International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 8 Issue 3, May-June 2024 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Sentiment Analysis of Bangladeshi E-Commerce Site Reviews Using Machine Learning Approaches

Mohammad Kasedullah[1], Nakib Aman[2], Md. Mehedi Hasan[3]

[1]Lecturer, Department of Computer Science and Engineering, Varendra University, Rajshahi, Bangladesh
[2]Lecturer, Department of Computer Science and Engineering,
Pabna University of Science and Technology, Pabna, Bangladesh
[3]Software Engineer I, Developer eXperience Hub, Rajshahi, Bangladesh

ABSTRACT

In the context of Bangladesh, the E-commerce sector is experiencing continuous growth, particularly during the global crisis. Amidst the plethora of available platforms, Daraz has emerged as the most successful marketplace, offering users a wide array of shopping options. However, the abundance of reviews and comments on this online platform presents a challenge for consumers trying to make optimal choices. This research focuses on systematically categorizing positive and negative reviews to enhance user decision-making. To achieve this objective, a range of classifiers, including Multinomial Naive Bayes, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K Neighbors Classifier, and Support Vector Machine with different kernels, were employed. The dataset underwent thorough cleaning, followed by the application of Term Frequency-Inverse Document Frequency (TF-IDF) with Principal Component Analysis (PCA) to enhance feature representation. The findings of this study indicate that the Multinomial Naive Bayes classifier, especially when utilizing Bigram and Trigram features, outperformed other classifiers, demonstrating superior accuracy. The implementation of this classifier holds significant promise for assisting businesses operating on various platforms, enabling them to distinguish between positive and negative reviews effectively. Consequently, this approach empowers businesses to furnish customers with valuable insights into the quality of products, contributing to a more informed and confident consumer base.

KEYWORDS: *E-commerce, Daraz, Reviews, Sentiment Analysis, Multinomial Naive Bayes, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, K Neighbors Classifier, Support Vector Machine, Data Cleaning, Term Frequency-Inverse Document Frequency (TF-IDF), Principal Component Analysis (PCA), Product Quality, User Decision-Making*

How to cite this paper: Mohammad Kasedullah | Nakib Aman | Md. Mehedi Hasan "Sentiment Analysis of Bangladeshi E-Commerce Site Reviews Using Machine Learning Approaches" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-8 | Issue-3, June 2024, pp.442-448, URL: www.ijtsrd.com/papers/ijtsrd64875.pdf

IJTSRD64875

INTRODUCTION

E-commerce, short for "electronic commerce," involves online transactions where currency is exchanged for goods and services. E-commerce businesses communicate their product offerings to customers, typically categorized into three segments: Business-to-Business (B2B), Business-to-Consumer (B2C), and Business-to-Government (B2G). Additionally, e-commerce is further classified based on the presentation of services, including branded e-commerce stores, e-commerce marketplaces, and conversational commerce.

In Bangladesh, popular e-commerce platforms include Othoba, Pickaboo, Ajkerdeal, ClickBD, Bagdoom, Bikroy, BoiMela, Rokomari, Food Panda, and the prominent Daraz, established in 2012, which has evolved into one of the country's leading e-commerce platforms. The e-commerce market in Bangladesh witnessed substantial growth, surging by an impressive 70% in 2017 compared to the preceding year, further accentuated during the COVID-19 pandemic.

This paper focuses on collecting and analyzing product reviews from Daraz, a pivotal online marketplace and logistics company operating in South Asian and Southeast Asian markets. A sentiment-based analysis was performed, categorizing each review as either positive or negative. A meticulously curated dataset of 1000 reviews from Daraz serves as the foundation for training a sentiment analyzer. Once trained, this analyzer proves effective in categorizing new reviews and comments, providing valuable insights into customer sentiments. As the e-commerce landscape continues to burgeon in Bangladesh, this study aims to forecast sales trends and evaluate customer satisfaction levels across five prominent e-commerce sites. The dataset, sourced from Daraz through web scraping, underwent manual labeling to ensure precision. This research contributes to an enhanced understanding of the evolving e-commerce market in Bangladesh, offering businesses tools to augment customer satisfaction and optimize their operational strategies.

PROBLEM STATEMENT

The e-commerce growth in Bangladesh, as shown by the success of Daraz, has presented consumers with a flood of evaluations, and confounding decision-making. This study aims to categorize reviews as good or negative to help users make informed decisions. The study uses classifiers such as Multinomial Naive Bayes, Logistic Regression, and others, as well as approaches such as Term Frequency-Inverse Document Frequency (TF-IDF) with Principal Component Analysis (PCA), to evaluate their effectiveness on a curated dataset of 1000 Daraz reviews. The results show that Multinomial Naive Bayes perform well, particularly with Bigram and Trigram features. By tackling this issue, companies on Daraz may get insights on improving product quality and operational strategies, cultivating a more informed consumer base in Bangladesh's expanding e-commerce ecosystem.

LITERATURE REVIEW

The study significantly enhances existing literature by showcasing the effectiveness of machine learning algorithms, particularly a hyperparameter-tuned SVM classifier, in accurately predicting sentiment polarity from Bangla texts within the context of E-commerce product reviews. It contributes valuable insights to sentiment analysis methodologies applied to the Bangla language, enriching the scholarly discourse on the subject. [1]

The exploration of Daraz reviews underscores a significant advancement in the domain of sentiment analysis, revealing a notable superiority of standard machine-learning techniques over human-produced baselines. In particular, the utilization of the Ridge classifier, and Logistic Regression has emerged as a robust methodology, showcasing remarkable performance in the realm of traditional topic-based categorization [2]

In the context of text sentiment analysis, the focus is on feature extraction, with algorithms playing a crucial role. A novel methodology is introduced for feature extraction, employing Generalized TF-IDF feature vectors by incorporating semantic similarity of synonyms. Local patterns of feature vectors are identified using the OPSM blustering algorithm [3].

To alleviate uncertainties regarding purchases, online reviews are utilized as a valuable resource for clearing doubts. A novel method is presented, integrating the Bass/Norton model and sentiment analysis with historical sales data and online review data to forecast product sales effectively [4].

Electronic commerce's popularity is on the rise due to a large number of product reviews. Opinion mining is employed to capture customer reviews and categorize them into subjective and objective expressions. A novel multi-dimensional model is proposed for opinion mining, integrating customer characteristics and opinions about any product [5].

An approach is described to identify and locate all occurrences of an object in a video using viewpoint invariant region descriptors. The efficient retrieval is achieved through methods from statistical text retrieval, including an inverted file system and weighting of text and document frequency [6].

Various neural network architectures for statistical language modeling are proposed and applied successfully in language modeling components of speech recognition systems. These architectures learn embeddings for words in a continuous space, contributing to smoother language and better generalization [7]. In the realm of IMDb movie reviews, Support Vector Machines emerge as frontrunners, showcasing superior performance and robust sentiment classification capabilities, underscoring their efficacy in discerning positive and negative sentiments. [8]

DATA COLLECTION AND PREPROCESSING

In this section, we outline the procedures followed for collecting the dataset from the Daraz website and preprocessing it for sentiment analysis.

A. Dataset

The dataset was obtained straight from the Daraz website and consisted of 1000 comments. Among these, 651 comments were classed as '0's, representing one sentiment category, and 349 as '1's, representing the other sentiment category.

- **Data Collection from Daraz Website:** Comments were systematically collected from relevant sections of the Daraz website, including product pages, review sections, and customer feedback segments. The collection process aimed to capture a diverse range of comments to ensure the model's effectiveness in understanding various sentiments expressed by customers.
- **Sentence Selection:** Specific sentences were meticulously chosen from the collected comments to form the basis of the dataset. Selection criteria likely included the comprehensiveness of the sentence, its representativeness of customer sentiments, and its freedom from irrelevant or misleading content.
- **Annotation:** The collected comments were annotated or labeled as either '0' or '1' to denote the sentiment category they represent. Annotation is crucial for supervised learning tasks like sentiment analysis as it provides the ground truth labels necessary for model training and evaluation.

B. Data Preprocessing:

Data preparation is a key step in ensuring the dataset's quality and efficacy for sentiment analysis. The following actions were done to preprocess the acquired dataset:

- **Stop Words Removal:** Common stop words, including articles, prepositions, and conjunctions, were detected and removed from each remark. Removing stop words improves the relevance and importance of the remaining words in the dataset, as stop words have minimal semantic value in sentiment analysis.
- **Term Frequency-Inverse Document Frequency (TF-IDF) Transformation:** The TF-IDF technology was used to transform textual comments into numerical data. TF-IDF adds weights to words depending on their value in a specific context, allowing comments to be represented numerically and used by machine learning algorithms.
- **Normalization:** The numerical data received from the TF-IDF transformation were normalized to guarantee that the input data was consistent. Normalization is the process of scaling variables to a standard range, such as [0, 1], to assist effective model training.

C. Model Hyperparameters and Environment Setup

Table I: Model Hyperparameters And Experimental Setup

Hyperparameter	Value
Machine Learning Library	scikit-learn, TensorFlow
Programming Language	Python 3.10.12
Development Environment	Google Colab
RAM Allocation	12 GB to 25 GB adaptable RAM
GPU Allocation	Nvidia T4 GPU
Storage Allocation	107.72 GB temporary storage (ROM)
Main Algorithm	Support Vector Machine(SVM), Multinomial Naive Bayes (MNB), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT)
SVM Hyperparameters	C=1.0, kernel='linear', gamma='auto'
MNB Hyperparameters	Alpha: 1.0, Fit Prior: True
Random Forest Hyperparameters	n_estimators=100, max_features='auto', max_depth=10
Logistic Regression Hyperparameters	C: 1.0 Penalty: L2 Solver: Liblinear
Decision Tree Hyperparameters	Max_depth: None Criterion: Gini
Model Evaluation Metric	Accuracy, F1-Score, Confusion Matrix
Data Preprocessing Techniques	TF-IDF, Stop Words Removal, Normalization

METHODOLOGY

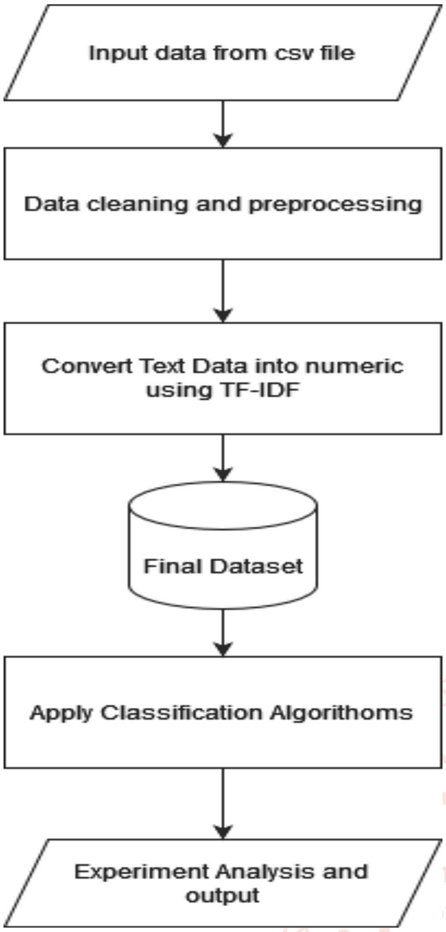

Fig. 1: Vertical Workflow of methodological steps

A. Term frequency-inverse document frequency

An analytical statistic expressed numerically or scientifically, seeks to encapsulate the essence of language within a document or dataset. This statistical methodology, commonly referred to as Term Frequency-Inverse Document Frequency (TF-IDF), goes beyond the mere consideration of isolated words. Instead, it encompasses unigrams, bigrams, and trigrams, aiming to comprehensively capture the significance of linguistic elements within a docket or corpus. The utilization of this approach is pivotal in information retrieval, text mining, and user modeling, providing valuable insights by thoroughly exploring this dataset.

➢ **Unigrams (1-grams):**

A unigram, also known as a 1-gram, is a single word in a line of text. The term frequency (TF) for a unigram i in a document j is defined as

$TF_{i,j}$ = count of i in jtotal number of words in j

The inverse document frequency (IDF) for a unigram i is determined as follows:

IDF(i) =log (total number of documentsnumber of documents containing i)

➢ **Bigrams(2-grams):**

Bigrams add dimension to TF-IDF calculations. A bigram is a literary sequence that includes two consecutive words. A bigram's term frequency (TF) and inverse document frequency (IDF) are calculated in the same way as unigrams are, but using word pairs instead of single words.

➢ **Trigrams(3-grams):**

Trigrams, which are sequences of three words, also have their own TF and IDF computations.

➢ **TF-IDF with Unigrams, Bigrams, and Trigrams:**

The TF-IDF score for unigrams, bigrams, and trigrams in a document j may be calculated by combining these elements:

$$\text{TF-IDF}_{i,j} = TF_{i,j} \times IDF(i)$$

$$\text{TF-IDF}_{bigram,\,j} = TF_{bigram,j} \times IDF^{bigram(i)}$$

$$\text{TF-IDF}_{trigram,j} = TF_{trigram,j} \times IDF_{trigram(i)}$$

where:

➢ $TF_{i,j}$, $TF_{bigram,j}$, $TF_{trigram,j}$ are the term frequencies for unigrams, bigrams, and trigrams in document j respectively.

➢ IDF(i), $IDF_{bigram(i)}$, $IDF_{trigram(i)}$ are the inverse document frequencies for unigrams, bigrams, and trigrams respectively.

➢ **Vectorization Technique:**

To quantitatively represent the textual material, we use the TF-IDF vectorization approach. We use unigrams, bigrams, and trigrams to represent various degrees of linguistic information. This method is critical in information retrieval, text mining, and user modelling, as it provides significant insights by extensively studying the dataset.

TF-IDF allows us to extract relevant keywords and phrases from a corpus for use in document categorization, information retrieval, and other natural language processing applications. This process enables us to comprehend the significance of various words and phrases within the corpus, resulting in more effective data analysis.

00	10	100	180	40	4k	4k হয়4k	...	৮৮৫ক ৬৫	৮৮৫ক ৬৫ ১০	৯০	৯০ কথ	৯০ কথ এগ	৯৩	৯৩ পড়ল	৯৩ পড়ল মচ	৯৫	৯৫ ৪৫
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 2: Vectorize data sample

RESULTS AND PERFORMANCE ANALYSIS

The study systematically assessed the performance of seven classifiers across various metrics, including accuracy, prediction speed, training time, and total misclassification. The outcomes, outlined in separate tables for the unigram, bigram, and trigram categories, focus on the top five classifiers determined by a composite measure.

Table II: Unigram: Top Five Classifiers Performance

Classifier Name	Acc.	Recall	Prec.	F1
MNB	79.00	90.00	78.26	83.72
RF	70.00	96.67	67.44	79.45
DT	69.00	85.00	69.86	76.69
LR	68.00	100.00	65.22	78.95
RBF SVM	63.00	100.00	61.86	76.43

Table III: Bigram: Top Five Classifiers Performance

Classifier Name	Acc.	Recall	Prec.	F1
MNB	80.00	88.33	80.30	84.13
RF	75.00	91.67	73.33	81.48
DT	64.00	98.33	62.77	76.62
LR	63.00	100.00	61.86	76.43
RBF SVM	61.00	100.00	60.61	75.47

Table IV: Trigram: Top Five Classifiers Performance

Classifier Name	Acc.	Recall	Prec.	F1
MNB	80.00	85.00	82.26	83.61
RF	69.00	86.67	69.33	77.04
DT	66.00	100.00	63.83	77.92
LR	62.00	100.00	61.22	75.95
RBF SVM	61.00	100.00	60.61	75.47

The Multinomial Naive Bayes (MNB) classifier consistently demonstrated the highest accuracy across all three categories (unigram, bigram, and trigram). Therefore, based on the composite measure used in this study, MNB appears to be the most effective classifier for the given task. The Random Forest (RF) classifier also showed competitive performance, particularly with bigram features, ranking second in accuracy in that category. Consideration should be given to the specific requirements and characteristics of the task at hand when choosing the most suitable classifier.

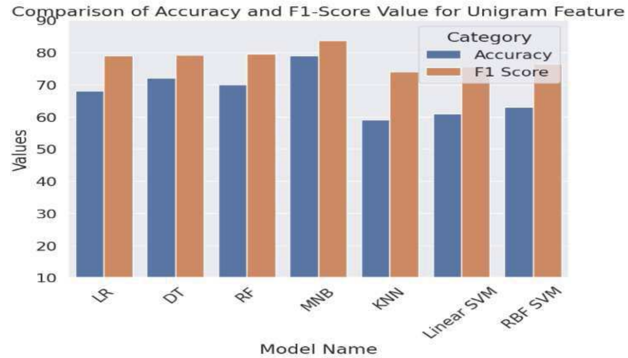

Fig. 3: Comparison of Accuracy and F1-Score for Unigram

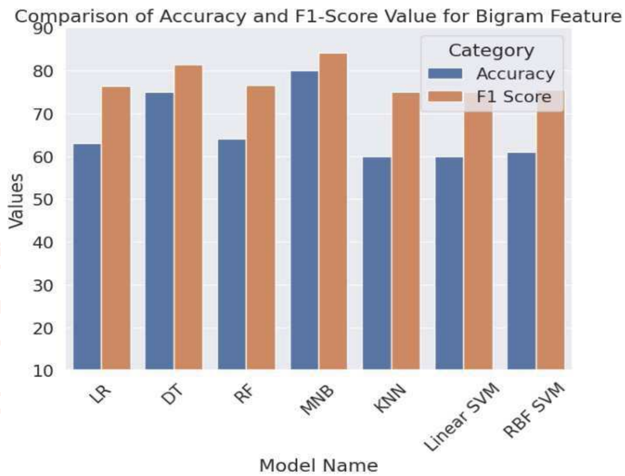

Fig. 4: Comparison of Accuracy and F1-Score for Bigram

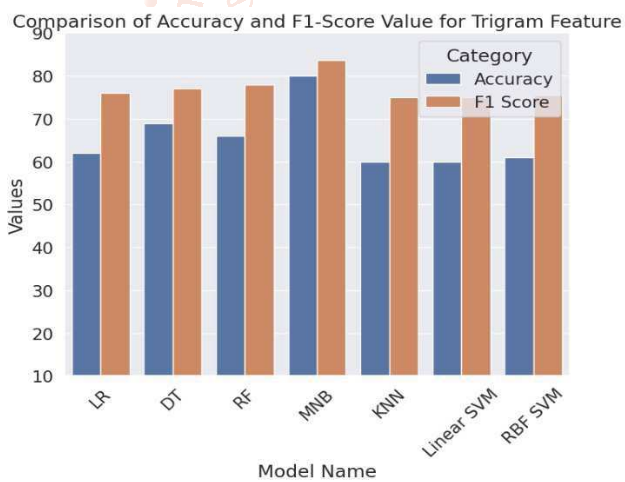

Fig. 5: Comparison of Accuracy and F1-Score for Trigram Multinomial Naive Bayes (MNB) Classifier gave the highest accuracy, followed by Random Forest.

A. Confusion Matrix

Visualizing algorithm performance through a discrete table format is a convenient and practical method, particularly beneficial for Machine Learning (ML) techniques. This representation employs a two-dimensional layout, showcasing sets for every class of data based on true class and predicted class combinations. The sets are defined as follows: TP (True Positive), FN (False Negative), FP (False Positive), and TN (True Negative).

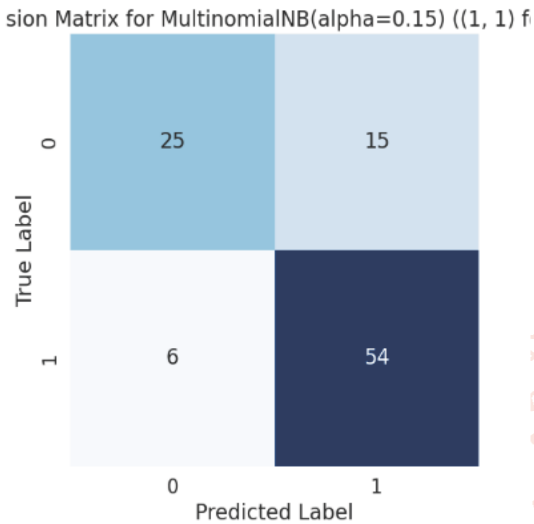

Fig. 6: Confusion Matrix for MultinomialNB for Unigram

Upon examining the confusion matrices in Figure 6-8, it becomes evident that both the Ridge classifier and the Extra Trees classifier exhibit high values. Specifically, the Ridge classifier stands out as having a notably low misclassification cost according to the provided confusion matrix. This observation emphasizes the effectiveness of the Ridge classifier in accurately predicting and classifying data, making it a promising choice for the ML task under consideration.

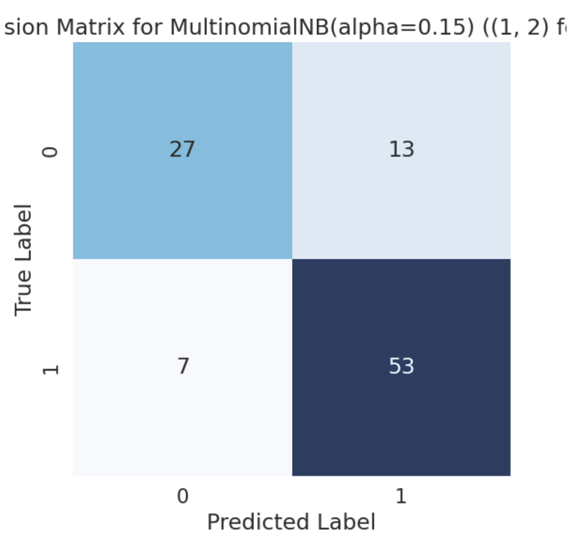

Fig. 7: Confusion Matrix for Multinomial NB for Bigram

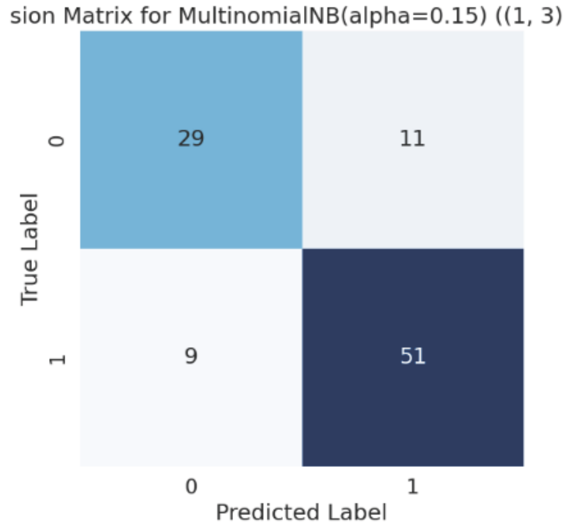

Fig. 8. Confusion Matrix for MultinomialNB for Trigram.

CONCLUSION

In this work evaluating the performance of seven classifiers across unigram, bigram, and trigram categories, we have gained valuable insights into the effectiveness for a given task. The results reveal that the Multinomial Naive Bayes (MNB) classifier consistently outperformed its counterparts, exhibiting the highest accuracy across all three feature categories. With accuracy scores of 79.00%, 80.00%, and 80.00% for unigram, bigram, and trigram respectively, MNB emerges as a robust choice for the classification task under consideration.

The Random Forest (RF) classifier also demonstrated noteworthy performance, securing the second position in accuracy, particularly excelling with bigram features. Its accuracy scores of 70.00%, 64.00%, and 66.00% in the unigram, bigram, and trigram categories respectively make RF a strong contender.

While decision trees (DT), logistic regression (LR), and radial basis function support vector machine (RBF SVM) exhibited competitive results, they consistently ranked below MNB and RF in terms of accuracy.

In conclusion, the choice of a classifier should align with the specific requirements and characteristics of the task at hand. The findings of this study provide a foundation for informed decision-making, with MNB emerging as the top-performing classifier, closely followed by RF. Further consideration of factors such as prediction speed, training time, and overall misclassification rates may be crucial in selecting the most suitable classifier for practical implementation.

References

[1] Zulfiker, S., Chowdhury, A., Roy, D., Datta, S., & Momen, S. (2022). Bangla E-Commerce Sentiment Analysis Using Machine Learning

Approach. In 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1–6). Dhaka. DOI: 10.1109/STI56238.2022.10103350

[2] Mst. Jannatul Ferdous, Pritom Sarker, and Nakib Aman Turzo, "Assortment of Bangladeshi E-commerce Site Reviews using Machine Learning Approaches", 2022.

[3] Y. X. ·. H. Z. ·. X. L. ·. X. H. ·. Z. M. Xin Chen1, "A novel feature extraction methodology for sentiment analysis of product reviews," Elsevier, 2017

[4] Y.-J. Zhi-Ping, "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis," Elsevier, pp. 90-100, 2017.

[5] R. Y. Y. L. Zhang, "Integration of Sentiment Analysis into Customer Relational Model: The Importance of Feature Ontology and Synonym," Elsevier, vol. 11, pp. 495-501, 2013.

[6] J. Sivic and A. Zisserman, "Efficient Visual Search of Videos Cast as Text Retrieval," IEEE, vol. 31, no. 4, pp. 591-606, 2008.

[7] F. Morin, "Hierarchical Probabilistic Neural Network Language Model," in Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics, 2005.

[8] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pages 79–86. Association for Computational Linguistics.