[1] Maysara Mazin Badr Alsaad

[2] Prof. Hiren Joshi

Transformer-Based Language Deep Learning Detection of Fake Reviews on Online Products

Journal of Electrical Systems

Abstract: - In e-commerce, spotting fake reviews is vital for ensuring trust among consumers. However, identifying them poses a challenge because fake reviews are often crafted to seem genuine, and the sheer volume makes thorough checks difficult. Prior methods involve basic strategies like grammar checks or pattern analysis, but they fell short due to fake reviews generation methods are becoming increasingly sophisticated. Even machine learning, while helpful, struggle to pinpoint subtle fake reviews accurately. This has led to a shift toward deep learning algorithms, which show promise in handling the complexities that traditional methods cannot manage accurately. Specifically, transformer models like BERT, RoBERTa, and XLNet have emerged as potential solutions. This study evaluates the effectiveness of these models in distinguishing between human-generated and computer-generated reviews. RoBERTa displays high accuracy but requires longer learning periods, while BERT and XLNet offer decent accuracy with varying error rates. The investigation delves into these deep learning models to ascertain their capability to spot fake reviews across different scenarios within online platforms. RoBERTa achieves the highest accuracy among the models, reaching 97.1%. It also demonstrates a lower Type I error rate at 2.2%, although its Type II error remains at a moderate level.

Keywords: RoBERTa, BERT, XLNet, BiLSTM, CNN, Transform Learning, Fake Reviews.

I. Introduction

The rise of online commerce has revolutionized how people buy and sell goods and services [1]. Customers increasingly rely on these platforms, sharing their thoughts and experiences through reviews after making purchases [2]. These reviews wield significant influence, shaping the shopping experiences of potential buyers. Positive feedback can drive more sales, while negative opinions can dampen interest in products or brands, impacting profits in the e-commerce realm [3]. Given that reviews have become integral to social media and e-commerce experiences, the quality and authenticity of these reviews hold immense importance for brands, e-commerce platforms, and other stakeholders [4]. Therefore, detecting fake reviews is a critical issue that demands urgent attention to safeguard the integrity of online businesses.

Merchants heavily consider public opinions to tweak their business strategies and product quality. However, amidst these authentic opinions, there's a growing concern about spam content infiltrating reviews. Spam content, essentially irrelevant or manipulative data, sneaks into reviews for advertising, promotion, or financial gain [5]. Consumers, aiming to make informed decisions, often turn to these reviews before buying, making it crucial to distinguish fake reviews from genuine ones [6]. Detecting fake reviews is a vital aspect of natural language processing, specifically in e-commerce settings. These fake opinions mislead consumers, leading to incorrect purchasing decisions and affecting product revenue [7]. Identifying such spam or fake reviews involves recognizing various patterns: from reviews lacking detail about the reviewer to those resembling duplicates, being overly brief, or displaying unusual uploading patterns. Additionally, these fake reviews often exaggerate emotions, using excessively positive or negative language that may seem not generated by human [8].

There are two primary methods to create fake reviews. First approach involves engagement of humans, where individuals are paid to craft seemingly genuine reviews for products they've never actually encountered. Second, there's the computer-generated method employing text-analysis algorithms to automate the creation of these deceptive reviews [9]. Historically, human-generated fake reviews were traded as commodities in a "market of fakes," but advancements in technology, particularly in natural language processing and machine learning, have encouraged automated fake review production, significantly reducing costs compared to human-generated ones [9]. Fake reviews are a big problem for online shopping [10]. They make it hard for people to trust what they read. Real reviews help buyers make good choices, and companies use them to improve their products [11]. But fake reviews can mess this up by tricking people and affecting which products show up first. It's not just about reputation; it can also hurt companies' money and buyers' trust. That's why fake reviews are a huge deal in online shopping [12].

[1*] Corresponding author: Department of Computer Science, Rollwala Computer Centre, Gujarat University, Navarangpura, Ahmedabad 380009, Gujarat, India. Email: maysara@gujaratuniversity.ac.in

[2] Professor, Department of Computer Science, Rollwala Computer Centre, Gujarat University, Navarangpura, Ahmedabad 380009, Gujarat, India. Email: hdjoshi@gujaratuniversity.ac.in

Numerous researchers have delved into studies concerning the identification of fake or spam reviews due to their profound impact on both consumers and e-commerce enterprises. The process of deriving crucial attributes from text data is known as feature engineering [13]. In the realm of fake review detection, studies typically focus on two approaches: a behavioral orientation, emphasizing characteristics of spamming reviewers, and a linguistic approach, centered on features within individual reviews.

Stylometric-based features play a pivotal role in discerning the writing style of reviewers and uncovering deceptive content. These features encompass two categories. The first includes lexical characters, like character count (N), numeric character ratio to N, letter proportion to N, uppercase letter ratio to N, space rate to N, tab ratio to N, alphabet occurrences (A-Z), and frequency of special characters [14]. The second type involves lexical word-based features, such as word count (T), ratio of words in the sentence, token length proportion, character rate in words to N, ratio of short words (1-3 characters), and word length ratio. Stylometric aspects reveal the reviewer's stylistic choices, including syntactic elements like punctuation frequency [15].

Another determinant is the maximum daily review count, with about 70% of fraudsters producing more than five reviews per day, a contrast to 90% of genuine users who typically submit only one review per purchase. This metric aids in identifying potential spammers [16].

Additionally, the proportion of positive reviews holds significance. Approximately 80% of fraudulent reviewers tend to compose 85% of their reviews as positive. Hence, a notably high proportion of positive reviews could signal deceitful behavior [17].

Review length also serves as a distinguishing factor. Research shows that 75% of spammers struggle to craft reviews surpassing 136 words, whereas over 90% of honest reviewers tend to write reviews with at least 200 words [2].

The deviation in reviewer ratings from the average rating of truthful reviewers is another noteworthy marker. Detecting variations in users' rating patterns can aid in the identification of potential spam activity. In summary, the challenges of fake review detection can be as follows:

- Variety of Fake Reviews: Fake reviews come in various forms, from overly positive or negative reviews to subtle manipulations that are challenging to differentiate from authentic feedback.
- Behavioral Changes: Fraudsters adapt their behavior based on detection methods. When specific patterns or criteria for identifying fake reviews are recognized, scammers alter their tactics to evade detection [18].
- Temporal Aspects: Timing and frequency of reviews can be indicative of fraudulent behavior, but this can also overlap with genuine reviewing behavior, making it challenging to establish clear-cut criteria.
- Subjectivity in Reviews: Reviews are inherently subjective, making it challenging to establish universal criteria for distinguishing genuine opinions from fake ones, especially when dealing with nuanced language.
- Constant Evolution: As detection methods improve, so do the tactics of those creating fake reviews. This ongoing arms race necessitates continual adaptation and innovation in detection techniques.

In light of the challenges, this paper aims to use several deep learning models and several Transformer-based Language models to detect fake reviews from a large dataset generated from [19]. This paper presents the use of three transform model which are Bidirectional Encoder Representations from Transformers (BERT), Robustly Optimized BERT-Pretraining Approach (RoBERTa), and Generalized Autoregressive Pretraining for Language Understanding (XLNet). These models are compared with two deep learning models which are Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Network (CNN). By examining simulation results and comparing them to existing literature for these different deep learning techniques, the contributions of this paper can be summarized as follows:

- Conduct a comprehensive benchmark analysis by employing multiple Transformer-based Language Deep Learning models, comparing their performance with other deep learning algorithms and traditional machine learning used in literature.
- Evaluating the advantages of using the RoBERTa model as one of the promising models in fake reviews detection in terms of accuracy.
- Address existing study limitations by utilizing a deceptive opinion dataset and propose future directions aimed at enhancing the filtration of spam reviews.

The remainder of the paper is organized as follows: Section 2 conducts a literature review of relevant articles, examining their approaches and discoveries. Section 3 outlines the methodology adopted in this study. Section 4 presents the simulation outcomes and analysis, focusing on accuracy and error rates. Finally, Section 5 offers conclusions and suggests future recommendations.

II. Literature Review

As e-commerce platforms expand, the volume of online reviews grows, but the rise in fake reviews is outpacing the improvement in review quality. Malicious false reviews are causing increasing harm to retailers and consumers, making it challenging for users to distinguish helpful reviews in a sea of information. Consequently, the reliability of online reviews, crucial in guiding purchase decisions, is diminishing, potentially eroding credibility and traffic on e-commerce platforms. There are a lot of works in the literature based on fake detection using transform-learning. This means that there is a big challenge in this topic and a high potential for more advancement. Table 2.1 shows the summary of the literature.

The work in [20] proposes a comprehensive Sentiment Analysis (SA) model that handles the scarcity of annotated data, leading to potential misclassification in sentiment analysis issues by introducing a Bi-directional Encoder Representation from Transformers (BERT) based Convolution Bi-directional Recurrent Neural Network (CBRNN). This model combines syntactic, semantic, sentimental, and contextual information. The process involves zero-shot classification for polarity scores, BERT for semantic understanding, and a neural network employing dilated convolution and BiLSTM to capture local and global context. The model's evaluation across diverse text datasets shows 0.97 results in accuracy, indicating its effectiveness in performing SA on social media reviews without losing information.

The work in [21] delves into the significant impact of online reviews on decision-making and emphasizes the need for trustworthy evaluations. Detecting fake reviews becomes crucial, prompting the exploration of effective methods. Using a labeled Deceptive Opinion dataset, the study employs semi-supervised language processing techniques. By merging sentiment analysis and readability, the research enhances fake review detection. Transformer models like BERT, RoBERTa, Transformer-XL (XLNet), and Cross-lingual Language model–RoBERTa (XLM- RoBERTa) are employed, elevating the screening process for fake reviews. This approach extracts and categorizes features from product reviews, thereby improving review filtering efficiency. The study demonstrates that the application of transformer models significantly enhances spam review filtering compared to existing machine learning and deep learning models. Simulation results show 0.912, 0.9713, and 0.982 accuracy results come from BERT, RoBERTa, and XLM- RoBERTa models.

Various strategies to detect fake news have been explored, spanning content-based, social context-based, image-based, sentiment-based, and hybrid context-based classifications. The work in [22] focuses on proposing a model for fake news classification, specifically centered on news headlines using a content-based classification approach. The model integrates a BERT model connected to an LSTM layer. Evaluation and training were conducted using the FakeNewsNet dataset, comprising PolitiFact and GossipCop sub-datasets. Comparative analysis with base classification models and a vanilla BERT model, trained under similar constraints as the proposed model but without an LSTM layer, revealed a performance of 0.8625 in accuracy.

The work in [23] developing a machine learning model that filters and classifies reviews, focusing on determining their authenticity and usefulness. Simulation results show that supervised learning for assessing review usefulness yielded 81-85% accuracy based on different analysis algorithms such as SVM, LightGBM Classifier, Random Forest Classifier, Logistic Regression, K-Nearest Neighbors Classifier, Gradient Boosting Classifier, XGBoost Classifier, Gaussian NB, Extra Trees Classifier, and Decision Tree Classifier.

The work in [19] delves into both the creation and identification of fake reviews. Initially, two language models, ULMFiT and GPT-2, were tested to generate fake product reviews using an Amazon e-commerce dataset. Subsequently, leveraging the superior performance of GPT-2, a dataset was fashioned for the classification of fake reviews. The research demonstrates that machine classifiers can remarkably excel at this task, outperforming human raters who displayed lower accuracy and consensus compared to the algorithms tested. Moreover, the model effectively identified human-generated fake reviews. These findings suggest that while fake review detection poses challenges for humans, employing "machines against machines" proves effective in this endeavor. The implications of this study span consumer protection, safeguarding firms against unfair competition, and emphasizing the responsibility of review platforms. 0.9664 accuracy result has been obtained using the fakeRoBERTa proposed model.

The work in [24] conducted a comprehensive assessment comparing various deep learning models like CNN, RNN, and Bi-directional-LSTM. These models were evaluated based on different word embedding methods, such as BERT, fastText, and Word2Vec. To enhance data, they utilized Easy Data Augmentation, resulting in two datasets: the original and an augmented version. They considered two setups: a 5-class and a compressed 3-class version. The results revealed that the most accurate prediction models stemmed from Neural Network-based approaches using Word2Vec. Specifically, the CNN-RNN-BiLSTM model displayed 0.96 as the highest accuracy obtained compared to other studied models.

A comparative analysis of BERT, Hybrid fastText-BiLSTM, and fastText Trigram models to address challenges in achieving more precise sentiment predictions of fake reviews has been performed [25]. Introducing fine-tuned BERT and Hybrid fastText-BiLSTM models for extensive datasets, the study demonstrates that the proposed fine-tuned BERT model outperforms other DL models and gives a 0.91 accuracy result.

The work in [26] conducts a comparative analysis between traditional machine learning models and newer transformer-based techniques using a dataset of customer reviews sourced from Trustpilot. The findings indicate that transformer-based models surpass traditional ones, achieving over 0.98 accuracy.

This work aims to use several transform-learning models to detect fake reviews and compare the results obtained with those available in the literature.

Table 2.1 Literature Review Summary

Ref.	Model	Highest Accuracy	Concerns
[20]	BERT, CBRNN	0.97	-Depends on two stages, sentiment and prediction. -Leads to higher complexity.
[21]	BERT, RoBERTa, XLNet, XLM-Roberta	0.982	Higher Learning time
[22]	BERT+LSTM	0.8625	Limited to fake news not fake reviews.
[23]	Machine Learning	0.85	Limited to machine learning algorithms
[19]	fakeRoBERTa	0.9664	Higher Learning time
[24]	CNN, RNN, BiLSTM	0.96	A hybrid of three models leads to higher complexity.
[25]	BERT, BiLSTM	0.91	A hybrid of three models leads to higher complexity.
[26]	RoBERTa	0.98	Higher Learning time

III. Methodology

The proposed analytical framework builds upon existing research by integrating Transformer models with distinct linguistic features like readability and sentiment mining. This approach aims to categorize reviews from deceptive sources, thereby enhancing the credibility of user-generated online content. The process involves two phases outlined as shown in "Fig. 3.1,". Each one of these phases has several operations that are performed to attain the desired outcomes of the study.

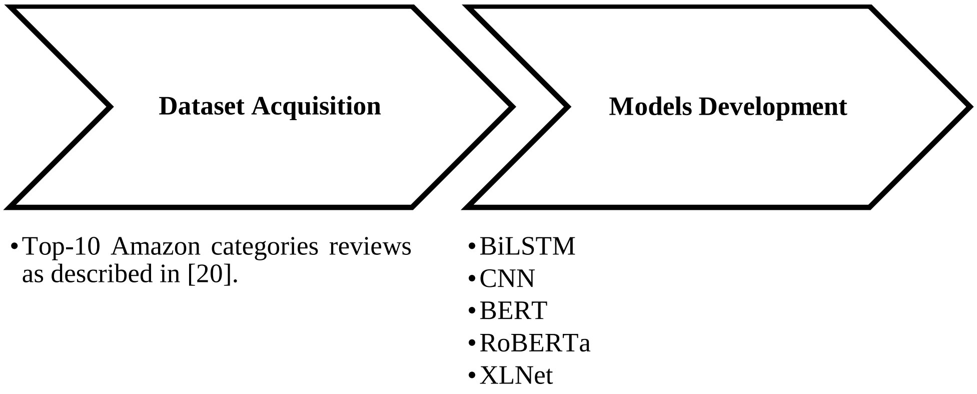


Fig. 3.1 Methodology Phases

3.1 Methodology Phases

There are two phases in this work:

3.1.1 Data acquisition

The dataset used in this work is obtained from [19]. It focuses on the Top-10 Amazon categories with the highest volume of product reviews, covering 88.4% of the baseline dataset. Each category sees the generation of 2000 reviews via a fine-tuned GPT-2 language model, aimed at diversifying sentence lengths. The process involves setting a starting word count and creating discrete length buckets based on the distribution of review lengths in the dataset. This ensures that generated reviews mirror the original dataset's length and category proportions.

The resulting dataset comprises 20,000 artificially generated (fake) reviews and 20,000 human-written (real) reviews from Amazon's dataset, totaling 40,000 reviews. This quantity is substantial for text classification tasks, even exceeding sizes used in binary text classification tasks. The dataset distinguishes between Computer-Generated reviews (CG) and Original Reviews (OR), enabling the task of detecting fake reviews.

3.1.2 Models development

There are five models used in this study as follows:

A. BERT model

BERT, a sophisticated deep-learning language processing model, surpasses previous language models by a significant margin. It operates on the principle of contextual understanding from both left and right contexts across all levels. BERT proves to be a fundamental yet powerful tool, demonstrating potential across various machine learning tasks. A fine-tuned BERT model requires only an additional layer to perform a wide array of functions. It employs a Masked Language Model (MLM) based on the masking of random words within input, predicting their context-based IDs. Unlike any other model, BERT comprehends contextual representations from both sentence ends, utilizing a 30K vocabulary of character-level Byte-Pair Encoding for tokenization. Special tokens ([CLS] and [SEP]) are added at the sequence's start and end, aiding in text categorization techniques like Next Sentence Prediction using the [CLS] token and providing separation with the [SEP] token.

B. RoBERTa model

The RoBERTa model is part of the broader Transformers family like BERT. This family of models aimed to address issues with long-range dependencies in sequence-to-sequence modeling. RoBERTa stands out with a larger vocabulary comprising 50K sub-word units and utilizes byte-level Byte-Pair Encoding. Its enhancements over BERT include training on more extensive data and longer sequences. The RoBERTa tokenizer incorporates special tokens like [CLS] and [SEP], marking sentence beginnings and endings, respectively. The [PAD] token aids in adjusting text length for optimal vector size. Through encoding raw text, RoBERTa's tokenizer generates input IDs representing token indices and numerical representations, along with an attention mask used to group sequences for optional input. Essentially, RoBERTa's base layers aim to provide meaningful word embedding, facilitating subsequent layers to extract valuable information efficiently.

C. XLNet model

A BERT-based autoregressive language model (XLNet) addresses the challenge of concurrent predictions faced by BERT. While BERT learns by predicting masked words simultaneously, it doesn't grasp relationships between these predictions. XLNet resolves this by integrating a permutational language model while retaining BERT's bidirectionality. It achieves word prediction by exploring all possible word permutations within a sequence, allowing for learning in a sequential and autoregressive manner within a random sequence framework. Consequently, XLNet consistently surpasses BERT's performance on the GLUE benchmark by 2–13 percent. It also uses similar tokens ([CLS] and [SEP]) for classification and separation purposes like BERT.

D. BiLSTM model

The Bidirectional Long Short-Term Memory (BiLSTM) model is a recurrent neural network (RNN) architecture used for fake review detection. Similar to other models mentioned earlier, it's designed to understand and process sequential data, making it well-suited for analyzing textual information in reviews. What sets BiLSTM apart is its bidirectional processing capability, allowing it to capture context from both past and future elements within a sequence simultaneously. This bidirectional aspect enables a more comprehensive understanding of textual context, particularly beneficial in discerning nuanced sentiments or deceptive content in reviews.

The BiLSTM model comprises multiple LSTM layers, which excel in capturing long-term dependencies within sequences while mitigating the vanishing gradient problem commonly encountered in traditional RNNs. By employing memory cells that selectively retain or forget information, LSTM units can effectively process and retain essential contextual information over longer sequences, thereby enhancing the model's ability to detect deceptive or misleading content within reviews.

E. CNN model

CNNs excel in feature extraction through convolutional layers, which operate by applying filters to small portions of the input data, capturing local patterns and features within the text. In the context of fake review detection, CNNs can identify specific textual patterns or phrases indicative of deceptive content. By leveraging various filter sizes and pooling layers, CNNs can efficiently learn hierarchical representations of textual information, distinguishing between genuine and fake reviews based on these learned features. Moreover, CNNs are adept at detecting spatial hierarchies within sequences, enabling them to capture intricate relationships among words and phrases. These hierarchies are pivotal in discerning deceptive language or sentiment cues in reviews.

3.2 Work flowchart

“Fig. 3.2,” illustrates the workflow, commencing with dataset collection and distinguishing between human-generated and computer-generated reviews. The dataset is split into an 80% training set and a 20% testing set. The process involves assessing whether the classification model belongs to deep learning models or transformer models. Finally, the performance metrics are computed and contrasted for comparison.

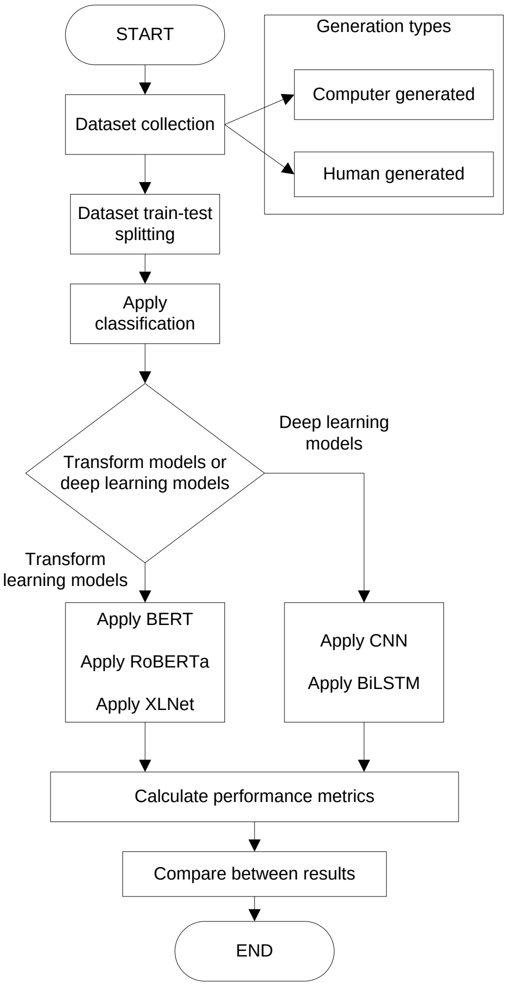


Fig. 3.2 Work Flowchart

3.3 Performance metrics and Simulation Parameters

In this section, a comprehensive review of the experimental outcomes is presented. The assessment of the proposed model’s performance involves deriving overall scores for recall, precision, f-score, and accuracy, computed using the confusion matrix. Additionally, the Receiver Operating Characteristic (ROC) and the Area

Under the Curve (AUC) are employed to gauge the model's effectiveness. These performance metrics hold significant importance in various text classification tasks, including Sentiment Analysis (SA).

The confusion matrix serves as a visual depiction of classification prediction results in any given problem. It summarizes the accurate and erroneous predictions within each class using statistical values, capturing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In this study, reviews are classified into two labels (fake or not). The paper outlines formulas used to calculate performance parameters, contributing to the evaluation metrics presented herein:

- Precision measures the ratio of accurately predicted positive outcomes among the total predicted positive results. This metric determines the classifier's accuracy or exactness in identifying positives. It is alternatively referred to as the positive prediction value and represented by Pre. The precision value can be calculated as:

$$P_{re} = \frac{T_p}{T_p+F_p} \qquad (1)$$

- Recall determines the quantity of accurately predicted positive samples in relation to the overall number of actual positive samples. Represented as Rec, it gauges the completeness of the classifier's positive predictions. The recall value can be as:

$$R_{ec} = \frac{T_p}{T_p+F_n} \qquad (2)$$

- The F-measure amalgamates both Pre and Rec to derive their harmonic mean, represented as Fme. Mathematically, it can be articulated as:

$$F_{me} = \frac{2\times P_{re}\times R_{ec}}{P_{re}+R_{ec}} \qquad (3)$$

- Accuracy represents the ratio of correctly predicted samples among the total samples. Denoted by Acc, it measures the overall correctness of the predictions.

$$A_{cc} = \frac{T_p+ T_n}{T_p+F_p+T_n+F_n} \qquad (4)$$

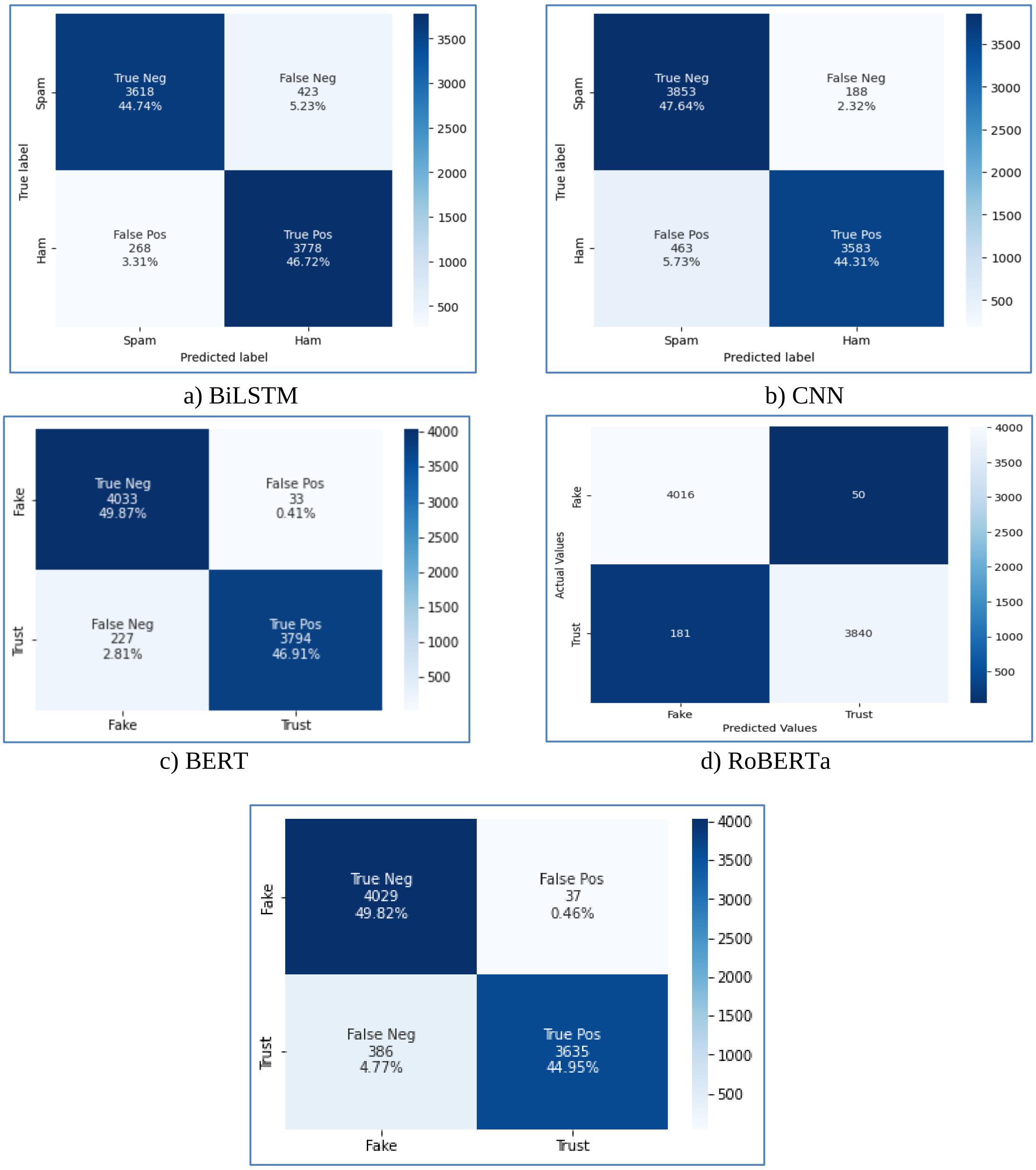
Table 3.1 shows the simulation parameters used in this work. The splitting rate is 80% for training and 20% for testing. This implies that 80% of the dataset (n=32,000) is utilized to train the model, while the remaining 20% (n=8,000) is kept aside specifically for evaluation purposes. Essentially, the test set comprises samples that the model did not encounter or train on during the model training phase. The learning rate used for the 5 models is 0.00005, CNN filters 64, Bi-LSTM units 128, batch size 64, and dense size 32.

Table 3.1 Simulation Parameters

Simulation Parameter	**Values**
Splitting rate	80/20 %
Learning rate	5e-5
CNN filters	64
BiLSTM units	128
Batch size	64
#of data	40000

IV. RESULT AND DISCUSSION

“Figs. 4.1 (a–e)” depict graphical representations of the confusion matrix corresponding to the five models used in this paper. The outcomes from these matrices suggest that the suggested model reduced the variance between the true and predicted labels. They showcase the acquired values for TP, Tn, FP, and Fn.



a) BiLSTM

b) CNN

c) BERT

d) RoBERTa

e) XLNet

Fig. 4.1 Confusion Matrices

Confusion matrices reflect on the values of Acc, Rec, Pre. This reflection is shown in Table 4.1. This table shows the comparison between the five modules used in this paper with respect to the accuracy results.

The RoBERTa model attained the top scores of 0.995 for training accuracy, 0.971 for testing accuracy, and 97 for the AUC value. Thus, it is evident that the proposed model outperformed the other models in the classification ability.

Table 4.1 Accuracy Results

Models	**Training Accuracy (%)**	**Testing Accuracy (%)**	**AUC (%)**
BiLSTM	98.7	91.4	91
CNN	99.3	91.9	92
BERT	99.0	96.7	97
RoBERTa	99.5	97.1	97
XLNet	98.9	94.7	95

Table 4.2 illustrates the f-measure, precision, and recall outcomes for the five models. Notably, RoBERTa exhibits a lower recall percentage of 95% for computer-generated reviews compared to the BERT model's highest value of 99%. This indicates that the RoBERTa model is comparatively less accurate in predicting positive values specifically for computer-generated reviews, implying reduced efficacy in identifying trustworthy reviews. However, it demonstrates significant strength in accurately predicting overall positive values within the entire dataset, achieving a precision of 99%.

Conversely, for original (human) reviews, RoBERTa yields a precision lower than that of BERT, while its recall is the highest. In both scenarios, the F-score remains consistent at around 97%.

Table 4.2 Performance Metrics Results

Models	**Computer Generated Reviews (Class == 1)**			**Original Reviews (Class ==0)**		
	Precision(%)	**Recall(%)**	**F-score(%)**	**Precision(%)**	**Recall(%)**	**F-score(%)**
BiLSTM	93	90	91	90	93	92
CNN	89	95	92	95	89	92
BERT	95	99	97	99	94	97
RoBERTa	99	95	97	96	99	97
XLNet	99	90	95	91	99	95

Table 4.3 displays the learning duration in hours required by each model. Despite the commendable accuracy achieved by the RoBERTa model, its learning duration surpasses that of the other models. It takes 1.94 hours to learn, while the CNN model, which stands as the second-highest accuracy model (99.3%), necessitates only 0.14 hours for learning. This prolonged learning duration signifies greater processing demands. It's important to note that this time is derived from learning using 40,000 samples, indicating that it would increase further with a larger sample size.

Table 4.3 Learning Time Results

Models	**Learning Time (Hour)**
BiLSTM	0.38
CNN	0.14
BERT	0.74
RoBERTa	1.94
XLNet	0.90

Table 4.4 displays the results concerning error rates in predictions. Type I error, known as a false positive, occurs when a model misclassifies a negative instance as positive, erroneously recognizing something as true when it's false, leading to an incorrect positive prediction. Specifically, the RoBERTa model demonstrates the lowest error rate at 2.2% for incorrect positive predictions, while the CNN model registers a Type I error rate of 5.73%, potentially influenced by its shorter learning time.

Additionally, Table 4.4 presents Type II error, where the model incorrectly identifies a positive instance as negative, representing a situation where something true is erroneously considered false, resulting in a missed positive prediction. In this aspect, BERT and XLNet show the lowest error rates at 0.4%, with RoBERTa slightly higher at 0.6%.

Table 4.4 Prediction Error Results

Models	**Type I Error (%)**	**Type II Error (%)**
BiLSTM	3.3	5.2
CNN	5.73	2.3
BERT	2.8	0.4
RoBERTa	2.2	0.6
XLNet	4.7	0.4

V. CONCLUSION

The findings in this paper underscore the significance of model performance in predictive tasks, particularly in differentiating between genuine and computer-generated reviews. Various deep learning models, including RoBERTa, BERT, XLNet, BiLSTM, and CNN, were employed to classify reviews, unveiling insightful outcomes. RoBERTa exhibited superior accuracy in differentiating between genuine and computer-generated reviews, showcasing a lower Type I error rate compared to the CNN model. Meanwhile, BERT and XLNet showcased commendable performance with the lowest Type II error rates, closely trailing RoBERTa.

Furthermore, examining the models' learning times revealed interesting insights. Although RoBERTa boasted the highest accuracy, it incurred a longer learning time compared to CNN, suggesting higher processing requirements. This observation holds significant implications for scalability and resource allocation when considering larger datasets. In summary, while RoBERTa emerged as the top performer in accuracy, the trade-off between accuracy and learning time is a crucial consideration. Models like BERT and XLNet, offering competitive accuracy with lower learning times, present promising alternatives for efficiency in handling sizable datasets. These results shed light on the nuanced performance aspects of transform deep learning models and emphasize the need to balance accuracy with computational demands in real-world applications. In future endeavors, exploring hybrid models integrating various transformer-based architectures promises to yield insightful outcomes by leveraging the unique strengths of each constituent model.

ACKNOWLEDGMENT

The experimental work described in this research was conducted in the lab of Department of Computer Science, Rollwala Computer Centre, Gujarat University.

REFERENCES

[1] G. Bathla, P. Singh, R. K. Singh, E. Cambria, and R. Tiwari, "Intelligent fake reviews detection based on aspect extraction and analysis using deep learning," Neural Computing and Applications, vol. 34, no. 22, pp. 20213-20229, 2022.

[2] S. N. Alsubari et al., "Data analytics for the identification of fake reviews using supervised learning," Computers, Materials & Continua, vol. 70, no. 2, pp. 3189-3204, 2022.

[3] C. Jing-Yu and W. Ya-Jun, "Semi-supervised fake reviews detection based on aspamgan," Journal of Artificial Intelligence, vol. 4, no. 1, pp. 17-36, 2022.

[4] A. Q. Mir, F. Y. Khan, and M. A. Chishti, "Online Fake Review Detection Using Supervised Machine Learning And Bert Model," arXiv preprint arXiv:2301.03225, 2023.

[5] H. Tufail, M. U. Ashraf, K. Alsubhi, and H. M. Aljahdali, "The effect of fake reviews on e-commerce during and after Covid-19 pandemic: SKL-based fake reviews detection," Ieee Access, vol. 10, pp. 25555-25564, 2022.

[6] Ş. Ö. Birim, I. Kazancoglu, S. K. Mangla, A. Kahraman, S. Kumar, and Y. Kazancoglu, "Detecting fake reviews through topic modelling," Journal of Business Research, vol. 149, pp. 884-900, 2022.

[7] W. Zhang, R. Xie, Q. Wang, Y. Yang, and J. Li, "A novel approach for fraudulent reviewer detection based on weighted topic modelling and nearest neighbors with asymmetric Kullback–Leibler divergence," Decision Support Systems, vol. 157, p. 113765, 2022.

[8] M. Lee, Y. H. Song, L. Li, K. Y. Lee, and S.-B. Yang, "Detecting fake reviews with supervised machine learning algorithms," The Service Industries Journal, vol. 42, no. 13-14, pp. 1101-1121, 2022.

[9] 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.

[10] Y. Liu, Z. Sun, and W. Zhang, "Improving fraud detection via hierarchical attention-based Graph Neural Network," Journal of Information Security and Applications, vol. 72, p. 103399, 2023.

[11] Z. Shunxiang, Z. Aoqiang, Z. Guangli, W. Zhongliang, and L. KuanChing, "Building fake review detection model based on sentiment intensity and PU learning," IEEE Transactions on Neural Networks and Learning Systems, 2023.

[12] M. Kumar and H. K. Sharma, "A GAN-based model of deepfake detection in social media," Procedia Computer Science, vol. 218, pp. 2153-2162, 2023.

[13] M. M. B. Alsaad and H. Joshi, "SUPERVISED MACHINE LEARNING-BASED FAKE REVIEW DETECTION: A COMPARATIVE EVALUATION OF FEATURE SELECTION APPROACHES."

[14] A. Iqbal, M. A. Rauf, M. Zubair, and T. Younis, "An Efficient Ensemble approach for Fake Reviews Detection," in 2023 3rd International Conference on Artificial Intelligence (ICAI), 2023, pp. 70-75: IEEE.

[15] V. Attri, I. Batra, and A. Malik, "Enhancement of Fake Reviews Classification Using Deep Learning Hybrid Models," Journal of Survey in Fisheries Sciences, vol. 10, no. 4S, pp. 3254-3272, 2023.

[16] F. A. Ramadhan, R. R. P. Ruslan, and A. Zahra, "Sentiment Analysis Of E-Commerce Product Reviews For Content Interaction Using[1] G. Bathla, P. Singh, R. K. Singh, E. Cambria, and R. Tiwari, "Intelligent fake reviews detection based on aspect extraction and analysis using deep learning," Neural Computing and Applications, vol. 34, no. 22, pp. 20213-20229, 2022.

[17] H. M. Alawadh, A. Alabrah, T. Meraj, and H. T. Rauf, "Semantic Features-Based Discourse Analysis Using Deceptive and Real Text Reviews," Information, vol. 14, no. 1, p. 34, 2023.

[18] G. Shahariar, S. Biswas, F. Omar, F. M. Shah, and S. B. Hassan, "Spam review detection using deep learning," in 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 0027-0033: IEEE.

[19] J. Salminen, C. Kandpal, A. M. Kamel, S.-g. Jung, and B. J. Jansen, "Creating and detecting fake reviews of online products," Journal of Retailing and Consumer Services, vol. 64, p. 102771, 2022.

[20] S. T. Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," Array, vol. 14, p. 100157, 2022.

[21] S. Kanmani and S. Balasubramanian, "Leveraging Readability and Sentiment in Spam Review Filtering Using Transformer Models," Computer Systems Science & Engineering, vol. 45, no. 2, 2023.

[22] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake News Classification using transformer based enhanced LSTM and BERT," International Journal of Cognitive Computing in Engineering, vol. 3, pp. 98-105, 2022.

[23] W. Choi, K. Nam, M. Park, S. Yang, S. Hwang, and H. Oh, "Fake review identification and utility evaluation model using machine learning," Frontiers in artificial intelligence, vol. 5, p. 1064371, 2023.

[24] V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, "A deep learning approach in predicting products' sentiment ratings: a comparative analysis," The Journal of Supercomputing, vol. 78, no. 5, pp. 7206-7226, 2022.

[25] A. Chinnalagu and A. K. Durairaj, "Comparative Analysis of BERT-base Transformers and Deep Learning Sentiment Prediction Models," in 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), 2022, pp. 874-879: IEEE.

[26] L. Davoodi and J. Mezei, "A Comparative Study of Machine Learning Models for Sentiment Analysis: Customer Reviews of E-Commerce Platforms," 2022.