Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language

Minhajul Abedin Shafin, Md. Mehedi Hasan, Md. Rejaul Alam,
Mosaddek Ali Mithu, Arafat Ulllah Nur, Md. Omar Faruk
Department of Computer Science and Engineering
Daffodil International University, Dhaka, Bangladesh
*minhajul15-8700@diu.edu.bd, mehedi15-8421@diu.edu.bd, rejaul15-9077@diu.edu.bd,
mosaddek15-8669@diu.edu.bd, arafat15-9372@diu.edu.bd, omar15-8709@diu.edu.bd

Abstract— **In this era of internet technology, in Bangladesh, online marketing or e-commerce businesses were already thriving. Due to the COVID-19 pandemic, as people are in lockdown, online shopping became the main platform for shopping as it is the safest way. It accelerated the businesses to come online. More online product service providers makes it better for people but also raises the question of product quality and services. So it is easy for new customers to get scammed while shopping online. Our goal is to make a system that will analyze the customer's feedback from online shopping and provide a ratio of the positive and negative feedback written in Bangla from the previous customers using Natural Language Processing (NLP). We have collected over 1000 feedback and comments on the product to conduct the research. We used sentiment analysis along with some classification algorithms like KNN, Decision Tree, Support Vector Machine (SVM), Random Forest and Logistic Regression. With the highest accuracy of 88.81%, SVM outperformed all the other algorithms.**

Keywords— *Machine Learning, Data Analysis, Sentiment Analysis, NLP, Classification, Prediction, SVM.*

I. INTRODUCTION

Bangladesh, being a developing country, had its fair share of technological improvement over the past decade. With the introduction of fast internet connectivity, a lot of businesses came online, which we know as e-commerce. E-commerce opened a door for offline businesses to expand all over the country. With more customers, comes great responsibility for maintaining quality and services. Started slowly in the late 90s [1], the growth of e-commerce started from early 2000 - 2008, with a drawback of proper online payment gateway. When in 2010, SSL COMMERZ launched its payment aggregator in Bangladesh as well as the addition of WiMAX which accelerated internet speed, e-commerce started to change rapidly in terms of user numbers. As a newly introduced marketing medium, the masses were not fully aware of the medium, even some did, they didn't have much faith in the system. But over time, people became more comfortable using the medium. Daraz, Bikroy, Evaly, Chaldal.com being a few of the popular e-commerce sites in Bangladesh. Worldwide pandemic made us stay home in lockdown which also increased the possibility of buying things online. More and more businesses are moving online, which also increase the probability of selling defect product by unscrupulous businessman across the country. In this case, public opinion plays a vital role to identify faulty products or ever unscrupulous businessmen. But in a product, there might be thousands of comments and reviews, it's hard for people to analyze all the comments and reviews to identify the best product. To be specific, the reviews that are in Bangla

language as people are more comfortable expressing their opinion in their mother tongue. So, our goal is to identify people's sentiment on products, to be precise, positive, and negative feedback on products based on Bangla Language comments and reviews. We wanted to establish a system that will enable the possibility to the users which will automatically show the ratio of negative and positive reviews by using sentiment analysis. Sentiment analysis enables us to identify the sentiment of a particular text. Sentiment analysis is one of the applications of NLP. We also used different supervised learning algorithms e.g. KNN, Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression to classify the data set and to see which algorithm performs best with the highest accuracy.

The paper includes the following sections, section II contains literature review of related work. The proposed methodology is provided in section III. In Section IV a complete discussion is held on Experimental result and we finalized the paper with a conclusion in Section V.

II. LITERATURE REVIEW

Sentiment analysis has become a major focus point in natural language processing. Review analysis is one of the well-studied subjects of Sentiment analysis.

Omar Sharif et. al. [2] proposed a system where they classify customer's restaurant reviews written in Bangla. They used sentiment analysis to classify reviews into positive and negative classes based on the feedback of the restaurants. They applied their proposed system with 1000 Bengali textual restaurant reviews and got the accuracy of 80.48% by using multinomial Na"ive Bayes.
Vallikannu et. al. [3] used sentiment analysis to catch people's opinion feedback about Oman tourism. For their research they collected people's opinions about tourism in their country from twitter. They analyzed the effect of four factors that are domain specific ontology, entity specific opinion extraction, combined lexicon based approach and conceptual semantic sentiment analysis to determine the sentiment analysis of tweets regarding Oman tourism. From the experiment they discovered that using conceptual semantic sentiment analysis improved the performance of the domain to 85.54%
Nayan Banik et. al. [4] worked on developing an automated polarity detection system from the textural information form movie reviews written in Bangla. They collected data from Facebook movie review related groups, pages, twitter posts

of famous reviewers and Bangla Movie Database (BMDb). Their dataset contained 800 movie reviews with 400 positive and 400 negative reviews. They used two classification algorithms, Naïve Bayes (NB) and Support Vector Machines (SVM) where SVM performed better with precision of 0.86. Nafis et. al.[5] Developed a deep learning based model to classify a Bangla sentence.

They classified them into three-class (positive, negative, neutral) and a five-class (strongly positive, positive, neutral, negative, strongly negative) sentiment label.. They also build models to extract the emotion from Bangla sentences as one of the six basic emotions (anger, disgust, fear, joy, sadness and surprise).They collected data from Bangla, English and Romanized Bangla comments from various YouTube videos. They implemented two approaches for the first model they used Short Term Memory (LSTM) and for The second method they applied Convolutional Neural Network (CNN).Their proposed model provided 65.97% and 54.24% accuracy in three and five labels sentiment, respectively.

Lutfun Nahar et. al. [6] developed a system to filter out Bangla political and sports news from social media form the textual information. 1000 long written posts were collected by them from different news portals as well as social media. They also used a self-made News Corpus for experimental purposes. In order to get better accuracy, stop words were removed from the dataset. Bengali stemming was done to minimize the size of the dictionar. They implemented Naïve Bayes Classifier, Support Vector Machine and Neural Networks to classify news as well as post or comments written in Bangla. Naive Bayes achieved better performance than other three algorithms

Paramita Ray et. al. [7] proposed a framework for sentiment analysis using R software which can study sentiment of users on twitter data using Twitter API. They categorized the tweets into positive, negative, or neutral opinion to evaluate the overall sentiment of customer or user regarding particular products or services. They used lexical method for classification and worked on a dictionary-based approach

K. Indhuja et. al. [8] their objective was to extract features from the product reviews and classify them into positive, negative and neutral categories. Their method used fuzzy functions to imitate the effect of modifiers, concentrators and dilators. Their research proved to be effective on sentiment analysis using fuzzy logic with efficiency of 85.58% accuracy.

Customer feedback is necessary to analyze the quality of the product. But people often forget to provide necessary review. MINARA P ANTO et. al. [9] proposed a framework for providing automatic feedback on the basis of data collected from twitter. They gathered data from twitter page using the crawler Twitter4J API that fetched most recently posted 100 tweets for analyzing. POS tagging was used to categorize data into various groups like nouns, adjectives, subjects etc. Also By using the unigram approach the frequency of the words were calculated.

From above discussion we figure out that our work is different in some ways from them as our dataset is quite large and the accuracy rate is quite high. Additionally, we analyzed ratio of all the reviews of a particular product to predict the positive and negative reviews.

III. METHODOLOGY

We have taken the commensurate seven steps to conclude our research final survey to collect our final data. The explanation is given bellow in Fig. 1.
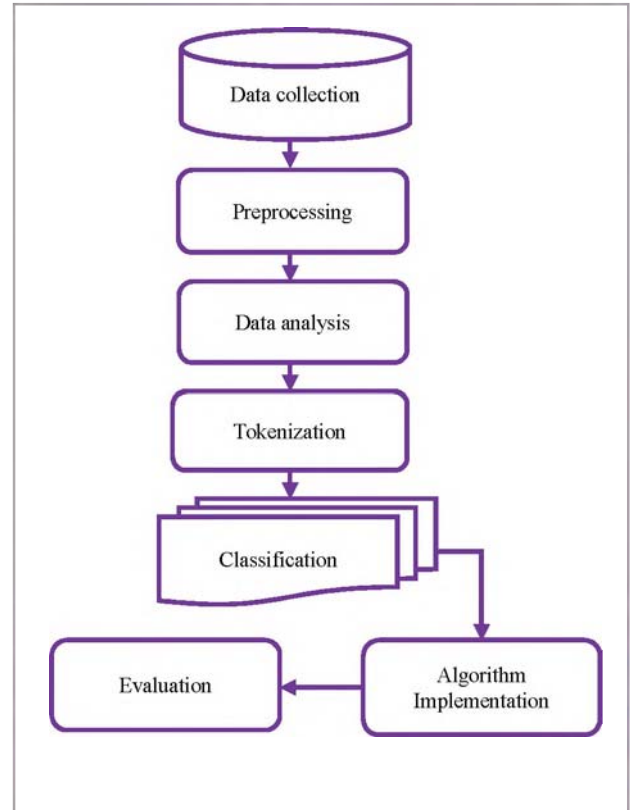

Fig. 1. Flowchart of Methodology

A. Data Collection:

E-commerce is a huge platform nowadays. Millions of comment reviews are shared by buyers or users every day for each product. Every review contains its individual sentiment. As we worked with only Bangla language, our focus was only Bangla reviews on e-commerce sites. Then we collected valid and understandable Bangla reviews from e-commerce sites. We have collected 1020 reviews for our research data.

B. Preprocessings:

Initially, we preprocessed data by excluding Bangla punctuations moreover Bangla stop words. After that, we cleared the emoji from inside of the reviews. As a portion of the sentiment, we cannot remove emoji from reviews constantly, therefore we included this at the end of the review so that the predilection of the sentence can evaluate accurately.

C. Data Analysis:

After preprocessing data we classified the dataset into two classes (positive and negative) based on reviews perspective. As algorithms never apprehend string data directly. Because of this, we are required to transform our processed data into numerical form. To achieve this goal, we employed TF-IDF algorithm.

$$W_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right) \qquad (1)$$

In the TF-IDF (1) $W_{i,j}$ is a score of TF-IDF. $tf_{i,j}$ is the number of incidents of i in j. The number of documents is signified

by N. And df_i representing the number of documents including i.

D. Tokenization:

TABLE I. TOKENIZATION

Raw Data	Type	Tokenized data
ডেলিভারিতে অনেক খুশি প্যাকেজিং ভালো ছিল	Positive	'ডেলিভারিতে', 'অনেক', 'খুশি', 'প্যাকেজিং', 'ভালো', 'ছিলো'
অরিজিনাল টা দেওয়ার জন্য ধন্যবাদ	Positive	'অরিজিনাল', 'টা', 'দেওয়ার', 'জন্য', 'ধন্যবাদ'
এত বাজে আর দেখিনি	Negative	'এত', 'বাজে', 'আর', 'দেখিনি'

By using tokenization, we separated single words as a token from a sentence which is crucial to scrutinize the sentence. Through Table I. we wanted to exhibit the tokenization procedure. Before utilizing TF-IDF algorithm, every string must be tokenized. Table I. illustrates that raw data form e-commerce sites that were taken as input. Punctuation, stop word were removed accordingly. Afterward, to tokenize our raw data, we applied porterstemmer which is a frequently used function for tokenization.
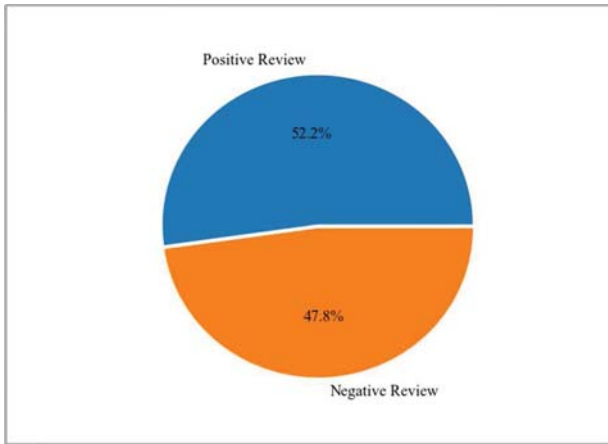
E. Classification:

Fig. 2. Database Representation

We established two categories of Bangla review, as a consequence, our whole database distributed into two classes represented positive and negative by corresponding 0,1. Every class included around five hundred Bangla review. Fig. 2 demonstrates the classification of the complete dataset. It also describes the percentage of data for individual classes.

F. Algorithm Implementaion:

After implementing five algorithms, we observed that SVM delivered the leading accuracy by using 30% of the test data. Other algorithms were similarly conducted pretty well. As SVM produces the most reliable performance, we determined to employ it to predict the sentiment of the reviews.

G. Evaluation:

In this part, we confirmed our thought and methodology manually. Initially, fifteen positive and nine negative Bangla reviews were adopted for each selected class form different e-commerce sites. Later we estimate it with the help of the SVM algorithm. Fig. 3 describes the evaluation of our project.
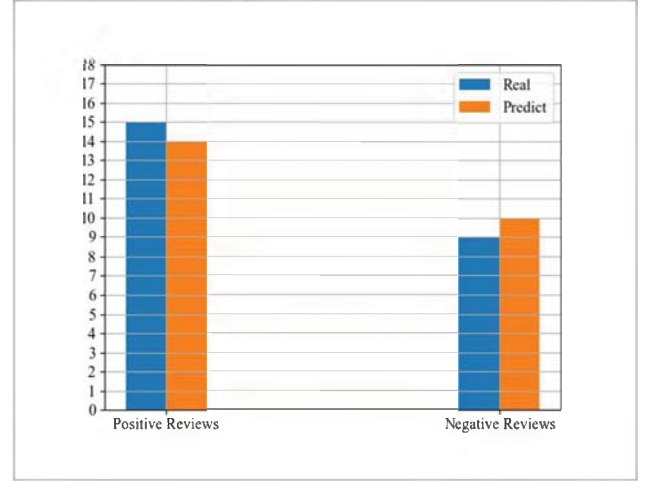

Fig. 3. Review prediction

We collected twenty four recent data from the e-commerce site, which was labeled as real data. Fig. 3. Displayed for positive and negative class, the algorithm executed accurately. The blue bar represents the amount of
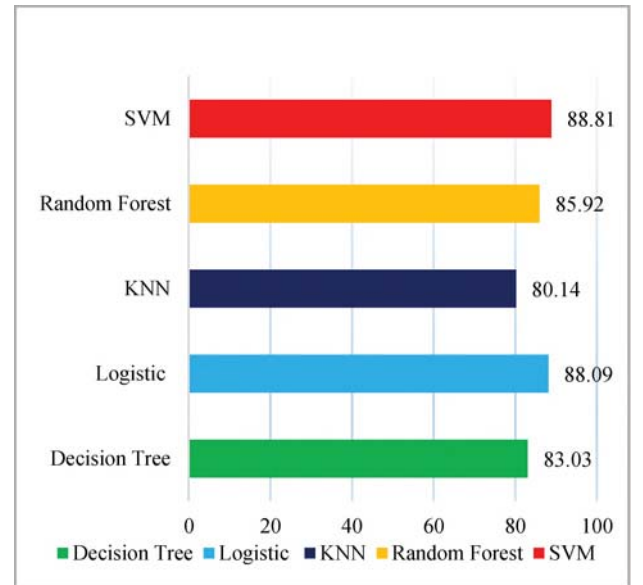

Fig. 4. Highest accuracy of applied algorithms

real data category and the orange bar represents the predicted amount of each class. From this graph, we can easily figure out that it gives exact outcomes excepting one error out of twenty four data. Fig. 4 represents the peak accuracy of each algorithms generated.

IV. Experimental Result

To find the preciseness of our work, we have employed five traditional ML algorithms in our preprocessed dataset. We entered the amount of performance that was produced by the algorithms in Table II. As a result, it will be understood, compare, and easy to decide on future research depending on their accomplishment. Five types of training and testing sets are applied to measure the efficiency which is 30%, 40%, 50%, 60%, and 70%. A stunning consequence was established by the dissolving of these five algorithms. In our research, all the algorithms are operated similarly with high-level accuracy. V. Vapnik [10] explained, In Machine Learning and pattern classification Support Vector Machine (SVM) procured enhance performance because of the minimization of recognized data From all the algorithms, SVM generated 88.81% accuracy with the 30% test data which is presented in Table II by red rectangular border-box. Every column contains a yellow mark that describes the most leading accuracy of those algorithms by various types of test data uses rate.

TABLE II.   ACCURACY TABLE

Test Data usage rate	Algorithms				
	KNN	Decision Tree	Logistic regression	Random Forest	SVM
30%	80.14%	83.03%	88.09%	85.92%	88.81%
40%	77.78%	79.95%	86.18%	83.74%	86.18%
50%	78.52%	77.66%	85.47%	82.21%	86.55%
60%	77.76%	76.13%	80.29%	80.29%	83.54%
70%	77.67%	77.67%	78.91%	75.66%	81.55%

R. E. Wright [11] explained that any inspection which is performed by employing logistic regression as like as using statistical techniques. The second highest accuracy was produced by Logistic Regression with 88.09% efficiency by using 70% training data and 30% test data. P. O. Gislason et al. [12] for classification of multisource, Random forest preserves high dimensional data and presides a large number of trees. In our work, the Random forest obtained 85.92% accuracy by 70% training data. C. Jin et. al. [13] Decision tree is one of the significant approaches for employing model classification and prediction for both induction research and data mining. The decision tree gained 83.03% by using 70% training data. Y. Yang et. al. [14] for text categorization application, KNN is appreciated broadly and it has the area of expertise on statistical and pattern reorganization. Lastly, KNN got 80.14% efficiency with 30% test data usage rate.
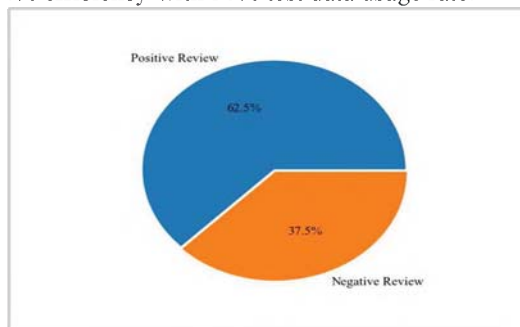

Fig. 5. Real Review Percentage

Fig. 5 and Fig. 6 illustrates the ratio of real and predicted class for a specific product. In Fig. 5 we can see that the percentage of positive and negative review in a specific product where positive is 62.5% and negative is 37.5%. And Fig. 6 shows the predicted output by our model for that particular product reviews by analyzing the review of that product. It also provides with a ratio of positive and
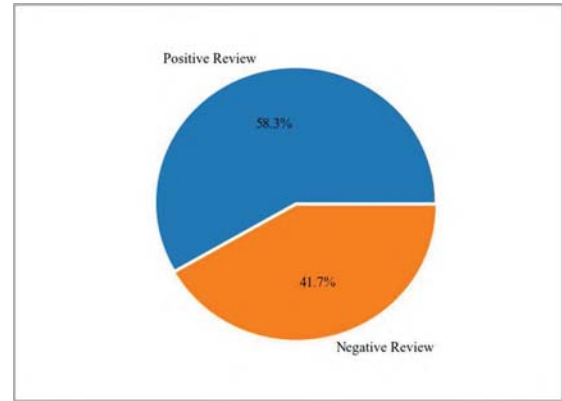

Fig. 6. Predicted Review Percentage

negative reviews which is very close to real ratio. So it can be said that our model works efficiently.

V. Conclusion and Future Work

With more than 1000 collective data, we analyzed product review in Bangla language. Further analysis of the data shows that our proposed system can detect proper sentiment off of Bangla review or comments effectively. From selected few popular classification algorithms KNN, Decision Tree, Support Vector Machine (SVM), Random Forest and Logistic Regression, SVM performed outstandingly with 88.81% accuracy. We believe that our proposed system can decrease customer ordeal while shopping online as they are able to see through the system for product reviews by ratio of previous customers' positive and negative feedback. It can be also proven helpful for the seller because he can identify defects of his products and provide better customer service.

Future work based on this particular research that we have done, we aim to develop a global-browser plugin that analyzes a particular webpage for reviews and comments to show the user real-time ratio of the previous customer's sentiment and also we want to accomplish the same task for a specific product.

References

[1] R. Moslem. (Apr 18, 2017). A Brief History of E Commerce in Bangladesh. Available: https://medium.com/@r_moslem/a-brief-history-of-e-commerce-in-bangladesh-e9ec27e29caf

[2] 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: IEEE.

[3] V. Ramanathan and T. Meyyappan, "Twitter text mining for sentiment analysis on people's feedback about oman tourism," in 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1-5: IEEE.

[4] N. Banik and M. H. H. Rahman, "Evaluation of naïve bayes and support vector machines on bangla textual movie reviews," in 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-6: IEEE.

[5] N. I. Tripto and M. E. Ali, "Detecting multilabel sentiment and emotions from bangla youtube comments," in 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018, pp. 1-6: IEEE.

[6] L. Nahar, Z. Sultana, N. Jahan, and U. Jannat, "Filtering Bengali Political and Sports News of Social Media from Textual Information," in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-6: IEEE.

[7] P. Ray and A. Chakrabarti, "Twitter sentiment analysis for product review using lexicon method," in 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), 2017, pp. 211-216: IEEE.

[8] K. Indhuja and R. P. Reghu, "Fuzzy logic based sentiment analysis of product review documents," in 2014 First International Conference on Computational Systems and Communications (ICCSC), 2014, pp. 18-22: IEEE.

[9] M. P. Anto, M. Antony, K. Muhsina, N. Johny, V. James, and A. Wilson, "Product rating using sentiment analysis," in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pp. 3458-3462: IEEE.

[10] V. Vapnik, The nature of statistical learning theory. Springer science & business media, 2013.

[11] R. E. Wright, "Logistic regression," 1995.

[12] P. O. Gislason, J. A. Benediktsson, and J. R. J. P. R. L. Sveinsson,"Random forests for land cover   classification," vol. 27, no. 4, pp.294-300, 2006

[13] C. Jin, L. De-Lin, and M. Fen-Xiang, "An improved ID3 decisiontree algorithm," in 2009 4th International Conference on ComputerScience & Education, 2009, pp. 127-130: IEEE.

[14] Y. Yang and X. Liu, "A re-examination of textcategorizationmethods," in Proceedings of the 22nd annual international ACMSIGIR conference on Research and development in informationretrieval,   1999, pp.   42-49