Sentiment Analysis of Restaurant Reviews from Bangladeshi Food Delivery Apps

***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. We collected a dataset of reviews from Foodpanda and Hungrynaki, two major food delivery apps in Bangladesh. We used RoBERTa, AFINN, and DistilBERT 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. Accuracy percentages for the three models tested were as follows: 74% for RoBERTa, 73% for AFINN, and 77% for DistilBERT.**

***Index Terms*—Sentiment analysis, Natural Language Processing, machine learning, 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. (2019) 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 [4], which provides a comprehensive overview of sentiment analysis, including techniques and applications. Furthermore, Hasan et al. (2018) conducted sentiment analysis for Twitter accounts using machine-learning techniques [5]. Rehman et al. (2019) developed a Hybrid CNN-LSTM Model to improve the precision of Sentiment Analysis for movie reviews, which outperforms traditional deep learning and machine learning techniques [6]. HarishRao et al. (2017) used VADER 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 (2021) 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].

According to our knowledge, no studies have focused specifically on sentiment analysis of restaurant reviews on Bangladeshi food delivery apps. Biswas et al. (2021) 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 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.

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-

createdAt	text	reviewerName	ratingvalue
2021-11-10T14:23:22.392Z	wow	MASUD RANA	5
2022-09-08T12:45:03.624Z	The quantity of food is less according to the ...	parvez Joy	2
2021-02-07T08:39:49.642Z	i pre ordered my food for feb 8th but the food...	Jihan Arafat Rahman	4
2021-12-01T08:01:53.684Z	Very good food	SAIF	5
2022-02-09T14:00:29.209Z	I'm satisfied	Turja	5
2022-10-29T15:40:17.395Z	cold food	Nilanjan	1
2022-04-24T15:01:39.534Z	The delivery man intentionally delivered late	Safin Rahman	5
2022-07-28T10:50:08.650Z	chilli chicken and rice order korechilam tara ...	Ayisha Khan	1
2021-09-27T17:39:38.288Z	i ordered fry box but they gave me pasta	Munira	2
2021-03-16T15:09:07.188Z	More you can improve	Sakib Al Hasan	2
2020-12-23T04:15:26.938Z	Unprofessional delivery person	Taufiq	5
2022-05-06T17:59:03.311Z	Wrong order delivered at my place	Farhan Ahmed	1
2022-10-01T15:20:05Z	Always good	kazi	5
2021-08-30T15:13:17.098Z	Go to the bathroom after eating 2 times. Bad t...	Zakir	1
2022-07-26T14:20:21Z	Not good, hair was found in the food.	Zillur	1
2022-05-21T13:13:37.240Z	small portion of chicken lollipop.	Rohan	2
2021-10-09T09:53:38.808Z	munch to munch..	Roy	5
2022-07-31T17:17:50Z	overpriced and deteriorating in quality	Sadat	3
2021-12-14T10:12:32.639Z	The bones were too much, the meat was not that...	Anis	4
2021-01-21T01:44:52.368Z	food was spoiled	asif	1
2022-03-22T10:49:50Z	Super!	Tori	5
2022-01-12T14:26:26.896Z	chicken was not cook enough, i found some raw ...	Suman	3

Fig. 1. Collected Dataset from Foodpanda and Hungrynaki

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 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 machines (SVMs), 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.

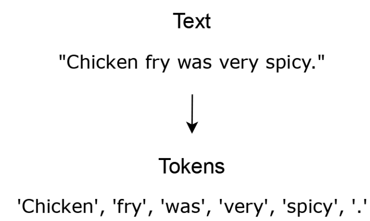

Fig. 2. Tokenizing text

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 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. In addition to the reviews, we collected data on the restaurants being reviewed (e.g., restaurant_id).

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.

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 [13].

V. Methodology

One of the difficulties in sentiment analysis is the scarcity of annotated data sets from which to train a model capable of responding to changes between domains which we can see in the AFINN model approach. Review Sentiment Analysis, or sentiment analysis of any text in general, is divided into two approaches, as we have seen earlier.

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

Fig. 3. Count of Reviews based on Rating

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 [14]. 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.

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 I
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

The above table 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 [15]. The model takes in a sequence of tokens (e.g., words or subwords) 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 II
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

The above table 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 [16]. 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.

Sentiment analysis uses natural language processing techniques to identify and extract subjective information from text, such as the sentiment or emotion expressed in the text. 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 III
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 FUTURE SCOPE

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 IV 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.

TABLE IV
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

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 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 restaurant review data from Bangladeshi food delivery apps. We compared the performance of DistilBERT to two widely used sentiment analysis tools, AFINN and RoBERTa, and found that DistilBERT outperformed both in terms of accuracy.

AFINN is a model that does sentiment analysis on some ground rules and, as a result, misses the human context of the reviews. While it is still perfect for short reviews, our primary focus in this paper is that this model may not be the most reliable if the reviews are larger. On the other hand, RoBERTa fixes this issue and is a great model for sentiment analysis of reviews.

On the other hand, the transformer architecture employed in DistilBERT is particularly well-suited for natural language processing tasks due to its ability to capture complex relationships between words and their contexts. 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.

Overall, our results suggest that DistilBERT is a valuable resource for extracting valuable insights and opinions from restaurant review data from Bangladeshi food delivery apps. Further research could explore the use of the transformer pipeline on other types of text data, as well as other applications for sentiment analysis in the business and consumer research domain. The DistilBERT model has the potential to be a valuable tool for researchers and practitioners interested in understanding the sentiments of consumers in the Bangladeshi food delivery market.

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] 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.

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

[15] 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.

[16] 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.