Journal of Global Research in Electronics and Communication
Volume 2, No. 01, January 2026
Available Online at: www.jgrec.info
RESEARCH PAPER

Enhanced Sentiment Analysis on Online Amazon Reviews Using RoBERTa with PSO-Based Hyperparameter Tuning

Saket Kotpalliwar
Research Scholar
Department of Computer Science and Engineering
Kalinga University
Raipur
Saketkotpalliwar94@gmail.com

Dr. Priya Vij
Assistant Professor
Computer Science and Engineering
ku.priyavij@kalingauniversity.ac.in

***Abstract*—Sentiment analysis stands vital for businesses that want to deliver superior customer satisfaction by optimizing their strategies. Rising dependence on digital platforms along with e-commerce leads consumers to use online reviews for their purchase choices. This work focuses on the customer product reviews and incorporates classical ML approaches, as well as DL models using the transformer coupling with optimization methods. The study creates a better sentiment analysis algorithm that uses RoBERTa to read product reviews on Amazon in Kaggle which has 34,000 records with rating information and user comments. The strategy to carry out this project necessitates an ample data preparation cycle as well as pre-lemmatization prior to the TF-IDF features operations. The dataset is organized into a training and testing partition approach and RoBERTa performs classification since it is a powerful self-attention mechanism and is pretrained using masked language modeling. Hyperparameter tuning that is implemented with Particle Swarm Optimization (PSO) is an advantage of the model optimization. Both confusion matrix and epoch-based performance plots display 85.31% accuracy and 92.26% precision and 85.31% recall and F-1 values. The results show that whereas RF, LR, Naive Bayes, and DT all perform poorly, RoBERTa beats the other fundamental models in terms of sentiment classification accuracy and F1 Score value. The evaluation finds that RoBERTa excels at generalizing and predicting sentiment which positions it as an effective model for sentiment analysis applications in practical use.**

Keywords—NLP, Sentiment analysis, TF-IDF, Particle Swarm Optimization (PSO), Amazon Product Reviews, deep learning.

I. INTRODUCTION

There will be 2.64 billion digital purchasers in 2023, according to recent study, indicating that online buying is expanding in popularity [1]. An explosion in user-generated content has been a major consequence of the proliferation of online spaces where people may freely share their thoughts and ideas via mediums like blogs, social media, online marketplaces, and message boards [2][3]. There is now an abundance of user-generated data due to this development [4][5]. Important information may be gleaned from these datasets, and individuals, groups, and governments must make use of it. Computational linguistic techniques are necessary for data analysis because to the increasing amount of these data, which makes it difficult to effectively and quickly gather important information [6][7]. Consumer evaluations are common when individuals make their choices during online shopping. Reviews may teach a lot about the quality of a service or a product, and the customer satisfaction of its users [8]. Positive evaluations enable companies to acquire new consumers and credibility. The data given to the businesses by the negative reviews can be utilized to improve their products and services. To summarize, customer feedback can be of great benefit both to the companies and to the customers [9]. Businesses can leverage it to enhance their products and services, and consumers can leverage it to make smart purchases [10]. This assists in monitoring and developing the image of a brand through digital channels and in making more effective decisions in general [11][12]. Sentiment analysis finds extensive use in many domains, such as marketing, customer service, and product development, because of the crucial role it plays in comprehending client views and responses [13][14][15].

Effective sentiment analysis methods enable businesses to proactively identify and manage such issues, which improves its brand image [16][17]. This study examines the practical applications of sentiment analysis for product-based businesses, emphasizing its value in customer relationship management via the analysis of product evaluations and feedback and the consequent guidance of strategic product enhancements [18][19]. Sentiment analysis AI-driven is becoming an important business instrument. It helps in comprehending the minds of the customers [20][21][22]. Improved scalability and reliability in automated sentiment analysis have resulted from developments in ML and NLP [23][24]. Deep learning approaches have gained a lot of traction and are now realistic for achieving acceptable accuracy [25][26]. A deep learning approach to review categorization exemplifies the general sanitization of consumer evaluations according to their accurate categorization into positive and negative feelings throughout this research [27][28][29].

Their primary objective is to develop a robust SA model capable of using predefined sentiment dimensions to identify positive, negative, or neutral product evaluations. To get there, it cleaned up the data thoroughly, compiled a dataset of Amazon product reviews, and got the text data ready for analysis. Tokenization and vectorization approaches, such TF-IDF, are part of their approach to feature extraction using a set of NLP-based algorithms. In addition, it has experimented with optimization methods and sophisticated transformer-based models based on DL, such as RoBERTs. Through this comprehensive process that ensures that a reliable and precise

SA system is created, businesses can enhance their products, make superior decisions and ultimately gain customer satisfaction. The research can significantly enhance the ability to analyze the sentiment of marketplaces on the Internet. This is important because understanding consumer mood improves customer service, marketing tactics, and product development. This study can help businesses since it contributes to their knowledge on the way they can locate good sentiment analysis models, which further permits us to make evidence-based decisions based on consumer input.

A. Motivation and Research contribution

The use of e-commerce has led to an explosion in consumer preference and business reputation as customers assess companies and their offerings during evaluations. However, analyzing large-scale textual reviews manually is inefficient and impractical. The common approaches to SA include two well-known techniques of sentiment analysis, contextual and lexicon-based, which have certain drawbacks due to their incapability to come across entities in line with contextual qualities and intricate linguistic profiling. Research requires an advanced sentiment classification model which successfully analyzes customer sentiments with high precision in processing and interpretation. The implementation of DL methods utilizes RoBERTa model technology.

- The research adopts RoBERTa as its state-of-the-art transformer-based framework to enhance sentiment classification accuracy within Amazon product reviews.
- The results of proper text preprocessing and feature extraction from the dataset followed by removal of stop words contribute to higher quality of input for the model.
- The PSO algorithm performs hyperparameter optimization of RoBERTa to enhance its efficiency and achieve better results.
- The investigation explains that RoBERTa outperforms traditional ML models RF and DT by demonstrating better performance in the experiments.
- The study results provide vital insights for the organization to understand consumer sentiment and enhance product quality and customer service.

B. Novelty and justification

The novel feature of the research is the introduction of Particle Swarm Optimization (PSO) to hyperparameter adjustment to the model of the RoBERTa transformer. This will enhance sentiment analysis on a huge amount of Amazon reviews. Unlike the traditional techniques, which involve manual selection of the features and fairly straightforward optimization tools, the current study utilizes sophisticated language representation learning and uses the metaheuristic strategy produced by the nature-inspired optimization algorithm. RoBERTa is used because it features a self-attention mechanism and PSO to fine-tune the model to achieve high performance, which makes it very adept at collecting semantic connections. This strategy demonstrates an enhanced improvement to the standard models that demonstrates its value and advantages to the evolution of SA in e-commerce.

C. Structure of the paper

The paper is organized as follows: Current sentiment analysis techniques are reviewed in Section II. Section III describes how to use RoBERTa with PSO for data preparation, feature extraction, and classification. The experimental results are presented and discussed in Section IV, and conclusions and recommendations for the future are provided in Section V.

II. Literature Review

An examination of the literature on transformer-based sentiment analysis models, including the DL and ML models that provide accurate classifications and forecasts, is presented in this section.

Krishna et al. (2025) analyzed how ML is used to analyze customer Amazon feedback, in particular by first splitting the reviews into product and service reviews and then sentiment analysing each type. The findings indicate the main tendencies of customer experiences and the usefulness of ML in the context of a multi-level sentiment analysis. By the use of Naïve Bayes Classifier, the model achieves the accuracy range of 80% to 90% [30].

Sinha (2024) sought to evaluate the efficacy of several LLMs, like as BERT and TF-IDF models, in classifying customer sentiment on an Amazon product review dataset. Analysis showed that an accuracy of TF-IDF model was enhanced by 15.59% as compared to the BERT model that yielded an accuracy of only 78. 91% [31].

Vanshika, Rani and Walia (2024) suggested method provides state-of-the-art performance in a variety of sentiment analysis tasks and successfully detects sentiment nuances. Using the Flipkart and Amazon datasets, the proposed method is compared to the standard models, achieving the best accuracy rates of 90.10% and 95.40%, respectively [32].

Kumar (2023) set out to ascertain whether product category correlates with consumer sentiment. The research results highlight that Multinomial NaiveBayes and TextBlob together with Logistic Regression and SVM demonstrate potential for advancing the accuracy rates of classification activities. With 4444 good ratings and 469 bad reviews, it achieved an accuracy of 91% using SVM, 84% using Logistic Regression, and 87% using Multinomial NaiveBayes [33].

Kumar et al. (2023) designed an automated model to categorize sentiment into three classes using the BERT, LSTM and bidirectional LSTM. This is done through metrics like; acc, prec, rec and F1score. The three models show a reasonable prediction accuracy; the best is the BERT model at 91%, followed by the bidirectional LSTM model at 90.7%, and the LSTM model at 88% [34].

Badal and Parmar (2023) included a 1DLSTM in hybrid model. The results from the Movie Review and Amazon Review datasets suggest that when applied to sentiment analysis, the network model might provide high classification results. The text preprocessing stream enables applications for both text mining operations and punctuation filtering and vocabulary generation. The findings demonstrate that when compared to the basic models SVM, KNN, and MNB, the hybrid model delivers superior performance [35].

Mehul et al. (2023) provide a good or negative rating to each review based on the customer's feedback. After that, they used LSTM, RBF, and a Sigmoid kernel as part of their ML model training. The model's performance was evaluated by computing its F1score alongside precision, acc, and recall measurements. According to their LSTM model, the accuracy rate for customer reviews on Amazon is 86% and for reviews on Yelp it is 85% [36].

Table I provides details about existing research which examines ML techniques for Amazon product review SA and their success rate in customer feedback classification.

TABLE I. SUMMARY OF THE RELATED WORK FOR SENTIMENT ANALYSIS USING MACHINE LEARNING TECHNIQUES

Author	Methods	Performance	Advantage	Limitations	Future Work
Krishna et al., 2025	Naïve Bayes Classifier	80%-90% Accuracy	Demonstrates effectiveness in multi-level sentiment analysis	Limited to a single classifier	Expanding to deep learning models
Sinha, 2024	BERT, TF-IDF	TF-IDF improved accuracy by 15.59% over BERT (BERT: 78.91%)	TF-IDF outperforms BERT for Amazon reviews	BERT underperforms in this context	Exploring advanced transformer models
Vanshika, Rani, and Walia, 2024	Standard ML models vs. Proposed method	Flipkart: 90.10% Accuracy, Amazon: 95.40% Accuracy	High accuracy in sentiment classification	Specific dataset dependency	Extending the approach to multilingual data
Kumar, 2023	TextBlob, LR, SVM, MNB	SVM: 91%, Logistic Regression: 91%, MNB: 87%	Identifies correlation between sentiment and product category	Limited dataset size	Applying DL techniques
Kumar et al., 2023	BERT, LSTM, Bidirectional LSTM	BERT: 91%, LSTM: 88%, BiLSTM: 90.7%	BERT shows the highest accuracy	Requires high computational power	Testing on real-time customer feedback
Badal and Parmar, 2023	Hybrid 1D-CNN + LSTM, SVM, KNN, MNB	Hybrid model outperforms baseline models	Effective for sentiment analysis	Lacks comparative analysis with transformers	Incorporating more advanced hybrid models
Mehul et al., 2023	SVM (Linear, RBF, Sigmoid), LSTM	Amazon: 86% (LSTM), Yelp: 85%	LSTM provides robust classification	LSTM is computationally expensive	Exploring other deep learning architectures

III. METHODOLOGY

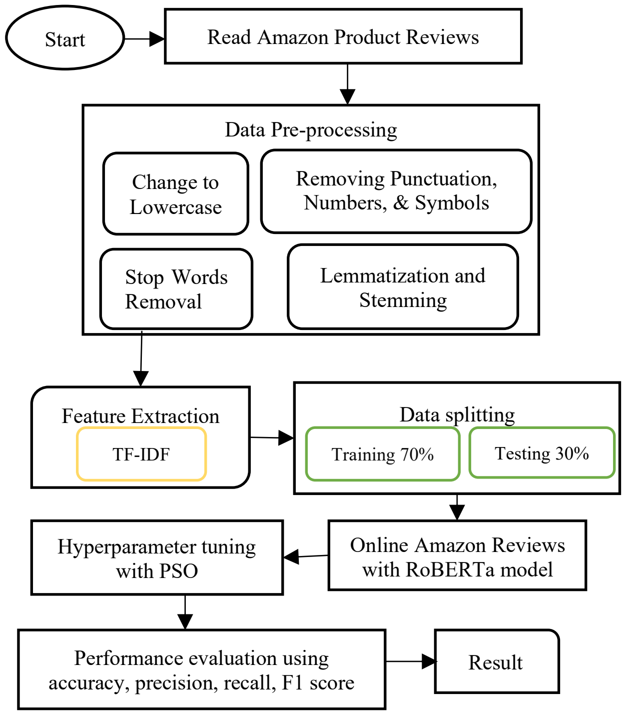

Fig. 1. Flowchart for Sentiment Analysis Using RoBERTa with PSO Optimization

The goal of SA is to computationally determine and classify the author's emotional state in a written work. The approach makes use of the Kaggle product reviews dataset for Amazon, which has more than 34,000 reviews with 21 parameters, including as star ratings and user comments. Data preparation involves lowering the case of the text, deleting digits, punctuation, and special characters, removing stop words, and normalizing the text using lemmatization and stemming. The TF-IDF is utilized for feature extraction in order to priorities words according to their relevance and frequency in the corpus. The dataset is split into 70% training and 30% testing data. The RoBERTa model, an optimized version of BERT, is employed for sentiment analysis, leveraging its self-attention mechanism and masked language modelling for pre-training. Particle Swarm Optimization (PSO) is the hyperparameter tuning algorithm, which optimizes the model parameters through a series of simulations of the model, where particles move around a search space. The effectiveness of the categorization is shown by a confusion matrix, and the model's efficacy is evaluated by means of F1score, rec, acc, and prec. Fig. 1 illustrates the subsequent implementation processes.

The proposed sentiment analysis of online review based on the advantage model based on deep learning.

A. Data Collection and Visualization

This study was done using the Kaggle.com dataset of Amazon product reviews. The dataset consists of more than 34,000 ratings of different Amazon products, such as electronics, furniture, and others. The collection also comprised a host of other information, including product ratings and reviews. The 21 available features in the dataset include product details, star ratings, customer reviews, and other attributes among others.

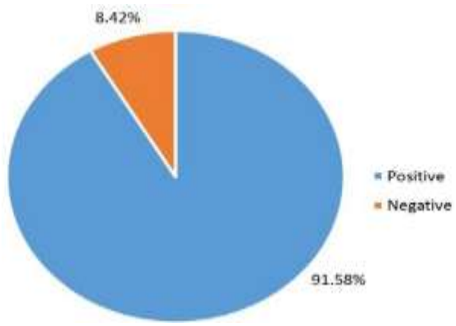

Fig. 2. Percentage of positive and negative sentiment analysis.

Fig. 2 is a pie chart describing the distribution of two categories. Positive and Negative. The category of Positive, denoted by blue color, will make up 91.58 percent of the total, whereas the Negative category, denoted by orange color, will make up 8.42 percent of the total. This graph shows that there is a strong disbalance between the two classes with the Positive one being predominant.

B. Data Preprocessing

Data preparation is an essential stage in sentiment analysis because it guarantees that the input text is in a format that the ML algorithms can analyze efficiently [37]. The proposed solution uses preprocessing to clean the raw text and remove any elements that may jeopardize the results of the sentiment analysis process. This review dataset has been preprocessed using the following methods:

1) Change to Lowercase:

The source has been made lowercase to make it standard and exclude discrepancies brought about by case sensitivity. This assists in enhancing the performance of classification and training.

2) Removal of Punctuation, Numbers, and Symbols:

It was eliminated as punctuations, numbers and special characters are not helpful in sentiment analysis. Their removal helps to decrease noise and increases the efficiency of the learning model.

3) Stop Words Removal

Stop words are low valued in sentiment analysis hence their removal is a necessary component of text preparation. The words such as a, an, and the were not used because sentiment analysis is not interested in such words. This move puts emphasis on meaningful content words that enhances the performance of the model.

4) Lemmatization and Stemming

Stemming deals with reducing words to their root forms, and lemmatization further simplifies the process by restoring words to their dictionary forms. Porter Stemmer and WordNet Lemmatizer were employed to make sure the text was normalized appropriately and improved sentiment classification accuracy.

5) Feature Extraction with TF-IDF

The feature extraction procedure in NLP is crucial for making numerical representations of unstructured text usable by ML models. The TF-IDF extraction method is used in this study. A method for retrieving information, TF-IDF takes both the term's frequency (TF) and the Inverse Document Frequency (IDF) into account. TF-IDF considers the frequency of a term in a single review in relation to its frequency in all reviews. This allows us to zero down on the most important terms for expressing emotions in every review. The term "TF-IDF" is just a clever method for extracting the most indicative terms from reviews. The following TF-IDF Equation (1):

$$TF - IDF(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (1)$$

Where:

The score for word t in document d related to the collection D of documents is denoted by TF-IDF (t, d, D).

tf(t, d) indicates the term frequency of term t in document d.

It is possible to express the IDF of word t in the document collection D as idf (t, D).

To get the Term Frequency tf (t, d), take the total word count of document d and divide it by the number of occurrences of term t. The IDF, or idf(t, D), is obtained by logarithmically dividing the total number of documents in D by the total number of documents that include the term t.

C. Data Splitting

The data is split into two parts in the ratio 7:3. One partition (70% of the data) is utilized to train a model, and a second partition (30%) is used to test its performance.

D. Online Amazon Reviews with RoBERTa Model

In this work for the sentiment analysis based on online Amazon reviews used RoBERTa model was used with PSO hyperparameter tunning. RoBERTa is an enhanced variant of BERT that removes the goal of predicting the next phrase and augments the original design with features such as increased batch size, longer training on a wider variety of data, and dynamic masking [38]. The model achieves advanced performance across natural language processing tasks by using masked language modelling for pre-training and the self-attention mechanism of the transformer architecture. It does this via its strong optimization techniques and upgraded training procedure [39]. The multi-head attention mechanism serves as the foundation for the suggested architecture's advanced attention and output calculation framework [40]. The following structure is followed by the suggested architecture for calculating output and attention. The mechanism for multi-head attention is shown by Equation (2).

$$\mathrm{MultiHead}\ (\mathrm{Q, K, V}) = \mathrm{Concat}(head_i, \dots, head_h)W^o \quad (2)$$

Equation (3) defines each unique attention head as follows:

$$head_i = Attention\left(QW_i, KW_{i,}VW_i\right) \quad (3)$$

Equation (4) describes how the attention function works:

$$Attention(Q, K, V) = softmax\left(\frac{QKT}{\sqrt{d_k}}\right)V \quad (4)$$

Q stands for the query matrix, K for the key matrix, and V for the value matrix. The QW, KW, VW and WO parameter matrices that can be learnt help in optimization of the model. This theory is connected with the dimensions of the key vectors. After applying layer normalization and a residual connection, the attention mechanism's final output is calculated as Equation (5):

$$FinalOutput = LayerNorm(MSA + FFN) \quad (5)$$

Where the results of the FFN and MSA are mixed. Equation (6) defines the masked language model goal, from which the loss function for the training procedure is obtained.

$$MaskedLanguageModelLoss = -\sum_{i \in M} log\ P(xi|\tilde{x}) \quad (6)$$

In this context, M stands for the collection of locations of masked tokens, xi for the original token, and for the corrupted input sequence. This framework guarantees successful contextualized representations learning coupled with strong training dynamics.

Hyperparameter tuning with PSO: The PSO meta-heuristic optimization algorithm is inspired by the foraging behaviour of swarms of birds. It efficiently explores the hyperparameter space by simulating the movement and interactions of particles, enabling the RoBERTa model to achieve optimal performance for analyzing Amazon reviews. Every swarm particle that travels across a search space stands in for a potential new solution. This optimization method repeatedly improves the candidate solutions [41]. Their optimal local position affects the particle transition in iteration i with respect to perfect p_{best}^{i-1} and the optimal search area location g_{best}^{i-1} in iteration i-1. Equations (7) and (8) define the velocity and position of particle t in s-dimensional optimization space as $(f_{avg}, f_{cur}, f_{min})$ and (x_{tj}^{i}), respectively.

$$v_t^i = wv_t^{i-1} + c_1 b_1^i\left(pbest_t^{i-1} - x_t^{i-1}\right) + c_2 b_2^i\left(gbest_t^{i-1} - x_t^{i-1}\right) \quad (7)$$

$$x_t^i = x_t^{i-1} + v_t^i \quad (8)$$

The neighboring particle moves closer to the optimal solution as the particle uses the best-researched position. PSO may be organized to handle feature selection issues with many dimensions. Algorithm 1 details the prepared method.

Algorithm 1: PSO algorithm

Generate the initial population
for number of iterations **do**
  for particle_i **do**
    φ_i evaluate particle_i using fitness function
    ζ_i compute value of fitness function for pbest of particle_i
    if $\varphi_i > \zeta_i$ **then** update pbest
    end if
    pbest > gbest **then** update gbest
    end if
  end for
  for each particle_i **do**
    for each dimension d **do**
      Compute new velocity according (7)
      Compute new position according (8)
    end for
  end for
end for
Return the value of gbest particle

E. Performance Evaluation

A proposed model was tested employing a number of popular performance measures, including F1measure, rec, acc, and prec. A confusion matrix produces all four measures and shows how well the DL model performed. The components of this matrix include FN, TN, TP, and FP. The following formulas are implemented to assess model performance.

Accuracy (Acc): It is a statistic for evaluating the overall accuracy of a ML model. Equation (9) is used to compute the accuracy:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

Precision (Prec): A frequency with which an ML model accurately forecasts the positive class is measured by this statistic. Equation (10) provides the calculation for the precision value.

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

Recall (Rec): It shows the fraction of positive instances (true positives) that a ML model gets right out of the whole collection of positive samples. Equation (11) is used to get the recall value.

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

F1 Score: The F1score is a helpful metric when rec and prec are equally significant. Good recall and precision are indicated by a high F1score. Equation (12) is used to compute the F-score value.

$$F1 - Score = 2 \times \left[\frac{(Recall*Precision)}{(Recall+Precision)}\right] \quad (12)$$

The proposed DL model for SA in online product evaluations is evaluated using these metrics to see how well it works with hyperparameter adjustment.

IV. Result Analysis and Discussion

The suggested RoBERTa model for sentiment classification is very promising, according to the results, for identifying favorable and unfavorable opinions expressed in the provided Amazon product reviews. The tests were conducted using Python as the main programming language and Google Colab as the IDE on a machine with an 8th-generation Intel Core i7 CPU, 8 GB of RAM, and GPU acceleration. Pandas, NumPy, transformers, TensorFlow, scikit-learn, and evaluation data were among the libraries used for data preparation, feature extraction, training, and evaluations. The outcomes shown in Table II shows that RoBERTa achieved superior performance than other models based on F1score, rec, acc, prec metrics.

TABLE II. Results of the RoBERTa for Amazon Reviews

Performance Matrix	RoBERTa
Accuracy	85.31
Precision	92.26
Recall	85.31
F1 Score	88.43

Table II shows the RoBERTa model's performance metrics for Enhanced Sentiment Analysis of Online Amazon Reviews. The model's accuracy rating of 85.31% demonstrates achievement in sentiment recognition. A high percentage of accurately detected positive evaluations among all anticipated positives is shown in the precision of 92.26%. The model's ability to identify genuine favorable reviews is shown by its 85.31% recall. The exceptional prediction capabilities of the model emerge from its F1 Score of 88.43% which maintains a balanced combination between accuracy and recall and achieves an acceptable ratio of false positives and negatives.

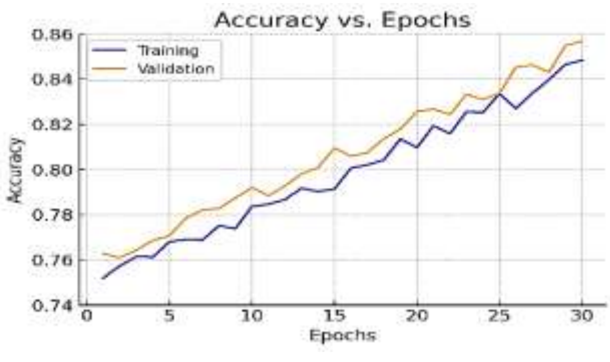

Fig. 3. Accuracy vs. Epochs for RoBERTa Model on Amazon Reviews

Fig. 3 presents the accuracy trends over 30 epochs for the RoBERTa model on Online Amazon Reviews. The accuracy curves exist individually for training set data (blue) and validation set data (orange) in this graph. Both curves indicate that there is a consistent increase in accuracy with the validation accuracy typically exceeding the training accuracy with the epoch. This trend points to high-quality learning and good generalization, and a small quantity of overfitting.

Fig. 4 shows the trend of losses in 30 epochs of the RoBERTa model in Enhanced Sentiment Analysis of Online Amazon Reviews. The training (blue) and validation (orange) data loss curves are shown separately in the graph. The curves are similar, as they both demonstrate a constant decrease, which means the optimality of the model. A strong indicator of the training model's generalizability and lack of overfitting is the fact that the validation loss is consistently lower than the training loss value. The reduction in the loss values attest to the fact that the model advances the sentiment classification performance across epochs.

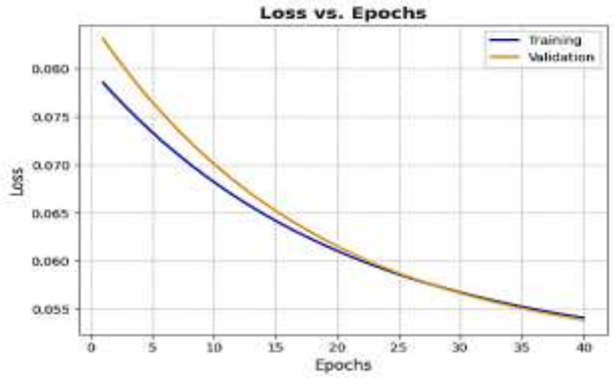

Fig. 4. Loss vs. Epochs for RoBERTa Model on Amazon Reviews

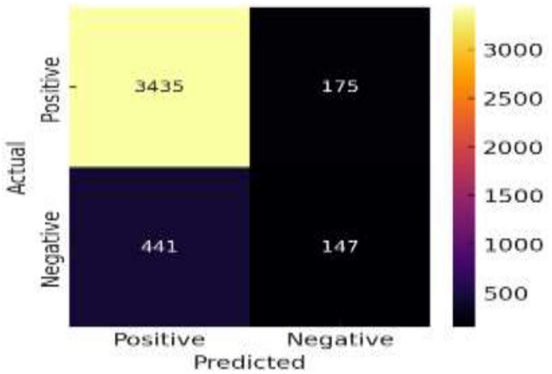

Fig. 5. Confusion Matrix of RoBERTa Model on Amazon Reviews

Fig. 5 is a Confusion Matrix of an Online Amazon Reviews (RoBERTa model) that illustrates its ability to categorize reviews as either Positive or Negative. The system accurately recognized 3435 positive and 147 negative reviews. It however misclassifies 441 negatives as positives and 175 positives as negatives. The results demonstrate that the performance is strong, but the existence of misclassifications implies that precision and recall can be improved.

A. Comparison and Discussion

Table III compares the proposed model, RoBERTa, to four baseline models (LR, NB, DT, and RF) in terms of vital performance metrics (Acc, Prec, Rec, and F1Score).The proposed model, RoBERTa, is found to be much better than the other baseline models (LR, NB, DT, and RF), with a high level of accuracy (85.31%), precision (92.26%), recall (85.31%), and F1 Score (88.43%). Logistic Regression performs better than the other baselines with 83.1% accuracy and F1 Score of 84.9. NB performs moderately. Reasonably low scores of Decision Tree and Random Forest, especially based on precision and F1Score, underline the extent to which the proposed RoBERTa model copes with the classification problem.

TABLE III. COMPARISON RESULTS ON THE AMAZON REVIEWS

Model	Accuracy	Precision	Recall	F1 Score
RoBERTa	85.31	92.26	85.31	88.43
Logistic Regression [42]	83.1	87.7	83.2	84.9
NaiveBayes [43]	80.12	79.32	73.57	69.32
Decision Tree[44]	78	62	57	60
Random forest[45]	67.9	59.3	67.9	59.6

The proposed RoBERTa model is accuracy-rich in performing sentiment analysis activities when analyzed on Amazon reviews. The transformer-based design with self-attention mechanism proves effective in contextual emotion collection through the reported findings. RoBERTa delivers superior performance than baseline models over LR, NB, DT and RF in all scoring dimensions. The model obtains better results through the combination of TF-IDF features and PSO hyperparameter optimization. RoBERTa delivers stronger capabilities for language processing and optimization which transforms it into a dependable tool for sentiment analysis tasks.

V. CONCLUSION AND FUTURE WORK

This study implemented RoBERTa along with PSO for tuning hyperparameters to create an effective system that analyzes Amazon reviews sentiment. The deep contextual relationships in user evaluations were successfully analyzed through RoBERTa's model because it utilizes transformer design alongside masked language modeling and multi-head self-attention. The experimental outcome showed that the model surpassed classical models by attaining 85.31% accuracy and 88.43% F1-Score as well as 85.31% recall and 92.26% precision. PSO improved the hyperparameter tuning part that helped in achieving better converge and generalization of the model. The dataset contains class imbalance that makes it difficult for the model to achieve generalization across sentiment categories due to its existing limitations. The performance could have been improved by applying more advanced data balancing techniques such as SMOTE or ADASYN although the study used TF-IDF techniques. Furthermore, although the RoBERTa model was superior to conventional ML models, future research can also focus on using the hybrid model of ML, DL (for instance, CNN, LSTM), and transformer for higher accuracy and better model interpretability. The use of ensemble methods or XAI will also be a part of future efforts to improve sentiment analysis system trustworthiness by making predictions easier to interpret.

REFERENCES

[1] S. S. S. Neeli, "Ensuring Data Quality: A Critical Aspect of Business Intelligence Success," *Int. J. Lead. Res. Publ.*, vol. 2, no. 9, p. 7, 2021.

[2] V. Verma, "LSTM-based to Predicting Stock Market Trends Using Machine Learning and Sentiment Analysis," vol. 8, no. 3, pp. 102–111, 2021.

[3] S. B. Shah, "Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.

[4] N. Raghunathan and K. Saravanakumar, "Challenges and Issues in Sentiment Analysis: A Comprehensive Survey," *IEEE Access*, vol.

11, pp. 69626–69642, 2023, doi: 10.1109/ACCESS.2023.3293041.

[5] J. K. Chaudhary, S. Tyagi, H. P. Sharma, S. V. Akram, D. R. Sisodia, and D. Kapila, "Machine Learning Model-Based Financial Market Sentiment Prediction and Application," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, May 2023, pp. 1456–1459. doi: 10.1109/ICACITE57410.2023.10183344.

[6] V. Dogra *et al.*, "A Complete Process of Text Classification System Using State-of-the-Art NLP Models," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–26, Jun. 2022, doi: 10.1155/2022/1883698.

[7] V. S. Thokala, "Scalable Cloud Deployment and Automation for E-Commerce Platforms Using AWS, Heroku, and Ruby on Rails," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 2, Oct. 2023, doi: 10.48175/IJARSCT-13555A.

[8] S. S. Synam Neeli, "Amazon Redshift Performance Tuning for Large-Scale Analytics," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 1, no. 4, pp. 2130–2134, Nov. 2023, doi: 10.51219/JAIMLD/sethu-sesha-synam-neeli/467.

[9] C. Ahmed, A. ElKorany, and E. ElSayed, "Prediction of customer's perception in social networks by integrating sentiment analysis and machine learning," *J. Intell. Inf. Syst.*, 2023, doi: 10.1007/s10844-022-00756-y.

[10] S. Arora and S. R. Thota, "Using Artificial Intelligence with Big Data Analytics for Targeted Marketing Campaigns," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 3, pp. 593–602, 2024, doi: 10.48175/IJARSCT-18967.

[11] K. M. R. Seetharaman, "Digital Transformation in Retail Sales: Analyzing the Impact of Omni-Channel Strategies on Customer Engagement," *J. Glob. Res. Math. Arch.*, vol. 10, no. 12, 2023, doi: 10.5281/zenodo.15280578.

[12] S. Singh, S. A. Pahune, P. Chatterjee, and R. Sura, "Advanced Machine Learning Methods for Churn Prediction and Classification in Telecom Sector," in *2025 IEEE 6th India Council International Subsections Conference (INDISCON)*, 2025, pp. 1–7. doi: 10.1109/INDISCON66021.2025.11252233.

[13] P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM," *Big Data Cogn. Comput.*, vol. 8, no. 12, 2024, doi: 10.3390/bdcc8120199.

[14] Y. Li, Z. S. Xu, and X. X. Wang, "Methods and applications of sentiment analysis with online reviews," 2023. doi: 10.13195/j.kzyjc.2022.1788.

[15] H. P. Kapadia and K. B. Thakkar, "Generative AI for Real-Time Customer Support Content Creation," *J. Emerg. Technol. Innov. Res.*, vol. 10, no. 12, pp. 36–43, 2023.

[16] S. J and K. U, "Sentiment analysis of amazon user reviews using a hybrid approach," *Meas. Sensors*, vol. 27, no. May, p. 100790, Jun. 2023, doi: 10.1016/j.measen.2023.100790.

[17] S. B. V. Naga, S. Thangavel, S. K. Kuchoor, N. Narukulla, and L. K. Yenduri, "Optimizing Online Marketing Strategies With Machine Learning and Deep Learning Innovations," in *Impact of Digital Transformation on Business Growth and Performance*, 2025, pp. 483–512. doi: 10.4018/979-8-3693-9783-1.ch018.

[18] S. Pandya, "Comparative Analysis of Large Language Models and Traditional Methods for Sentiment Analysis of Tweets Dataset," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 12, 2024, doi: 10.5281/zenodo.14575886.

[19] S. Pahune and M. Chandrasekharan, "Several Categories of Large Language Models (LLMs): A Short Survey," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 7, pp. 615–633, 2023, doi: 10.22214/ijraset.2023.54677.

[20] S. Nokhwal, P. Chilakalapudi, P. Donekal, S. Nokhwal, S. Pahune, and A. Chaudhary, "Accelerating Neural Network Training: A Brief Review," in *8th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (2024)*, 2024.

[21] P. Choudhary, R. Choudhary, and S. Garaga, "Enhancing Training by Incorporating ChatGPT in Learning Modules : An Exploration of Benefits , Challenges , and Best Practices," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 11, 2024, doi: 10.5281/zenodo.14272874.

[22] H. Xia, Y. Yang, X. Pan, Z. Zhang, and W. An, "Sentiment analysis for online reviews using conditional random fields and support vector machines," *Electron. Commer. Res.*, 2020, doi: 10.1007/s10660-019-09354-7.

[23] S. Rashiq Nazar and T. Bhattasali, "Sentiment Analysis Of Customer Reviews," *Azerbaijan J. High Perform. Comput.*, 2021, doi: 10.32010/26166127.2021.4.1.113.125.

[24] J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[25] M. E. Alzahrani, T. H. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3840071.

[26] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, "Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews," *Math. Probl. Eng.*, vol. 2021, pp. 1–10, May 2021, doi: 10.1155/2021/5536560.

[27] T. Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," *Int. J. Sci. Res. Arch.*, vol. 15, no. 3, pp. 618–627, Jun. 2025, doi: 10.30574/ijsra.2025.15.3.1767.

[28] C. Patel, "Generative AI for Personalized Marketing and Customer Experience in E-Commerce," *Int. J. Emerg. Res. Eng. Technol.*, vol. 7, no. 1, pp. 12–19, 2026, doi: 10.63282/3050-922X.IJERET-V7I1P103.

[29] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[30] J. Krishna, A. L. Narayana, S. M. K. Juber, Y. M. Yadav, G. S. Karthik, and V. P. Kumar, "Analyzing Customer Feedback on Amazon to Reveal Sentiment Trends using Machine Learning," in *2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, IEEE, Feb. 2025, pp. 1–7. doi: 10.1109/ICCECE61355.2025.10941446.

[31] H. Sinha, "A Sentiment Analysis of Product Review Data Through Large Language Models (LLMs)," in *2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/IC-SIT63503.2024.10862104.

[32] Vanshika, N. Rani, and R. Walia, "HSA: A Hybrid Sentiment Analysis Technique using Machine Learning," in *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*, Jun. 2024, pp. 1–6. doi: 10.1109/OTCON60325.2024.10687412.

[33] A. Kumar, "A Machine Learning-based Automated Approach for Mining Customer Opinion," in *2023 4th International Conference on Electronics and Sustainable Communication Systems, ICESC 2023 - Proceedings*, 2023. doi: 10.1109/ICESC57686.2023.10193256.

[34] R. P. Kumar, G. Bharathi Mohan, R. Elakkiya, M. Charan Kumar, and M. Rithani, "Automated Sentiment Classification of Amazon Product Reviews using LSTM and Bidirectional LSTM," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques, EASCT 2023*, 2023. doi: 10.1109/EASCT59475.2023.10393514.

[35] A. Badal and M. Parmar, "Sentiment Analysis Using Deep Neural Network 1D Convolutional with Long Short Term Memory," in *2023 2nd International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2023*, 2023. doi: 10.1109/ICEEICT56924.2023.10157420.

[36] A. R. Mehul, S. M. Mahmood, T. Tabassum, and P. Chakraborty, "Sentiment Polarity Detection Using Machine Learning and Deep Learning," in *3rd International Conference on Electrical, Computer and Communication Engineering, ECCE 2023*, 2023. doi: 10.1109/ECCE57851.2023.10101494.

[37] S. Murri, M. Bhoyar, G. P. Selvarajan, and M. Malaga, "Transforming Decision-Making with Big Data Analytics: Advanced Approaches to Real-Time Insights, Predictive Modeling, and Scalable Data Integration," *Int. J. Commun. Networks Inf. Secur.*, vol. 16, no. 5, 2024.

[38] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019.

[39] P. Pookduang, R. Klangbunrueang, W. Chansanam, and T. Lunrasri, "Advancing Sentiment Analysis: Evaluating RoBERTa against Traditional and Deep Learning Models," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, pp. 20167–20174, Feb. 2025, doi: 10.48084/etasr.9703.

[40] Z. Sun, H. Yu, X. Song, R. Liu, Y. Yang, and D. Zhou, "MobileBERT: a Compact Task-Agnostic BERT for Resource-Limited Devices," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 2158–2170. doi: 10.18653/v1/2020.acl-main.195.

[41] D. Tiwari *et al.*, "SPSO-EFVM: A Particle Swarm Optimization-Based Ensemble Fusion Voting Model for Sentence-Level Sentiment Analysis," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3363158.

[42] A. S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews," *Int. J. Comput. Sci. Inf. Technol.*, vol. 13, no. 3, pp. 15–30, Jun. 2021, doi: 10.5121/ijcsit.2021.13302.

[43] A. Purohit, "Product Based Sentiment Analysis Based on Deep Learning for Amazon Review," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 9, pp. 68–77, Sep. 2024, doi: 10.22214/ijraset.2024.64089.

[44] R. S. Monsoor *et al.*, "Sentiment Analysis of Amazon Reviews Using Machine Learning Classifier," in *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*, 2023. doi: 10.1109/ICCIT60459.2023.10441259.

[45] H. Ali, E. Hashmi, S. Yayilgan Yildirim, and S. Shaikh, "Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques," *Electronics*, vol. 13, no. 7, 2024, doi: 10.3390/electronics13071305.