World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

E-commerce clothing review analysis by advanced ML Algorithms

Nayem Uddin Prince [1,*], Mohammed Nazmul Hoque Shawon [2], Zakaria Kamal Shahed [3], Riya Islam Nupur [4], Farzana Akther Mele [5] and Md Abdullah Al Mamun [3]

[1] *Information Technology (2022), Washington University of Science and Technology, USA.*
[2] *Department of Computer Science and Engineering, Daffodil International University, Bangladesh.*
[3] *Department of Software Engineering, Daffodil International University, Bangladesh.*
[4] *Department of Computer Science and Engineering, Southeast University, Bangladesh.*
[5] *Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh.*

World Journal of Advanced Research and Reviews, 2024, 24(01), 1680–1690

Publication history: Received on 27 August 2024; revised on 04 October 2024; accepted on 07 October 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.1.3044

Abstract

A significant aspect in the rapid ascent of Bangladesh's e-commerce sector in recent years has been the significance of consumer evaluations. By examining these reviews, readers can gain enhanced insight into consumer happiness and product quality. This research analyses the sentiment of online clothes reviews with advanced machine-learning techniques. The fundamental purpose of evaluating several machine learning models, such as KNN, RF, XGB, Multi, LSTM, and CNN, is to identify the most effective method for sentiment classification. Following the compilation of an extensive dataset of clothing assessments and the execution of data preparation tasks, including cleaning, tokenization, and stop word elimination, we utilised a combination of deep learning and machine learning models for classification purposes. KNN had the highest accuracy in forecasting the future attitudes of the models evaluated. The findings indicate in the 5000 datasets that KNN is the most effective algorithm for assessing the sentiment of online purchasing reviews. KNN reached at 0.91% Accuracy. This project's automated and scalable methodology may enable online businesses to evaluate client feedback and make more informed decisions. Future research initiatives encompass augmenting the dataset, experimenting with various methods, and incorporating these models into AI-driven mobile and online applications for real-time sentiment analysis.

Keywords: Sentiment; Advance ML; Tokenization; Classification; Deep learning; Stop word; CNN; LSTM; KNN; Split; Review.

1. Introduction

In recent years, Bangladesh's e-commerce sector has shown remarkable growth due to high Internet penetration, a substantial middle class, and the ubiquitous availability of smartphones. Forecasts show that the e-commerce market in Bangladesh will yield over $3 billion in revenue by 2023, with an anticipated CAGR of 17.6% from 2023 to 2028 [1]. Due to its flourishing digital economy, Bangladesh possesses one of the fastest-growing e-commerce markets in South Asia, with over 35 million individuals engaging in online purchases each year [2]. Due to improvements in logistical infrastructure and the convenience of mobile shopping, online marketplaces have experienced a significant surge in popularity, especially in the retail and apparel industries. As e-commerce proliferates, sustaining client satisfaction and fostering brand loyalty in a fiercely competitive market becomes increasingly difficult. Reviews serve as an essential resource for consumers in the expansive and dynamic realm of online shopping. [3] reports that an impressive 87% of shoppers consult online reviews before purchasing. Key topics addressed in these evaluations encompass customer happiness, delivery challenges, and product quality. Review data offers critical insights into customer preferences and product performance, which organisations may leverage to enhance their services, refine marketing efforts, and resolve

* Corresponding author: Nayem Uddin Prince

Copyright © 2024 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

difficulties. The subjective nature of criteria like fit, quality, and style renders customer reviews significantly impactful on sales in the clothes industry. Manually sifting through the multitude of daily assessments on e-commerce platforms is an arduous and inefficient endeavour. This is the domain where machine learning excels. Employing machine learning techniques, which offer automated and scalable solutions for evaluating extensive datasets, enables firms to derive significant insights from client input. By using sentiment analysis and text categorisation facilitated by sophisticated machine learning models, firms may discern customer sentiment trends, classify reviews based on positive or negative feedback, and pinpoint areas of concern. This study employed sophisticated machine learning techniques to analyse online clothes reviews from Bangladesh. We seek to ascertain the satisfaction and happiness of our clients; therefore, we require an efficient and precise method to categorise their feedback. Machine learning models like KNN, RF, XGB, Multi, LSTM, and CNN were analysed. After comprehensive testing across several data splits and studies with differing test sizes, the KNN model was identified as the most accurate method. KNN proved effective for assessing client input and providing insights due to its superior categorisation accuracy. Our project underscores the significance of user reviews in the online retail sector by applying machine learning methodologies. Furthermore, it illustrates how sophisticated machine learning models may optimise sentiment analysis, offering organisations valuable insights to improve product quality and customer satisfaction. This research demonstrates how e-commerce platforms in Bangladesh's rapidly expanding digital sector could benefit from machine learning-driven review analysis.

This work discusses several of the following ideas. The second section comprises summaries of significant publications. The techniques were delineated in Section III. Section IV emphasises the experimental results, whereas Section V evaluates our model. Section VI examines these mechanisms.

2. Literature review

Employing machine learning and deep learning methodologies is incredibly productive when handling sensitive data. Enhancing individuals' quality of life is a fundamental objective. While we provide a novel viewpoint, previous research has employed similar methodologies. To elucidate the distinctions, two studies contrasted the two.:

Hamsagayathri et al. [4] utilised regression tree classifiers and SVM to assess the recommendation level of online reviews for women's clothing. Precision, accuracy, and recall metrics evaluate the classifiers' efficacy. REPTree achieved the maximum classification accuracy of 91.43% using SVM, with a recall of 96.04% and a precision of 93.75%.

Mahmud et al. [5] utilised many methodologies, including word clouds, topic modelling, data summaries, and sentiment distribution visualisations. The dataset employed in this study comprises the following components: product names, product reviews, prices, seller ratings, shipment timeliness, chat response rate, star ratings (out of 5), total star count, 4-star rating, 3-star rating, 2-star rating, 1-star rating, discount percentage, price, seller information, payment method, brand name, and sentiment analysis. Numerous decision-support algorithms exist, including SVM, RF, LR, NN, and GBoosting. The Random Forest Classifier attained the highest performance, with an accuracy of 96.51%, precision of 96.51%, recall of 96.50%, and F1-score of 96.50%.

M. A. Shafin et al. [6] employed natural language processing to develop a system that quantifies prior Bangla-speaking customers' positive and negative feedback regarding their online buying experiences. They gathered approximately one thousand product comments and reviews for this study. We employed various categorization techniques in sentiment analysis, including SVM, DT, RF, LR, and KNN. The SVM achieved the most excellent accuracy rate of 88.81% among all the methods.

M. H. Rahman et al. [7] aimed to aid bibliophiles in locating and purchasing desired books by analyzing evaluations in Bangla while offering significant insights to online bookstores. A total of 6,281 distinct data points were amassed to train the computer. This study uses Machine Learning and Deep Learning techniques to categorize emotions as positive or negative. Natural language processing is employed in the text preprocessing phase. Algorithms, including SVM, DT, RF, LR, KNN, and LSTM, are commonly utilized for sentiment classification. Utilizing LSTM, they attained a peak accuracy of 97.49%.

P. J. et al. [8] comprehensively examined DT, RM, KNN, and Size prediction. Their approach exhibits significant promise for improving sizing precision, with the DT and RF models attaining remarkable accuracy scores of 0.99 in shirt size prediction. These models remain effective for predicting pant sizes, with accuracy ratings of 0.89 and 0.82, respectively.

Lin et al. [9] employed the Women's E-Commerce Clothing Reviews dataset for their research on sentiment analysis of consumer recommendations. They possess logistic regression, support vector machines, random forests, XGBoost, and LightGBM in their arsenal. These systems employed natural language processing to derive insights from reviews and utilize them for product promotion. The LightGBM method attained superior accuracy and area under the curve (AUC) relative to the other algorithms. The accuracy, recall, and F1 score yielded a total of 0.97. Three distinct support vector machines—Linear Kernel, XGBoost, and Ridge Regression—achieved results nearing perfection (0.94).

Manikiran et al. [10] classified each review as good, harmful, or neutral while examining the correlation among several attributes in customer surveys regarding women's clothes e-commerce data. To achieve these objectives, they utilized NLP for emotional analysis and recommendation generation, alongside univariate and multivariate analysis of survey responses. The findings indicate that both endorsements and perspectives can enhance scores. Nonetheless, the assessment of item survey sentiment scores is inaccurate. The F1-score for sentiment classification was 0.9596, while for opinion arrangement, it was 0.928355, both achieved using the multi-NB method.

Agarap et al. [11] requested reviewers on an online women's clothes store to determine the variables most frequently referenced in positive, negative, or neutral assessments. They employed a bidirectional (RNN with an LSTM and conducted univariate and multivariate analyses on dataset features, excluding review titles and contents, for sentiment classification and recommendation objectives. The results indicate that suggestions are a robust predictor of elevated sentiment ratings, and conversely, the opposite holds. Review scores, conversely, do not consistently reflect genuine sentiments. The bidirectional LSTM attained an F1-score of 0.88 for recommendation classification and an F1-score of 0.93 for sentiment classification.

Y. Hong et al. [12] conducted sentiment analysis on evaluations of an online fashion retailer utilizing machine learning techniques. In this instance, they collected commentary data with the data acquisition program. Data preprocessing, word segmentation, and sentiment annotation were later steps. They develop a sentiment classifier to autonomously classify unlabeled data by modifying various parameters for training purposes. Their experimental findings illustrate the effectiveness of employing machine learning methods for sentiment classification in product reviews.

Oueslati et al. [13] focused on the degrees of positivity and negativity and the basic metrics derived directly from reviews concerning emotional variables. Furthermore, it derives the emotional aspect via an emotion lexicon. To address issues such as orthographic errors, domain-specific dependencies, and neologisms, they suggested the development of an internal emotion lexicon. They implemented the proposed method on six distinct medical product Facebook pages. The SVM algorithm attains a prediction accuracy of 97.95%.

Pujari et al. [14] developed a system utilizing Python standard libraries to aggregate and classify customer reviews by sentiment, aiming to aid prospective purchasers in making informed product assessments. The reviews are categorized utilizing three classification algorithms: Support Vector Machine, Naïve Bayes Algorithm, and Maximum Entropy Classifier. Subsequently, the efficacy of these algorithms is evaluated. The methodology outlined in this study is applicable across various domains.

2.1. Comparison with other work

Table 1 compares other works with our work. After analyzing this table, it can be said that existing works contain fewer algorithms and low accuracy. No advanced techniques are applied. Our research is rich in advanced techniques with brilliant accuracy [15]. Massive advanced techniques were not used before. So, it can be said that our research is innovative and outstanding.

Table 1 Comparison Table with other work

<table>
<tr><th colspan="2">Other work</th><th>Our work</th></tr>
<tr><td>Author Name</td><td>Algorithm & Accuracy</td><td>Algorithm & Accuracy</td></tr>
<tr><td>Hamsagayathri el. At.</td><td>SVM
91.43%</td><td rowspan="2">RF,
CNN,
XGB,
KNN,
LSTM,
Multi
&
91.00%</td></tr>
<tr><td>Mahmud el. At.</td><td>SVM,
RF,
LR,
NN,
GBoosting
96.51%</td></tr>
</table>

Oueslati el. At.	SVM 97.95%	
Agarap el. At.	LSTM 88.00%	

3. Material and methods

This analysis of E-commerce clothing reviews employed cutting-edge machine learning and deep learning methodologies, as illustrated in the approach diagram. We have divided our methodology into five steps: data collection, data Preprocessing, applied algorithms, and evaluation. All of the parts have covered our project suitably.

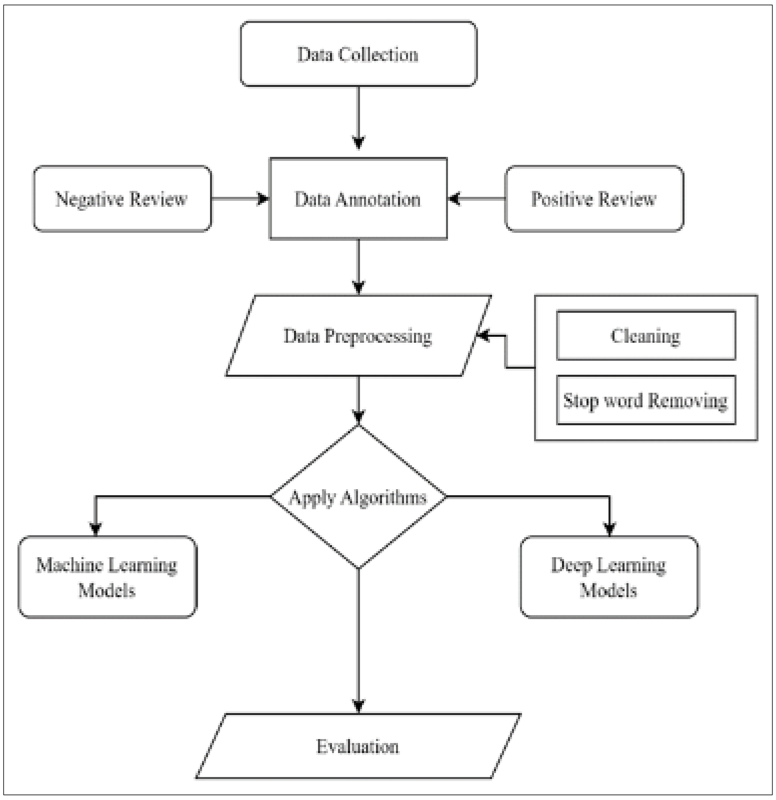

Figure 1 Methodology Diagram**.**

3.1. Data collection

Data collection is paramount in constructing a sentiment analysis model for online clothes reviews. We obtain review data from many online apparel marketplaces during the process. This encompasses e-commerce platforms (such as Daraz and Bikroy.com) and social media networks where consumers regularly provide reviews [16]. These reviews frequently encompass favourable and unfavourable user-generated content regarding garment acquisition experiences. We now possess 4,500 reviews.

3.2. Dataset pre-pressing and representation

Data preparation is an essential phase in the sentiment analysis pipeline, as it facilitates machine learning algorithms' comprehension and processing of raw, unstructured data. Natural language online store reviews are notoriously chaotic and replete with inconsistencies, including colloquialisms, typographical errors, and misspellings. Inadequate preprocessing can hinder machine learning models from identifying the patterns necessary for precise predictions [17]. The subsequent section comprehensively explains the many procedures involved in the data preparation phase.

Table 2 shows the sample of our applied dataset. This dataset had two classes, positive and negative, as we labelled our data classes 0 and 1. Our data set is text data, so it is essential to convert our class as a numerical value for machine understanding [18].

Table 2 Sample Data representation and labelling

Raw Data	Type	Label
যা অর্ডার করেছিলাম তাই দিয়েছে	Positive	1
ছবি দেখায় এক দেয় আর এক কাপড়	Negative	0

3.2.1. Tokenization & Stop word removing

Tokenization is the procedure of deconstructing the review content into its constituent words. Tokenization enables the model to more effectively identify text patterns by considering each word as an independent unit [19]. Stop words, which are popular but superfluous terminology, can diminish the impact of a phrase. We omitted these terms as they do not contribute to sentiment recognition [20]. To assist the model in focusing on significant phrases associated with consumer sentiment, stop words are eliminated from the dataset, reducing its dimensionality. Tools like NLTK, spaCy, or Scikit-learn facilitate the removal of stop words from datasets.

Table 3 The tokenization table for our dataset.

Raw Data	Type	Tokenization Data
যা অর্ডার করেছিলাম তাই দিয়েছে	Positive	“যা”,” অর্ডার”,” করেছিলাম”,” তাই”,” দিয়েছে”
ছবি দেখায় এক দেয় আর এক কাপড়	Negative	“ছবি”,” দেখায়”,” এক”,” আর”, “দেয়” এক”,” কাপড়”
কাপড়টি অনেক ভালো ছিলো	Positive	“কাপড়টি”,” অনেক”,” মজা”,” ভালো”
ছবিতে কপড়ের রং ছিলো উজ্জ্বল সবুজ এখন জামা হাতে পেয়ে দেখি অন্ধকার সবুজ	Negative	“ছবিতে”,” কপড়ের”,” রং”,” ছিলো”,“উজ্জ্বল”, “সবুজ”, “এখন”, “জামা”,” হাতে”, “পেয়ে”, “দেখি”,” অন্ধকার”, “সবুজ”

3.3. Data Classification

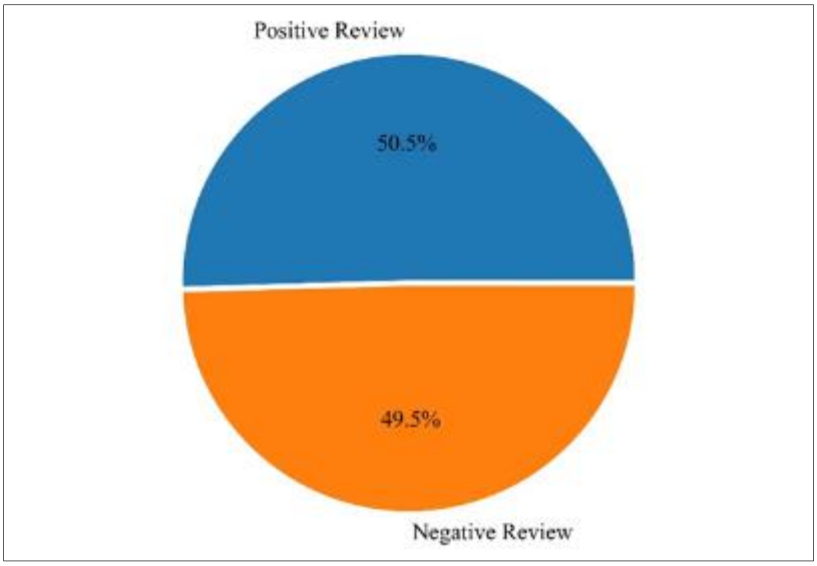

Figure 2 Methodology Diagram

The pie chart of figure 2 illustrates the distribution of positive and negative assessments within the e-commerce clothes review dataset. The data indicates a predominance of good ratings over negative ones. This equal distribution suggests that a balanced dataset is essential for training and assessing sentiment analysis machine learning models [21]. When the dataset is uniformly distributed, algorithms may more efficiently discern positive and negative sentiment patterns, mitigating bias towards either class. This equilibrium is essential for attaining robust classification efficacy and ensuring precise predictions in practical situations. The need to comprehend positive and negative client feedback to enhance products and services is especially emphasized in e-commerce platforms.

3.4. Model selection and algorithms

For this research, we have used traditional and advanced machine learning algorithms. Traditional algorithms are KNN, RF, XGB, and Multi. Advanced algorithms refer to deep learning algorithms, such as LSTM and CNN. This concept has never been used before, so it can be said that our contribution is precious.

3.5. Evaluation

Various performance metrics are utilized to assess the model's post-application, including accuracy, precision, recall, and F1-score. The evaluation findings demonstrate the efficacy of the models in classifying reviews. This project comprehensively explains the techniques for constructing a deep learning or machine learning sentiment analysis model. It emphasizes the importance of comprehensive data processing and model evaluation using assessment metrics.

We systematically acquire diverse review data from multiple sources to ensure a comprehensive dataset for the sentiment analysis project. This will enable more accurate training and evaluation of models. Collecting ample data is essential to effectively represent both affirmative and negative concepts, enhancing the model's generalizability.

4. Results and discussion

Table 4 shows the accuracy of all algorithms applied in our project. In this project, we selected five traditional machine learning algorithms: RF, LSTM, Multi, KNN, and XGB. Here, also applied custom, an advanced machine learning technique that refers to deep learning. This is a combined technique applied to the project. It is called an out-stunting and innovation method. That has never been seen before. For this project, we split our dataset 40%, 50%, 60%, and 70%. Most of the test data, including the traditional and deep learning algorithms, scored well [22]. All algorithms gained the best result when the test data was 40%. RF and CNN achieved 91.55 and 91.0. KNN gained the best value at 91.68%. Our split data set was text. The text dataset is suitable for traditional machine learning algorithms.

Table 4 Accuracy table all algorithms

Test size	RF	LSTM	CNN	Multi	KNN	XGB
40	91.55	0.88	0.91	87.52	91.68	
50	91.22	0.88	0.91	87.29	90.92	
60	89.16	0.91	0.92	86.72	90.17	
70	88.26	0.90	0.91	86.10	88.90	

4.1. Classification Report

In this stage, we take three matrixes to purify our result further. We wanted to determine which algorithm was more accurate. Our data set had two classes: positive and negative. Since the machine does not understand text data, we labeled class two into numeric values. In that case, '0' means negative, and '1' indicates positive reviews. A good project is the first step in making the data machine-readable.

According to the performance criteria, the LSTM model attained a remarkable(table 5) overall classification accuracy of 89%. The model's performance is enhanced when predicting Label 1 compared to Label 0. For Label 1, the accuracy is 0.89, the recall is 0.92, and the F1-score is 0.90; for Label 0, the corresponding metrics are 0.87. Label reliability is consistent across the two data sets, as all three metrics exhibit comparable averages: recall (0.88), precision (0.89), and F1-score (0.88). Although there are 304 occurrences of Label 0 and 291 occurrences of Label 5, the weighted average, which considers the class imbalance, is almost equivalent to the macro average. The model performs effectively when recall and precision are regarded as equally significant.

Table 5 Classification table for LSTM

Label	precision	Recall	F1-score	Support
0	0.88	0.85	0.87	304
1	0.89	0.92	0.90	291
Accuracy	0.89			595
Macro avg	0.89	0.88	0.88	595
Weight avg	0.89	0.89	0.89	595

The CNN model has an overall accuracy of 92%. It identified 91% of actual cases and 92% of correct predictions for Label 0, with a recall of 0.91 and a precision of 0.92. Label 0 was assigned an F1-score of 0.92. Label 1 exhibits a commendable F1 score, recall, and precision, precisely 0.91. Despite a minor data imbalance, the model demonstrates consistent performance (table 6) across both labels, achieving a weighted average of 0.92 for precision, recall, and F1-score. The CNN model indicates that forecasts for both groups are typically accurate and equitable.

Table 6 Classification table for CNN

Label	precision	Recall	F1-score	Support
0	0.88	0.85	0.87	304
1	0.89	0.92	0.90	291
Accuracy	0.89			595
Macro avg	0.89	0.88	0.88	595
Weight avg	0.89	0.89	0.89	595

According to the performance metrics table 7, the XGB model attained an exceptional overall classification accuracy of 92%. The model excels in predicting Label 6, in contrast to Label 0. Label 0 exhibits metrics of 0.94, but Label 1 demonstrates an accuracy of 0.91, a recall of 0.94, and an F1-score of 0.92. Recall (0.92), accuracy (0.92), and F1-score exhibit comparable averages, signifying that label reliability is consistent across the two datasets (0.92). Although Label 0 occurs 304 times and Label 1 291 times, the weighted average, which considers class imbalance, is roughly equivalent to the macro average. If recall and precision are deemed equally significant, the method is effective.

Table 7 Classification table for XGB

Label	precision	Recall	F1-score	Support
0	0.94	0.91	0.92	304
1	0.91	0.94	0.92	291
Accuracy	0.92			595
Macro avg	0.92	0.92	0.92	595
Weight avg	0.92	0.92	0.92	595

The Multi-model achieved an impressive (table 8) total classification accuracy of 89%, meeting the performance criteria. Unlike Label 0, the model demonstrates superior performance in predicting Label 1. The accuracy, recall, and F1-score for Label 0 and Label 1 are each 0.90. The reliability of labels is uniform across the two datasets, demonstrated by similar averages for recall, precision, and F1-score (0.89%). The weighted average, which considers class imbalance, closely resembles the macro average despite Label 0 occurring 304 times and Label 1 291 times. If accuracy and recall are of equal significance, the approach is productive.

Table 8 Classification table for Multi

Label	precision	Recall	F1-score	Support
0	0.90	0.88	0.89	304
1	0.87	0.90	0.88	291
Accuracy	0.89			595
Macro avg	0.89	0.89	0.89	595
Weight avg	0.89	0.89	0.89	595

According to the performance measures table 9, the KNN model achieved an impressive 93% classification accuracy. Unlike Label 0, the model demonstrates superior performance in predicting Label 1. The metrics for Label 0 are 0.95, 92, and 0.945, whereas those for Label 1 are 0.92, 0.95 for recall, and 0.93 for F1-score. The two datasets demonstrate uniform label reliability, with mean values of 0.93 across all three criteria. The weighted average, which considers class imbalance, closely resembles the macro average despite Label 0 occurring 304 times and Label 1 291 times. If accuracy and recall are of equal significance, the strategy proves to be effective.

Table 9 Classification table for KNN

Label	precision	Recall	F1-score	Support
0	0.95	0.91	0.93	304
1	0.91	0.95	0.93	291
Accuracy	0.93			595
Macro avg	0.93	0.93	0.93	595
Weight avg	0.93	0.93	0.93	595

The RF model attains an impressive accuracy of 93%, indicating strong performance (table 10). The model achieved an accuracy of 0.95 for Label 0, correctly predicting 95% of instances, and a recall of 0.91, accurately identifying 91% of actual cases. The F1-score for Label 0 is 0.93. The precision, recall, and F1-score for Label 1 are 0.91, 0.95, and 0.93, respectively. The precision, recall, and F1-score average is 0.93 for both macro and weighted variants, indicating that the two labels operate comparably. Ultimately, the Random Forest model consistently and precisely forecasts results for both groups.

Table 10 Classification table for RF

Label	precision	Recall	F1-score	Support
0	0.95	0.92	0.94	304
1	0.92	0.95	0.93	291
Accuracy	0.93			595
Macro avg	0.93	0.93	0.93	595
Weight avg	0.93	0.93	0.93	595

5. Evaluation

Figure 3 presents a bar chart juxtaposing the anticipated and actual values of Hate Speech and Normal Speech. The blue bars represent the current values, and the orange bars denote the expected values. The tight alignment between both groups anticipated and actual values demonstrates the model's efficacy in accurately identifying speech [23]. The technology effectively detects harmful hate speech, as its predictions closely align with the actual data. Correspondingly, with closely aligned anticipated and actual values, the model effectively classifies non-hate speech as Normal Speech.

The disparity between both categories' anticipated and actual values is negligible, indicating that the model effectively differentiates between hate speech and everyday speech.

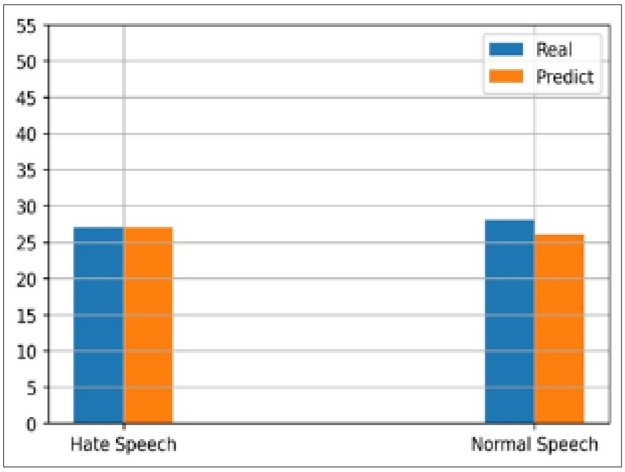

Figure 3 Evaluation Graph for KNN Algorithm**.**

Figure 4 presents a confusion matrix elucidating the model's classification performance. Based on the data, the model detected 26 occurrences of typical speech (26 TN) and 27 occurrences of hate speech (27 TP). The model identified the absence of hate speech, evidenced by the lack of false negatives (FN) in the matrix [24], underscoring its superior recall. Nonetheless, two false positives occurred when non-hate speech was erroneously categorized as hate speech. Despite occasionally overestimating hate speech, the system's low false positive rate demonstrates its accuracy. The confusion matrix indicates that the model effectively and equitably identifies hate speech while maintaining precision in misclassifications, which is essential for sensitive applications requiring detecting harmful content. It is, therefore, highly reliable.

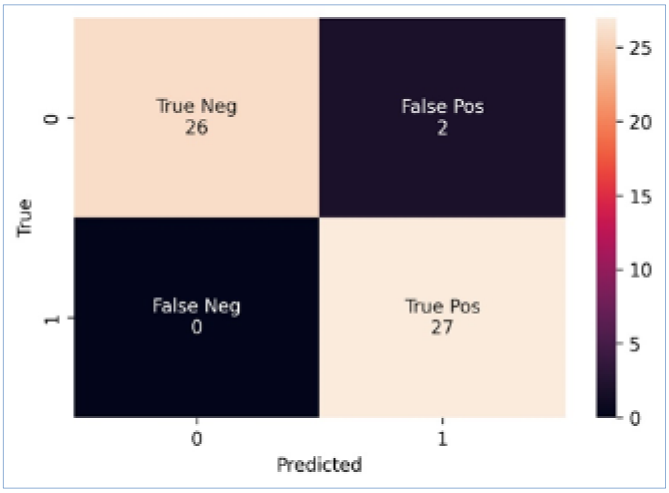

Figure 4 Confusion Matrix.

5.1. Decision

Comparing the two numbers reveals that the model effectively distinguishes between hate speech and everyday speech with few errors. The bar chart illustrates that planned and actual values align in both categories, demonstrating the reliable performance of KNN. The confusion matrix unequivocally indicates that the model correctly detects instances of hate speech, exhibiting no false negatives. Occasionally, the algorithm erroneously identifies two everyday speech cases as hate speech. These findings offer compelling evidence that this system can proficiently identify hate speech in practical situations. It is appropriate for scenarios where the repercussions of overlooking hate speech may be graver than those of erroneously flagging benign content due to its elevated recall and minimal false positive rate [25]. Altering this model to diminish false positives could enhance its applicability for content filtering or identifying sensitive hate speech.

6. Conclusion

This experiment highlights the essential role of customer review research in improving e-commerce strategies, particularly within Bangladesh's burgeoning online clothes industry. Employing advanced machine learning methodologies, we successfully categorized customer assessments, revealing the degree of satisfaction and emotion conveyed by the reviewers. This effort automates the examination of consumer feedback to assist businesses in managing substantial amounts of reviews. They must evaluate the advantages and disadvantages of their products to remain competitive in the evolving digital market. KNN had the highest classification accuracy of the methods evaluated. Sentiment analysis utilizing KNN demonstrates potential, especially for e-commerce enterprises managing substantial volumes of textual data [26]. Online enterprises can gain from the project's streamlining of review assessment, data-informed decision-making, and enhanced customer satisfaction. Customer loyalty increases, and business success is enhanced when difficulties are resolved and products are refined [27]. We intend to expand this initiative by creating online and mobile applications utilizing AI for real-time review analysis. These applications enable organizations to efficiently monitor client input, assess shifts in sentiment, and modify their strategies accordingly. Incorporating a more diverse array of evaluations from other e-commerce platforms into the dataset can boost model resilience and generalizability [28][29]. We will explore advanced methodologies to improve classification accuracy and processing speed, including deep learning architectures and transformer models. We aim to provide a comprehensive, scalable solution for web-based enterprises seeking to leverage consumer evaluations to optimize their operations through continuous data and algorithm refinement.

References

[1] Statista. (2023). Bangladesh E-commerce Market Size 2017-2027. Retrieved from https://www.statista.com/statistics/1194707/bangladesh-ecommerce-market-size/

[2] GlobalData. (2023). Bangladesh E-commerce Market Analysis. Retrieved from https://store.globaldata.com/report/bangladesh-ecommerce-market-analysis/

[3] Statista. (2022). Influence of Online Reviews on Purchase Decisions Worldwide. Retrieved from https://www.statista.com/statistics/944085/influence-of-online-reviews-on-purchase-decisions-worldwide/

[4] Hamsagayathri, P., & Rajakumari, K. (2020, January). Machine learning algorithms to empower Indian women entrepreneur in E-commerce clothing. In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.

[5] Mahmud, F. A. M., Mullick, S. B. R. A., & Anas, T. C. M. Sentiment Analysis of Women's Clothing Reviews on E-commerce Platforms: A Machine Learning Approach.

[6] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur and M. O. Faruk, "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-5,

[7] M. H. Rahman, M. S. Islam, M. M. U. Jowel, M. M. Hasan and M. S. Latif, "Classification of Book Review Sentiment in Bangla Language Using NLP, Machine Learning and LSTM," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India,

[8] P. J, N. V, S. S. S, S. S. V, S. K and A. V, "Enhancing Clothing Size Prediction in E-Commerce: A Machine Learning Approach for Improved Customer Experience," 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 2024, pp. 375-380,

[9] Lin, X. (2020, April). Sentiment analysis of e-commerce customer reviews based on natural language processing. In Proceedings of the 2020 2nd international conference on big data and artificial intelligence (pp. 32-36).

[10] Manikiran, G., Greeshma, S., Vishnu Teja, P., Sreehari Rao, Y., Sardar, T. H., & Mallik, M. A. (2022, February). E-commerce clothing review analysis and model building. In Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2021 (pp. 491-497). Singapore: Springer Nature Singapore.

[11] Agarap, A. F. (2018). Statistical analysis on E-commerce reviews, with sentiment classification using bidirectional recurrent neural network (RNN). arXiv preprint arXiv:1805.03687.

[12] Y. Hong and X. Shao, "Emotional Analysis of Clothing Product Reviews Based on Machine Learning," 2021 3rd International Conference on Applied Machine Learning (ICAML), Changsha, China, 2021, pp. 398-401, doi: 10.1109/ICAML54311.2021.00090.

[13] Oueslati, O., Khalil, A. I. S., & Ounelli, H. (2018). Sentiment analysis for helpful reviews prediction. International Journal of Advanced Trends in Computer Science and Engineering.

[14] Pujari, C., Aiswarya, & Shetty, N. P. (2018). Comparison of classification techniques for feature oriented sentiment analysis of product review data. In Data Engineering and Intelligent Computing: Proceedings of IC3T 2016 (pp. 149-158). Springer Singapore.

[15] Bhavsar, H., & Ganatra, A. (2012). A comparative study of training algorithms for supervised machine learning. International Journal of Soft Computing and Engineering (IJSCE), 2(4), 2231-2307.

[16] M. H. Rahman, M. S. Islam, M. M. U. Jowel, M. M. Hasan and M. S. Latif, "Classification of Book Review Sentiment in Bangla Language Using NLP, Machine Learning and LSTM," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021,

[17] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur and M. O. Faruk, "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-5, doi: 10.1109/ICCIT51783.2020.9392733.

[18] Thapa, R. B., Shrestha, S., Prince, N. U., & Karki, S. (2024). Knowledge of practicing drug dispensers about medication safety. European Journal of Biomedical and Pharmaceutical sciences, 11.

[19] Setya Rintyarna, B., Sarno, R., & Fatichah, C. (2019). Semantic features for optimizing supervised approach of sentiment analysis on product reviews. Computers, 8(3), 55.

[20] Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Information Processing & Management, 58(5), 102656.

[21] Arobi, R., Rijon, R. H., Haque, N., Safkat, T. I. M., & Safoan, S. (2022). Sentiment analysis on E-commerce based product reviews using machine learning algorithms (Doctoral dissertation, Brac University).

[22] Fayaz, M., Khan, A., Rahman, J. U., Alharbi, A., Uddin, M. I., & Alouffi, B. (2020). Ensemble machine learning model for classification of spam product reviews. Complexity, 2020(1), 8857570.

[23] Akter, M. T., Begum, M., & Mustafa, R. (2021, February). Bengali sentiment analysis of e-commerce product reviews using k-nearest neighbors. In 2021 International conference on information and communication technology for sustainable development (ICICT4SD) (pp. 40-44). IEEE.

[24] Gaye, B., & Wulamu, A. (2019). Sentiment analysis of text classification algorithms using confusion matrix. In Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health: International 2019 Cyberspace Congress, CyberDI and CyberLife, Beijing, China, December 16–18, 2019, Proceedings, Part I 3 (pp. 231-241). Springer Singapore.

[25] Shah, A. (2021). Sentiment analysis of product reviews using supervised learning. Reliability: Theory & Applications, 16(SI 1 (60)), 243-253.

[26] Kirana, Y. D., & Al Faraby, S. (2021). Sentiment analysis of beauty product reviews using the K-nearest neighbor (KNN) and TF-IDF methods with chi-square feature selection. Journal of Data Science and Its Applications, 4(1), 31-42.

[27] Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. CIRP Journal of Manufacturing Science and Technology, 23, 128-144.

[28] Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. CIRP Journal of Manufacturing Science and Technology, 23, 128-144.

[29] Rahman, Md Hamidur, Md Saiful Islam, Md Mine Uddin Jowel, Md Mehedi Hasan, and Ms Subhenur Latif. "Classification of Book Review Sentiment in Bangla Language Using NLP, Machine Learning and LSTM." In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2021.