International Journal of Development Mathematics Vol 1 Issue 3, Page No. 209 - 217

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

Predicting Consumer Behaviour from E-Commerce Platform Product Reviews: A Machine Learning Approach to Sentiment Analysis

Akisani A. Enos[a*] and Etemi J. Garba[a]

[a]*Department of Computer Science Modibbo Adama University, Yola, Nigeria.*

ARTICLE INFO

Article history:
Received 17 June, 2024
Received in revised form 05 August, 2024
Accepted 19 August, 2024

Keywords:
Consumer behaviour, E-Commerce platform, Machine learning, Product review, Sentiment analysis, Support vector machine.

MSC 2020 Subject classification:
90B50, 68T30

ABSTRACT

Given the recent advancements in web technology, users make use of the online services and offer their feedback thereby creating a vast amount of data available on the internet. To enable improved decision-making, it is essential to study, analyze, and organize this vast amount of user opinion, perspectives, feedback, and suggestions available through online resources. Nevertheless, the majority of e-commerce websites encounter challenges in comprehensively analyzing each individual review owing to the sheer volume of user feedback they receive. The aim of this study is to conduct sentiment analysis of product review from e-commerce platforms using machine learning to predict consumer behaviour so as to make informed decision. The methodology used was a sentiment analysis using models developed in Python with machine learning algorithms (Support Vector Machine and Naive Bayes) in order to compare the performance accuracy. The model was trained and tested on a dataset, and the performance of the two supervised machine learning algorithms for sentiment classification was compared. Accordingly, the results of the experiment showed that the Support Vector Machine technique outperformed Naive Bayes, obtaining accuracy levels of 97%. The study also demonstrates the significance of carrying out further research by exploring different machine-learning techniques and evaluating their performance on local market datasets. Additionally, it was observed that utilizing more data might improve the model's classification and prediction accuracy.

1. Introduction

Companies have embraced business models that integrate globalization and the use of the internet as an advertising medium for their products and services as new technologies have emerged. These business transaction models have evolved to incorporate new procedures and social changes (Saura *et al.*, 2018). The widespread adoption of internet technology has fundamentally disrupted the conventional ways for doing business, leading to the rise of e-commerce and influencing the trajectory of global economic growth. E-commerce refers to a website or mobile application platform that facilitates the buying and selling of goods and services (Willianto & Wibowo, 2020). Since its inception, it has grown exponentially by exploiting the opportunity offered by the technological advancement in information technology which offers the advantage of cross border trading, having a wide choice of products, quick retrieval, affordable prices, and the ability to customize and personalize services to customers (Chen, 2021).

The activities of users on these ecommerce platforms generate huge data that can be used by business organizations to find insight. User Generated Content (UGC) is the expression of users' opinions on digital platforms (Saura *et al*., 2019). These contents can be used by e-commerce platforms to evaluate and improve product and services to serve the customers better and to retain them. UGC also, helps e-commerce platforms to customize and personalize products and services for specific group with a strong target market orientation, which the market success is quite often high (Chen, 2021). Companies have leverage on these business strategies by utilizing technology that enables them better understand the perception and opinion of their customers using historical data. Additionally, consumers can use the reviews available on these e-commerce platforms as a guide to help them make informed choices. On retail websites like Amazon.com, reviewers have a number of options for posting their reviews. For example, the consumer can rate the product on a scale of one to five stars or write comments about it (Singla *et al*., 2017).

[*] Corresponding author. Tel.: +234 8060097016
E-mail address: 1enoshz@gmail.com (Akisani A. E.)
https://doi.org/10.62054/ijdm/0103.16

In recent years, machine learning has emerged as an important area of data science that has gained popularity because of its greater predictive and forecasting abilities, and as a result, it has replaced other equally essential fields as a standard approach various prediction task. Predictive models can be build using machine learning algorithms to determine the views and opinion of customers., patterns, forecast future sales trends, and provides useful information (Arunraj & Ahrens 2015).

This study aims to improve online store decision-making and enhance machine learning-driven e-commerce solutions by using the capabilities of machine learning in categorization and prediction tasks. The study is significant because machine learning can improve operational efficiency and organizational effectiveness through automation of the workforce, enhancing predictive intelligence of the decision-makers, and creating better competitive advantages. In addition, it helps improve recommender systems by comparing product parameters such as price, ratings, and special offers on e-commerce platforms to customize and personalize products to target certain customers. Furthermore, customers may find it easy to make informed choices based on the available information from previous reviews, and e-commerce website administrators will find it helpful in determining the inventory of products with high ratings.

2. Literature Review

Sentiment analysis, also known as opinion mining, is a machine learning technique for classifying and analyzing texts or documents containing information about human sentiments, emotions, and views. According to the amount of label from data training, it classifies texts into some classes (Willianto & Wibowo, 2020). Opinionated text in a review can be classified as "positive" or "negative" or "neutral" using Natural Language Processing (NLP). The reviews generated from e-commerce platforms and the hug nature of these opinionated texts makes it difficult to analyze and summarize. This is why sentiment analysis becomes increasingly important in analyzing and summarizing the opinions expressed in these huge data.

This study employed sentiment analysis to analyze product reviews from e-commerce platforms and classify them into either positive or negative and summarize such review to enable businesses and e-commerce administrator to improve the standard of their e-commerce website and the results could be processed further and be used to summarize customers' opinions about a certain product without having to read through every single review. The study utilized a pair of supervised machine learning algorithms, namely Support Vector Machine and Naïve Bayes. These algorithms were used to build machine learning models using Natural Language Processing (NLP) to accurately classify the identified textual content into negative, positive, and neutral respectively. The model's accuracy was analyzed using various metrics and recommendations were made.

Numerous researches have been carried out using sentiment analysis on data from customer reviews on various e-commerce platforms. It has been observed that Machine learning algorithms, deep learning algorithms, and sentiment lexicon have been used in the majority of these studies (Gondhi *et al*., 2022). Jiang (2022) introduced a method that involves using the combination of support vector machine (SVM) and enhanced particle swarm optimization (IPSO) to classify the sentiments expressed in reviews collected from an e-commerce platform. The study shows that the proposed model performed the best for classifying the sentiment of text data and the accuracy of classification has increased with the SVM.

In a study conducted by Olagunju *et al*. in (2020), they analyzed user reviews from mobile e-commerce applications in Africa to determine the sentiment expressed by the users. Two different approaches, Linguistic Inquiry Word Count (LIWC) and Machine Learning (ML), were used to classify the reviews as either positive or negative. The findings of the study indicate that LIWC yielded the highest performance, achieving an F1-score of 86.7%. Sentiment analysis can be used on stock markets, news articles, and political discussions, in addition to reviews and Twitter data. Consumer products-related businesses can benefit from sentiment analysis. It employs a rule-based sentiment analysis approach to extract subject phrases from unfavorable opinion statements, hence promoting the competitors of the products receiving negative feedback. Similarly, on several blogging sites aimed at bloggers, appropriate adverts depending on a person's liking or disliking are displayed. The Blogger-Centric Contextual Advertising Framework was created to identify users' specific interests and display adverts that are relevant to them (Singla *et al*., 2017).

3. Materials and Methods

The study focused on machine learning techniques to analyze the sentiment expressed in product reviews from an e-commerce platform to help ecommerce website administrator make informed decisions. To be more precise, machine learning algorithms were built and trained using Python to perform the task.

3.1 Method of data collection and study area

The data used in this study was extracted from Jumia Nigeria e-commerce platform. The data was ethically scraped from the e-commerce platform over a two-month period (May to June 2023) using the Python library Beautiful Soup. The scraping adhered to the platform's terms of service, protected user privacy by anonymizing data, and was solely for academic research. The dataset, comprising 1,028 samples with six features, was collected from diverse product categories and included a range of ratings, ensuring representativeness of customer sentiment. The unstructured data was carefully labeled, preprocessed, and formatted to create a well-structured dataset suitable for sentiment analysis. Table 1 shows the features in the scrapped dataset.

Table 1: Dataset Description

Data feature	Description	Data Type
Serial Number (SN)	Unique identifier for the review	Integer
Summary	Brief summary of the review	Object
Review	Detailed review content	Object
Rating	Numeric rating given in the review	Int64
Date	Date when the review was written	Object
Sentiment	Positive, negative, or neutral	Int64

3.2 Data preprocessing and cleaning

To ensure high-quality input for sentiment analysis, the dataset from Jumia Nigeria underwent thorough preprocessing using Python libraries including Pandas, NumPy, NLTK, and BeautifulSoup. The preprocessing steps were designed to enhance data quality by addressing various challenges such as noisy data, incompleteness, and inconsistencies. Key techniques included removing special characters, punctuation, and extra whitespaces from the review text; converting all text to lowercase to maintain consistency; eliminating stop words that do not contribute significant meaning; and removing duplicate reviews to prevent bias and overfitting. Missing values were handled by either imputing them with the most frequent value or removing records as necessary. Challenges such as inconsistent review dates were resolved by standardizing date formats, while ambiguous text was managed using lemmatization to ensure uniformity and accuracy in sentiment classification.

4. Model Approach and Modeling Tools

The study employed supervised machine learning method considering the fact that we are dealing with classification problem. This is because Machine learning is focused on developing an algorithm to enhance the performance of the system utilizing sample data or prior knowledge. Although there are numerous machine learning classification techniques, we used only two of them: Naive Bayes and support vector machines to classify reviews into positive and negative respectively to ascertain consumer behaviour. The outcomes of these classification algorithms could be almost similar to those obtained by other classification methods but our aim is to examine the performance of this models. The tools utilized for the model building and evaluation was Google Colab, a cloud-based Python development environment that makes use of Google software, CPU and other hardware for effective performance. It provides all the necessary required platforms and application for Machine Learning experiments.

4.1 The proposed model

This study aims to use machine learning for sentiment analysis of product reviews from e-commerce platforms to predict consumer behaviour. It addresses a classification problem using Natural Language Processing (NLP) to classify consumer sentiments in the reviews. The process involves several steps: Import and install necessary Python libraries and packages, Load the CSV dataset into a Colab notebook and read the data, Pre-process the dataset to structure the data and remove noisy information, Extract features from the dataset, Split the dataset into an 80:20 ratio for training and testing, Build the model using Machine Learning algorithms, specifically Support Vector Machine (SVM) and Naïve Bayes, chosen for their accuracy and industry acceptance, Evaluate the model's performance by comparing the two techniques and presenting metrics and recommendations.

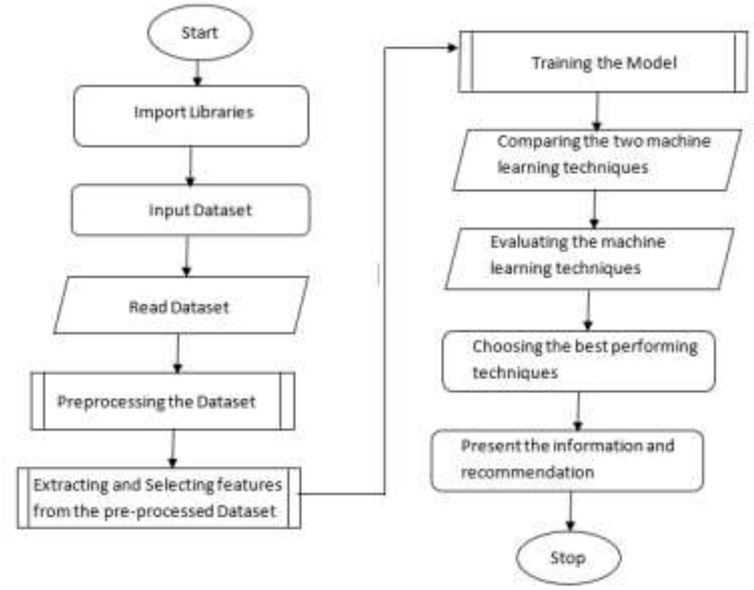

Figure 1 demonstrates the proposed model for sentiment analysis.

4.2 Performance Evaluation Metrics

Evaluation metrics play a crucial role in assessing and determining the effectiveness of a machine learning model. The study employed precision, recall, F-score and support to efficiently evaluate whether the model is correctly classifying the product reviews into "positive" and negative reviews effectively. The study employed the confusion matrix library which is a table representation of how well the model performs in terms of classification of the product reviews. A confusion matrix is crucial for evaluating the performance of a classification model. In order to provide a better understanding of a model's recall, accuracy, precision, and overall efficacy in class differentiation, it provides a comprehensive analysis of true positive, true negative, false positive, and false negative predictions.

Accuracy is used to measure the model's performance. It is used in multi class classification problems to show how well the model can differentiate between different classes in to evaluating the ratio of correct predictions to the total predictions.

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN +FP + FN}} \tag{1}$$

Precision is a measure of how accurate a model's prediction of positive the observations commensurate with the total number of positive observations predicted.

$$Precision = \frac{\text{TP}}{\text{TP+ FP}} \tag{2}$$

Recall measures how well a classification model predicted positive observations to the total numbers of actual positive observations. $Recall = \frac{\text{TP}}{\text{TP+ FP}}$ (3)

F1-score is used to measure the overall performance of a classification model. It is the harmonic mean of precision and recall, combining both metrics into a single value, providing a balanced assessment of the model's predictive capability.

$$F1 - Score = \frac{2*Recall*Precision}{\text{Precision+Recall}} \tag{4}$$

5. Results

The details of the experiments carried out during this study are described in this section, along with discussion of the findings. We proposed a set of machine learning techniques for sentiment analysis of product reviews to classify and predict sentiment in an opinionated text. The aim of employing these machine learning techniques which include Support Vector Machine and Naïve Bayes models was to analyze and classified sentiment from ecommerce website and to compare the outcome of the two models. We utilized Google Colab, a cloud-based Python development environment that makes use of Google software, CPU and other hardware for effective performance. The criteria employed to evaluate the models' effectiveness throughout each experiment was accuracy and F-score. We also display recall and precision because they are the inputs used to calculate F-score.

The performance of the proposed techniques was evaluated on the dataset after preprocessing with an efficient text processing methods to enhance accuracy in the text classification. Figures 4.1 shows the result of the visualized extracted dataset, displaying the distribution of the sentiments using bar chat. The x-axis represents the number of sentiments, while the y-axis represents the levels of sentiment (Zero is indicates negative and One indicates positive).

Table 2: Sample of the dataset

Rating	Review	Sentiment
4	started using far looks cool	1
4	tv nice sounds good video quality also good re...	1
5	good love	1
5	itel really awesome job recommend anybody	1
5	looks really sleek low energy consuming	1

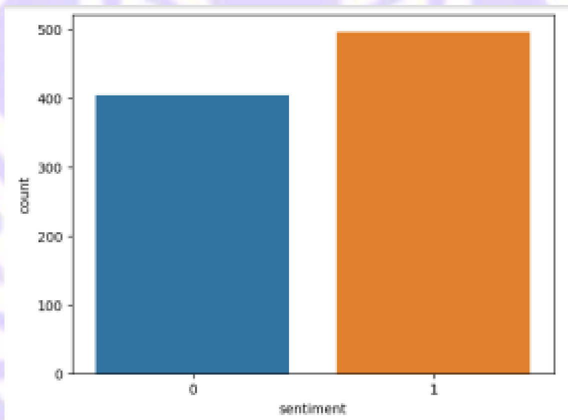

Figure 2: Bar Chat showing the negative and positive Sentiment Distribution

In order to build, test and evaluate the proposed model using the machine learning techniques, we split the dataset into 80% training and 20% testing ratio. Each experiment is run several times, and the final performance is taken.

Figure 3: Word cloud for the most frequent positive reviews

5.1 Performance evaluation

All of the experiments demonstrated that support vector machines performed more accurately on the dataset. Due to the size of the working dataset and the fact that support vector machines perform better when given huge datasets without being over fit. The highest accuracy score, accordingly using the support vector machine was 97% while the Naïve Bayes has the lowest 85%. The accuracy performance serves as a measure of the recall and accuracy of each category's sample sets. This performance reflects the success of the sentiment analysis performed by the Support Vector Machine (SVM) algorithm, a machine learning-based algorithm which indicates the algorithm's overall average performance when classifying the reviews. Table 2 shows the performance of the naive Bayes and support vector machine (SVM) classifiers for the sentiment analysis while utilizing 80:20 training-to-testing ratio.

Evaluation Metrics	**Support Vector Machine**	**Naïve Bayes**
Accuracy	97%	85%

Table 2: Performance accuracy results for sentiment analyzed on the first dataset.

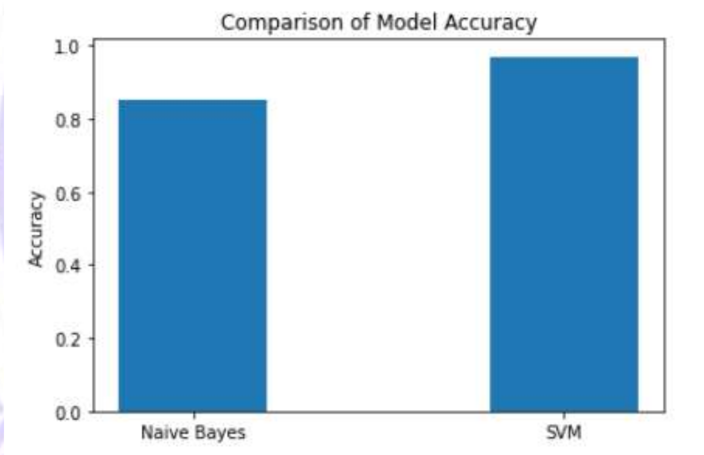

Figure 4: Comparison of performances accuracy for SVM and Naïve Bayes

Based on the bar chart presented in Figure 4.3, it can be observed that the performance of two distinct machine learning algorithms, namely Support Vector Machine (SVM) and Naïve Bayes, differs significantly. The accuracy of SVM is reported to be 97%, whereas Naïve Bayes achieves an accuracy of 85%. This indicates that SVM outperforms Naïve Bayes slightly when considering overall accuracy.

Table 3: Confusion matrix for the Support Vector Machine classifier

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.89	0.57	0.70	14
1	0.97	0.99	0.98	192
Accuracy			0.97	206

Macro Avg	0.93	0.78	0.84	206
Weighted Avg	0.96	0.97	0.96	206

Table 3: presents the confusion matrix for the Support Vector Machine (SVM) classifier, providing an analysis of its prediction accuracy. The true positive rate for positive predictions is 97%, indicating that 97% of the instances predicted as positive are indeed positive. Conversely, the precision for the negative class is 89%, meaning that 89% of the predicted negatives are actually negative (true negatives). Regarding recall, the model accurately identifies 57% of the examples that are truly negative. For the positive class, the recall is 0.99, indicating that the model correctly identifies 99% of the instances that are true positive. This high precision results in a moderately high F1-score of 70% for the positive class. Additionally, due to the high precision and recall, the F1-score for the positive class is 98%, signifying an excellent performance. These findings support the conclusion that the classifier accurately assigns the sensitive dataset to the correct class. The overall accuracy of the model is 97%, implying that it correctly classifies instances 97% of the time. The model performs well in classifying the positive class, exhibiting high precision, recall, and F1-score. It also demonstrates reasonable performance for the negative class, with high precision but low recall. This suggests that the model may struggle to identify instances of the true negative, but when it does, it is often correct. Overall, the high accuracy of 97% indicates that the model performs well on the dataset as a whole.

Table 4: Confusion Matrix for the Naïve Bayes Classifier

Class	**Precision**	**Recall**	**F1-Score**
0	0.27	0.71	0.39
1	0.98	0.86	0.91
Accuracy			0.85
Macro Avg	0.62	0.79	0.65
Weighted Avg	0.93	0.85	0.88

Table 4 presents the confusion matrix for the Support Vector Machine (SVM) classifier, which evaluates the performance of a binary classification model that predicts whether a data point belongs to a specific class. Regarding the precision of the model, the true negative precision is 0.27, indicating that only 27% of the predicted negative instances are actually negative. On the other hand, the true positive precision is 98%, implying that nearly all positive predictions are true positives. In terms of recall, the model achieves a recall of 0.71 for class 0 (negative), meaning that 71% of the instances are correctly identified as true negatives. For class 1 (positive), the recall is 0.86, indicating that 86% of the positive examples are properly identified by the model. However, due to the low precision for negative predictions, the F1-score for class 0 is only 0.39. Conversely, the F1-score for class 1 is 0.91, reflecting a high number of instances correctly classified as positive. The accuracy of the model, which represents the proportion of correct predictions out of all predictions, is 0.85. This implies that the model accurately predicts the class of 85% of the instances. Overall, the model demonstrates good performance in classifying the positive class, with high precision, recall, and F1-score. However, it performs poorly for the negative class, exhibiting low precision and F1-score. The weighted average metrics indicate that the model is accurate overall, considering the number of instances in each class. Nonetheless, the macro average metrics suggest that there is room for improvement in the model's performance. Overall, the high accuracy of 85% indicates that the model performs well on the dataset as a whole

Table 5: Comparative Confusion Matrices of the two models

	SVM	Naive Bayes
Precision (class 0)	0.89	0.27
Precision (class 1)	0.97	0.98
Recall (class 0)	0.57	0.71
Recall (class 1)	0.99	0.86
F1-score (class 0)	0.70	0.39
F1-score (class 1)	0.98	0.91
Accuracy	0.97	0.85
Macro Avg	0.93	0.62
Weighted Avg	0.96	0.93

In general, the SVM technique performs better than the Naive Bayes model. The SVM model has higher precision and recall scores for both the positive and negative with higher F1-scores overall. However, the Naive Bayes model has a higher recall score for the classification of the negative class 0, meaning it correctly identifies more of the true negative cases.

Support Vector Machine and Nave Bayes classifier were the two fundamental classifiers that we used to categorize sentiment using train-test-split ratio 80:20. Table 2 shows the results of the two experiments and their outcome. From the dataset, Support Vector Machine techniques performed better with the accuracy of 97% and 77% respectively while the Naïve Bayes classifier had 85% and 74% respectively. Table 3 and 4 displays the Support Vector Machine and Naive Bayes models' confusion matrices. The rates of True Positive, False Positive, True Negative, and False Negative in the sample are displayed using the confusion matrix. The assessment metrics (accuracy, recall, precision, and F1-score) were computed based on these rates to assess the models' ability to use datasets to predict consumer sentiment. Table 3 and 4 displays the overall accuracy of the Support Vector Machine model as high as 97%, indicating that the model performs well on the dataset as a whole. Even though, the models were trained on a relatively small amount of dataset, which may have led to the likely prediction under performance rates in the negative class in Naïve Bayes but overall, the Naïve Bayes performs well for the positive class (class 1) but poorly for the negative class (class 0), and we discovered there is potential for improvement in its performance if more data is added. Additionally, given the variety of options for selecting initial parameters, our outcome might be as a result of the parameter settings used that resulted in Naïve Bayes' poor performance given that numerous studies have demonstrated it good performance in classification problem. Figure 4 and Table 5 shows the comparative analysis for the two classifiers.

5.2 Findings

In our analysis, we found that a proficiently trained machine learning algorithm can achieve highly accurate classifications on product review datasets, providing valuable insights for ecommerce platform administrators in predicting consumer behaviour. Although machine learning is widely accepted for classification tasks, its effectiveness is influenced by the specific domain it is applied to. Furthermore, it presents additional challenges such as acquiring appropriate training datasets and correctly labeling documents, selecting the most suitable machine learning technique, achieving high accuracy, and effectively managing imbalanced datasets. Despite these limitations and challenges, machine learning techniques remain highly effective and extensively utilized for classification purposes.

6. Conclusion

The aim of this study is to conduct sentiment analysis of product review from ecommerce platforms using machine learning to predict consumer behaviour. We were able to achieve this goal by demonstrating the importance of sentiment analysis in the ecommerce domain. The study achieved a high accuracy score using Support Vector Machine (SVM) technique. Overall, sentiment analysis is a valuable tool for analyzing and understanding customers' opinions and sentiments towards products, services, or organizations. In order to have a deeper understanding, future research in sentiment analysis could be enhanced by integrating multimodal data, such as text combined with images and videos. Further research into advanced deep learning models like BERT and GPT may improve the extraction of complex opinions, while addressing code-switching and mixed-language content could improve general applicability, Furthermore, developing methods for identifying irony and sarcasm and customizing models for certain sectors like E-commerce, finance or healthcare could produce more precise and specialized insights. The aim of these approaches is to make sentiment analysis more accurate and applicable in a wider range of real-world situations.

References

Chen, S. (2021). Analysis of Customization Strategy for Ecommerce Operation Based on Big Data. *Wireless Communications and Mobile Computing*, *2021*. https://doi.org/10.1155/2021/6626480

Gondhi, N. K., Chaahat, Sharma, E., Alharbi, A. H., Verma, R., & Shah, M. A. (2022). Efficient Long Short-Term Memory-Based Sentiment Analysis of Ecommerce Reviews. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/3464524

Gondhi, N. K., Sharma, E., Alharbi, A. H., Verma, R., & Shah, M. A. (2022). Efficient Long Short-Term Memory-Based Sentiment Analysis of E-Commerce Reviews. *Computational Intelligence and Neuroscience*, *2022*.

Olagunju, T., Oyebode, O., & Orji, R. (2020). Exploring Key Issues Affecting African Mobile Ecommerce Applications Using Sentiment and Thematic Analysis. *IEEE Access*, *8*, 114475–114486. https://doi.org/10.1109/ACCESS.2020.3000093

Saura, J. R., Palos-Sanchez, P., & Grilo, A. (2019). Detecting indicators for startup business success: Sentiment analysis using text data mining. *Sustainability (Switzerland)*, *11*(3), 1–14. https://doi.org/10.3390/su11030917

Saura, J. R., Reyes-Menendez, A., & Alvarez-Alonso, C. (2018). Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability (Switzerland)*, *10*(9). https://doi.org/10.3390/su10093016

Singla, Z., Randhawa, S., & Jain, S. (2017). Sentiment analysis on product reviews using machine learning techniques. *2017 International Conference on Intelligent Computing and Control (I2C2)*.

Willianto, T., & Wibowo, A. (2020). Sentiment Analysis on Eccommerce Product using Machine Learning and Combination of TF-IDF and Backward Elimination. *International Journal of Recent Technology and Engineering*, *8*(6), 2862–2867. https://doi.org/10.35940/ijrte.f7889.038620.