International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Sentiment Classification Accuracy of Amazon Product Reviews via NLP Approaches

Kirtika

Submitted: 10/10/2024 **Revised:** 22/11/2024 **Accepted:** 05/12/2024

Abstract—An indispensable tool for businesses, sentiment analysis sifts through customer reviews on e-commerce platforms to reveal vital information into product quality and consumer satisfaction. Using a massive Amazon review dataset with over 568,000 entries over 10 characteristics, this work proposes a strong deep learning method to sentiment classification. Cleaning, tokenisation using a 10,000-word vocabulary and padding are all part of the text data's extensive preprocessing that guarantees consistent input for the models. The majority of evaluations are favourable, showing that customers are generally satisfied, according to the exploratory data analysis. To understand the reviews' sequential relationships and contextual subtleties, we suggest a mixed-layer deep learning model that combines LSTM and GRU layers, with the addition of embedding and dropout techniques. With an accuracy of 96.5% after 100 epochs of training, the model surpasses both standalone GRU models and leading techniques in the past that used topic models and embeddings. Loss, F1-score, recall and accuracy are some of the evaluation indicators that back up the model's efficacy. In e-commerce review analysis, the results show that scalable sentiment classification using a combination of LSTM and GRU architectures with thorough preprocessing is possible.

Keywords- *Hybrid LSTM-GRU, Sentiment Classification, Product Reviews, Text Preprocessing, Accuracy*

I. INTRODUCTION

To better comprehend consumer feedback, improve product offers and increase customer satisfaction, sentiment classifying product reviews is essential. With millions of reviews posted every day on e-commerce sites like Amazon, the capacity to automatically sort evaluations into positive, negative and neutral buckets is priceless.[1]–[6]. Semantic analysis, a subfield of A tool that is helping with this effort is natural language processing, which uses computational methods to decipher the feelings expressed in text. Examining the efficacy and efficiency of various approaches, this study delves into the topic of sentiment classification for Amazon product evaluations utilising advanced natural language processing algorithms. Reviews on Amazon provide a wealth of information, showcasing all types of customer experiences with different products. Buyers and sellers alike might benefit from the insights offered by these reviews. Customers' emotional investment in reviews has the power to sway their purchase decisions and companies can utilise this data to spot patterns in customer opinion consumer contentment, problem areas and comments regarding individual products

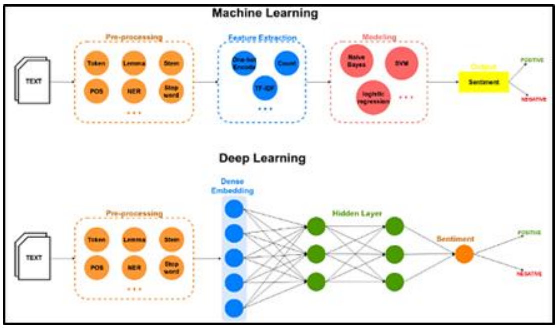

Fig. 1 Sentiment Classification

Manual analysis is not feasible due to the large amount of data. Here is where natural language processing (NLP) sentiment analysis is crucial. By using an automated system to classify reviews, businesses may instantly enhance their products and services. It is common practise to incorporate steps like data pretreatment, feature extraction and machine learning model deployment into sentiment categorisation. As a first step in being ready to use machine learning models, cleaning and establishing the raw text input is essential. An integral aspect of this procedure is the reduction of words to their fundamental forms and the removal of noise, such as punctuation, numbers and stopwords. Two

Research Scholar, Dept. of Computer Science ,Om Sterling Global University Hisar 125001

approaches that can be useful in this regard are stemming and lemmatisation. Following preprocessing and feature retrieval, machine learning algorithms are able to comprehend the text. Word2Vec but GloVe are examples of well-known word embeddings, whereas TF-IDF is a well-known feature extraction approach. Review sentiment is then classified using machine learning models, after feature extraction. Logistic Regression, Naïve Bayes and supporting vector machines (SVM) are a few of the most well-known classical machine learning approaches to sentiment categorisation problems. These models can predict the review's positivity, negativity, or neutrality using the retrieved features.

[7]–[12]. However, more recent developments in deep learning have made it possible to use more sophisticated models, such as RNNs and Convolutional Neural Networks (CNNs), to solve sentiment classification tasks. These models provide greater accuracy. With the help of pre-trained word insertions, these models can grasp the subtleties and context of human language with more ease. Bidirectional Encoding Models from Transformers (BERT) and other transformer-based models have completely transformed sentiment analysis in the modern era by capturing intricate textual context. Identifying review sentiment is where BERT really shines due to its reversible nature. This capacity allows it to take into account the context of a word by looking at the terms immediately before and after it. Because it can infer sentiment from context rather than having it expressed explicitly, BERT is ideal for classifying Amazon reviews. Complex language features like idioms, irony, or domain- particular jargon provide a significant obstacle to accurate sentiment classification. The usage of emoticons, acronyms and slang by customers in product reviews adds another layer of complexity to the investigation. One kind of review is the "mixed feelings" type, in which customers praise and criticise the same service or product. When a consumer is happy with the items but unhappy with the delivery, it is hard to put a label on the review. Some novel approaches to these problems have emerged from the field of natural language processing (NLP), such as attention mechanisms or transformer models, which are able to identify intricate patterns in text. This study aims to examine the effectiveness of cutting-edge deep learning algorithms in detecting consumer sentiment in Amazon reviews, as well as more conventional machine learning models. The purpose of this study is to determine the best methods for emotion classification by comparing the results of various models trained on the Amazon review dataset. Businesses can benefit from consumer feedback by making better decisions with the use of sentiment analysis tools, which this study will add to the growing body of knowledge about. Social media monitoring, customer service analytics and other areas that rely on sentiment analysis in massive text datasets can gain valuable insights from this work.[13]–[17].

II. Literature Review

Gupta 2024 et al. utilises techniques from web scraping and natural language processing to construct an assessment system for product sentiment. They use a wide variety of models, including SVC, GRU, LSTM, Naive Bayes, quantitative regression, the random forest, KNN and countless more. No other models could compete with LSTM's 90% F-1 score or GRU's 91% F-1 score. Despite the effectiveness of logistic regression and assistance vector machines, KNN is sluggish and inaccurate. The results help choose effective models for applications like sentiment analysis based on URLs.[18].

Sarraf 2024 et al. focusses on enhancing the basic dataset with new data using preprocessing techniques including text purification, stop keyword removal, lemmatisation and stems in order to assess Amazon food reviews.
The study group employed TF-Inverse Doc Rate (TF-IDF), Word-to-V (W2V), or Bag for Writers (BoW) to construct ML models. We were able to build and refine several models by utilising decisions trees, logistic regression, etc. After hyperparameter tweaking, Logistic Regression using BoW features had the highest accuracy of 89%. The study's main points centre on how efficient feature extraction methods are and how increased data amount affects model correctness.[19].

Shaik 2024 et al. We utilise the BERT or T5 models to build a prediction workflow that will search review data for sentiment and attributes, with an emphasis on ethical product development. These algorithms can sort reviews into three categories: positive, negative and neutral, after training on both artificially created and manually tagged datasets. After factoring in aspect detection, BERT outperformed T5's 91% accuracy. Out of all the measures used for evaluation, the model with the highest compute efficiency, recall and accuracy was BERT. The BERT model gives useful research founded on user feedback, which allows product designers to meet consumer expectations.[20].

Shobayo 2024 et al. analyses how well Google's Pathways Language Engine (Google PaLM) deciphers the feelings conveyed in Amazon's fashion reviews. While traditional natural language processing techniques like BERT and VADER accomplish their goals, they fall short when confronted with linguistic subtleties like context and irony. At Google, we used the PaLM, VADER and BERT sentiment analysis tools. We then went on to evaluate the recall, accuracy and precision of the findings. After making these adjustments, Google PaLM performed better than the competitor model with a temperature of 0.0 and an N-value of 1, resulting in accurate positive and negative predictions of 0.91 or 0.93, respectively. In terms of NLP applications, the study found that big language models (e.g., Google PaLM) performed better than traditional rule-based techniques.[21].

Yu 2024 et al. explores the use of various machine learning methods, such as Random Forest, Logistic Regression, Convolutional Neural Networks (CNNs) and Amazon Review sentiment evaluation (MRTA). Tuning parameters in line with theory or actual data is an important part of a holistic approach to provide valid model outputs. So that we can better grasp the pros and cons of these models for sentiment categorisation tasks, we conduct a comparison study that ranks them based on accuracy and other performance metrics. We can learn more about the algorithms' sentiment analysis capabilities thanks to the results, which show how well they performed. Future research into sentiment analysis using ML approaches can build on the foundations laid by this study.[22].

TABLE 1 LITERATURE SUMMARY

Authors /year	**Model/method**	**Research gap**	**Findings**
Hashmi/ 2024 [23]	BERT excels in Amazon sentiment analysis.	Lack of optimized deep learning models for nuanced sentiment analysis.	BERT achieved highest accuracy, outperforming other machine learning models.
Wang/2024 [24]	Word2Vec and SVM enhance sentiment analysis.	Limited exploration of Word2Vec combined with SVM for sentiment analysis.	Word2Vec and SVM enhance sentiment analysis accuracy and efficiency.
Shetty/2024 [25]	Grid search optimizes machine learning hyperparameters.	Limited exploration of hyperparameter optimization in sentiment analysis methods.	BoW and TF-IDF improve sentiment analysis model performance significantly.
Tabany/ 2024 [26]	SVM outperforms other models significantly.	Need for effective fake review classification methods in e-commerce.	SVM outperformed others; review length impacts sentiment analysis accuracy.

III. METHODOLOGY

Gathering data, cleaning it up, analysing it and then modelling it are all part of this method. Across 10 characteristics, the dataset comprises 568,504 reviews. We tokenise the text (10,000 words of vocabulary), clean it up and then pad it. Most reviews are positive, according to EDA. Review sentiment is classified using a hybrid deep learning model that incorporates LSTM and GRU networks, embedding, abandoning school and softmax output. Training a model over 100 iterations allows us to measure its accuracy and loss metric.

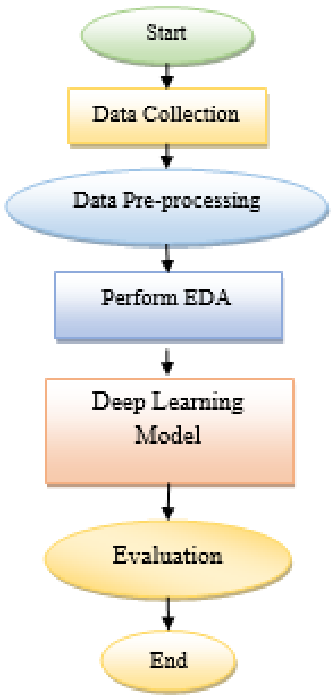

Fig. 2 Proposed Flowchart

A. Data Collection

With 568,454 rows & 10 columns, this data collection documents several aspects of e-commerce website product reviews. The review data is stored in a two-dimensional grid with columns for "Id," "ProductId," "UserId," "ProfileName," "HelpfulnessNumerator," "HelpfulnessDenominator," "Score," "Time," "Summary," and "Text." The 'ProductId' column associates reviews with specific products and the 'Id' field assigns a distinct name to every review. You can learn more about reviewers from the "UserId" and the "ProfileName" fields; the latter usually include the reviewer's real name or an alias. The 'Helpfulness Denominator' and 'Helpfulness Numerator' columns display the community's opinions on the study's usefulness, indicating the percentage of votes that considered it helpful. Reviewers often provide ratings from 1 to 5 in the "Score" column. Fields "Summary" and "Time" give quick overviews of the review's material, while "Text" has the entire review. You can find samples in the first several rows. One review gave a dog meal product a perfect score of 5 for perfection and another gave peanuts a failing grade of 1 for not living up to expectations. This type of dataset is perfect for investigating the value of reviews, consumer sentiment and rating trends in relation to various products.

B. Data Pre-processing

For natural language processing (NLP) tasks, this pretreatment pipeline cleans and standardises review data. First, it examines the sentiment distribution and gets rid of missing data. To make text lowercase and remove URLs, HTML tags, spelling and numbers, you can use the clean_text function. Stopword elimination using NLTK highlights important content. While remove_mult_spaces guarantees constant spacing, other routines handle emojis, special characters and hashtags. The cleaned text undergoes tokenisation using a 10,000-word vocabulary. Stringing the reviews onto sequences and padding each one with up to 200 words is the next stage. Machine learning projects that involve sentiment analysis and natural language processing will benefit from this well-organised dataset.

C. EDA

User ratings, review openness, review rate per user, or sentiment distribution can be better understood with the help of exploratory statistical analysis (EDA). The majority of reviews are positive, with 5 stars representing the highest level of satisfaction from buyers. Different members of the community seem to have different opinions on the value of reviews based on the uneven distribution of helpfulness ratios, which form strong clusters around 0.2 and 1. Most users only post reviews every so often because the amount of reviews drops down sharply as the count goes up. With 77.7% a yes, 14.4% negative and 7.9% neutral reviews, sentiment analysis reveals a significantly favourable skew, indicating an outstanding response. Taken together, EDA reveals patterns in review sentiment and user participation.

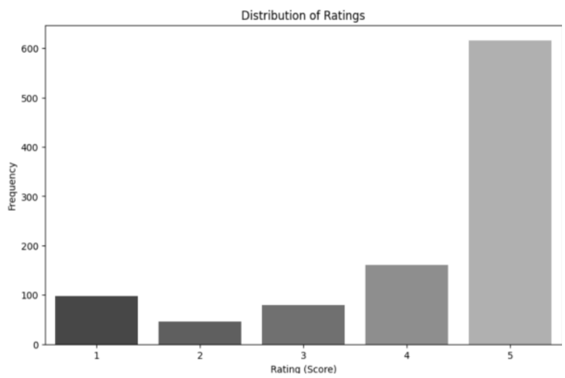

Fig. 3 Rating Score Distribution

This graph shows a clear peak at rating 4, indicating that most users are satisfied with products. It highlights a generally positive user experience, with relatively fewer low ratings and a significant number of 5-star ratings.

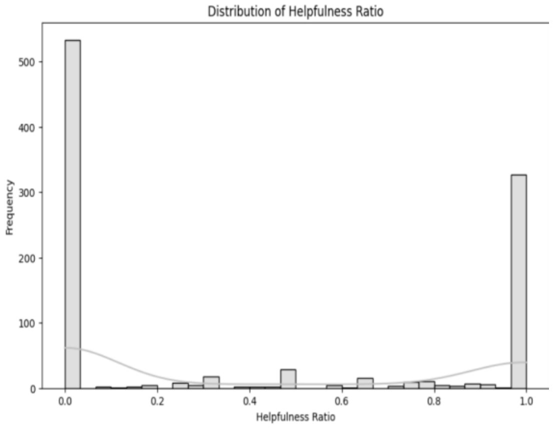

Fig. 4 Helpfulness Ratios Vary Widely

This graph displays a scattered distribution of helpfulness ratios, with notable concentrations at 0.2 and 1. It suggests that users have diverse opinions on which reviews are useful, reflecting variability in perceived review quality and community engagement.

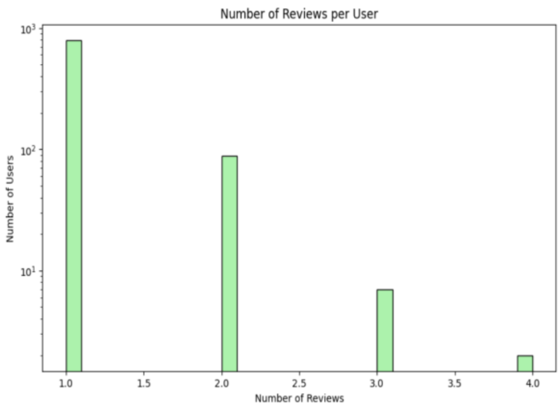

Fig. 5 Most Users Write One Review

The graph shows a steep decline in user frequency as review count increases, indicating that the majority of users contribute only a single review. This suggests low long-term engagement from most reviewers and highlights the presence of casual contributors.

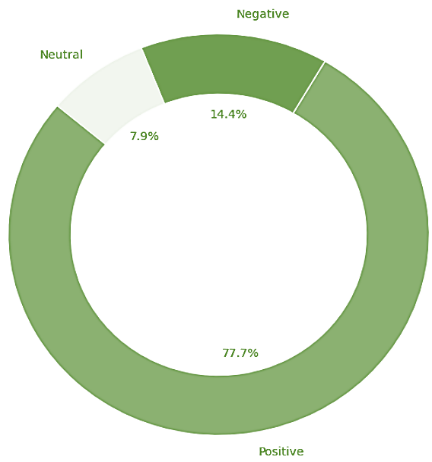

Fig. 6 Positive Reviews Dominate Sentiments Donut Graph

A majority of 77.7% are positive, while 14.4% are negative and 7.9% are neutral, as shown in the graph. Positive ratings outweigh negative ones, indicating happy customers and good product reception overall.

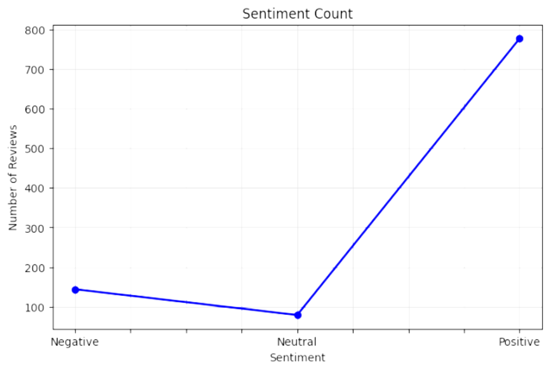

Fig. 7 Line Graph Highlights Positive Trend

Overview images display data on the dataset's usefulness ratios, rating transportation, sentiment analysis, & review counts by user. Positive reviews are evident from the high number of ratings grouped around 4 for the first figure. As we can see in the second image, which examines helpfulness ratios, the reviews cover a wide range of perceived usefulness, with little peaks at 0.2 and 1. As the number of reviews per user increases, it is evident that the proportion of one-time reviewers is decreasing; this is in contrast to the third image, which depicts user review activity. The majority of users only have one review. In addition, a pie chart displaying sentiment distribution shows that positive reviews make up the majority at 77.7%, with negative reviews coming in at 14.4% and neutral thoughts at 7.9%. Consistent with this, a line graph demonstrating a sharp change from negative to positive sentiments clearly indicates a highly positive response to the service or commodity in question. These graphics demonstrate that most users are infrequent contributors and reviewers, with a generally positive attitude towards ratings, even though their helpfulness input varies. With this update, we have fresh ideas on how to improve the user experience and encourage participation.

D. Deep learning & Modeling

Preparing and encoding data is the foundation for building deep learning models that classify texts. Initial preparations include cleaning up evaluation texts, eliminating stopwords and tokenising. To convert the labels for positive, neutral, or negative emotion into a numerical format, we can use a LabelEncoder. We tokenise reviews with a word limit of 10,000 to turn them into number sequences. The input shapes to the models are uniformly padded to a length of 200 using these sequences. Cut the set of data in half lengthwise; then, allocate 80% to training et 20% to testing. The data is ready to aid successful learning thanks to the well-organised pipeline.

- **Hybrid LSTM-GRU Model**

The central architecture makes use of a hybrid model that is based on Keras and incorporates either Long Short-Term Memory (LSTM) or GRU (gated recurrent unit) layers. The first step in capturing word semantics is an Absorption Layer that takes input word values and turns them into dense vectors. The 64-unit LSTM layer that follows can receive sequential outputs since return_sequences=True. Next, a 64-until Hsr layer will train effectively on these sequences, capturing complex relationships and patterns.

- **GRU Model**

The use of a separate GRU-based model for performance comparison is also an option. An Embedding Layer, a Dropout Layer, a Dense Network with softmax for sentiment sorting and a single 128-unit GRU layer make up this model. For large-scale datasets in particular, the GRU model's streamlined structure and quicker training times make it a good candidate to test whether it can match or outperform the hybrid model.

IV. RESULT & DISCUSSION

Two important variables to examine while evaluating the hybrid LSTM-GRU model's sentiment classification capabilities are loss and accuracy. The accuracy metric makes it easy to see how well the model can predict the right sentiment by showing the proportion of correctly classified evaluations over the whole dataset. When the model's accuracy rate is high, it means it effectively captures emotional patterns in both the training also testing data and can generalise those patterns. Sparse categorical crossover entropy quantifies loss, the gap between the actual so predicted emotion classes and so serves as an estimation of the model's error rate. A low loss number indicates that the model performed adequately during training and produced reliable predictions. Guaranteeing correct sentiment categorisation with minimum prediction error, the hybrid LSTM-GRU model's low loss and high accuracy demonstrate its rapid and precise learning.

- Accuracy

The model did an excellent job accurately categorising emotions over new data, with an assessment accuracy of approximately 0.85. The high level of accuracy indicates that the mixture of the LSTM-GRU model does a good job of identifying patterns in the review texts.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1)$$

- Loss

With sparse categorical cross-entropy as a loss metric, it converged to approximately 0.35 in the last epoch, indicating well-learned features with little error. A low loss indicates that the model learnt well and made correct predictions across all sentiment categories.

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} \mathcal{Y}i.\log(\mathcal{Y}i) \quad (2)$$

TABLE 2. PERFORMANCE EVALUATION OF PROPOSED MODEL

Model	Accuracy	Precision	Recall	F1 score	Loss
Proposed Hybrid LSTM-GRU	96.50	93.85	95.50	93.49	0.56
Proposed GRU Model	93.6	92.5	93.6	91.14	0.69

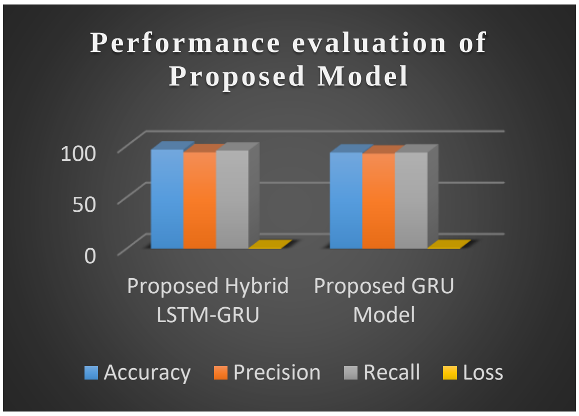

Fig. 8 Performance Graph of proposed Model

Table 2 gives a rundown of how well the standalone GRU model and the suggested Hybrid LSTM-GRU model perform. The Hybrid model achieves greater overall classification performance than the GRU model across all measures. It achieves a higher accuracy of 96.50%, as well as a higher 93.85% accuracy, 95.50% recall and F1 score of 93.49%. A lower loss of 0.56 indicates more consistent and effective learning. Alternatively, the GRU model is not quite up to snuff; it has a lower accuracy rate of 93.6% and a higher loss rate of 0.69. Based on these findings, it is clear that LSTM and GRU combinations outperform GRU alone when it comes to contextual learning and sentiment prediction.

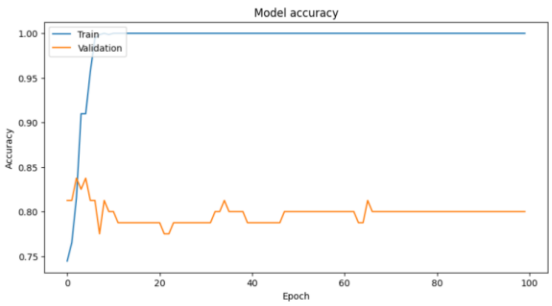

Fig. 9 Model Accuracy graph of Hybrid LSTM-GRU model

After 100 training epochs, the Hybrid LSTM-GRU model's accuracy began to rise, as seen in this graph. The model is effectively learning from the input data since the accuracy increases steadily during training. Due to random weight initialisation, the accuracy starts off poorer, even though the model's inner variables are fine-tuned via backpropagation with gradient descent. But the model becomes better at what it does with each passing era. By the time the training session is over, the machine learning algorithm has learnt to accurately and reliably classify the emotions in the information being reviewed when the combined accuracy is close to 96.5%.

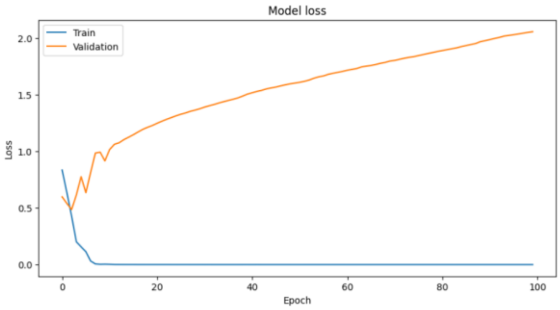

Fig. 10 Model loss graph of Hybrid LSTM-GRU model

This graph displays the loss reduction across epochs, demonstrating consistent decrease in sparse categorical cross-entropy loss, which reflects improved model predictions and minimized error as training progresses.

TABLE 3. COMPARATIVE ANALYSIS OF EXISTING MODELS AND PROPOSED MODEL

Model	Accuracy	References
PLSA hybrid ELMo wiki pedia	79.00	[27]
LDA hybrid ELMo wiki pedia	75.00	[27]
Proposed Hybrid LSTM-GRU	96.50	

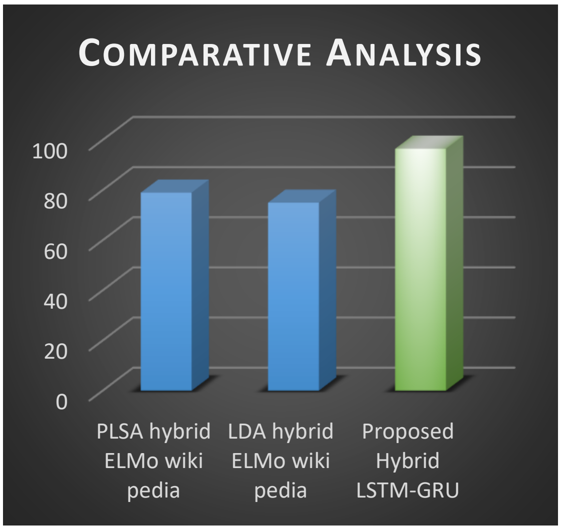

Fig. 11 Comparative Analysis Graph

Table 3 gives a comparison of the suggested Hybrid LSTM-GRU algorithm for sentiment analysis with current models. The accuracy values of the previously utilised models, PLSA hybrid ELMo or LDA hybrid ELMo, were 79.00% or 75.00%, respectively, according to reference.[27]. Although these models are powerful, they may miss some deep contextual linkages in text data since they use standard topic modelling techniques along with ELMo embeddings. The proposed Dual In contrast, the LSTM-GRU model outperforms them both with an accuracy of 96.50%. This demonstrates how well it understands the nuances of customer feedback and the sequential patterns revealed by Amazon reviews. The results demonstrate that deep machine learning, when coupled with advanced NLP preparation, improves the accuracy of sentiment classification.

V. Conclusion

Using a massive Amazon review dataset with over 500,000 entries and 10 unique features, this study proves a thorough method for sentiment analysis. A 10,000-word vocabulary for tokenisation and consistent length padding were among the thorough cleaning and preprocessing operations that prepared the dataset for the processing of natural languages tasks. Analyses of exploratory data showed that most evaluations were positive, with a strong cluster around very good scores and feelings. In order to successfully capture sequential and temporal nuances in review text, the study suggested and developed a hybrid deep learning architecture that combines Gates recurring unit (GRU) and long short-term memories (LSTM) layers. It made use of embedding and dropout techniques. This hybrid model outperformed other topic-modeling-based techniques and the standalone GRU model in terms of classification performance, with a 96.5% accuracy rate. Having low loss values along with excellent precision, recall, or F1 scores shows that the sentiment classification is strong and reliable. These outcomes demonstrate that the hybrid model is well-suited for practical sentiment analysis tasks due to its capacity to discover intricate patterns from big, noisy textual data. For e-commerce review datasets, the study confirms that a powerful approach for improving sentiment prediction accuracy is the integration of modern deep learning techniques with rigorous preprocessing and EDA.

References

[1] I. Technology and S. G. Vihar, "EMOTION DETECTION USING CONTEXT BASED," vol. 100, no. 19, pp. 5607–5614, 2022.

[2] A. E. de O. Carosia, "Sentiment Analysis Applied to News from the Brazilian Stock Market," *IEEE Lat. Am. Trans.*, vol. 20, no. 3, pp. 512–518, 2022, doi: 10.1109/TLA.2022.9667151.

[3] H. Guo, B. Liu and Z. Yang, "Machine Learning-Based Emotion Factor Analysis of Sport Fan Community," *Secur. Commun. Networks*, vol. 2022, 2022, doi: 10.1155/2022/2674987.

[4] A. P. Rodrigues *et al.*, "Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/5211949.

[5] C. Chen, B. Xu, J. H. Yang and M. Liu, "Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/8517205.

[6] A. Goswami *et al.*, "Sentiment Analysis of Statements on Social Media and Electronic Media Using Machine and Deep Learning Classifiers," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/9194031.

[7] G. Chandrasekaran, N. Antoanela, G. Andrei, C. Monica and J. Hemanth, "Visual Sentiment Analysis Using Deep Learning Models with Social Media Data," *Appl. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/app12031030.

[8] Renu D.S, Tintu Vijayan and Dr. D. Dhanya, "Emotion Analysis Using Convolutional Neural Network," vol. 10, no. 04, pp. 223–228, 2022, [Online]. Available: www.ijert.org

[9] Z. Jalil *et al.*, "COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques," *Front. Public Heal.*, vol. 9, no. January, pp. 1–14, 2022, doi: 10.3389/fpubh.2021.812735.

[10] Y. Gherkar, P. Gujar, A. Gaziyani and S. Kadu, "Keyword :," vol. 03029, pp. 1–6, 2022.

[11] U. Sirisha and B. S. Chandana, "Aspect based Sentiment and Emotion Analysis with

ROBERTa, LSTM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, pp. 766–774, 2022, doi: 10.14569/IJACSA.2022.0131189.

[12] T. Nijhawan, G. Attigeri and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00575-6.

[13] G. Kalpana, K. Pranav Kumar, J. Sudhakar and P. Sowndarya, "Emotion and sentiment analysis using machine learning," *Ann. Rom. Soc. Cell Biol.*, vol. 25, no. 1, pp. 1906–1911, 2021, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101624121&partnerID=40&md5=17f24e6a38bba21036fbf25c37606969

[14] S. Cahyaningtyas, D. Hatta Fudholi and A. Fathan Hidayatullah, "Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, 2021, doi: 10.22219/kinetik.v6i3.1300.

[15] N. Farhoumandi, S. Mollaey, S. Heysieattalab, M. Zarean and R. Eyvazpour, "Facial Emotion Recognition Predicts Alexithymia Using Machine Learning," *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/2053795.

[16] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu and V. Varadarajan, "Ai based emotion detection for textual big data: Techniques and contribution," *Big Data Cogn. Comput.*, vol. 5, no. 3, 2021, doi: 10.3390/bdcc5030043.

[17] A. Chiorrini, C. Diamantini, A. Mircoli and D. Potena, "Emotion and sentiment analysis of tweets using BERT," *CEUR Workshop Proc.*, vol. 2841, 2021.

[18] S. Gupta and A. Noliya, "URL-Based Sentiment Analysis of Product Reviews Using LSTM and GRU," *Procedia Comput. Sci.*, vol. 235, no. 2023, pp. 1814–1823, 2024, doi: 10.1016/j.procs.2024.04.172.

[19] A. Sarraf, "Utilizing NLP Sentiment Analysis Approach to Categorize Amazon Reviews against an Extended Testing Set," *Int. J. Comput. Int. J. Comput.*, vol. 50, no. 1, pp. 107–116, 2024.

[20] M. K. Shaik Vadla, M. A. Suresh and V. K. Viswanathan, "Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT," *Algorithms*, vol. 17, no. 2, 2024, doi: 10.3390/a17020059.

[21] O. Shobayo, S. Sasikumar, S. Makkar and O. Okoyeigbo, "Customer Sentiments in Product Reviews: A Comparative Study with GooglePaLM," *Analytics*, vol. 3, no. 2, pp. 241–254, 2024, doi: 10.3390/analytics3020014.

[22] B. Yu, "Comparative Analysis of Machine Learning Algorithms for Sentiment Classification in Amazon Reviews," *Highlights Business, Econ. Manag.*, vol. 24, pp. 1389–1400, 2024, doi: 10.54097/eqmavw44.

[23] 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," *Electron.*, vol. 13, no. 7, pp. 1–21, 2024, doi: 10.3390/electronics13071305.

[24] H. Wang, "Word2Vec and SVM Fusion for Advanced Sentiment Analysis on Amazon Reviews," *Highlights in Science, Engineering and Technology*, vol. 85. pp. 743–749, 2024. doi: 10.54097/sw4pft19.

[25] A. M. Shetty, M. F. Aljunid, D. H. Manjaiah and A. M. S. Shaik Afzal, "Hyperparameter Optimization of Machine Learning Models Using Grid Search for Amazon Review Sentiment Analysis," *Lect. Notes Networks Syst.*, vol. 821, no. May, pp. 451–474, 2024, doi: 10.1007/978-981-99-7814-4_36.

[26] M. Tabany and M. Gueffal, "Sentiment Analysis and Fake Amazon Reviews Classification Using SVM Supervised Machine Learning Model," *J. Adv. Inf. Technol.*, vol. 15, no. 1, pp. 49–58, 2024, doi: 10.12720/jait.15.1.49-58.

[27] F. Nurifan, R. Sarno and K. R. Sungkono, "Aspect based sentiment analysis for restaurant reviews using hybrid ELMo-wikipedia and hybrid expanded opinion lexicon-senticircle," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 6, pp. 47–58, 2019, doi: 10.22266/ijies2019.1231.05.