International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 2, Mar-Apr 2025

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Fake Reviews Identification Using Deep Learning Techniques

Omkar Singh*, Swati Singh**, Sweety Rawat***, Sravani Nirati****

*(Data Science, Thakur College of Science & Commerce, Mumbai, India
Email: omkarsingh@tcsc.edu.in)

**(Data Science,Thakur College of Science & Commerce, Mumbai, India
Email: swatisingh2466@gmail.com)

***(Data Science, Thakur College of Science & Commerce, Mumbai, India
Email: rawatsweety177@gmail.com)

****(Data Science, Thakur College of Science & Commerce, Mumbai, India
Email: sravaninirati07@gmail.com)

----------**********----------

Abstract:

Fake online reviews pose a significant challenge to e-commerce platforms by misleading consumers and damaging business credibility. These reviews distort market perceptions, influence purchasing decisions, and manipulate product ratings. Deep learning techniques have emerged as powerful tools to detect fraudulent reviews with higher accuracy and efficiency.

This study explores the application of advanced deep learning models—Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Bidirectional Encoder Representations from Transformers (BERT)—for fake review detection. Using the "Fake Reviews Dataset" from Kaggle, consisting of 40,433 reviews, we evaluate the effectiveness of these models with enhanced text preprocessing techniques, including tokenization, stop word removal, and lemmatization.

The LSTM model achieves superior accuracy, effectively capturing sequential dependencies in text. achieving higher accuracy and effectively capturing sequential dependencies in textual data. CNN is utilized for pattern recognition, RNN processes sequential text-based dependencies, and BERT is explored for its contextual understanding of language. The evaluation metrics—precision, recall, and F1-score—further validate the performance of each deep learning model.

The findings of this research emphasize the potential of deep learning in strengthening the reliability of online reviews, fostering transparency in e-commerce, and protecting consumer trust. This study contributes to the advancement of fake review detection methodologies, highlighting the significance of deep learning in combating digital misinformation.

Keywords — **Fake Review Detection, Deep Learning, LSTM, CNN, RNN, BERT, Natural Language Processing, Text Classification, E-commerce Fraud.**

----------**********----------

I. INTRODUCTION

The rapid growth of e-commerce has made online reviews a critical factor in consumer decision-making, influencing trust, business credibility, and market competition. However, the increasing prevalence of fake reviews threatens the integrity of online platforms, misleading consumers and

distorting fair competition. Traditional rule-based and machine learning approaches struggle to detect fake reviews due to evolving deceptive techniques and language complexity.

Deep learning has emerged as a powerful solution, enabling models to automatically learn intricate patterns in textual data. This research explores the effectiveness of deep learning models—CNN, RNN, LSTM, and BERT—in detecting fake reviews by capturing contextual and semantic nuances. By leveraging these advanced techniques, we aim to enhance the accuracy and reliability of fake review detection, ultimately preserving trust in online platforms.

II. Literature Review

Fake Reviews Detection Using Supervised Machine Learning paper proposes a machine learning approach to detect fake reviews using supervised machine learning. The paper applies several machine learning classifiers to identify fake reviews based on the content of the reviews and extracted features from the reviewers. The paper compares the performance of several experiments done on a real Yelp dataset of restaurant reviews with and without features extracted from users' behaviors.

In both cases, the performance of several classifiers (KNN, Naive Bayes (NB), SVM, Logistic Regression, and Random Forest) is compared. The results reveal that KNN outperforms the rest of the classifiers in terms of f-score, achieving the best f-score of 82.40%. The f-score has increased by 3.80% when considering the extracted reviewers' behavioral features. The paper also compares the results in the absence and the presence of the extracted features in two different language models namely TF-IDF with bi-grams and TF-IDF with tri-grams. The results indicate that the engineered features increase the performance of the fake review detection process.[1]

Detecting Opinion Spams and Fake News Using Text Classification this paper introduces a new n-gram model to detect automatically fake content, focusing on fake reviews and fake news. The model uses two different feature extraction techniques and six machine-learning classification techniques. The paper highlights the growing problem of opinion spam, which involves writing and spreading false information or beliefs on the web. Fake news can be categorized into three groups: false news, fake satire news, and poorly written news articles.

A new dataset for fake news was collected, containing 12,600 fake news articles and 12,600 legitimate news articles, achieving an accuracy of 92%. A unified approach is proposed for automatic detection of fake content, combining text analysis, n-gram features, terms frequency metrics, and machine learning classification. The model outperforms existing methods, achieving 90% accuracy on Ott et al.'s dataset and 87% on Horne and Adali's dataset. [2]

The study by Paweł Gryka and Artur Janicki focuses on detecting fake reviews in Google Maps places in Poland. They construct a dataset containing 18 thousands of counterfeit and genuine reviews in Polish and use it to train machine learning models to detect fake reviews and their authors.They propose a novel metric for measuring account name typicality and geographical dispersion of reviewed places. Initial recognition results were promising, with an F1 score of 0.92 and 0.74 for fake accounts and reviews, respectively. Researchers have used machine learning (ML) methods to detect fake reviews.

However, to build a binary classifier, a dataset must be trained on. Existing datasets include Jindal and Liu's 5.8 million Amazon reviews, Ott et al.'s 400 real and 400 fake reviews, Yoo and Gretzel's 42 deceptive reviews, Sandulescu and Ester's 9000 reviews, Amazon's updated version, Joni S et al.'s 20,000 fake reviews, and Yelp's 7 million reviews. These datasets provide insights into how fake reviewers may think and construct their reviews. [3]

Online shopping stores have experienced significant growth in recent years, leading to the detection of fake reviews. Fake reviews are used to mislead customers

and undermine the honesty and authenticity of online shopping environments. To improve the accuracy of existing fake review classification or detection approaches, the researchers propose using the BERT model to extract word embeddings from texts (i.e. reviews). Word embeddings are obtained using various

basic methods such as Support Vector Machine (SVM), Random Forests, Naive Bayes, and others. The confusion matrix method was also taken into account to evaluate and graphically represent the results.

The results indicate that the SVM classifiers outperform the others in terms of accuracy and f1-score with an accuracy of 87.81%, which is 7.6% higher than the classifier used in the previous study. This study mainly focuses on review content to detect fake reviews, using Natural Language Processing (NLP) to create review features not directly related to the data or text provided. In conclusion, the study proposes using the BERT model and machine learning techniques to improve the accuracy of fake review detection in online shopping environments.[4]

The study proposes a fake review detection model using Text Classification and Machine Learning techniques, including Support Vector Machine, K-nearest neighbor, and logistic regression (SKL). The model detects fraudulent reviews based on pronouns, verbs, and sentiments, outperforming other techniques with 95% and 89.03% accuracy on the Yelp and TripAdvisor datasets.

Techniques such as Data Mining and Natural Language Processing (NLP) have been utilized to deal with raw data, but they have not efficiently solved the spam review problem. As the popularity of social media increases, it is essential to identify indicators of fraudulent reviews based on fraudsters'

behavior. A supervised learning technique, such as K-Nearest Neighbor, Support Vector Machine, and Logistic Regression (SKL), has been proposed to detect fake reviews and extract genuine emotion in opinions. SKL algorithms outperform state-of-the-art methodologies in detecting fake reviews and improving decision-making in the online marketplace. [5]

The article "Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets" has been retracted due to concerns about the compromised peer review process. The study focuses on the identification of fake reviews in e-commerce using multidomain datasets such as hotels, restaurants, Yelp, and Amazon. The proposed methodology uses various input word-embedding matrices, convolutional and max-pooling layers of the CNN technique, LSTM layer, and a sigmoid activation function. The model was evaluated in both in-domain and cross-domain experiments, with the results showing 77%, 85%, 86%, and 87% accuracy for restaurant, hotel, Yelp, and Amazon datasets, respectively. The cross-domain experiment achieved 89% accuracy.

Web 4.0 has increased internet shopping through E-commerce platforms, with online reviews playing a significant role in influencing customer-buying decisions. However, fake reviews, which can be untrusted, product name-only, or nonreviews, can manipulate brand reputations and affect businesses. Companies hire fake reviewers or spammers to generate low-quality reviews, making it difficult to extract accurate findings and handle the issue effectively.[6]

This paper aims to develop an intelligent system that can detect fake reviews on e- commerce platforms using n-grams of review text and sentiment scores given by the reviewer. The proposed methodology used a standard fake hotel review dataset for experimenting and data preprocessing methods, and a term frequency-inverse document frequency (TF-IDF) approach for extracting features and their representation. For detection and classification, n-grams of review texts were inputted into the constructed models to be classified as fake or truthful. The experiments were carried out using four different supervised machine-learning techniques and were trained and tested on a dataset collected

from the Trip Advisor website. The classification results showed that naïve Bayes (NB), support vector machine (SVM), adaptive boosting (AB), and random forest (RF) received 88%, 93%, 94%, and 95%, respectively, based on testing accuracy and the F1-score.

The obtained results were compared with existing works that used the same dataset, and the proposed methods outperformed the comparable methods in terms of accuracy. With the rapid development of e-commerce, customers are increasingly using online marketing websites for purchasing and selling products and services. Positive reviews entice more customers to purchase specific products or brands, providing considerable financial gain, while negative opinions often cause sales losses in e-commerce. Most merchants depend primarily on public opinion to reshape their business plans by improving the quality of products. Typically, opinions are the key to any online blog, post, or review. Spam content can be defined as meaningless or unsolicited data that are merged into opinions and are used for advertising, promoting, disseminating information, and financial profit purposes.[7]

The increasing use of social media has led to the need to combat the spread of false information and reduce reliance on information retrieval from such sources. This paper presents a novel approach to detect fake news using deep learning models. By collecting 1356 news instances from various users via Twitter and media sources like PolitiFact, the study compares multiple state-of-the-art approaches such as convolutional neural networks (CNNs), long short-term memories (LSTMs), ensemble methods, and attention mechanisms.

The study concludes that CNN + bidirectional LSTM ensembled network with attention mechanism achieved the highest accuracy of 88.78%, while Ko et al tackled the fake news identification problem and achieved a detection rate of 85%. Fake news is written with the intent to spread information disguised as propaganda or a hoax that results in financial or political gain, which may be used to sway public opinion toward falsity.

As a result, fake news has received significant research attention from organizations such as Facebook, Google, Twitter, and many researchers. In a study,researchers at Stanford found that students have difficulty in determining the credibility of information online. Poor journalism leads to misinterpretation of the actual news itself due to false connections, misleading content, and false information. In conclusion, this study provides a novel approach to detect fake news using deep learning models, addressing the growing issue of fake news on social media platforms.[8]

Online review platforms are becoming increasingly popular, encouraging dishonest merchants and service providers to deceive customers by creating fake reviews for their goods or services. This paper adopts a semi-supervised machine learning method to detect fake reviews on any website, among other things. Online reviews are classified using a semi-supervised approach (PU-learning) since there is a shortage of labeled data and they are dynamic. Then, classification is performed using the machine learning techniques Support Vector Machine (SVM) and Nave Bayes.

Consumers sometimes depend significantly on reviews because they can't see the goods before buying them. Many techniques have been developed to detect these fake reviews, mainly relying on supervised Machine Learning (ML) methods and distinctive features. Implementing ML for fake review detection efficiently distinguishes between fake and genuine content. Shan et al. provided a technique based on internet customer reviews for detecting false reviews (OCR). The authors of detected bogus reviews using supervised machine learning techniques, employing five classifiers: SVM, Naive Bayes, KNN, k-star, and decision trees. To identify fraudulent reviews on the dataset they had gathered, the authors employed Naive

Bayes, Decision trees, SVM, Random forests, and Maximum Entropy classifiers. The yield dataset, which comprises of a total of 5,000 reviews, was used to train the proposed system.[9]

The International Research Journal of Engineering and Technology (IRJET) has published a study on improving the performance of fake reviews detection in online reviews using semi-supervised learning. The study focuses on the importance of online reviews in e-commerce, as they are crucial for consumers to make informed decisions about products or services. Online reviews not only help users understand the product or service thoroughly but also affect their decision-making abilities and can divert sentiments about the product positively or negatively.

The study will examine cumulative studies of this work and how the addition of unlabeled data improves the accuracy in identifying fake reviews using three different base learner algorithms: Naïve Bayes, Decision Tree, and Logistic Regression. Sentiment analysis is a vast domain that deals with information retrieval and knowledge discovery from text using data mining, natural language processing, and machine learning techniques. The knowledge of this analysis can be used for recommendation systems, government intelligence, citation analysis, human-computer interaction, and computer-assisted interactivity. The study focuses on the field of online reviews and its analysis. Opinion spamming is a common practice in the e-commerce industry, where companies hire people to write fake reviews about products to promote their own products or demote competitors' products. According to Bing Liu, an expert in opinion mining, there are an estimated 33% fake reviews in consumer review databases.

The study aims to improve the accuracy of detecting fake reviews using semi-supervised learning and other methods to protect consumers' genuine business experiences.[10]

III. Methodology

The Fake Review Detection system is developed using deep learning techniques to classify online reviews as genuine or fake. The process begins with data collection and preprocessing, where the dataset is cleaned by removing special characters, and stop words, and applying tokenization. The text is then converted into numerical form using word embeddings for CNN, RNN, and LSTM models, while BERT tokenization is applied for the transformer-based model.

For model implementation, four deep learning architectures—CNN, RNN, LSTM, and BERT—were trained separately. The CNN model extracts spatial features from text using convolutional layers, identifying patterns that distinguish fake and genuine reviews. The RNN model processes sequential data by maintaining hidden states, allowing it to capture dependencies in the input sequence. The LSTM model, an advanced version of RNN, preserves long-term dependencies, making it effective for detecting subtle linguistic patterns in fake reviews. Lastly, the BERT model, a transformer-based architecture, leverages pre-trained contextual embeddings to provide a deeper semantic analysis of reviews.

Each model was trained and evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score. The results showed varying degrees of effectiveness. To enhance accessibility, a Streamlit-based user interface was developed, enabling users to input reviews and receive real-time classification results.

The UI was designed for simplicity, responsiveness, and ease of interpretation, ensuring a user-friendly experience

a. Recurrent Neural Network (RNN):

The RNN model is a sequential data processing technique used for tasks involving ordered

dependencies within data, such as word sequences in sentences. It includes an embedding layer, an RNN layer, and a final dense layer for classification. The textual data is tokenized and padded, then passed through an embedding layer to map each word to a high-dimensional vector space. The RNN layer processes the input sequentially, retaining context through internal states. The output layer produces probability scores for genuine and fake reviews. The model was trained over 10 epochs using cross-entropy loss and the Adam optimizer for gradient-based optimization.

b. Convolutional Neural Network (CNN):

CNNs are useful in text classification tasks due to their ability to recognize local patterns. They capture n-gram structures and short-term dependencies within text data. The CNN model tokenizes and embeds each review into a fixed-length vector format, using a Conv1D layer to scan the text for local features. A global max-pooling layer condenses feature maps, retaining only critical information while reducing dimensionality. A dense layer with ReLU activation and a dropout layer improves generalization ability, while a final dense layer outputs the probability distribution for each class. The model was trained with sparse categorical cross-entropy loss and Adam optimization over 10 epochs, allowing adjustments to mitigate overfitting or underfitting.

c. Long Short-Term Memory (LSTM) Network:

The Long Short-Term Memory (LSTM) Network is a model that is ideal for identifying fake reviews due to its ability to capture nuanced linguistic patterns over extended word sequences. The model architecture includes an embedding layer with 100-dimensional word embeddings, an LSTM layer with 128 units and a dropout rate of 0.2, a dense layer with 64 units and ReLU activation, and an output layer for binary classification of reviews as genuine or fake. The model is trained using 80% of the dataset, with the remaining 20% reserved for testing. The Adam optimizer is used with a learning rate of 0.001, and binary cross-entropy serves as the loss function. The training process consists of 5 epochs, and the methodology is designed to ensure reliability, validity, and generalizability of the findings. The model was trained with categorical cross-entropy as the loss function and validated on a set-aside portion of the data to monitor generalization.

d. Bidirectional Encoder Representations from Transformers (BERT):

The proposed solution uses the Bidirectional Encoder Representations from Transformers (BERT) model for fake review detection. BERT is a deep learning model that understands contextual relationships between words in text. It has been pretrained on large corpora. The workflow involves data collection, preprocessing, feature extraction, training, and validation. A pretrained BERT model is used for binary classification, and a custom dataset and DataLoader are created for efficient batch processing. The model predicts whether a review is fake or genuine, and its performance is evaluated using Precision, Recall, F1-score, and Accuracy

e. User Interface:

The user interface of a deep learning-based review analysis tool, developed using Streamlit, allows users to detect fake reviews using one of three deep learning models: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), or Long Short-Term Memory (LSTM). The interface includes header information, model selection tabs, model-specific detection, a user input field, and a prediction button. The tool leverages the strengths of CNN, RNN, and LSTM, demonstrating the practical implementation of deep learning architectures in natural language processing tasks. The interface contains the following key elements:

- Header Information: A confirmation message at the top indicates that all models and tokenizers have been successfully loaded, ensuring the system is ready for analysis.
- Model Selection Tabs: Users can choose between the CNN, RNN, or LSTM models via tab navigation. This feature makes the application versatile, allowing experimentation with different deep-learning architectures.
- Model-Specific Detection: Depending on the selected model, the interface dynamically updates to reflect the chosen approach:
- Title and Description: The title specifies the selected model (e.g., "CNN Fake Review Detection"), while a brief description highlights its purpose.
- User Input Field: A large text box allows users to input a review they wish to analyze.
- Prediction Button: A button labeled with the selected model's name (e.g., "Predict with CNN Model") triggers the analysis.

The application leverages the strengths of the three models:

- **LSTM**: An advanced RNN variant, adept at learning long-term dependencies and mitigating the vanishing gradient problem.

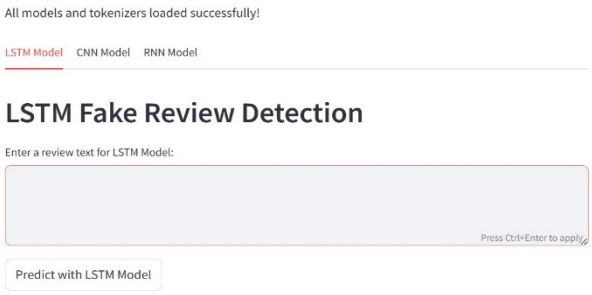

Fig 1.1: LSTM Fake Review Detection Interface

- **CNN**: Captures spatial hierarchies in text data using convolutional layers.

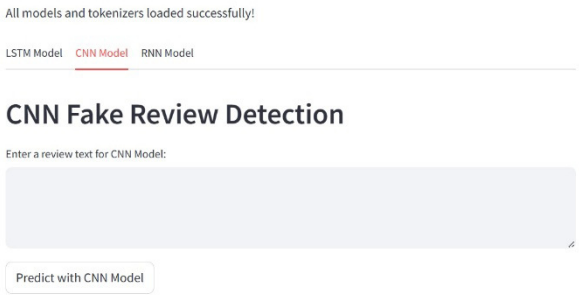

Fig 1.2: CNN Fake Review Detection Interface

- **RNN**: Processes sequential information effectively by maintaining a memory of previous inputs.

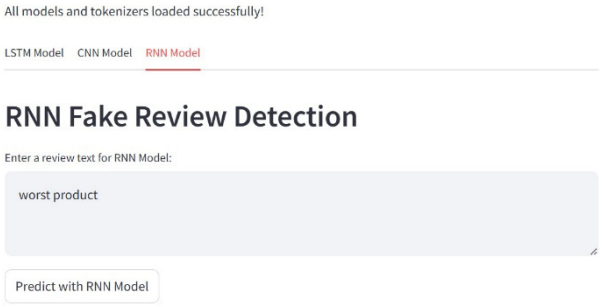

Fig 1.3: RNN Fake Review Detection Interface

This interactive tool demonstrates the practical implementation of deep learning architectures in natural language processing tasks, offering a flexible platform for fake review detection.

IV. CONCLUSIONS

Deep learning techniques are revolutionizing fake review detection, enhancing trust and credibility in online platforms. These models, including CNN, RNN, LSTM, and BERT, help mitigate the impact of deceptive reviews, ensuring a more transparent and reliable online experience for consumers worldwide. CNN excels at identifying local patterns and fraudulent keywords, while RNNs, particularly LSTMs, handle sequential dependencies and inconsistencies in writing style. BERT, a transformer-based model, enhances detection by capturing contextual meaning and understanding subtle variations in text. Each model was trained and evaluated separately, providing independent

insights into the nature of fake reviews. The ability of these models to automatically extract and learn features from large datasets improves efficiency and scalability. As fake review tactics evolve, leveraging deep learning ensures detection systems remain adaptive and resilient in identifying fraudulent content. The Fake Review Detection Project presents a powerful, automated solution for combating review manipulation, offering significant benefits for e-commerce platforms, consumers, and businesses.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

[1] Elmogy, Ahmed M., Usman Tariq, Ammar Mohammed and Atef Ibrahim. "Fake Reviews Detection using Supervised Machine Learning." International Journal of Advanced Computer Science and Applications (2021): n. pag.

[2] Ahmed, H., Traore, I. and Saad, S., 2018. Detecting opinion spam and fake news using text classification. Security and Privacy, 1(1), p.e9.

[3] Gryka, P. and Janicki, A., 2023. Detecting Fake Reviews in Google Maps—A Case Study. Applied Sciences, 13(10), p.6331.

[4] Mir, A.Q., Khan, F.Y. and Chishti, M.A., 2023. Online fake review detection using supervised machine learning and BERT model. arXiv preprint arXiv:2301.03225.

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

[6] Alsubari, S. N., Deshmukh, S. N., Al-Adhaileh, M. H., Alsaade, F. W., & Aldhyani, T. H. H. (2021). Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets. Applied Bionics and Biomechanics, 2021, 1–11. doi:10.1155/2021/5522574.

[7] Alsubari, S.N., Deshmukh, S.N., Alqarni, A.A., Alsharif, N., Aldhyani, T.H., Alsaade, F.W. and Khalaf, O.I., 2022. Data analytics for the identification of fake reviews using supervised learning. Computers, Materials & Continua, 70(2), pp.3189-3204.

[8] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N. and Akbar, M., 2020. Fake news detection using deep learning models: A novel approach. Transactions on Emerging Telecommunications Technologies, 31(2), p.e3767.

[9] Alshehri, A.H., 2024. An Online Fake Review Detection Approach Using Famous Machine Learning Algorithms. Computers, Materials & Continua, 78(2).

[10] Jadhav, A.B., Rathod, V.U. and Jadhav, H.B., 2019. Improving Performance of Fake Reviews Detection in Online Reviews using Semi-Supervised Learning. International Research Journal of Engineering and Technology (IRJET), 6(06).

