2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), 8-9 March, Dhaka, Bangladesh

Customer Sentiments Towards Delivery Services in Bangladesh: A Machine Learning-based Sentiment Analysis

Abu Kowshir Bitto
Department of Software Engineering
Daffodil International University
Dhaka-1207, Bangladesh
abu.kowshir777@gmail.com

Md. Shohel Arman
Department of Software Engineering
Daffodil International University
Dhaka-1207, Bangladesh
sshuvo27@gmail.com

Rakhi Moni Saha
Department of Software Engineering
Daffodil International University
Dhaka-1207, Bangladesh
rakhimonisaha02@gmail.com

Hasnur Jahan
Department of Software Engineering
Daffodil International University
Dhaka-1207, Bangladesh
hasnurjahan1997@gmail.com

Imran Mahmud
Department of Software Engineering
Daffodil International University
Dhaka-1207, Bangladesh
imranmahmud@daffodilvarsity.edu.bd

Aka Das
Department of Computer Science
Premier University
Chattogram-4203, Bangladesh
akadas.ctg.premier@gmail.com

***Abstract*—Bangladesh has a large population, which is causing the delivery system to grow up, day by day. Therefore, several companies that provide these delivery services usually referred to as "Currier Service", are growing gradually. In this paper, we focus on the top delivery service companies in Bangladesh. We analyze customer sentiment based on reviews and comments collected from their social media pages. We collected data of customers' comments from the verified social media pages of Sundorban Currier Services, Redx Currier Services, and Pathao Delivery System. We used a combination of Natural Language Processing and machine learning models to categorize Bangla sentiment reviews based on N-gram features. We used the Unigram, Bigram, and Trigram methods to identify the number of negative and positive reviews on real-life social media platforms and to predict whether new reviews were negative or positive. We have found that the Bigram feature is the best for this analysis because it has the highest accuracy of 90.72%. Using the Bangla NLP approach, the machine learning models are categorized in negative and positive reviews with the sentiment analysis method.**

Keywords*—*Bangla Language Processing, Delivery Service, Machine Learning, N-Gram.

I. INTRODUCTION

A "courier service" is a service that transports objects or documents from one area to another. They might be either local or international in scale. A courier company specializes on offering delivery services. Courier firms hire personnel to perform delivery services. The courier company charges a flat rate or a rate directly proportionate to the weight of the good or dependent on the urgency of delivery of the good or both. Courier services differ from regular mail and other types of delivery in terms of speed, security, tracking capabilities, and quick delivery, among other things.

The first courier service in Bangladesh was established in the 1990s. However, it was not as widely used then as it is today. Sundarban Courier Service has been a leading and reliable courier and parcel service provider in Bangladesh since 1983. In addition, courier services such as SA Paribahan, Janani Express Parcel Service, USB Express, DHL Bangladesh, FedEx Bangladesh, Karatoa Courier Service, Redx, Pathao, and others are now operating in Bangladesh. Customer satisfaction with courier services hinges on various factors, including accurate address information, secure parcel packaging, courteous staff behavior, competitive pricing, and effective communication. Therefore, understanding customer sentiment regarding courier services is crucial for improving their systems. Consequently, they have created online platforms where customers can express their opinions through comments. As a result, courier company can understand their fault and customers need which are actually impacting in business. Numerous courier service websites and Facebook pages exist in Bangladesh, but Sundarban, Redx, and Pathao are among the most popular based on user feedback. In our study, we will analyze customers' sentiment about the delivery system based on three courier online activities. For that, we have collected customer comments as data from the Facebook pages of Sundarban, Redx, and Pathao.

II. RELATED WORKS

Here are some instances of similar work publications that align with our research. Below is a synopsis of those editions:

Rachmawan Adi Laksono et al. [1] examined restaurant customer reviews dependent on TripAdvisor information using web crawling methodology, and was examined utilizing the Naive Bayes strategy on WEKA. A relationship with the TextBlob thought examination. WebHarvy devices have been utilized for web information slithering in there. This paper attempted to fix consumer loyalty through negative and positive suppositions in online media.

Qing Sun et al. [2] proposed a unique strategy to explore the eWOM of things dependent on opinion investigation at a fine-grained level from a gigantic volume of online client overviews. Component-based and setting delicate notion examination instrument was utilized there. Thus, utilising the vast amount of customer reviews through social media. Utilizing information from genuine web clients. To examine a programmed fluffy item philosophy mining calculation that removes semantic information from positive and negative online client surveys.

979-8-3503-5028-9/24/$31.00 ©2024 IEEE

2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) | 979-8-3503-5028-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICACCESS61735.2024.10499618

Dietmar Grabnera et al. [3] used lexicon-based sentiment analysis and a classification-based technique to look at customer reviews of hotels. This study followed 3 steps created a semantically oriented dictionary of those text components, using lexicon-based sentiment analysis to create a classification system for customer reviews, and classification findings are compared against a set of unpublished reviews with numerical ratings. The analysis part has been made on basis of TripAdvisor data.

Puspita Kencana Saril et al. [4] investigated the internet review comments on Tokopedia and its service quality dimensions. To comprehend the assistance quality level, sentiment test methods are used to order the audits into positive and negative estimation for five elements of electronic help quality. They used dataset collection comes from an online survey site. As an outcome, they got absolute 609 information audits.

Warrant Songpan [5] proposed the analysis and forecast rating from client audits who remarked as assessment using a likelihood classifier model. The categorization models used in a contextual investigation of client audit in open remarks for preparing information to order remarks as good or bad are exclaimed assessment mining. Many types of research in sentiment analysis and opinion mining have been many languages. Second, the papers recommend and audit a classifier model for categorizing texts as good or bad, and then attempt to apply it to other case studies.

Hyun-Jeong Ban et al. [6] evaluated the essential characteristics in hotel customer experience and satisfaction. 6596 restaurant observations were gathered from Internet platform via web crawling, including hotel brands, writer's identity code, written date, overall score, and review. Then, for text analysis, linear regression technique, text mining techniques, such as semantic network analysis, CONCOR analysis, and factor analysis were employed. Further study of positives and negatives, as well as emotive analysis, are expected to be carried out in future research in order to better comprehend the customer's experience and happiness.

Hyun-Jeong Ban et al. [7] examined airline passengers' sentiment analysis through vast number of Skytrax reviews data. Web crawling was used to collect information. The analysis is split into two sections. One was to use qualitative analysis to examine the meaning of words extracted from the review data through semantic network analysis. The other used the quantitative analytic method to investigate the links between six assessment elements, as well as customer satisfaction and suggestion.

Babak Maleki Shoja et al. [8] developed a glossary for every item category and eliminated unessential terms utilizing Latent Dirichlet Allocation (LDA). They implemented matrix factorization as the collaborative sifting strategy to provide proposals. Their technique moves forward the execution of the recommender framework by joining data from audits and producing proposals with giant status in phrase of rating forecast exactness differentiated to the pattern strategies.

S. K. Lakshmanaprabu et al. [9] prepared a crude dataset from various online websites which are model-based. The approach highlighted in their paper appears and illustrates the most popular social network-based shopping platforms and how they conveyed on. This investigation situating strategy positions the perfect tall lights based on recommending E-trade districts with RPN investigation.

C.J. Hutto et al. [10] proposed a paper about the development, validation, and evolution of VADER. they utilized a combination of subjective and quantitative strategies to deliver, and after that observationally validate, a gold-standard assumption dictionary tailored to microblog-like settings. They discover that joining these heuristics makes strides in the exactness of the estimation examination engine over a few space settings. Interestingly, the VADER dictionary performs outstandingly well within the social media space.

Arghya Ray and et al. [11] find out customer comment section as great source of information of customer in any social media pages. These studies used an multi-method approach combining traditional and NLP techniques. With semi structured interview which is qualitative the study has been one through. With NLP t has been extant. They tested the survey data by NLP which have been gathered by reviews. SEM (Structural-Equation-Modeling) also used here for test path model. In the paper it got that price and benefit in trust have major impact in using customer's OTA and OFD.

III. PROPOSED METHODOLOGY

The main goal of our research is to develop a model that can understand Bengali sentiment phrases of customer satisfaction with Bangladeshi delivery services using an N-gram feature. Data processing and validation are crucial steps [12] before proceeding with model creation. To achieve our goal, we undertook several steps, including data collection, data processing, model create, and others. In Fig. 1, the working procedure is outlined.

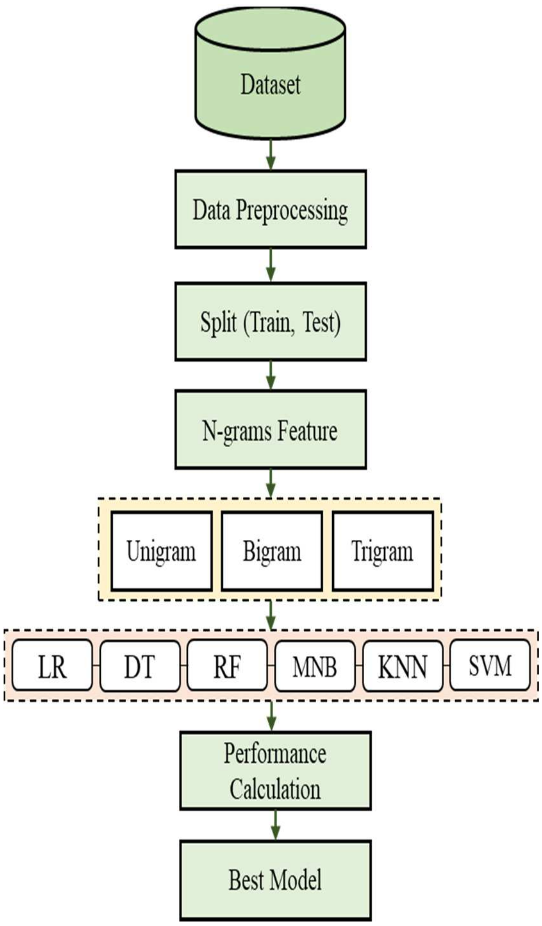

Fig. 1. Step by step procedure diagram.

A. Data Collection

The information we included in our analysis was from Sundarbon, Redx and Pathao courier service Facebook pages comment section. We used the web scraping [13] method to get reviews from the customers. The dataset have 2 features. One is customers' comments, and the other is comment class type. The total number of reviews collected is 822. In Table 1, the sample of those data is shown. As we collect Bangla data, to ensure understanding for everyone, we manually added English phrases to the table.

TABLE I. DATA SAMPLE.

Customers Comment	Class
ময়মনসিংহ এ ডেলিভারি সিস্টেম কোথায়? Where is the delivery system in Mymensingh?	Neutral
আমি এদের থেকে যে সার্ভিস পেয়েছিলেন তা ভালোই I received the service from them and it was good.	Good
আপনাদের মধ্যে আন্তরিকতার অভাব ভাই।আগে ব্যবহার ঠিক করেন There is a lack of transparency among you. Please improve.	Bad
আপনাদের উচিত গ্রাহকদের সাথে ভালো আচরণ করা Treat your customers well, it's your responsibility.	Bad
আমি সময়ের মধ্যে ডেলিভারি পেয়েছি. সার্ভিসে সন্তুষ্ট। I have received the delivery on time. Satisfied with the service.	Good

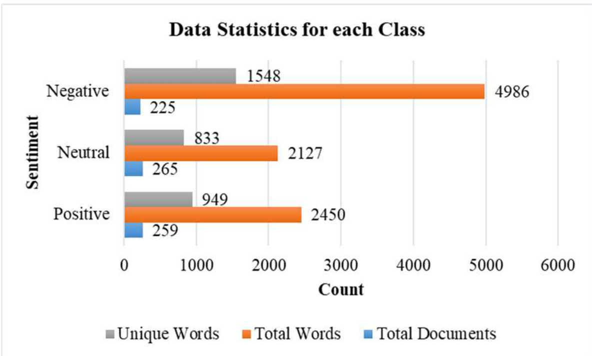

Fig. 2. Sentiment Class for Review.

We manually input and ensured that all positive reviews are labeled as '0', negative reviews as '1', and neutral reviews as '2' manually. Specifically, there are 321 positive reviews, 227 negative reviews, and 274 neutral reviews "Fig. 2". The comments were written entirely in Bengali, with some incorporating a mix of Bengali and English.

B. Data Cleaning

We used a Python script to remove unnecessary data. Next, we eliminated 73 data points with a length of less than three words. After cleaning, we were left with 749 records. The entire dataset was cleaned by applying a condition to remove comments and reviews with a minimum length of two, followed by the same process for the remaining data.

C. Data Preprocessing

After cleaning the data, we created a summary of negative, positive, and unique data. For the neutral class, there are 265 sentences, totaling 2,127 words with 833 unique words. The top ten most frequently used words are also shown. For the positive class, there are 259 sentences, totaling 2,450 words with 949 unique words. The 10 most frequently used words are also shown. For the negative class, there are 225 sentences, totaling 4,986 words with 1,548 unique words. The 10 most frequently used words are also shown "Fig. 3".

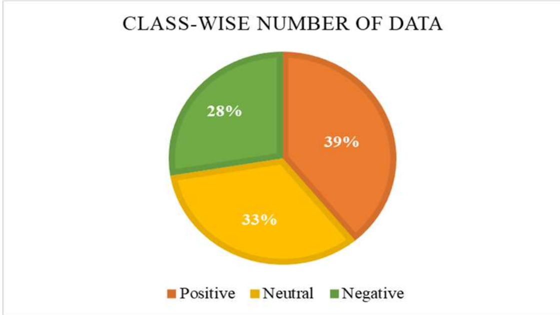

Fig. 3. Data Statistics.

The length of the comment is then displayed. There are 253 comments with a maximum length of 253 characters, 2 comments with a minimum length of 2 characters, and 13 comments with an average length of 13 characters "Fig. 4".

Next, we created a new dataset using only the positive and negative response. There are 259 positive and 225 negative data. Label encoding has been applied to this dataset. After that, our data was divided into two main categories: training and testing. We then utilized the holdout method of cross-validation to separate our dataset, resulting in a training and testing data ratio of 80:20.

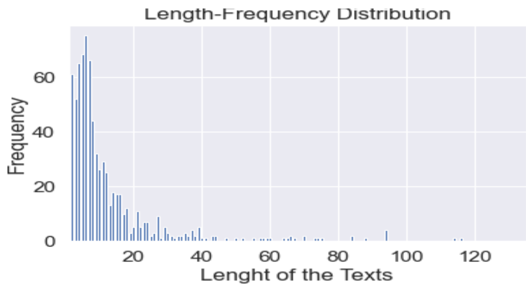

Fig. 4. Length-Frequency Distribution.

D. Feature Extraction

We implement the TF-IDF function to extract the feature, which transformed our perspective.

E. Model Implementation

We implement six machine-learning classifiers (MLCs) and their relevant theory is given below.

Logistic Regression (LR) is a classification algorithm that predicts discrete values such as true or false, yes or no, 0 or 1, and so on, based on a given input set of independent variables. In essence, it estimates the probability of an event occurring from the data. Since it estimates probability, the values obtained will always fall within the range of 0 and 1. The Sigmoid function is the function used in Logistic Regression.Formula is.

$$Z = W^T M = \mu(Z) = \mu = \left(\frac{1}{e^z}\right) \tag{1}$$

In Eq. (1), W is defined as a weight vector, M is defined as data vector and μ is the function of Sigmoid.

Decision Tree (DT) analysis could be a common, prescient displaying tool for these solutions. This model refers to the non-performed learning method that may be used for both classification and regression.

$$E(x) = \sum_{n=1}^{n} -p_i \log_2 p_i \tag{2}$$

Random Forest (RF) analysis builds decision trees based on data sets and then averages the results from each tree. It selects the best model by means of voting.

Multinomial Naive Bayes (MNB) is the method that is generally utilized in Natural Language Processing (NLP), which is a probabilistic method. The algorithm is made up based on the Bayes theorem and guessing the articles. The Eq. (3) of MLB is.

$$P(M|N) = P(M) \times \frac{P(N|M)}{P(N)} \quad (3)$$

Where *M* is prior probability = *P(M), N* is prior probability = *P(N),* and predictor *B* is incident for the given class *A* likelihood = *P(N|M).*

K-Nearest Neighbors (KNN) is a supervised learning algorithm commonly used in data mining and artificial intelligence. The classifier computation involves learning based on the comparison nature of information vectors from others. The k-closest neighbor calculation is a type of "lazy learner" that does not construct a model using the training set until a query is made about the assembly process.

Linear Support Vector Machines (SVM) utilized for classification and regression tasks. It is commonly utilized for limited datasets due to the high computational cost of training.

$$f\ (m) = \ w_t m + n = \sum_{k=0}^{k} w_k n_k + n = 0 \quad (4)$$

In Eq. (4) the M-dimensional vector is w and hyperplane is denoted by m which is scalar.

F. N-gram Feature

This is defined that the N number of tokens or language word sequences for the NLP model processing [15]. There are three types of N-grams features in NLP processing those are:

1) *Unigram:* This is by one-word sequence.
2) *Bigram:* This is two by two sequences of words.
3) *Trigram:* This is the three-word sequence of words in the whole sentence.

G. Performance Measurement

We computed precision, recall, F1-Score and accuracy to compare the model performance for each of N-gram (N=1, 2, 3) feature, showed in Eq. (5), Eq. (6), Eq. (7) and Eq. (8).

Precision: This is the percentage of all favorable remarks that can be expected and correctly predicted.

$$Precision = \frac{TP+FP}{TP} \quad (5)$$

Recall: The fraction of all comments, not just the good ones, that is accurately anticipated is known as Recall.

$$Recall = \frac{TP+FN}{TP} \quad (6)$$

F1-Score: This is the estimated value of precision and recall equation.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

Accuracy: It refers to the number of comments properly predicted.

$$Accuracy = \frac{TP+TN}{Total\ no.\ of\ Comments} \quad (8)$$

IV. Analysis and Discussions

To determine our model's optimum performance, we ran multiple models on our split data. Given the classification nature of the problem, we utilized a variety of classification algorithms. We applied LR, DR, RF, MNB, KNN, and SVM algorithms. These algorithms are used for text mining and natural language processing. Unigram, Bigram, and Trigram have been used here. We evaluated the performance of each model for Unigram, Bigram, and Trigram features after fitting our data to the model. We calculated the model's accuracy, precision, recall, and F1 score to assess performance.

TABLE II. Performance in Unigram Feature.

Model	Accuracy	Precision	Recall	F1 Score
LR	88.66	83.33	98.04	90.09
DT	88.66	88.46	90.20	89.32
RF	85.57	81.36	94.12	87.27
MNB	89.69	90.20	90.20	90.20
KNN	88.66	91.67	86.27	88.89
Linear SVM	83.51	76.92	98.04	86.21

TABLE III. Performance in Bigram Feature.

Model	Accuracy	Precision	Recall	F1 Score
LR	85.57	79.37	98.04	87.72
DT	84.54	82.14	90.20	85.98
RF	85.57	79.37	98.04	87.72
MNB	90.72	90.38	92.16	91.26
KNN	84.54	82.14	90.20	85.98
Linear SVM	54.64	53.68	100.00	69.86

TABLE IV. Performance in Trigram Feature.

Model	Accuracy	Precision	Recall	F1 Score
LR	84.54	78.12	98.04	86.96
DT	83.51	81.82	88.24	84.91
RF	86.60	80.65	98.04	88.50
MNB	88.66	84.48	96.08	89.91
KNN	87.63	84.21	94.12	88.89
Linear SVM	52.58	52.58	100.00	68.92

From Table 2 we can see each model's performance in the Unigram feature, where MNB performed with the highest accuracy at 89.69 percent. Table 3 shows that SVM exhibited the lowest accuracy among all models, with 54.64%. Table 4 illustrates each model's performance for the Trigram feature, with MNB achieving the highest accuracy of 88.66% and Linear SVM again demonstrating the lowest accuracy at 52.58%.

Fig. 5 shows the accuracy comparison of the N-gram feature with all six models' performance. In every N-gram feature, MNB got the highest accuracy, and Linear SVM got the lowest accuracy. Among Unigram, Bigram, and Trigram features, Bigram got the highest accuracy of 90.72 with MNB models.

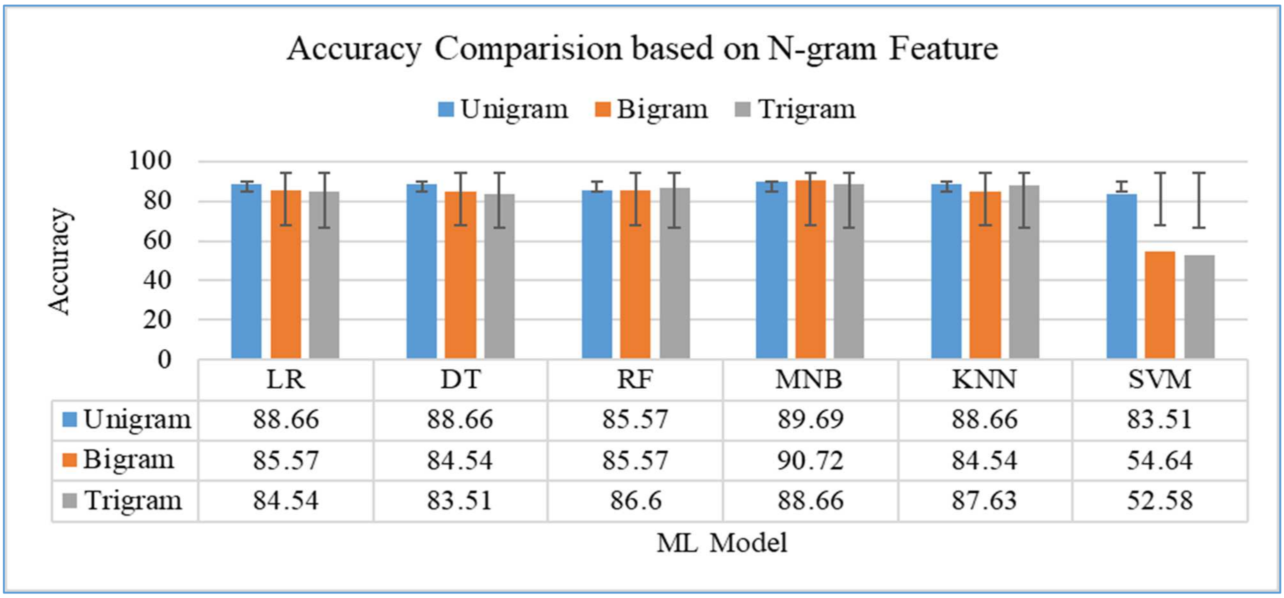

Fig. 5. Comparison of Accuracy.

V. CONCLUSION

In this study, we employed Natural Language Processing (NLP) in Bangla to examine consumer feedback and assess opinions regarding three well-known courier services in Bangladesh: Sundarban, Redx, and Pathao. From these delivery service providers' Facebook sites, we gathered ratings and comments from previous customers. Using Unigram, Bigram, and Trigram features in a variety of models, including LR, DR, RF, MNB, KNN, and SVM, we discovered that the Bigram feature worked particularly well in conjunction with the Multinomial Naive Bayes (MNB) model, yielding the highest accuracy at 90.72%. Based on consumer input, this study improves our understanding of these courier services' overall performance by offering insightful information about how customers feel about them. In the ahead, we will work with neutral comments, as well will collect more data and we will also include deep learning models, and also will try to integrate ensemble models.

REFERENCES

[1] R. A. Laksono, "Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019," *Proc. 2019 Int. Conf. Inf. Commun. Technol. Syst. ICTS 2019*, pp. 49–54, 2019, doi: 10.1109/icts.2019.8850988.

[2] Q. Sun, J. Niu, Z. Yao, and H. Yan, "Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level," *Eng. Appl. Artif. Intell.*, vol. 81, no. December 2018, pp. 68–78, 2019, doi: 10.1016/j.engappai.2019.02.004.

[3] D. Gräbner, M. Zanker, G. Fliedl, and M. Fuchs, "Classification of Customer Reviews based on Sentiment Analysis," in *Information and Communication Technologies in Tourism 2012*, M. Fuchs, F. Ricci, and L. Cantoni, Eds., Vienna: Springer, 2012, pp. 460–470. doi: 10.1007/978-3-7091-1142-0_40.

[4] P. K. Sari, A. Alamsyah, and S. Wibowo, "Measuring e-Commerce service quality from online customer review using sentiment analysis," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018, doi: 10.1088/1742-6596/971/1/012053.

[5] W. Songpan, "The analysis and prediction of customer review rating using opinion mining," *Proc. - 2017 15th IEEE/ACIS Int. Conf. Softw. Eng. Res. Manag. Appl. SERA 2017*, pp. 71–77, 2017, doi: 10.1109/SERA.2017.7965709.

[6] H. J. Ban, H. Choi, E. K. Choi, S. Lee, and H. S. Kim, "Investigating key attributes in experience and satisfaction of hotel customer using online review data," *Sustain.*, vol. 11, no. 23, 2019, doi: 10.3390/su11236570.

[7] H. Ban and H. Kim, "Ban, H. J., _ Kim, H. S. (2019). Understanding customer experience and satisfaction through airline passengers' online review. Sustainability, 11(15), 4066..pdf," *Sustain.*, vol. 11, 2019.

[8] B. M. Shoja and N. Tabrizi, "Customer Reviews Analysis with Deep Neural Networks for E-Commerce Recommender Systems," *IEEE Access*, vol. 7, pp. 119121–119130, 2019, doi: 10.1109/ACCESS.2019.2937518.

[9] S. K. Lakshmanaprabu *et al.*, "Ranking analysis for online customer reviews of products using opinion mining with clustering," *Complexity*, vol. 2018, 2018, doi: 10.1155/2018/3569351.

[10] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for," *Eighth Int. AAAI Conf. Weblogs Soc. Media*, pp. 216–225, 2014.

[11] A. Ray and P. K. Bala, "User generated content for exploring factors affecting intention to use travel and food delivery services," *Int. J. Hosp. Manag.*, vol. 92, p. 102730, Jan. 2021, doi: 10.1016/J.IJHM.2020.102730.

[12] A. K. Bitto and I. Mahmud, "Multi categorical of common eye disease detect using convolutional neural network: a transfer learning approach," *Bull. Electr. Eng. Informatics*, vol. 11, no. 4, pp. 2378–2387, Aug. 2022, doi: 10.11591/eei.v11i4.3834.

[13] A. K. Bitto, M. H. I. Bijoy, M. S. Arman, I. Mahmud, A. Das, and J. Majumder, "Sentiment analysis from Bangladeshi food delivery startup based on user reviews using machine learning and deep learning," *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 2282–2291, 2023, doi: 10.11591/eei.v12i4.4135.

[14] A. K. Bitto, M. H. I. Bijoy, S. Khan, I. Mahmud, and K. B. M. B. Biplob, "Approach of Different Classification Algorithms to Compare in N-gram Feature Between Bangla Good and Bad Text Discourses," in *Machine Intelligence Techniques for Data Analysis and Signal Processing. Lecture Notes in Electrical Engineering.*, D. S. Sisodia, L. Garg, R. B. Pachori, and M. Tanveer, Eds., Singapore: Springer, 2023, p. Vol 997. doi: https://doi.org/10.1007/978-981-99-0085-5_9.

[15] H. Jahan *et al.*, "Sentiment Analysis of E-commerce Consumer Based on Product Delivery Time Using Machine Learning," in *Machine Intelligence and Data Science Applications. Lecture Notes on Data Engineering and Communications Technologies.*, V. Skala, T. P. Singh, T. Choudhury, R. Tomar, and M. A. Bashar, Eds., Singapore: Springer, 2022, p. Vol 132. doi: https://doi.org/10.1007/978-981-19-2347-0_51.