ISSN: 2634 - 8853

# Journal of Engineering and Applied Sciences Technology

SCIENTIFIC Research and Community

Research Article Open Access

# Enhancing Marketing Analytics in Online Retailing through Machine Learning Classification Techniques

**Bhumeka Narra[1]*, Navya Vattikonda[2], Anuj Kumar Gupta[3], Dheeraj Varun Kumar Reddy Buddula[4], Hari Hara Sudheer Patchipulusu[5] and Achuthananda Reddy Polu[6]**

[1]Sr Software Developer, Statefarm, USA

[2]Business Intelligence Engineer, International Medical Group Inc, USA

[3]Oracle ERP Senior Business Analyst ,Genesis Alkali, USA

[4]Software Engineer, Anthem Inc, USA

[5]Software Engineer, Iheartmedia, USA

[6]SDE3, Goldman Sachs, USA

**ABSTRACT**
In the fiercely competitive retail industry, satisfying consumer expectations while optimizing company processes is more important than ever. Therefore, it is crucial to handle and channel data in a way that both seeks to delight consumers and generates healthy revenues if you want to survive and prosper. Data—or more specifically, Big data analytics is being utilized by large retailers at every stage of the process, participants in the global and Indian retail markets, including tracking new, popular items and predicting sales. The use of machine learning classification approaches for sentiment analysis in online shopping is examined in this research, utilizing a publicly available Amazon review dataset. The text-cleaning techniques processed the dataset before converting texts into numerical representations by implementing TF-IDF measures. The assessment concentrated on the three machine learning models' F1-score, accuracy, and precision-recall: Bidirectional Encoder Representations from Transformers (BERT), Support Vector Machine (SVM), and Gradient Boosting (GB). BERT ended up outperforming all other models by demonstrating 89% accuracy, which proves its extraordinary capability to detect customer sentiments. The research results show how transformer-based models work for improving sentiment analysis procedures in marketing analytics applications.

***Corresponding author**
Bhumeka Narra, Sr Software Developer, Statefarm, USA.

**Received:** August 12, 2021; **Accepted:** August 17, 2021; **Published:** August 28, 2021

**Introduction**
The digital period requires data-based decision-making because it stands fundamentally important for business achievements. Business organizations operating in many industries apply analytics tools to analyze customer behavior while simultaneously optimizing their operations and improving customer interaction [1]. Businesses now use improved AI coupled with ML to transition from traditional analytics into more complex predictive as well as prescriptive decision-making. Modern technologies furnish organizations with the ability to handle enormous datasets to discover hidden patterns which helps them make strategic decisions. Online retailing has become one of the main areas where such technological progress has delivered major benefits through adapting to transforming consumer interests and purchasing patterns [2].

Online retailing has displayed explosive development across the world because e-commerce sales exceeded $4.2 trillion. Digital buying through the Internet has developed to become a key aspect of customer lifestyle because of ongoing increases in internet use and mobile commerce trends. Businesses operating online continue to encounter difficulties in analyzing customer actions and enhancing their marketing plans along with boosting converting potential customers [3]. Effective decision-making gaps emerge from this limitation which requires online retailers to adopt advanced analytics solutions [4]. Business performance and customer targeting along with campaign effectiveness can be better reached through data-driven techniques of marketing analytics which bridge the existing gap [5]. Here is where machine learning classification algorithms provide a game-changing answer, allowing marketers to more precisely categorize client categories, forecast purchase trends, and optimize promotional efforts.

One subset of AI is machine learning categorization, which may be used to provide marketing statistics for online retailers [6,7]. It also uses to widespread for anomaly detection for online retailing because it helps businesses detect frauds, suspicious activities and other forms of consumer behavior shifts real time. The Machine learning classification in marketing analytics for online retail plays a great role, which is to integrate the capabilities of AIpowered classification models in marketing analytics to provide more personalization, more accurate digital advertising, and more customer retention.

**Motivation and Contributions of the Study**

To meet the demands of exponential retailing growth, a point, customer sentiment plays an important part in creating product offerings and strategies to make marketing. The analysis of Amazon product reviews offers valuable insights into customer opinions, which can significantly impact business decision-making. This study enhances sentiment analysis in the retail sector by utilizing DL and advanced ML techniques.

**The main findings of this study are:**

- Utilization of a large Amazon product reviews dataset, providing a diverse foundation for sentiment analysis across various product categories.
- The research implements a robust preprocessing framework, including missing value handling, text normalization, stop word removal, and tokenization, ensuring high-quality input data for machine learning models.
- By utilizing TF-IDF, the study effectively converts unstructured textual reviews into meaningful numerical representations, enhancing model interpretability and performance.
- Evaluation and comparison of multiple models, Using BERT, SVM, and GB for sentiment analysis.
- In-depth assessment of model performance using important parameters such the confusion matrix, F1 score, recall, accuracy, and precision.

**Organization of the Paper**

The paper is structured in the following manner: Sections II provide existing work on Online Retailing. The mechanism for this is then provided in Section III. Section IV then presents the results and debate, while Section V presents the conclusion and next steps. Conclusion and next steps are provided in the final section.

**Literature Review**

In the past, most researchers were interested in online retailing using machine learning techniques. There is various previous research are as:

Mondragon et al. proposed a technology paradigm that makes it possible to record, assess, and display the customer experience throughout the awareness, consideration, purchase, usage, and loyalty phases of the customer journey. The assessment is used to track the pre-purchase, purchase, and post-purchase processes. Within two months, a retail company's use of the concept enabled a more than 50% increase in sales [8].

Riantini examined factors from the Theory of Planned Behaviour for omnichannel customers and examined the characteristics of Jakarta online shoppers. The study involved 125 online retail consumers, including young women under 30 years old, workers, and employers. The findings demonstrated that customers' intention to buy the goods, which is impacted by subjective norms, perceptions, and varying attitudes towards behaviour, influences their online buying behaviour. The study used the Structural Equation Model to test the hypothesis [9].

Ngamjarussrivichai, Jeemali and Panitsettakorn designed a technique for evaluating internet marketing that displays customer journey details, ROMI and ROAS in real-time. The system uses internal and external databases, including customer requirement and sale databases, and Google Analytics. The study finds that historical data analysis reveals a maximum ROMI increase of 6.56% and an ROAS increase of 20% [10].

Cheng and Jiang characterized the platform's and sellers' best tactics and earnings across various competitive conditions using the Stackelberg and Nash game models. According to the numerical analysis, the platform's commission rate is quite low, thus the two sellers will be in a prisoner's dilemma and may compete fiercely on price. Among other marketing situations, the platform will currently make more money without providing a platform rebate [11].

Lei, Guo and Liang examined how the growth of E-commerce had an impact on the locations that global retail corporations chose in China, concentrating on six major retailers: Carrefour, Auchan, Metro, Wal-Mart, RT-Mart, and Lotus. These firms' site selections are evaluated in relation to e-commerce development using statistical data, conditional logistic regression analysis, and the China e-commerce development index. Based on size, penetration, and supporting indices, the results show a favorable relationship between these merchants' geographical preferences and the extent of e-commerce development. Furthermore, elements like competitiveness, market potential, infrastructure, and entry obstacles have significant impacts [12].

Table 1 shows prior studies on Evaluating Machine Learning Approaches for Online Retailing using deep learning and machine learning methods, comparing their effectiveness.

**Table 1: Summary of the Related Work on Online Retailing using Machine learning**

| Authors | Dataset | Methodology | Finding | Limitations | Future Work |
|---|---|---|---|---|---|
| Mondragon, et al. | Retail company data | Technological model for capturing, evaluating, and visualizing customer experience | Sales improved by more than 50% in two months | Limited to a single retail company | Extend model to multiple industries for validation |
| Riantini | 125 online retail consumers in Jakarta | The structural equation model (SEM) and the theory of planned behavior (TPB) | Online buying behavior influenced by intention, attitudes, subjective norms, and perceptions | Small sample size, focused on a specific demographic | Expand study to include diverse consumer groups and locations |
| Ngamjarussrivichai, Jeemali, and Panitsettakorn | Internal and external databases, including Google Analytics | Online marketing evaluation system analyzing ROMI and ROAS in real time | ROMI increased by 6.56%, ROAS by 20% | Lack of generalizability to industries outside retail | Extend analysis to different business sectors |

| | | | | | |
|---|---|---|---|---|---|
| Cheng and Jiang | Numerical study based on game theory | Nash Game and Stackelberg Game Model | Sellers face a Prisoner's Dilemma in price competition, platform profits increase without rebates. | Limited to theoretical and numerical analysis | Empirical validation with real-world data |
| Lei, Guo, and Liang | China e-commerce development index, statistical data from six major retailers | Conditional logistic regression analysis | E-commerce development positively impacts retailer location choices | Focused only on six major retailers in China | Broaden study to include more retailers and different markets |

**Methodology**

This study investigates marketing analytics in online retailing by applying machine learning techniques to a publicly available Amazon review dataset. The overall implementation process is shown in Figure 1. The dataset, encompassing reviews across five product categories, was preprocessed using techniques like missing value handling, lowercase conversion, stop word removal, and tokenization. Textual data was transformed into numerical representations using TF-IDF for feature extraction. The dataset was then split between 80% training sets and 20% testing sets. To ascertain how well three classification models—SVM, GB), and BERT—classify customer reviews and inform marketing strategies, they were trained and assessed using metrics taken from the confusion matrix, like F1-score, recall, accuracy, and precision.

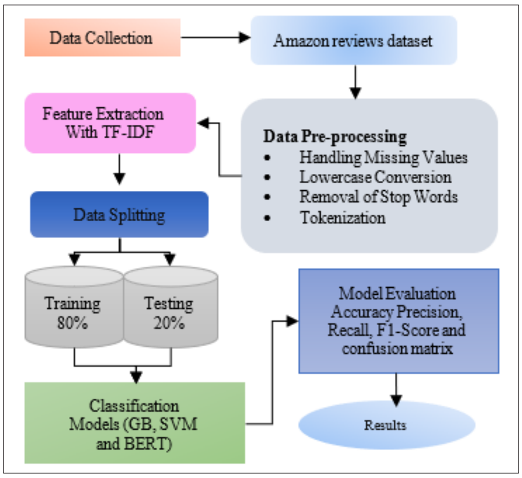


**Figure 1:** Flowchart for Online Retailing

The following steps of flowchart are listed in next section.

Data Collection

A publicly accessible Amazon Review dataset that can be accessed through Kaggle has been utilized. This dataset includes reviews for five different product categories: watches, cameras, groceries, furniture, and mobile gadgets. Features including marketplace details, customer and review IDs, product IDs, product parent, product titles, product categories, star ratings, helpful votes, total votes, buy verification status, review headlines, and review bodies are all included in each record. The following visualization of review product category are illustrated in Figure 2.

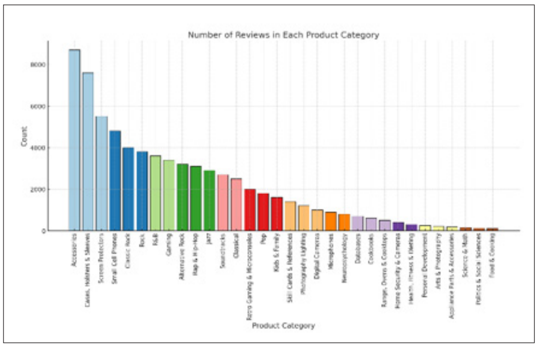


**Figure 2:** Distribution of Number of Reviews

Figure 2 illustrates the number of reviews across various product categories in an online retail dataset. The x-axis displays the different product categories, while the y-axis displays the quantity of reviews for each product category. Accessories, Cases & Accessories, and Small Cell Phones have the highest number of reviews, indicating their popularity among consumers. Other categories, such as Classic Rock, R&B, and Gaming, also receive significant attention. In contrast, categories like Language & Literature, Computers & Software, and Science & Disasters have the fewest reviews, suggesting lower consumer engagement. The visualization highlights disparities in customer interest across different product segments, which can inform targeted marketing and inventory decisions.

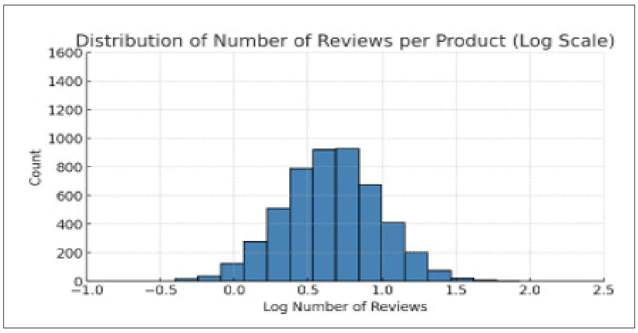


**Figure 3:** Distribution of Number of Reviews in Log Scale

The distribution of product reviews on a logarithmic scale is shown in Figure 3, where the x-axis ranges from -1.0 to 2.5 and the y-axis is from 0 to 1600. The right-skewed distribution shows most products receive fewer reviews while a few dominate. The highest concentration is in the mid-range, reflecting uneven customer engagement. The color change improves clarity without altering axis values.

**Data Preprocessing**
A crucial stage in NLP activities is data preparation, which maximizes the effectiveness of knowledge discovery. In order to prepare and enhance the dataset for more efficient analysis, it uses strategies such data cleansing, integration, transformation, and reduction. The preprocessing steps are listed below:
- **Handling Missing Values:** To supply the null values for features with an object data type, they have used Python's fillna() function.
- **Lowercase Conversion:** In this step, change all of the review words to lowercase. Through case-insensitive word treatment, lowercasing decreases the dimensionality of the data and helps standardize the language.

**Removal of Stop Words**
In the discipline of text mining, stop words are phrase components that are irrelevant in all contexts [13]. Removed every HTML element, punctuation mark, and stop word from the reviews in a corpus. This preparation stage increases computing performance and lowers data noise.

**Tokenization**
Applied both word and phrase tokenization were used. The technique of tokenization divides a text sequence into discrete parts called tokens. Tokens can range in length from a single word to a sentence or phrase. Following that, these tokens are fed into procedures like text mining and parsing. Instead of processing a full text as a single sequence, it enables models to focus on the significance of individual components.

**Feature Extraction with TF-IDF**
In text analytics and NLP, one of the frequently utilized feature extraction techniques is TF-IDF [14]. It allows us to represent textual data as numerical data that is suitable to feed to ML model. The TF-IDF is calculated as Equation (1) for a given term t from document $d$ of corpus D:

$$\text{TF-IDF}(t,d)=tf(t,d).idf(t,D) \quad (1)$$

**Where:**
- TF-IDF(t,d) is Document D's TF-IDF score for word t in relation to document D.
- tf(t,d) represents the phrase t's frequency in document d.
- idf(t,D) is the document set D's inverse document frequency for term t.

**Data Splitting**
There is a 20% testing set and an 80% training set in the dataset. Models are tested on the testing set after being taught and optimized on the training set.

Bidirectional Encoder Representations from Transformers (BERT)
In NLP, Bidirectional Encoder Representations of Transformers is a hidden language model [15]. Machine translation, named entity identification, question answering, text categorization, and sentiment analysis are just a few of the natural language processing tasks that BERT has been used to. Transformers are the foundation of it. Based on their relationship, the weight between the encoder and decoder, two entities in the bidirectional self-supervised model, is dynamically determined. Transformers are used to carry out the translation process as both encoders and decoders.

This technique captures the contextual links between words by allowing word order to be represented in the matrix of sentences. AE and RE are the two types of position embedding. AE is used to map the space element of representation, whereas RE is used to map the positional distance between words. In the attention head, the weighted value $W_k$, weighted key $W_k$, and weighted query W_q are taken into consideration while calculating the BERT transformer. Consider, for illustration purposes, that two locations are xϵN and yϵN. Thus, the embeddings are represented by $E_x$, and the x positional word vector by $WVx$. Moreover, $E_{x-y}$ is defined as the relative position's embedding. The formula below may be used to calculate the query (q), value (v), and key (k) vectors for the x positional word Equation (2 and 3):

$$AE\colon \begin{bmatrix} q_x \\ k_x \\ v_x \end{bmatrix} = (WV_x + E_x)x \begin{bmatrix} W_q \\ W_k \\ W_v \end{bmatrix} \quad (2)$$

$$RE\colon \begin{bmatrix} q_x \\ k_x \\ v_x \end{bmatrix} = (WV_x + E_x)x \begin{bmatrix} 0 \\ E_{x-y} \\ E_{x-y} \end{bmatrix} \quad (3)$$

The total of all attention heads' values is the output at the end (a).

**Model Evaluation**
Performance evaluation of classification methods depends on accuracy combined with F-score and recall and precision metrics. Supervised ML algorithms receive critical assessments by using these performance parameters. Performance evaluation of models becomes possible through in order to provide comprehensive model assessment data, the confusion matrix shows both true positive and true negative classifications in addition to false positive and false negative findings:
- **True Positive (TP):** The analysis requires actual and predicted values to be positive in both cases.
- **False Positive (FP):** A positive predicted value gets contradicted by an existing negative actual value.
- **True Negative (TN):** A negative real value that already exists contradicts a positive forecasted value.
- **False Negative (FN):** A positive real observation matches the projected negative value.

**Accuracy:** Accuracy serves as a predictive measure for the correct number of classifications produced by the system. The accuracy ratio stands for the true prediction count divided by the complete prediction count. Equation (4) defines accuracy (A).

$$A = \frac{(True\ Positive + True\ Negative)}{(Total\ Number\ of\ Predictions)} \quad (4)$$

**Precision:** There are fewer false positives when the precision is higher, and more false positives when the precision is lower. The precision (P) ratio is the proportion of correctly recognized cases to all instances. It may be described as. Calculating P involves using Equation (5):

$$P = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \quad (5)$$

**Recall:** The sensitivity of a classifier and the percentage of positive data it produces are determined by recall. There are fewer false negatives when recall is higher. The ratio of correctly identified instances to all anticipated events is known as recall. This may be demonstrated as. Equation (6) is used to calculate the recall:

$$R = \frac{TP}{(TP+FN)} \tag{6}$$

**F1 Score:** Precision and recall are combined to create the F-measure, a single statistic that is the weighted harmonic mean of accuracy and recall. It can be defined as. Equation (7) is used to obtain the F1 score:

$$F1 = 2 \times \frac{precision \times recall}{(precision+recall)} \tag{7}$$

This matrix is used for model evaluation and predict the highest classification performance of models.

## Result Analysis and Discussion

This section pertains to the examination and interpretation of the results, as well as the subsequent discussion. The whole experiment is performed on the Python programming language and various devices such as Intel core i7 8th generation, CPU, and personal computer. This section presents the BERT model, as summarized in Table 2.

**Table 2: Results of the BERT Model for Online Retailing on Amazon Review Data**

| Performance Matrix | Bidirectional Encoder Representations from Transformers (BERT) |
|---|---|
| Accuracy | 89% |
| Precision | 88% |
| Recall | 88% |
| F1 Score | 89% |

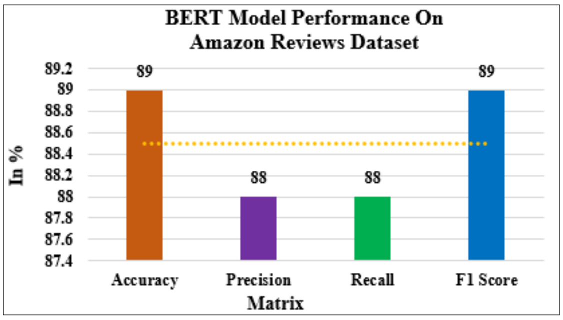


**Figure 4:** Bar Graph for BERT Model Performance

Table 2 and Figure 4 present the evaluation results for the BERT model. All of the assessment measures show that the BERT model performs well; its accuracy is 89%, precision is 88%, recall is 88%, and F1-score is 89%. The model's dependable detection of genuine positives with a low proportion of false negatives and false positives is demonstrated by these numbers, resulting in a reliable and balanced classification capability. This performance highlights that the BERT model is well-suited for the task, delivering consistent and robust results.

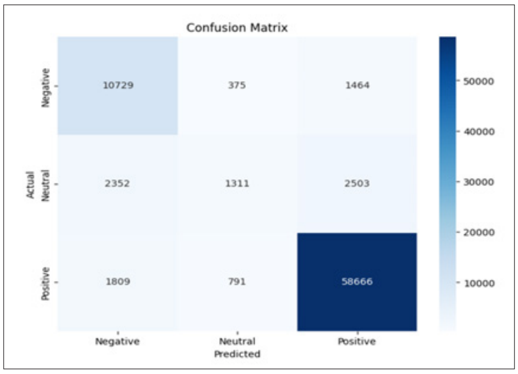


**Figure 5:** Confusion Matrix For BERT

Figure 5 shows the confusion matrix that depicts performance of a BERT model in a job involving multi-class classification. The classes include "Negative," "Neutral," and "Positive." The model correctly classified 10,729 "Negative," 1,311 "Neutral," and 58,666 "Positive" instances. However, there are some misclassifications: 375 "Negative" cases as "Neutral," 2,352 "Neutral" cases as "Negative," and 2,503 "Neutral" cases as "Positive." Additionally, 1,464 "Negative" cases and 791 "Positive" cases have been misclassified. The matrix highlights that the model performs best for the "Positive" class, performance of a BERT model in a job involving multi-class classification.

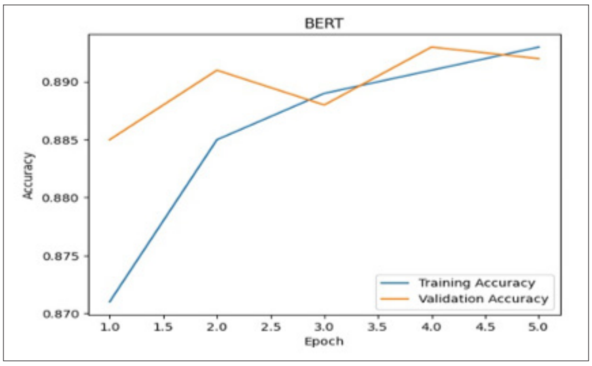


**Figure 6:** Training and Validation Accuracy for BERT

Figure 6 illustrates This graph displays a BERT model's accuracy during five epochs of training and validation. Training accuracy is shown by the blue line, and validation accuracy by the orange line. Starting at around 87% in the first epoch and rising to about 89% by the fifth, the model shows a steady increase in training accuracy. Validation accuracy fluctuates slightly but remains close to the training accuracy, peaking around the third epoch before converging near 89% in the final epoch. This suggests that there is little overfitting, and the model generalizes effectively.

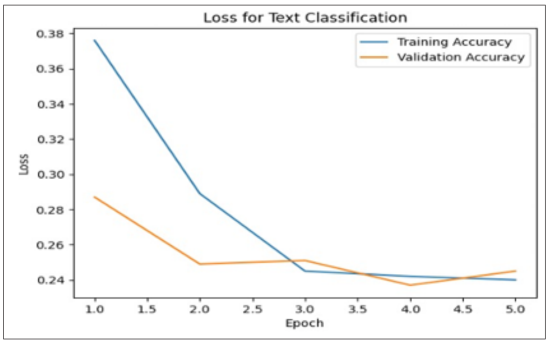


**Figure 7:** Training and Validation Loss for BERT

A BERT model's loss reduction during training and validation throughout five epochs is depicted in Figure 7's loss graph. The blue line shows training loss, whereas the orange line shows validation loss. Starting at around 0.38 for training loss and 0.28 for validation loss in the first epoch, both losses gradually decline. By the fifth epoch, the losses converge near 0.24, indicating improved model optimization and reduced error.

**Table 3: Comparison between BERT and Existing Model Performance for Online Retailing**

| Models | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| SVM [16] | 81.82% | 59.45% | 42.52% | 49.57% |
| GB [17] | 87% | 88% | 98% | 92% |
| BERT | 89% | 88% | 88% | 89% |

Table III shows the outcomes of both the suggested and existing models. In this comparison, BERT's 89% accuracy was the best, thanks to its 89% F1 score, 88% recall, and 88% precision. In addition, GB did very well, with an F1 score of 92%, accuracy of 87%, precision of 88%, and recall of 98%. With a recall of 42.52%, precision of 59.45%, and accuracy of 81.82%, the SVM model performed admirably. The F1 score was 49.57%. According to the results, BERT is the best model to use for this job.

## Conclusion and Future Scope

The marketing mix in promoting products and services via web and mobile apps is changing as modern marketing moves away from traditional platforms and towards digital ones. To ensure ethical online consumption, companies are using marketing analytics to fine-tune their online models while addressing consumer privacy, artificial intelligence, and the marketing mix. Technology and data handling are crucial for improving the effectiveness of marketing analytics with a customized approach. This research used an Amazon review dataset to investigate the capability of ML models for sentiment classification online. The accuracy and f1-score (89%), recall (88%), and precision (88%) of BERT were better than those of SVM, GB, and the other models examined. However, limitations such as dataset scope, high computational requirements, class imbalance, and limited sentiment context awareness were identified. To enhance sentiment analysis, future research should explore advanced transformer models like RoBERTa and XLNet, real-time classification, cross-platform generalization, multi-modal sentiment analysis, and improved interpretability techniques such as SHAP. Addressing these challenges will further optimize AI-driven marketing strategies, enabling e-commerce businesses to better analyze customer feedback and make data-driven decisions [18].

## References

1. Nawi NBC, Mamun AA (2017) Customer Satisfaction of Online Apparel Businesses in Malaysia: Point-Purchase and Post-Purchase Comparison. Int J Bus Innov Res 12: 386.
2. Ibrahim NF, Wang X (2019) Decoding the Sentiment Dynamics of Online Retailing Customers: Time Series Analysis of Social Media. Comput Human Behav 96: 32-45.
3. Delen D, Ram S (2018) Research Challenges and Opportunities in Business Analytics. J Bus Anal 1: 2-12.
4. Haque TU, Saber NN, Shah FM (2018) Sentiment analysis on large scale Amazon product reviews. IEEE International Conference on Innovative Research and Development https://ieeexplore.ieee.org/document/8376299.
5. Wedel M, Kannan PK (2016) Marketing analytics for data-rich environments. J Mark 80.
6. Akkus IE, Chen R, Hardt M, Francis P, Gehrke J (2012) Non-tracking web analytics. Proceedings of the 2012 ACM conference on Computer and communications security 687-698.
7. Kakatkar C, Spann M (2019) Marketing analytics using anonymized and fragmented tracking data. Int. J. Res. Mark 36: 117-136.
8. Mondragon G, Mendez K, Mauricio D, Diaz E (2019) Evaluation model of the digital experience in the retail sector using customer journey. Proceedings of the 2019 IEEE 26th International Conference on Electronics, Electrical Engineering and Computing, INTERCON https://ieeexplore.ieee.org/document/8853635.
9. Riantini RE (2019) The Effect of Omni Channel Marketing on the Online Search Behavior of Jakarta Retail Consumers with Theory of Planned Behavior (TPB) Approach. Proceedings of 2019 International Conference on Information Management and Technology, ICIMTech https://www.researchgate.net/publication/335946104_The_Effect_of_Omni_Channel_Marketing_on_the_Online_Search_Behavior_of_Jakarta_Retail_Consumers_with_Theory_of_Planned_Behavior_TPB_Approach.
10. Ngamjarussrivichai P, Jeemali P, Panitsettakorn W (2018) Designing an online marketing investment evaluation system case study of cosmetics manufacturing company. Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business and Social Science, ICBIR https://ieeexplore.ieee.org/document/8391243.
11. Cheng X, Jiang X (2018) Online Retailing Promotion Strategy under Platform Scheme with Considering Seller Competition. 15th International Conference on Service Systems and Service Management, ICSSSM 2018 https://ieeexplore.ieee.org/document/8465033.
12. Lei Q, Guo J, Liang C (2017) The Impact of e-Commerce Development Level on Location Choice of Foreign Retail Companies in China. Proceedings - 13th IEEE International Conference on E-Business Engineering, ICEBE 2016 - Including 12th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC.
13. Raulji JK, Saini JR (2016) Stop-Word Removal Algorithm and its Implementation for Sanskrit Language. Int J Comput Appl.
14. Kadhim AI (2019) Term Weighting for Feature Extraction on Twitter: A Comparison between BM25 and TF-IDF. International Conference on Advanced Science and Engineering, ICOASE.
15. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.
16. Shrestha N, Nasoz F (2019) Deep Learning Sentiment Analysis of Amazon.Com Reviews and Ratings. Int J Soft Comput Artif Intell Appl.
17. Veluchamy A, Nguyen H, Diop ML, Iqbal R (2018) Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. SMU Data Sci. Rev 1: 1-22.
18. Routhu K, Bodepudi V, Jha KM, Chinta PCR (2020) A Deep Learning Architectures for Enhancing Cyber Security Protocols in Big Data Integrated ERP Systems. Available at SSRN 5102662.