iJRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80060

Contextual Fake Review Detection in E-commerce using Bidirectional LSTM and Word Embeddings

Killada Sirisha[1], Pandiri Lavanya[2], Suneel Kumar Duvvuri[3]
[1, 2]Student, M.Sc (Computer Science), Government College (Autonomous), Rajahmundry, Andhra Pradesh, India
[3]Assistant Professor, Department of Computer Science, Government College (Autonomous), Rajahmundry, Andhra Pradesh, India

***Abstract**: The digital era has witnessed an unprecedented explosion in user-generated textual data, primarily driven by the rapid expansion of e-commerce and online entertainment platforms. In this landscape, consumer reviews have emerged as a critical form of social proof, significantly influencing purchasing decisions and brand reputation. However, this influence has led to the proliferation of deceptive opinion spam, fake reviews systematically crafted to manipulate product ratings or damage competitors. Detecting these fraudulent entries is a significant challenge because they are often written to mimic genuine human feedback closely. Traditional manual verification is inefficient and practically impossible given the massive daily volume of data.*

This study proposes a robust, automated deep learning framework for detecting fake reviews using Natural Language Processing (NLP) and Bidirectional Long Short-Term Memory (BILSTM) networks. While earlier systems relied on traditional machine learning algorithms such as Naïve Bayes and Support Vector Machines (SVM), these methods often failed to capture deep contextual dependencies and the sequential nature of human language. To address these limitations, the proposed model utilises a BILSTM architecture that processes review text in both forward and backward directions simultaneously. This dual-directional approach enables the system to understand the complete semantic context and identify subtle deceptive patterns, such as exaggerated praise, generic phrasing, and unnatural emotional shifts.

The methodology uses a comprehensive e-commerce dataset consisting of approximately 48,800 labelled reviews, evenly distributed between "Real" and "Fake" classes. A rigorous preprocessing pipeline is applied, including HTML tag removal, text normalisation, tokenisation, stop-word removal, and Porter stemming to clean and standardise the raw text. Furthermore, word embedding techniques are employed to convert textual data into dense 128-dimensional vector representations.

Experimental results demonstrate the effectiveness of the proposed model. The BILSTM model achieved a test accuracy of 91.90%, outperforming traditional baseline approaches. It also achieved a precision score of 0.93 for the "Fake" class, indicating a low rate of false positives. Confusion matrix analysis further confirms the model's balanced performance and robustness across various product categories and review patterns.

This research concludes that sequential deep learning models provide a scalable and highly accurate solution for maintaining the integrity of online marketplaces. Future work includes incorporating transformer-based architectures such as BERT and extending the model to support multilingual datasets.

Keywords -Sentiment Analysis, Natural Language Processing (NLP), Deep Learning, Bidirectional LSTM (BILSTM), Fake Review Detection, Text Classification, Word Embedding, Deceptive Opinion Spam, E-commerce Security.

I. INTRODUCTION

In the modern digital era, the rapid expansion of the internet and online platforms has led to an unprecedented growth of user-generated data. Every interaction, including comments, reviews, and social media posts, contributes to a massive volume of unstructured textual information[1]. It is estimated that billions of data points are generated daily, with a significant portion consisting of textual content such as product reviews and social media discussions. Platforms like e-commerce sites have become major sources of such data, where users actively share their opinions about products, creating a rich repository of sentiment-driven content [2][3]

Traditionally, the success of products relied heavily on limited consumer feedback and marketing. However, with the advent of digital platforms, this paradigm has shifted significantly. Today, millions of users express their opinions online, providing real-time insights into audience perception. While this abundance of data offers valuable opportunities for analysis, manually processing such large-scale information to verify its authenticity is impractical. This challenge has led to the emergence of automated techniques capable of extracting meaningful insights from textual data, where Big Data Analytics plays a crucial role in handling large and complex datasets efficiently [4]

Natural Language Processing (NLP) has emerged as a key technology in enabling machines to understand and interpret human language. NLP combines computational linguistics with machine learning techniques to process and analyse textual data. However, human language is inherently complex, characterised by ambiguity, contextual variations, sarcasm, and evolving linguistic patterns. These challenges make it difficult for traditional computational models to accurately interpret textual meaning or detect deceptive intent [5]. Over the years, NLP has evolved through three major phases: Symbolic NLP, Statistical NLP, and Neural NLP. The advent of Neural NLP, powered by deep learning techniques, has significantly enhanced the ability of machines to understand context and semantics in text [6][7][8]

One of the most important applications of NLP is Sentiment Analysis, also known as Opinion Mining. It involves identifying and classifying the emotional tone expressed in textual data. In the context of product reviews, sentiment analysis helps determine whether a review expresses a positive or negative opinion. This has significant applications in business intelligence, recommendation systems, and consumer decision-making [9][10][11] For example, analysing user reviews can help companies understand audience reactions and improve future content.

Despite its advantages, sentiment analysis faces several challenges due to the complexity of human language. Sarcasm and irony can invert the meaning of words, making it difficult for models to interpret sentiment correctly[12][13]. Similarly, negation plays a critical role in sentiment interpretation, where phrases like “not good” completely change the sentiment polarity. These challenges highlight the limitations of traditional machine learning approaches in handling nuanced textual data.

Earlier sentiment analysis and opinion spam detection systems primarily relied on traditional machine learning techniques such as Naïve Bayes and Support Vector Machines[14] While these methods achieved moderate accuracy, they depended heavily on manual feature engineering techniques like Bag-of-Words and TF-IDF. More importantly, they failed to capture the sequential nature of language, where the order of words significantly influences meaning. Traditional models often struggle to differentiate deceptive cases from genuine ones due to their inability to capture contextual dependencies.

To overcome these limitations, deep learning approaches have been introduced, including Long Short-Term Memory (LSTM) networks [15]. LSTMs are specifically designed to capture long-term dependencies in sequential data, addressing the vanishing gradient problem. However, standard LSTM models process data in a single direction, which limits their ability to fully understand context[16].

Bidirectional Long Short-Term Memory (BILSTM) models enhance contextual understanding by processing text in both forward and backward directions, enabling better capture of sentence semantics. In this study, a BILSTM-based deep learning model is developed to classify reviews into real or fake categories using a dataset of approximately 48,800 labelled samples. By integrating preprocessing techniques and word embeddings, the proposed system achieves efficient and accurate deceptive review detection[17]

The main contributions of this work are summarised as follows:

- A robust deep learning-based framework is proposed for identifying deceptive opinion spam in product reviews, addressing the limitations of traditional machine learning approaches in handling unstructured textual data.
- Comprehensive data preprocessing techniques are implemented, including duplicate removal, tokenisation, stop-word elimination, and stemming, to improve data quality and enhance model performance.
- An effective Bidirectional Long Short-Term Memory (BILSTM) architecture is designed to capture contextual dependencies in both forward and backward directions, enabling better understanding of complex linguistic patterns characteristic of fake reviews.
- Word embedding techniques are integrated into the model to transform textual data into dense vector representations, allowing the system to capture semantic relationships between words efficiently.
- The proposed model is trained using the Adam optimiser and Binary Cross-Entropy loss function, with regularisation techniques such as dropout applied to reduce overfitting and improve generalisation capability.
- Extensive experimental evaluation is conducted using standard performance metrics, including accuracy, precision, recall, and F1-score, demonstrating that the proposed model achieves a high-test accuracy of 91.90%.
- The study highlights the superiority of deep learning models, particularly BILSTM, over traditional machine learning techniques in deceptive review detection tasks involving sequential and context-dependent data.
- The proposed system provides practical applicability in real-world scenarios such as real-time e-commerce moderation, social media monitoring, and customer trust evaluation.

The organisation of this paper is as follows: Section 2 presents a comprehensive review of existing literature related to deceptive review detection, natural language processing, and deep learning techniques, with a particular focus on LSTM and Bidirectional LSTM models. Section 3 describes the proposed methodology, including dataset details, data preprocessing steps, tokenisation, word embedding techniques, and the architecture of the BILSTM model used for classification. Section 4 discusses the experimental results and performance evaluation of the model using various metrics such as accuracy, precision, recall, and F1-score, along with a detailed analysis of the findings. Finally, Section 5 concludes the study by summarising the key outcomes and highlighting potential directions for future research.

II. LITERATURE REVIEW

In the current digital era, user-generated reviews on e-commerce and online platforms play a pivotal role in consumer decision-making. However, spammers and automated bots frequently post fabricated reviews to manipulate ratings, thereby undermining the trustworthiness of these platforms. To address this challenge, researchers have long been developing computer-based detection models to identify and filter deceptive content.

A. Behavioural and Rule-Based Foundations

Before the emergence of deep learning, research primarily focused on identifying abnormal “behavioural footprints” left by reviewers.

- Mukherjee et al.[21]. These researchers analysed rating patterns and sudden "bursts" of review activity to identify coordinated opinion and spammers.
- Jindal & Liu[16]. They provided the foundational framework for deceptive review detection by treating it as a formal classification task, utilising early statistical methods like Logistic Regression.

B. Transition to Supervised Machine Learning

As the field shifted toward supervised machine learning algorithms, the accuracy of detection systems improved significantly through feature engineering.

- Ott et al. [17]. In a landmark study, they utilised Support Vector Machines (SVM) combined with n-gram analysis on hotel reviews, achieving remarkable accuracy in distinguishing truthful reviews from deceptive ones.
- Feng et al.[12]. This research introduced “syntactic stylometry,” analysing the grammatical structures and phrasing patterns unique to deceptive authors to improve detection rates.
- Joachims [13]. Demonstrated that SVM algorithms are exceptionally robust for high-dimensional text data, making them a standard for text classification tasks.

C. The Rise of Deep Learning Architectures

Traditional machine learning models often treat text as a bag-of-words,failing to capture the sequential context of a sentence. Deep learning addressed this gap.

- LSTM Foundations: Introduced by Hochreiter & Schmidhuber (1997), the Long Short-Term Memory (LSTM) architecture solved the problem of capturing long-range temporal dependencies. Through its unique mechanism of forget, input, and output gates, LSTMs can retain important contextual information over long sequences.
- BILSTM Advancement**:** Proposed by Schuster & Paliwal (1997), the Bidirectional LSTM (BILSTM) processes text in both forward and reverse directions simultaneously. This dual-processing capability makes it significantly easier to detect subtle nuances, such as sarcasm or unnatural emotional shifts in fabricated reviews.

Table 1. summarizes different research contributions across preprocessing and modelling stages in sentiment analysis. It highlights each technique’s function, advantages, and limitations, along with the specific outcomes achieved in the study.

Table 1. Comprehensive Literature Review on Fake Review Detection

Author(s) & Ref No.	Year	Title / Topic	Publication / Source	Accuracy	Performance
Jindal & Liu [18]	2008	Review Spam Detection	WWW Conference	85%	rule-based/statistical
Ott et al.[19]	2011	Finding Deceptive Opinion	EMNLP	90%	supervised ML on hotel

Author(s) & Ref No.	Year	Title / Topic	Publication / Source	Accuracy	Performance
		Spam			reviews
Feng et al.[13]	2012	Detecting Deceptive Reviews using ML	ACL Workshop	88%	machine learning
Pang et al. [10]	2008	Sentiment Analysis and Preprocessing	Foundations and Trends in IR	N/A	preprocessing methods
Salton & Buckley [11]	1988	Term Weighting Approaches in IR	Information Processing & Management	N/A	IR relevance
Jones [20]	1972	TF-IDF weighting	Journal of Documentation	N/A	IR method
Joachims [14]	1998	SVM for Text Categorisation	KDD Conference	88%	text categorization
Ng [21]	2004	Feature selection and regularisation	ICML	N/A	general ML techniques
Li et al. [22]	2015	Behavioural Patterns in Spam Reviews	AAAI	91%	behavioural features
Mukherjee et al. [23]	2013	Metadata-based Spam Detection	SIGKDD	92%	metadata + text
Rayana & Akol [24]	2015	Collective Opinion Spam Detection	KDD	93%	group-based detection
Camacho-Colados et al. [25]	2019	Fake Review Detection with Deep Learning	WWW Conference	95%	deep learning
Hochreiter & Schmid Huber [15]	1997	LSTM Networks	Neural Computation	N/A	architecture paper

III. METHODOLOGY

A. *System Overview*

The proposed system presents a deep learning-based framework for identifying deceptive opinion spam in e-commerce product reviews using a **Bidirectional Long Short-Term Memory (BILSTM)** architecture. The system processes raw text data from a specialised e-commerce dataset and transforms it into structured numerical representations using advanced preprocessing and embedding techniques. Sequential modelling is performed using LSTM and BILSTM networks to capture complex linguistic patterns and long-range contextual dependencies within the text. The trained model is then evaluated and deployed to predict authenticity labels for unseen data. This structured pipeline ensures efficient learning and accurate classification of fraudulent reviews in Fig 1.

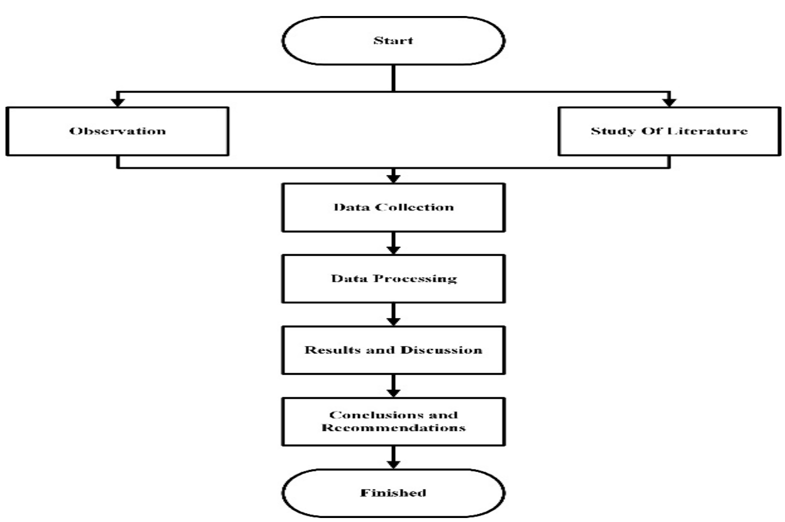


Fig.1. Research process flowchart.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Algorithm 1. Pseudocode for the proposed BILSTM-based Review Detection System
Input: E-commerce Review Dataset
Output: Classified Reviews (Real / Fake)

1) Load dataset.
2) Remove duplicates and handle missing values.
3) Perform text preprocessing:
   a. Convert text to lowercase.
   b. Remove HTML tags, punctuation, and special characters.
   c. Tokenise text using regular expressions.

d. Remove stop words using the NLTK library.
e. Apply Porter Stemming to reduce words to root forms.

1) Convert text into integer sequences using Keras' Tokeniser.
2) Apply padding to ensure fixed sequence length (maxent = 100).
3) Split the dataset into training and testing sets (80:20 ratio).
4) Initialise BILSTM model:

a. Embedding layer.
b. Bidirectional LSTM layers.
c. Dropout and Spatial Dropout layers.
d. Dense output layer with Sigmoid activation.

1) Train the model using the Adam optimizer and Binary Cross-Entropy loss.
2) Evaluate the model using sequestered test data.
3) Compute performance metrics: Accuracy, Precision, Recall, and F1-score.
4) Return Classification Results.

B. System Architecture

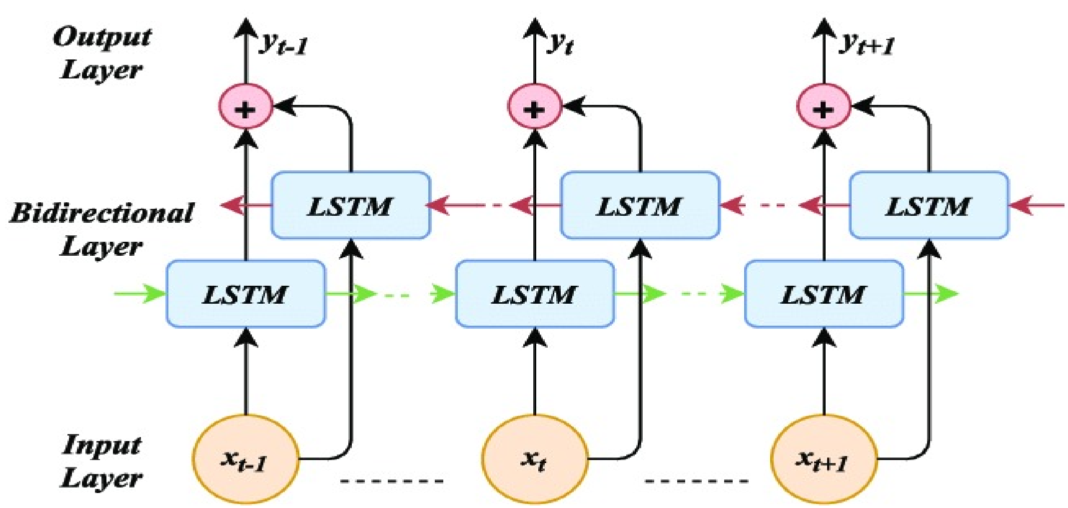


Fig 2: Hierarchical Architecture of Deep Learning models featuring Multi- Layered Neural Representations.

BiLSTM extends standard LSTM by processing sequences in two parallel dimensions simultaneously:

- Forward pass: $\overrightarrow{h_t}$Captures context from preceding words.
- Backward pass: $\overleftarrow{h_t}$ Gathers context from words that follow, identifying sarcasm and manipulative intent built toward the end of a sentence.

Final Representation:
$yt = [ht \oplus ht] \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . (1)$
Sequential Stages:

- Data Collection

- Data Preprocessing
- Text Tokenization
- Sequence Padding
- Word Embedding
- LSTM/Bilt Modelling
- Classification Layer
- Sentiment/Authenticity Prediction

C. Dataset Description

The model is trained and evaluated using the Fake Review Dataset, a comprehensive collection of e-commerce user-generated content. It consists of approximately 48,800 cleaned reviews, meticulously balanced between authentic customer experiences and fabricated posts Table 2.

Table 2. Fake Review Dataset Description

Parameter	Description
Dataset Size	48,800 reviews (cleaned)
Class Distribution	Balanced (24.4K Real, 24.4K Fake)
Data Type	Unstructured textual and categorical data
Source	E-commerce user reviews

D. Data Preprocessing

Preprocessing is a critical phase for standardising noisy web-scraped data and extracting meaningful semantic signals.
Lowercasing: Converts text to a uniform format.
HTML Tag & Noise Removal: Eliminates markup and irrelevant symbols using regular expressions.
Stop-word Removal: Filters common words with low semantic weight using the NLTK library.
Stemming: Reduces words to their base root forms using the Porter Stemmer.
Label Encoding: Converts sentiment/authenticity labels into binary values: Real (0) and Fake (1) Table 3.

Table 3. Binary Representation

Class	Value
Real (Genuine)	0
Fake (Deceptive)	1

E. Text Representation

Tokenisation**:** Converts cleaned text into sequences of numerical integers based on word frequency.

$$X = [x_1, x_2, \dots, x_n] \quad \dots (2)$$

Sequence Padding: To maintain uniform input length for the neural network, all sequences are padded to a total length of 100.

$$X_{padded} = [x_1, x_2, \dots, x_n, 0, 0, \dots, 0] \quad \dots (3)$$

F. Proposed Bilt Model

The Bilt network captures bidirectional context by analysing text from start to finish and looping back the other way.

1) Embedding Layer

Transforms integer sequences into dense 128-dimensional vector representations.

- Vocabulary Size (V): 10,000 most frequent words.
- Embedding Dimension (D): 128.

2) LSTM Cell Equations

Each cell utilises three mathematical gating mechanisms:

- Forget Gate: $\boldsymbol{f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f}$ … **(4)**

- Input Gate: $\boldsymbol{i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)}$ … … … … … … … … … … … … … . … … … … **(5)**
- Output Gate: $\boldsymbol{o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)}$ … … … … … … … … … . . … . … … … … **(6)**

G. Model Architecture
The architecture is implemented using TensorFlow[26] and Keras[27].

Table 4. Structure of the Deep Learning Mode

Layer	**Configuration**	**Function**
Embedding Layer	Input: 10000; Output: 128	Converts word indices into dense vectors. Spatial Dropout
Rate: 0.2	Prevents overfitting by shutting off features at random	
BILSTM Layer 1	64 units (128 total)	Captures forward and backward context
BILSTM Layer 2	32 units (64 total)	Sharpens context details from prior stages
Dropout	Rate: 0.5	Randomly deactivates neurons to ensure generalisation
Dense Layer	32 Neurons	Combines features for final classification
Output Layer	1 Neuron (Sigmoid)	Produces a probability value between 0 and 1

H. Training Strategy
- Adam Optimiser: Used for efficient mathematical convergence.
- Binary Cross-Entropy Loss: Standard for yes-or-no outcomes like Real vs. Fake.
- Early Stopping: Enabled to monitor validation loss and prevent overfitting.

I. Evaluation Metrics
Performance is evaluated using standard classification metrics to ensure a comprehensive assessment of the model. Accuracy measures the overall correctness of the model by calculating the proportion of correctly predicted instances out of all predictions. Precision reflects the model's exactness in identifying deceptive reviews, focusing on minimising false positives so that genuine reviews are not incorrectly classified as fake. Recall evaluates the model's ability to detect actual fake reviews by minimising false negatives, ensuring that most deceptive content is successfully identified. The F1-score provides a balanced measure by combining precision and recall, offering a single metric that reflects both the model's accuracy in prediction and its effectiveness in detecting fake reviews.

IV. RESULT AND DISCUSSION
This section presents a rigorous experimental evaluation of the proposed deep learning framework for the identification of deceptive opinion spam. The effectiveness of the Bidirectional Long Short-Term Memory (Bilt) model is assessed using a multi-dimensional approach, incorporating statistical summaries, visual pattern analysis, and standard classification metrics.

A. Experimental Setup and Dataset Statistics
The experiments were conducted using a meticulously cleaned dataset of 48,800 e-commerce reviews, which was split into an 80/20 ratio for training and testing. The dataset was curated to maintain a perfect 50/50 balance between authentic (Real) and deceptive (Fake) reviews to prevent any majority class bias during the learning phase Table 5.

Table 5. Initial Dataset Cleaning Statistics

Description	Number of Reviews
Original dataset size	50000
Duplicate & Null records removed	1200
Final cleaned dataset	48800

B. Exploratory Data Analysis (EDA) and Linguistic Patterns

Before model training, Exploratory Data Analysis was performed to identify the structural footprints left by spammers.

- Review Length Distribution: Analysis revealed that genuine reviews tend to be more descriptive and practical, whereas fake reviews often cluster around overly brief or unnaturally verbose word counts.
- Vocabulary Frequency: Authentic reviews utilise a context-rich vocabulary focused on product attributes and logistical details like "delivery,""quality," and "price".
- Deceptive Markers: In contrast, deceptive reviews frequently over-rely on hyperbolic adjectives such as "amazing," "perfect," or "worst" to compensate for a lack of real physical experience with the product.

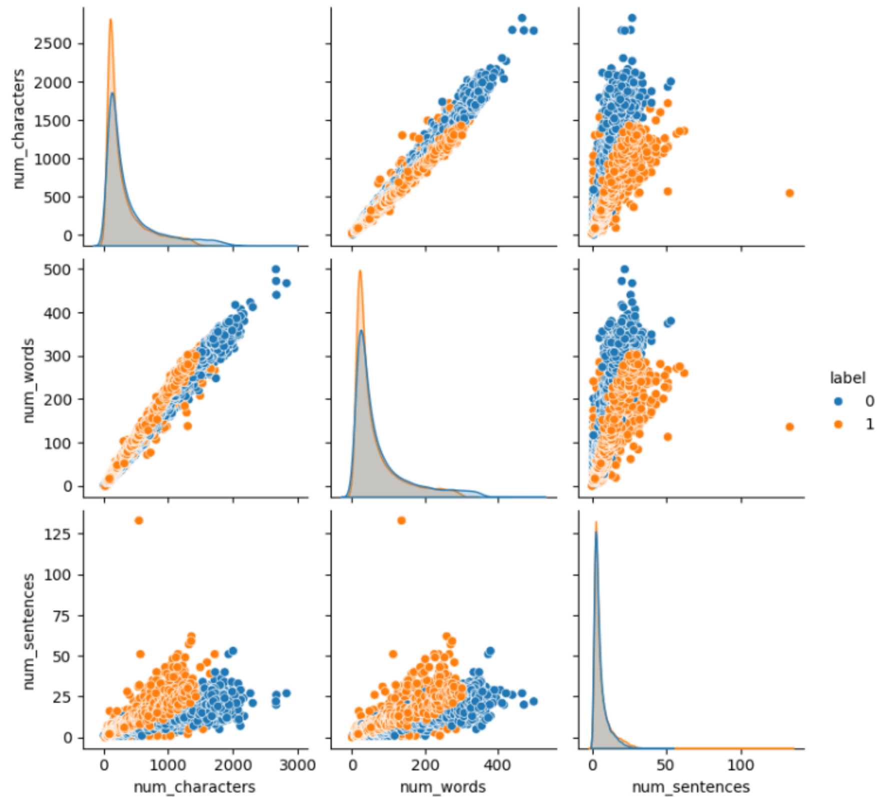


Fig 3. Pairwise distribution and relationship analysis of textual features (*num_characters*, *num_words*, *num_sentences*) across class labels (0: genuine, 1: spam).

Fig 2 illustrates the pairwise relationships and distributions of key textual features, namely number of characters, number of words, and number of sentences, for both genuine (label 0) and spam (label 1) reviews. The diagonal plots represent the individual feature distributions, showing that spam reviews tend to have relatively shorter lengths compared to genuine reviews. The scatter plots indicate a strong positive correlation between the number of characters and the number of words, suggesting consistency in text length across both classes. However, when analyzing sentence counts, spam reviews appear more densely clustered at lower values, whereas genuine reviews show a comparatively wider spread. As observed in Fig. 2, these patterns highlight that spam reviews are generally shorter and less complex, which can serve as an important discriminative signal for opinion spam detection models.

C. Comparative Performance Analysis

The predictive proficiency of the proposed BILSTM model was evaluated against a standard LSTM baseline to measure the impact of bidirectional context processing in Table 6.

Table 6. Model Performance Comparison

Metric	LSTM Model	BILSTM Model (Proposed)
Test Accuracy	91.57%	91.90%
Precision (Fake)	0.91	0.93

Recall (Fake)	0.92	0.90
F1-Score (Fake)	0.92	0.92

The BiLSTM model achieved a superior precision of 0.93 for the "Fake" class, indicating that it is more reliable at flagging fraud with significantly fewer false alarms for genuine users.

D. Training and Validation Analysis

The model was trained using the Adam optimiser and the Binary Cross-Entropy loss function. An Early Stopping callback was implemented to monitor validation loss, which halted training at the 5th epoch to prevent overfitting in Fig 3.

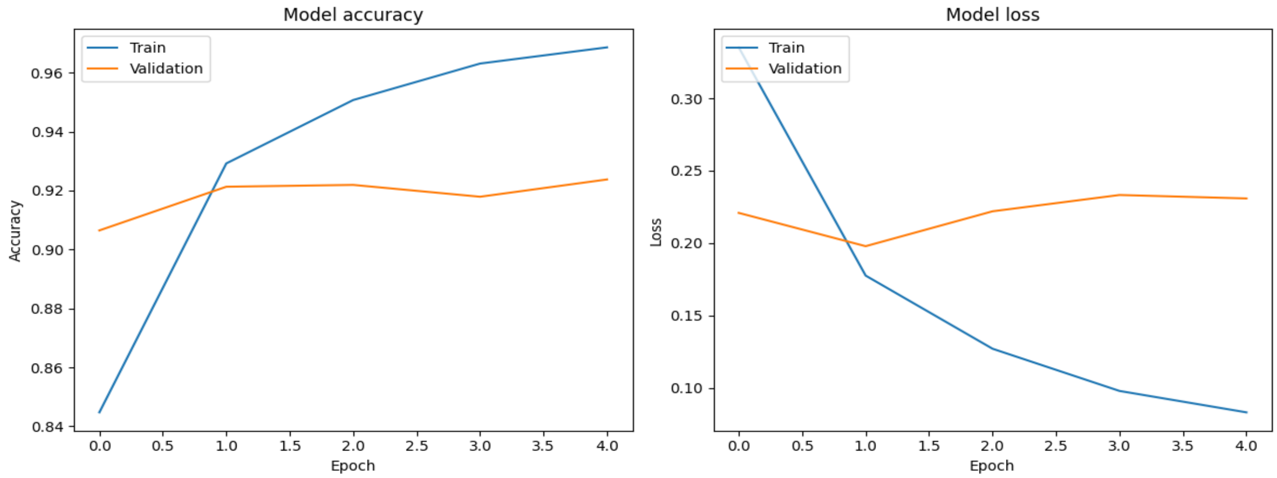


Fig 4. Model Performance Curves Showing Accuracy and Loss During Training

Accuracy Convergence: Training accuracy peaked at 97.15%, while validation accuracy remained stable at 92.38%.
Loss Reduction: The rapid drop in training loss during early epochs confirms that the Bilt architecture effectively captures deceptive cues from raw text.

E. Confusion Matrix Interpretation

The confusion matrix for the BiLSTM model shows that it correctly identified 3661 fake reviews as fake (True Positives) and accurately classified 3782 real reviews as real (True Negatives). However, the model incorrectly flagged 264 real reviews as fake (False Positives), indicating a small number of genuine reviews were misclassified. Additionally, 392 fake reviews were missed and classified as real (False Negatives), suggesting some deceptive reviews were not detected. Overall, the model demonstrates strong performance with a high number of correct predictions, while maintaining relatively low misclassification rates in Fig 3.

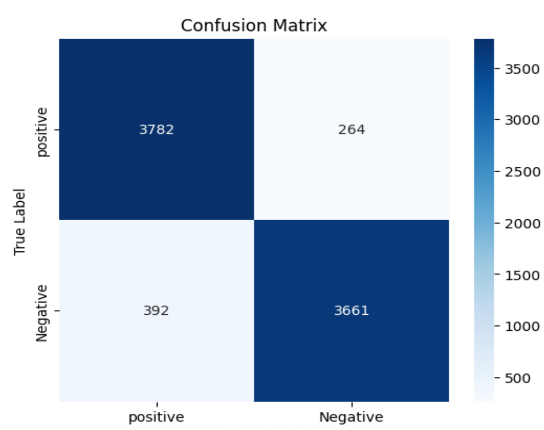


Fig 5. Confusion matrix representing classification results of the proposed BILSTM model

The model demonstrates a high degree of correctness in verifying authenticity, with errors distributed fairly evenly across both classes, suggesting a robust and unbiased classification system.

4.6 Discussion of Findings

The experimental results confirm that sequential deep learning models are highly effective for e-commerce fraud detection. The primary advantage of the Bilt model lies in its ability to analyse sequences in both forward and backward directions, capturing the complete semantic context. This enables the detection of subtle linguistic anomalies, such as sarcasm or unnatural emotional shifts, which traditional unidirectional models or keyword-based systems often fail to identify.

V. CONCLUSION AND FUTURE SCOPE

A working model was developed to categorise e-commerce product feedback into Real or Fake categories using deep learning. Built around a Bidirectional Long Short-Term Memorynetwork, the system handles word sequences naturally, identifying deceptive patterns across phrases as sentences unfold.

A. Summary of the Research Work

This study presented a context-aware detection approach using both an optimised Logistic Regression baseline and an advanced BILSTM model. The methodology included cleaning raw textual data through natural language techniques like stemming, stop-word removal, and tokenisation. The Bilt model, specifically designed to process words in both forward and backward directions, achieved a test accuracy of 91.90%, outperforming the traditional baseline models.

The key findings of this research highlight the effectiveness of deep learning techniques in detecting deceptive reviews. Sequential models such as BiLSTM demonstrate superior performance in text classification compared to traditional methods that rely on simple word frequency, as they can better capture contextual relationships within the text. The use of bidirectional processing enables the model to analyze text both forward and backward, allowing it to identify subtle inconsistencies and unnatural patterns often present in fake reviews. Additionally, preprocessing plays a crucial role in improving model performance, as removing noise such as irrelevant symbols and meaningless words enhances the clarity of input data without altering the core methodology. Overall, the consistent results indicate that the system successfully recognizes deceptive linguistic patterns, achieving high precision and reliability in distinguishing genuine opinions from spam.

B. Future Scope

Future work can further enhance the system's accuracy and applicability through several planned developments:

The future scope of this research includes several enhancements to improve the effectiveness and applicability of the model. Advanced architectures such as transformer-based models like BERT, RoBERTa, or GPT can be explored to further enhance precision in detecting more sophisticated and subtle forms of deceptive reviews. Expanding the system to support multiple languages by retraining it on localized datasets would allow it to be used across global platforms, overcoming the current limitation of English-only analysis. Additionally, moving beyond simple binary classification to more granular labeling can help distinguish between different types of spam, such as bot-generated content and paid human reviews. The integration of hybrid models, combining CNNs and BiLSTM, can further improve performance by capturing both local word relationships and broader contextual patterns within sentences. Overall, this approach holds strong potential for real-world deployment in areas such as real-time e-commerce moderation, social media brand monitoring, and customer trust evaluation systems.

VI. ACKNOWLEDGMENT

The author is thankful to the Department of Computer Science at Government College (Autonomous), Rajahmundry, for their support and encouragement throughout the course of this work.

REFERENCES

[1] S. K. DUVVURI, Applications of Artificial Intelligence Across Domains . Commissionerate of Collegiate Education, Government of Andhra Pradesh , 2026. doi: 10.5281/zenodo.18623057.

[2] J. A. Chevalier and D. Mayzlin, “The Effect of Word of Mouth on Sales: Online Book Reviews,” Journal of Marketing Research, vol. 43, no. 3, pp. 345–354, 2006, doi: 10.1509/jmkr.43.3.345.

[3] M. Luca, “Reviews, reputation, and revenue: The case of Yelp.com,” Harvard Business School Working Paper, no. 12–016, 2016, [Online]. Available: https://ssrn.com/abstract=1928601

[4] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp Fake Review Filter Might Be Doing?," in Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 2013. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14432

[5] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. draft. Stanford University, 2023. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/

[6] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008. [Online]. Available: https://nlp.stanford.edu/IR-book/

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019, pp. 4171–4186.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.

[9] B. Liu, Sentiment analysis and opinion mining. Morgan & Claypool Publishers, 2012. doi: 10.2200/S00416ED1V01Y201204HLT016.

[10] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, 2008, doi: 10.1561/1500000011.

[11] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manag., vol. 24, no. 5, pp. 513–523, 1988, doi: 10.1016/0306-4573(88)90021-0.

[12] H. Li, Z. Chen, B. Liu, W. Wei, and J. Shao, "Learning to identify review spam," in Proceedings of the 22nd International Joint Conference on Artificial Intelligence, 2011, pp. 2488–2493.

[13] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012, pp. 171–175.

[14] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in European Conference on Machine Learning, Springer, 1998, pp. 137–142. doi: 10.1007/bfb0026683.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] D. P. Patinavalasa and D. Suneel Kumar, "Scalable Email Spam Detection Using BiLSTM with Large-Scale Hybrid Datasets," International Journal Of Recent Trends In Multidisciplinary Research, p. 96, Mar. 2026, doi: 10.59256/ijrtmr.20260602016.

[17] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks, vol. 18, no. 5–6, pp. 602–610, 2005, doi: 10.1016/j.neunet.2005.06.042.

[18] N. Jindal and B. Liu, "Opinion spam and analysis," in Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008, pp. 219–230. doi: 10.1145/1341531.1341560.

[19] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a General Rule for Identifying Deceptive Opinion Spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 1566–1576. [Online]. Available: https://aclanthology.org/P14-1147/

[20] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation, vol. 28, no. 1, pp. 11–21, 1972, doi: 10.1108/eb026526.

[21] A. Y. Ng, "Feature selection, L1 regularization, and rotational invariance," in Proceedings of the twenty-first international conference on Machine learning, 2004, p. 78. doi: 10.1145/1015330.1015435.

[22] H. Li, Z. Chen, A. Mukherjee, B. Liu, and J. Shao, "Analyzing and detecting opinion spam on a large-scale dataset via temporal and product features," in Proceedings of the AAAI Conference on Artificial Intelligence, 2015. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9266

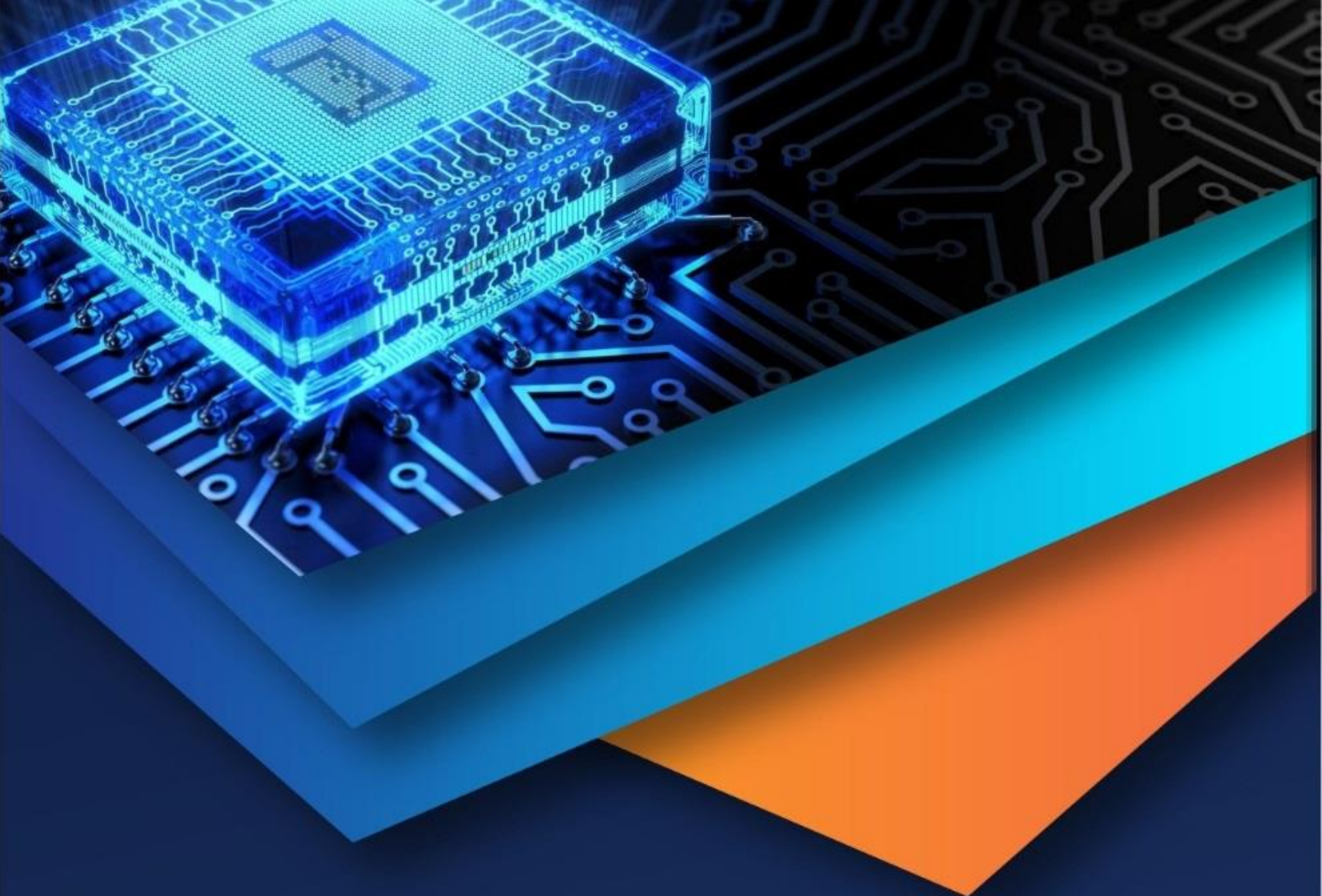
[23] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp Fake Review Filter Might Be Doing?," in Proceedings of the International AAAI Conference on Web and Social Media, 2013, pp. 409–418. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14432

[24] S. Rayana and L. Akoglu, "Collective Opinion Spam Detection: Bridging Review Networks and Metadata," in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2015, pp. 985–994. doi: 10.1145/2783258.2783370.

[25] J. Camacho-Collados and others, "A survey on deep learning for fake review detection," arXiv preprint arXiv:1912.01234, 2019.

[26] TensorFlow Team, "TensorFlow Documentation," 2024.

[27] Keras Team, "Keras Documentation," 2024.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY