduate Medical Journal*, 2021.

[10] V. S and J. R, "Text mining: Open source tokenization tools – an analysis," *Adv. Comput. Intell. Int. J. (ACII)*, vol. 3, no. 1, pp. 37–47, Jan. 2016.

[11] J. Biswas, M. M. Rahman, A. A. Biswas, M. A. Z. Khan, A. Rajbongshi, and H. A. Niloy, "Sentiment analysis on user reaction for online food delivery services using bert model," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 2021, pp. 1019–1023.

[12] S. B. Moralwar and S. N. Deshmukh, "Different approaches of sentiment analysis," in *International Journal of Computer Sciences and Engineering*, 2015.

[13] Q. A. R. Adib, M. H. K. Mehedi, M. S. Sakib, K. K. Patwary, M. S. Hossain, and A. A. Rasel, "A deep hybrid learning approach to detect bangla fake news," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2021, pp. 442–447.

[14] M. S. Hossain, N. Nayla, and A. A. Rassel, "Product market demand analysis using nlp in banglish text with sentiment analysis and named entity recognition," in *2022 56th Annual Conference on Information Sciences and Systems (CISS)*, 2022, pp. 166–171.

[15] F. Å. Nielsen, "A new anew: Evaluation of a word list for sentiment analysis in microblogs," *arXiv preprint arXiv:1103.2903*, 2011.

[16] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[18] M. Emon, I. Haque, K. N. Iqbal, M. Mehedi, H. Kabir, M. J. A. Mahbub, and A. A. Rasel, "Detection of bangla hate comments and cyberbullying in social media using nlp and transformer models," in *International Conference on Advances in Computing and Data Sciences*. Springer, 2022, pp. 86–96.