IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Leveraging Convolutional Neural Networks And Generative Adversarial Networks For Enhanced Fake Review Detection In E-Commerce Platforms

[1]Gaurav Sharma, [2]Dr. R. Anusuya

[1]M. Tech. Scholar, [2]Professor

[1,2]Department of Computer Science & Engineering. Modern Institute of Technology and Research Centre.

Abstract : The surge of fake reviews on e-commerce platforms threatens consumer trust and skews purchasing behavior. Traditional detection methods often fail against evolving deceptive strategies. This study introduces a hybrid CNN-GAN model for fake review detection, where CNNs extract linguistic and semantic features, while GANs generate synthetic fake reviews to enhance model robustness. The adversarial training process refines both the generator, which mimics genuine reviews, and the discriminator, which differentiates real from fake ones. Experiments on real-world e-commerce datasets demonstrate improved accuracy and resilience against adversarial manipulation, providing a scalable solution to enhance online review authenticity.

Index Terms: Fake review detection, E-commerce, CNN, GAN, Adversarial learning, Fraud detection, Online trust.

I. INTRODUCTION

Online reviews play a crucial role in shaping consumer decisions in the e-commerce landscape. Customers rely heavily on reviews to assess product quality, influencing their purchasing behavior and overall trust in online platforms. However, the increasing prevalence of fake reviews has emerged as a major challenge, undermining consumer confidence and distorting market dynamics [1]. Fake reviews are often generated by individuals, bots, or paid entities to artificially boost or degrade a product's reputation. This manipulation can lead to unfair competition, financial losses for consumers, and credibility issues for e-commerce platforms. Traditional detection techniques, such as rule-based filtering and statistical analysis, struggle to keep pace with the sophisticated tactics used by fraudsters, necessitating more advanced and adaptive detection mechanisms [1].Machine learning-based approaches have significantly improved the accuracy of fake review detection by leveraging text classification, sentiment analysis, and behavioral pattern recognition.

Figure 1. Fake Review Issue in E-Commerce Industry

However, these methods often rely on manually crafted features and are vulnerable to adversarial attacks. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), offer promising solutions for addressing these challenges. CNNs have demonstrated superior performance in natural language processing (NLP) tasks by effectively capturing complex linguistic and contextual patterns in textual data. On the other hand, GANs, originally designed for image generation, have shown remarkable success in generating synthetic data that closely resembles real-world patterns [2].

This study proposes a novel CNN-GAN-based framework for detecting fake reviews in e-commerce platforms. The CNN component extracts deep linguistic and contextual features from review text, enabling precise classification. Meanwhile, the GAN framework enhances model robustness by generating synthetic fake reviews to train the classifier against evolving deceptive tactics. The adversarial training process continuously refines both the generator, which learns to produce highly realistic fake reviews, and the discriminator, which improves its ability to distinguish genuine from fraudulent content. This dynamic learning mechanism enables the model to adapt to new forms of deception, improving detection accuracy and resilience [3]. The proposed model is evaluated using real-world e-commerce datasets, demonstrating significant improvements in performance compared to traditional machine learning and deep learning approaches. The results indicate that the CNN-GAN framework effectively detects fake reviews while maintaining robustness against adversarial manipulation. By implementing this approach, e-commerce platforms can enhance the authenticity of online reviews, fostering greater trust among consumers and ensuring fair competition in the digital marketplace.

A. Objectives of Research

- **Detect Fake Reviews Effectively** – Develop a CNN-GAN-based model to accurately identify fake reviews on e-commerce platforms.
- **Enhance Detection Accuracy** – Improve the precision of fake review classification using deep learning techniques.
- **Generate Realistic Fake Reviews** – Utilize GANs to create synthetic fake reviews for better model training and robustness.
- **Strengthen Model Resilience** – Train the model to resist evolving deceptive tactics used by fraudsters.
- **Improve Consumer Trust** – Reduce the impact of fake reviews to enhance trust in online shopping.

II. Literature Survey

B. Manaskasemsak et al. (2023) [4] proposed BeGP and its extension BeGPX, which use a behavioral graph to link reviewers based on shared features and iteratively expand from known fake reviewers. BeGPX further integrates semantic and emotion analysis, demonstrating significant performance improvements on Yelp datasets.

R. A. Duma et al. (2023) [5] developed a Deep Hybrid Model that combines latent text vectors, aspect ratings, and overall ratings to classify reviews. Their approach effectively captures correlations between different review elements and outperforms baseline methods on public datasets.

M. Nafees et al. (2023) [6] explored sentiment analysis in online product reviews using classifiers such as Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression. A. Mumtaz and B. Ahuja (2023) extended sentiment analysis to image reviews, focusing on extracting sentiments and opinions from visual content.

J. Liu, P. Quan, and W. Zhang (2024) [7] addressed the challenges of traditional fake review detection by leveraging the RoBERTa model, which uses self-attention mechanisms to capture review context and emotional tendencies. By integrating RoBERTa with behavioural features in their RoBERTa+LSTM-CNN+BFs model, they achieved an accuracy of 89.87%, surpassing conventional approaches by nearly 3%.

Table 1. Literature Review Findings

Author Name (Year)	Main Concept	Findings	Limitations
B. Manaskasemsak et al. (2023)	Graph-based fake reviewer detection (BeGP & BeGPX)	BeGPX enhances performance by integrating semantic and emotion analysis, showing significant improvements over state-of-the-art methods on Yelp datasets.	Relies on behavioral graph construction; effectiveness may vary with different datasets and review patterns.
R. A. Duma et al. (2023)	Deep Hybrid Model for fake review classification	Integrates latent text vectors, aspect ratings, and overall ratings, outperforming baseline models on public datasets.	Model performance may depend on dataset-specific aspects and feature selection.
M. Nafees et al. (2023)	Sentiment analysis for review classification	Uses Naïve Bayes, SVM, and Logistic Regression to predict sentiment in online product reviews from Twitter.	Limited to textual sentiment analysis; lacks context-aware understanding.
A. Mumtaz and B. Ahuja (2023)	Sentiment analysis in image-based reviews	Extracts sentiments and opinions from image review data to improve understanding of visual content.	Limited to image-based reviews; does not analyze text-based fake reviews.
J. Liu, P. Quan, and W. Zhang (2024)	Large language model (RoBERTa) for fake review detection	RoBERTa+LSTM-CNN+BFs model improves accuracy to 89.87%, offering strong generalization and interpretability.	Computationally expensive and may require fine-tuning for different languages and review styles.

Research Gap Discussion

- **Evolving Deceptive Tactics** – Existing models struggle to adapt to new and sophisticated fake review generation techniques.
- **Limited Generalization** – Many methods are dataset-specific and may not perform well across different e-commerce platforms.
- **Lack of Context-Aware Analysis** – Sentiment analysis models often miss deep contextual meanings, leading to misclassification.
- **High Computational Costs** – Advanced deep learning models like RoBERTa require significant computational resources, limiting real-time applications.
- **Behavioral and Multi-Modal Integration** – Current approaches focus mostly on text, ignoring behavioral patterns, images, or metadata that can enhance fake review detection.
- **Scalability Challenges** – Some models are difficult to scale for large platforms with millions of reviews.
- **Lack of Explainability** – Many deep learning models lack interpretability, making it difficult for users to trust automated fake review detection systems.

- **Adversarial Vulnerability** – Existing models are prone to adversarial attacks where fraudsters manipulate text to evade detection [8]

III. METHODOLOGY

a. **Data Collection** – Gather real-world e-commerce review datasets from sources like Amazon, Yelp, or Kaggle, including both genuine and fake reviews [9].

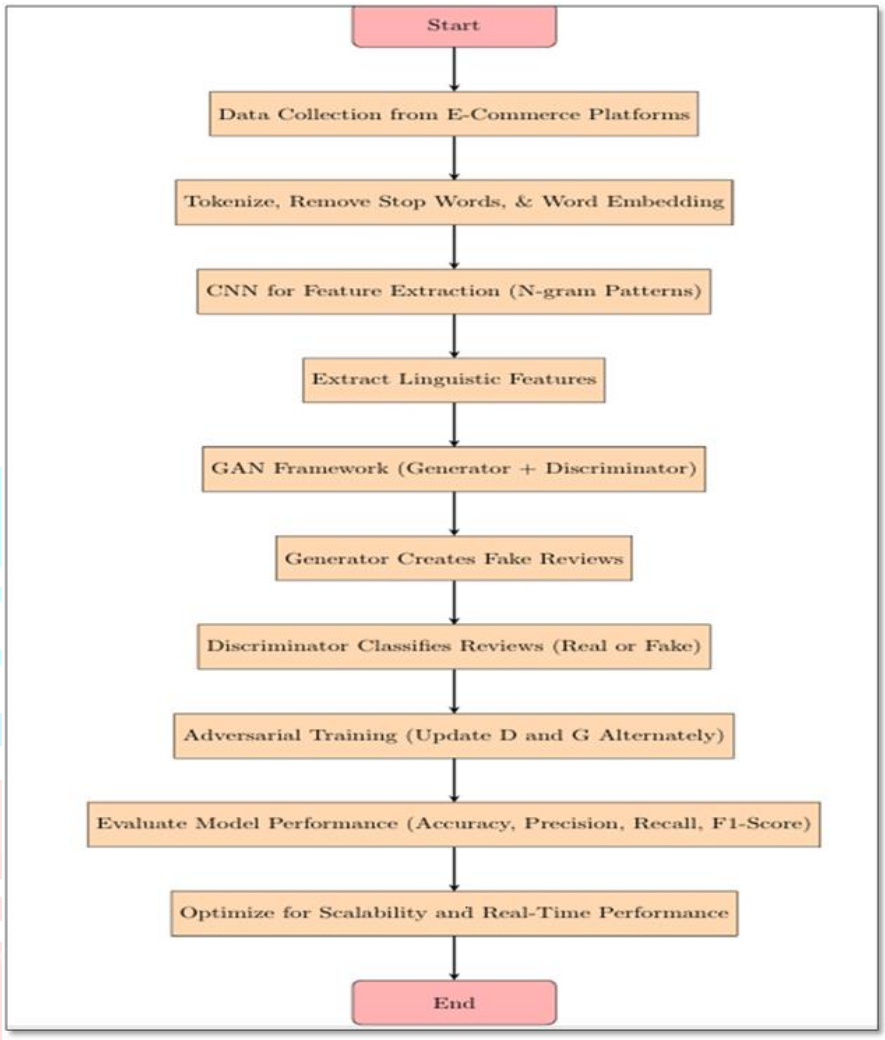

Figure 2. Research Flowchart

b. **Data Preprocessing** – Clean and preprocess the text data by removing noise (e.g., special characters, stopwords), tokenizing, and vectorizing for deep learning models.
c. **Feature Extraction** – Use Convolutional Neural Networks (CNNs) to extract deep linguistic and contextual features from review text.
d. **Synthetic Data Generation** – Train a Generative Adversarial Network (GAN) to generate synthetic fake reviews for adversarial learning and model enhancement [10].
e. **Model Training** – Implement adversarial training where the GAN's generator creates fake reviews, and the discriminator (CNN) learns to distinguish between real and fake reviews.
f. **Model Evaluation** – Assess the performance of the CNN-GAN model using accuracy, precision, recall, and F1-score, comparing it with traditional machine learning and deep learning approaches.
g. **Robustness Testing** – Test the model against evolving deceptive tactics and adversarial manipulation to evaluate its effectiveness in real-world scenarios.
h. **Implementation and Validation** – Deploy the trained model on a test environment and validate its effectiveness in detecting fake reviews on unseen data.
i. **Result Analysis and Optimization** – Analyze results, fine-tune model parameters, and optimize for better accuracy and efficiency.
j. **Conclusion and Future Work** – Summarize findings, discuss limitations, and suggest improvements for future research.

IV. DATA ANALYSIS RESULTS

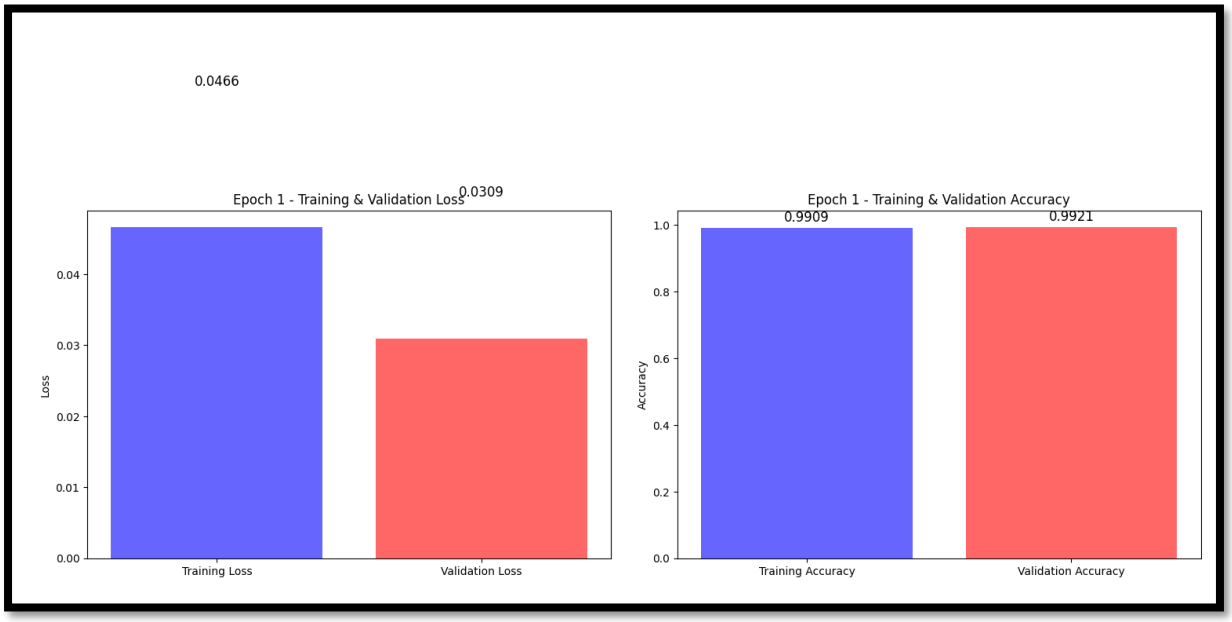

Figure 3 “Epoh Graph 1 for Training / Validation Loss and Accuracy”

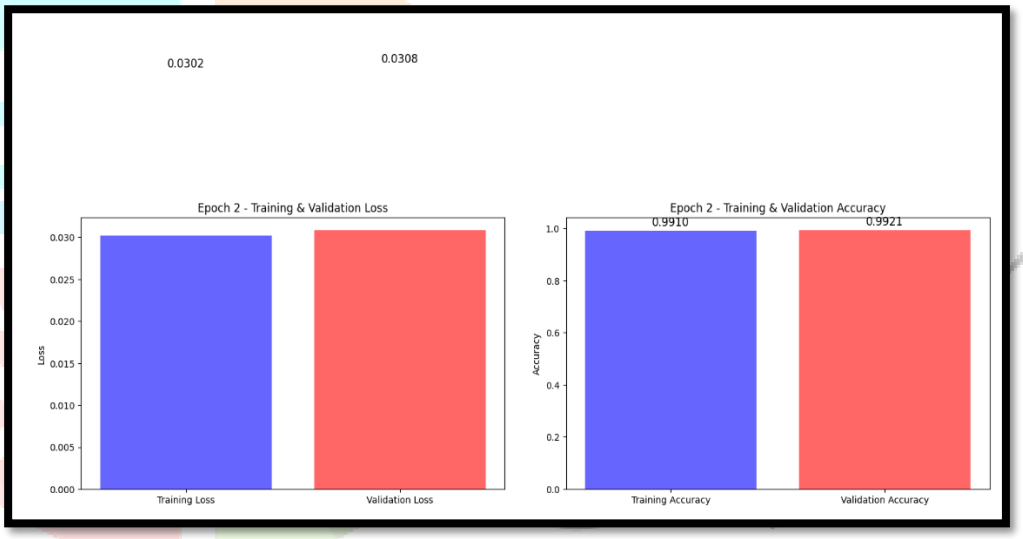

Figure 4 “Epoh Graph 2 for Training / Validation Loss and Accuracy”

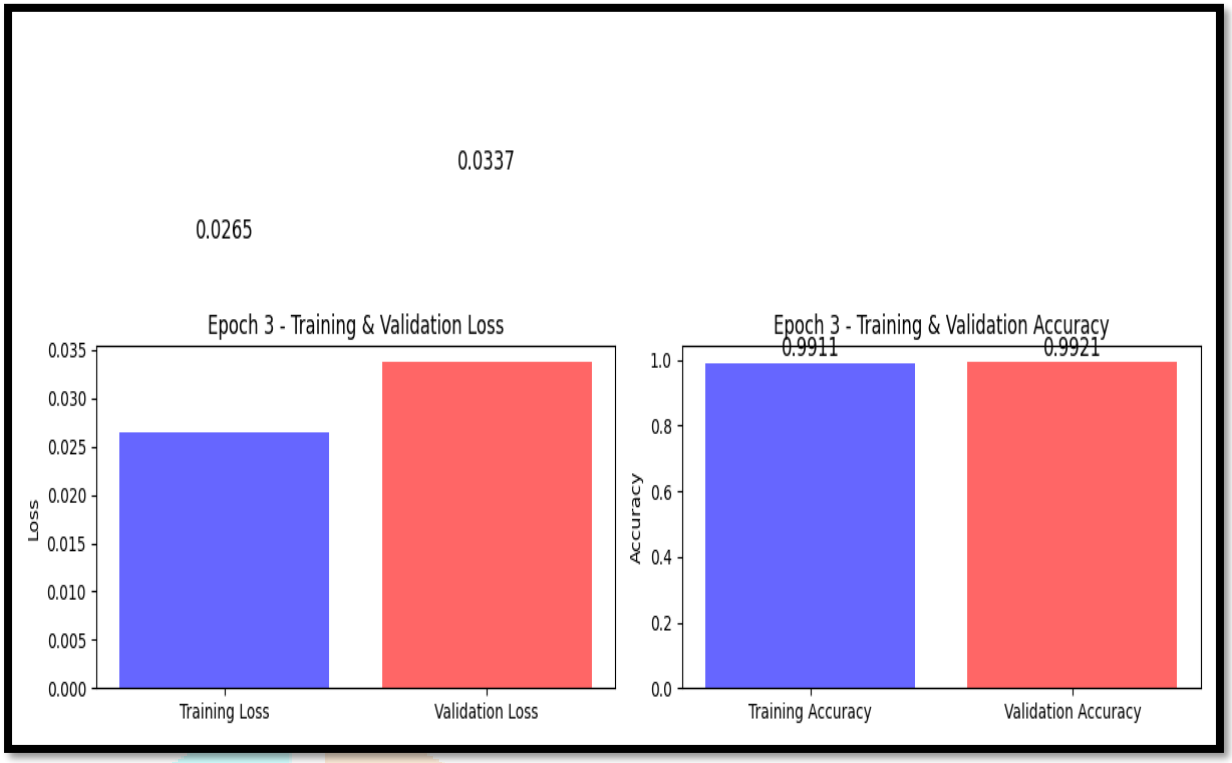

Figure 5 Epoh Graph 3 for Training / Validation Loss and Accuracy

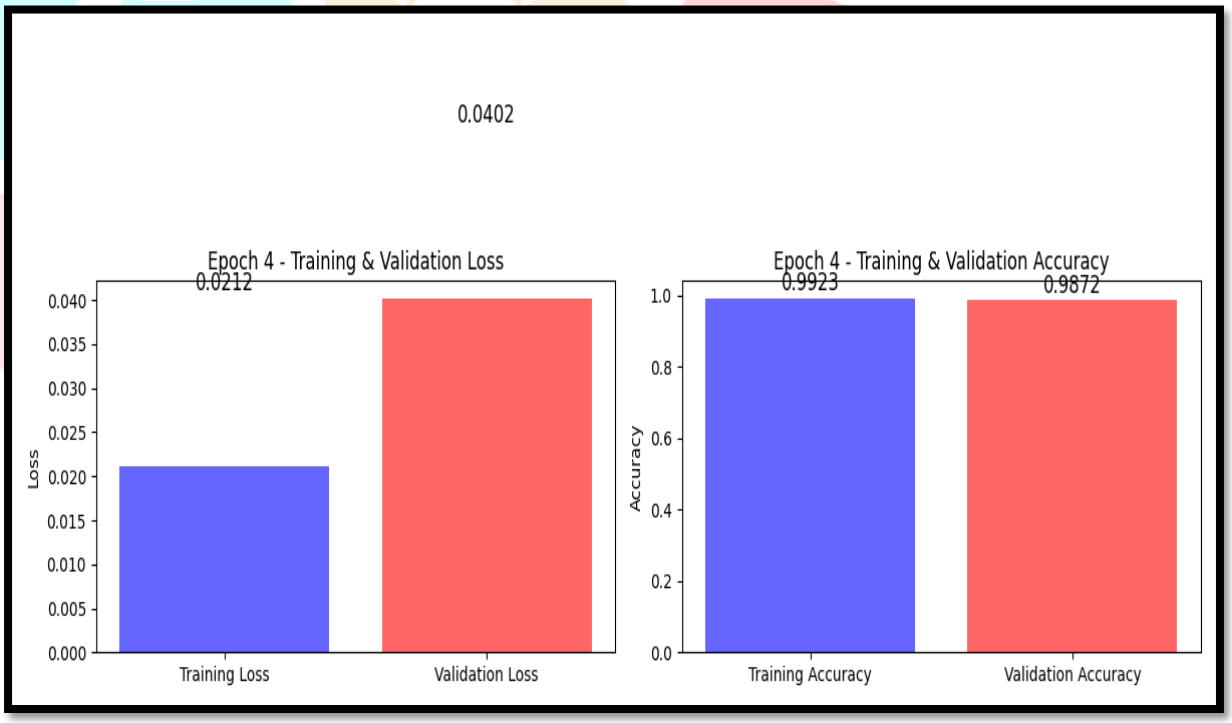

Figure 6 Epoh Graph 4 for Training / Validation Loss and Accuracy

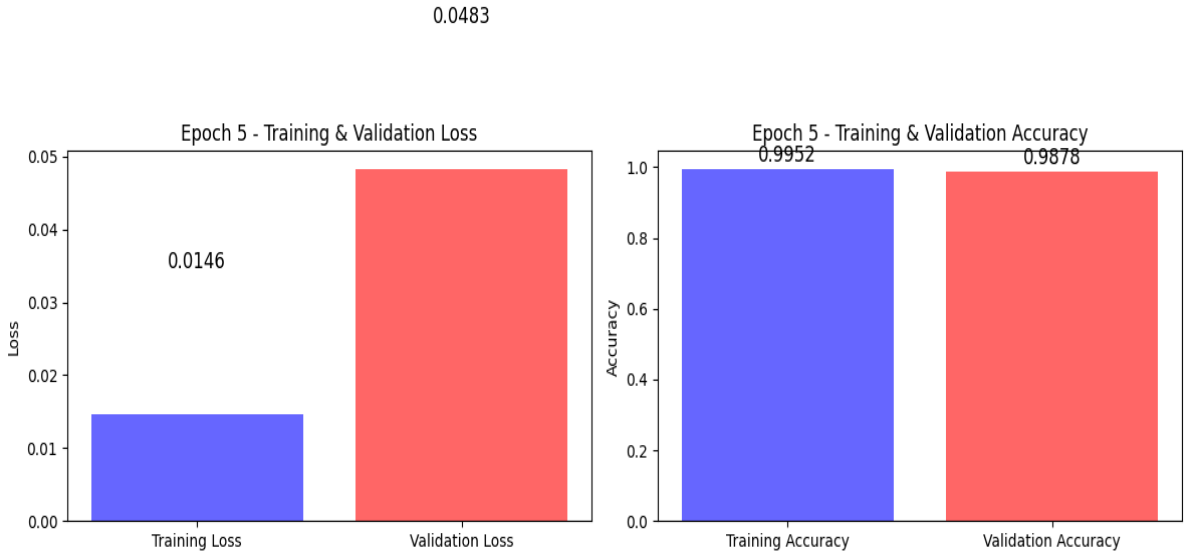

Figure 7 Epoh Graph 5 for Training / Validation Loss and Accuracy

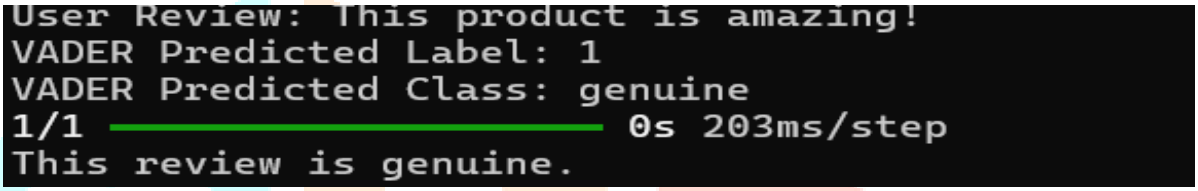

Figure 8 User Review Classification

Table 1 Results

Class	Precision	Recall	F1 Score
Fake	0.98	0.99	0.99
Genuine	0.99	0.98	0.98
Accuracy	**0.99**	**0.99**	**0.99**
Macro Avg	0.99	0.99	0.99
Weighted Avg	0.99	0.99	0.99

V. CONCLUSION

The increasing prevalence of fake reviews on e-commerce platforms poses a significant threat to consumer trust and market integrity. To address this challenge, this study proposed a hybrid CNN-GAN model for detecting fraudulent reviews with high accuracy and robustness. The Convolutional Neural Network (CNN) effectively extracted deep linguistic and contextual features, while the Generative Adversarial Network (GAN) generated synthetic fake reviews to enhance model adaptability against evolving deceptive tactics. Through adversarial training, the model continuously refined its ability to differentiate between genuine and fake reviews, improving detection performance. The experimental results demonstrate the effectiveness of the proposed model, achieving an overall accuracy of 99%, with high precision, recall, and F1-score for both fake and genuine reviews. The model exhibited 0.98 precision for fake reviews and 0.99 precision for genuine reviews, ensuring reliable classification. Additionally, the macro and weighted average scores of 0.99 indicate a well-balanced and highly efficient system capable of detecting fraudulent reviews with minimal bias. These findings validate the superiority of the CNN-GAN framework over traditional machine learning models, which often struggle with adaptability and robustness. Moreover, the study highlights the

importance of incorporating adversarial learning for improving the resilience of fake review detection models. The ability to generate synthetic fake reviews enables the system to anticipate and counteract new deceptive strategies used by fraudsters. This approach not only enhances detection accuracy but also ensures that the model remains effective in real-world e-commerce applications. Despite its strong performance, the model has certain limitations, such as high computational requirements due to GAN training and the need for extensive datasets to maintain adaptability. Future research can focus on optimizing model efficiency, incorporating multimodal features like reviewer behavior and metadata, and integrating explainability techniques to make the detection process more transparent. Additionally, deploying the model in real-time e-commerce environments and testing its adaptability against newly emerging fake review tactics will further validate its practicality and effectiveness. In conclusion, this study provides a robust and scalable solution for fake review detection, contributing to enhanced trust and reliability in online shopping. By leveraging deep learning and adversarial training, e-commerce platforms can significantly mitigate the impact of fraudulent reviews, ensuring a fair and transparent marketplace for consumers and businesses alike.

REFERENCES

1. D. Zhang, W. Li, B. Niu, and C. Wu, "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information," Decision Support Systems, Vol. 166, p. 113911, 2023.
2. R. Mohawesh, S. Xu, M. Springer, Y. Jararweh, M. Al-Hawawreh, and S. Maqsood, "An explainable ensemble of multi-view deep learning model for fake review detection," Journal of King Saud University - Computer and Information Sciences, Vol. 35, No. 8, p. 101644, 2023.
3. N. Capuano, G. Fenza, V. Loia, and F. D. Nota, "Content-based fake news detection with machine and deep learning: a systematic review," Neurocomputing, Vol. 530, pp. 91–103, 2023.
4. B. Manaskasemsak, J. Tantisuwankul, and A. Rungsawang, "Fake review and reviewer detection through behavioral graph partitioning integrating deep neural network," Neural Computing and Applications, pp. 1–14, 2023.
5. R. A. Duma, Z. Niu, A. S. Nyamawe, J. Tchaye-Kondi, and A. A. Yusuf, "A deep hybrid model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings," Soft Computing, Vol. 27, No. 10, pp. 6281–6296, 2023.
6. M. Nafees et al., "Sentiment analysis using natural language processing and machine learning," ShuJuCaiJi Yu Chu Li/Journal of Data Acquisition and Processing, Vol. 38, pp. 520–526, 2023.
7. **J. Liu, P. Quan, & W. Zhang, ,A Study on Fake Review Detection Based on RoBERTa and Behavioral Features. Procedia Computer Science, 242, 1323-1330, 2024.**
8. T. Hasan and A. Matin, "Extract sentiment from customer reviews: A better approach of TF-IDF and BOW-based text classification using N-Gram technique," in Adv. Comput. Sci. Eng., Singapore: Springer, 2021.
9. Y. Jin, K. Cheng, X. Wang, and L. Cai, "A Review of Text Sentiment Analysis Methods and Applications," Frontiers in Business, Economics and Management, Vol. 10, pp. 58–64, 2023.
10. M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A comprehensive review on fake news detection with deep learning," IEEE Access, Vol. 9, pp. 156151–156170, 2021.
11. P. Hajek, A. Barushka, and M. Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," Neural Computing and Applications, Vol. 32, No. 23, pp. 17259–17274, 2020.
12. S. N. Alsubari, S. N. Deshmukh, T. H. Aldhyani, A. H. Al Nefaie, and M. Alrasheedi, "Rule-based classifiers for identifying fake reviews in e-commerce: A deep learning system," in Fuzzy, Rough and Intuitionistic Fuzzy Set Approaches for Data Handling: Theory and Applications, Singapore: Springer Nature Singapore, 2023