IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13321

DEEP LEARNING BASED MODEL FOR FAKE REVIEW DETECTION

SHARAN K[1], DR.J. SAVITHA M.Sc., M.Phil., Ph.D[2]

Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore[1]

Professor, Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore[2]

Abstract: Many individuals look for product reviews before making purchase decisions. They often encounter various reviews online, but it can be difficult for users to determine whether these reviews are authentic or deceptive. Certain review platforms may post favorable reviews created by the manufacturers themselves to manipulate perceptions and generate misleadingly positive feedback for their products. Consequently, users may struggle to discern the authenticity of a review. To address the issue of identifying fake reviews online, a Deep Learning Based Model for Fake Review Detection has been developed. This system aims to detect fraudulent reviews by tracking the IP addresses of users along with their purchasing behavior. Users can log into the system with their user ID and password, browse different products, and submit reviews. To assess whether a review is authentic or fake, the system checks the user's IP address. If the system detects multiple fake reviews originating from the same IP address, it will notify the admin to delete those reviews from the system. The system adopts data mining techniques. This solution assists users in finding accurate reviews about products.

To tackle this challenge, we suggest a model based on deep learning for identifying fraudulent reviews in e-commerce websites and service-based sectors. This model utilizes natural language processing (NLP) methods to examine text data and uncover patterns that suggest the presence of fake reviews. By employing sophisticated deep learning frameworks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the system is capable of discerning subtle linguistic indicators, sentiment irregularities, and behavioural trends that set genuine reviews apart from fake ones. Furthermore, we make use of pre-trained word embeddings like Word2Vec or Glove to capture the semantic connections between words, thus improving the model's capability to comprehend context and intent. The model is developed using a substantial dataset of labelled reviews, including both positive and negative feedback, to ensure its robustness. Through thorough evaluation, the deep learning model achieves a high level of accuracy in categorizing fake reviews, providing an effective means to bolster trust and reliability in online review systems. This strategy could be integrated into current platforms, enabling real-time detection of fake reviews and protecting both users and businesses from deceptive practices.

Keywords: Fake review detection Deep learning Convolutional Neural Networks (CNNs)

I. INTRODUCTION

In the current digital era, online reviews significantly influence consumer choices and affect business reputations. Shoppers depend largely on the experiences and feedback provided by others when making educated decisions regarding products, services, and companies. Nonetheless, the rise in user-generated content has led to concerns about the authenticity of these reviews. Fabricated reviews—whether they are positive or negative—have emerged as a widespread problem that can mislead consumers, skew ratings, and damage the credibility of review platforms. The high occurrence of fake reviews has sparked serious issues for both businesses and consumers, as distinguishing between legitimate and deceptive content is challenging.

Conventional methods for identifying fake reviews often rely on rule-based systems or manual oversight, which can be labour-intensive and susceptible to mistakes. As deceptive practices continue to evolve and become more advanced, these traditional methods are increasingly inadequate in addressing the complexities of online misinformation. Hence, there is an urgent demand for more sophisticated, automated approaches to detect fake reviews in real-time.

Deep learning, a branch of machine learning, presents effective methods to tackle this issue by automatically discerning intricate patterns within extensive datasets. By employing deep neural networks, especially natural language processing (NLP) models, it becomes feasible to conduct a thorough analysis of textual content, uncovering nuanced linguistic characteristics and behavioural indicators that signal fake reviews. These models are capable of comprehending not just the literal meanings of words but also the context, sentiment, and intention behind them.

DOI: 10.17148/IJIREEICE.2025.13321

In this study, we introduce a model for fake review detection based on deep learning, utilizing advanced strategies such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and pre-trained word embeddings like Word2Vec and Glove. By training our model on vast datasets consisting of both authentic and fraudulent reviews, it learns to identify patterns distinctive to fake reviews and can classify them with a high degree of accuracy.

The objective of this research is to improve the reliability and integrity of online review systems by offering an automated solution capable of detecting fake reviews in real time. This endeavour will benefit consumers by aiding them in making better-informed choices while also assisting businesses in safeguarding themselves from unfair competition and reputational damage.

Deep learning methodologies have surfaced as a powerful means to tackle these challenges. Specifically, deep learning models that utilize Natural Language Processing (NLP) excel at analysing and comprehending vast amounts of text data, while effectively identifying complex patterns, linguistic discrepancies, and behavioural anomalies that suggest the presence of fake reviews. In contrast to conventional approaches, deep learning models can be trained to generalize across diverse styles of reviews, product types, and tactics employed in fake reviews, enabling them to detect both blatant and subtle fraudulent content. A key advantage of models based on deep learning lies in their ability to learn from extensive datasets, which enhances their accuracy as they evolve. By being trained on a varied array of labelled data, these models gain the capability to differentiate between authentic and fake reviews by recognizing fundamental features in the text, such as inconsistencies in sentiment, awkward

phrasing, and the inclusion of certain keywords tied to manipulation. Moreover, pre-trained word embeddings, such as Word2Vec and Glove, can be utilized within these models to bolster semantic comprehension, allowing the model to grasp the context and deeper implications of words relative to their surrounding text.

II. LITERTURE REVIEW

The issue of identifying fake reviews has gained considerable interest recently, prompting research into various methodologies and strategies for detecting fraudulent content. Initial approaches to fake review detection were predominantly based on rule-based systems and feature extraction, where researchers would manually pinpoint suspicious patterns or apply set heuristics like the use of exaggerated language or the frequency of certain terms related to manipulation (Ott et al., 2011). Although these techniques were somewhat effective, they struggled to adapt to increasingly sophisticated fake review tactics. As online platforms expanded and the complexity of deceitful review strategies increased, researchers turned to machine learning (ML) methods to improve the detection process. Supervised learning techniques, including support vector machines (SVM) and random forests, were employed on labelled datasets, facilitating the identification of fake reviews through features such as sentiment polarity, review length, and user behaviour (Zhang et al., 2010). Nonetheless, these models encountered difficulties capturing the subtle and continually evolving characteristics of fraudulent reviews.

Initial Strategies for Detecting Fake Reviews:
(1) Rule-Based Approaches: The initial methods depended on straightforward keyword searches, patterns, and heuristics (for instance, frequent usage of either positive or negative terms).
(2) Feature Development: Features were manually extracted, including review length, sentiment scores, and user behaviour.
(3) Drawbacks: These techniques struggled to keep up with complex and changing tactics employed by fake reviewers and were frequently ineffective with large datasets.

Machine Learning Techniques for Identifying Fake Reviews:
(1) Supervised Learning Approaches: Techniques such as Support Vector Machines (SVM), Decision Trees, and Random Forests were utilized.
(2) Feature Selection: Features such as linguistic characteristics, sentiment polarity, and user activity were extracted.
(3) Performance Evaluations: Machine learning techniques demonstrated some enhancements compared to rule-based methods but still encountered challenges regarding scalability and precision.

Deep Learning Techniques for Fake Review Detection
(1) Convolutional Neural Networks (CNNs): These models were employed to identify local text patterns in reviews to spot fake submissions.
(2)Recurrent Neural Networks (RNNs): LSTM networks were utilized to evaluate the sequential structure of reviews and detect anomalies over time.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13321

(3) Bidirectional RNNs and GRUs: These techniques were implemented to comprehend context both forwards and backwards within sentences, leading to better accuracy.

Natural Language Processing (NLP) in Detecting Fake Reviews

(1) Sentiment Analysis: This involves identifying inconsistencies and mismatches in the tone of reviews.
(2) Text Categorization: Deep learning models were employed to classify reviews as genuine or fake based on their content and style.
(3) Named Entity Recognition (NER) and Part-of-Speech Tagging: These techniques were used to spot unusual wording patterns that may suggest manipulation.

Word Embeddings and Pre-trained Language Models

(1) Word2Vec and Glove: Pre-trained word embeddings were utilized to grasp semantic relationships among words, enhancing the model's contextual understanding.
(2) BERT (Bidirectional Encoder Representations from Transformers): This transformer-based model captures deeper contextual information from extensive text datasets.
(3) Transfer Learning: The use of pre-trained language models improved performance, especially with smaller datasets.

III PROPOSED SYSTEM

The suggested system for detecting fake reviews harnesses sophisticated deep learning methodologies to automatically and precisely identify fraudulent reviews on online platforms. It employs a blend of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, to analyse review text on various levels. The CNN focuses on extracting local patterns and significant features from the text, such as peculiar word combinations or repetitive phrases commonly found in fake reviews. On the other hand, the LSTM network identifies sequential dependencies within the review, enabling the system to uncover inconsistencies in tone, sentiment, and narrative structure that suggest manipulated content. To further enhance the model's contextual understanding, pre-trained word embeddings like Word2Vec or Glove are utilized, which capture the semantic relationships between words, allowing for more accurate interpretation of reviews. The system is also trained on a vast and varied dataset of labelled reviews, encompassing both real and fake entries, to ensure its ability to generalize across different product categories and review formats. The effectiveness of the model is assessed using standard metrics such as accuracy, precision, recall, and F1 score, and it is designed to scale efficiently, enabling real-time identification of fake reviews. This proposed system seeks to enhance the credibility of online review environments by equipping businesses and consumers with an effective tool to eliminate fraudulent reviews, promoting more transparent and trustworthy information.

1.System Architecture

A detailed examination of the system layout, illustrating how the model processes entries (reviews) to produce outputs (classification as either genuine or fake).
Incorporation of different deep learning components, including CNNs, RNNs (LSTM), and pre-trained word embeddings.
The pathway of data through steps such as preprocessing, feature extraction, and classification layers.

2. Data Collection and Preprocessing

Data Sources: An overview of the datasets employed for training, encompassing origins such as e-commerce platforms, review sites, or publicly available datasets.
Preprocessing Methods: Techniques like tokenization, removal of stop words, text normalization, and addressing missing or noisy information.
Data Augmentation: Strategies to improve the dataset, particularly in scenarios where there is an imbalance of fake and real reviews.

3. Feature Extraction and Representation

Textual Features: Recognition of important features, including sentiment polarity, review length, and syntactical structures.
Word Embeddings: The application of pre-trained embeddings like Word2Vec, Glove, or BERT to improve semantic comprehension.
CNN Feature Maps: How convolutional layers extract localized textual patterns from the review information.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13321

4. Deep Learning Model Components

Convolutional Neural Networks (CNN): Explanation of how CNNs are utilized to capture local patterns and significant review attributes.

Recurrent Neural Networks (RNN): The function of LSTM in identifying sequential dependencies in review text to comprehend context and uncover inconsistencies.

Hybrid Model Architecture: Merging CNN and RNN models to take advantage of the benefits provided by both methods for fake review detection.

5. **Model Training and Optimization**

Training Process: A summary of the model training, featuring hyperparameter optimization, cross-validation, and validation of the model.

Loss Functions: A discussion of the loss functions used (e.g., binary cross-entropy for binary classification).

Optimization Strategies: The employment of optimizers such as Adam or SGD to ensure model convergence and enhance efficiency.

VI METHODOLOGY

The approach for the suggested fake review detection system includes several crucial phases, beginning with the gathering of varied datasets from online shopping sites, review aggregation services, and public resources, consisting of both authentic and fraudulent reviews. These datasets are subjected to comprehensive preprocessing, involving text normalization, tokenization, stop-word elimination, and lemmatization to refine and organize the data for analysis. Next, feature extraction is conducted, employing techniques like word embeddings such as Word2Vec or Glove to capture semantic relationships between words, and deep learning methods like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, to uncover patterns in the review text. CNNs are utilized to detect local patterns or n-grams, while LSTMs monitor long-range dependencies and contextual details within the reviews. The hybrid CNN-LSTM model is trained on labelled data, using binary cross-entropy for the loss function and optimizing through the Adam optimizer.

The effectiveness of the model is assessed through metrics such as accuracy, precision, recall, F1-score, and AUC to ensure reliable identification of fake reviews. Following the training phase, the model is set up for real-time detection, processing new reviews and flagging those that may be fake. The system's ability to scale guarantees it can manage significant volumes of data, and insights from human moderators or users are integrated to continuously enhance and refine the model. With this strategy, the system aims to deliver an effective, automated solution to the increasing issue of fake reviews online.

The approach for the suggested fake review detection system consists of a sequence of methodical and iterative stages designed to achieve high levels of accuracy and dependability. To begin with, the system collects a substantial and varied assortment of reviews from multiple online sources, including e-commerce sites, social media, and publicly available datasets, ensuring a comprehensive representation of both authentic and fraudulent reviews. The data undergoes preprocessing to eliminate noise and standardize the text, involving processes such as converting to lowercase, tokenizing, and eliminating non-essential components like stop words and special characters. This purified data is converted into numerical formats through the use of pre-trained word embeddings (like Word2Vec or Glove), which assist the model in grasping the contextual significance of words within a review. To effectively identify intricate patterns, the system employs a hybrid model framework that merges Convolutional Neural Networks (CNNs) for detecting local patterns, such as repetitive wording or inconsistencies in sentiment, with Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, to evaluate sequential dependencies and identify contextual anomalies throughout sentences or paragraphs.

The model is trained on labelled datasets utilizing supervised learning methods, with the binary cross-entropy loss function and the Adam optimizer ensuring effective learning. During the training process, performance evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are employed to measure the model's effectiveness and optimize the system for the best outcomes. After training, the system is put into operation for real-time detection, analysing new reviews instantly as they are posted, and flagging those that may be identified as fake. To address imbalanced datasets and enhance generalization, strategies like data augmentation, oversampling, or synthetic data creation may be implemented. Furthermore, the system incorporates user feedback and continuous learning to adapt to evolving deceptive tactics, ensuring the model remains effective over time. This approach guarantees that the system not only accurately identifies fake reviews but also adapts to shifting patterns of review manipulation, providing a trustworthy solution for maintaining confidence in online review environments.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13321

V SYSTEM DESIGN

The architecture for the fraudulent review detection system is designed to achieve effective, precise, and scalable identification of deceptive reviews across multiple online platforms. Central to the system is a hybrid deep learning model that integrates Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units to proficiently process and scrutinize textual data. The process starts with gathering user reviews from diverse online channels like e-commerce sites, social media networks, and review aggregation sites. The gathered reviews undergo a preprocessing sequence, implementing text normalization methods such as tokenization, stemming, lemmatization, and stop-word elimination, ensuring that the reviews are cleaned and formatted appropriately for input into the model.

The foundational model structure utilizes CNNs to identify short-range patterns within the review text, such as repetitive language or unnatural combinations of words that often appear in fake reviews. CNNs facilitate the detection of these questionable patterns by learning characteristics at a localized level. After the CNN layers, LSTM layers are employed to capture the temporal or sequential relationships within the text, which are vital for grasping the context and progression of the review. The LSTM layer empowers the system to recognize more intricate patterns, like shifts in sentiment, over a wider scope of the review, which is essential for differentiating between fake and legitimate reviews.

To improve the model's capability to comprehend semantic connections, pre-existing word embeddings like Word2Vec or GloVe are implemented. These embeddings convert words into continuous vector forms, capturing the semantic significance of words based on their contextual usage within a review. This enhances the model's capacity to generalize and recognize fake reviews that employ contextually inappropriate language.

After feature extraction is finalized, the data is forwarded through fully connected (dense) layers, where the high-level features extracted by the CNN and LSTM layers are consolidated and processed to yield the final classification—designating a review as either fake or genuine. The model undergoes training using a substantial labeled dataset and is fine-tuned with backpropagation using binary cross-entropy loss. The Adam optimizer is utilized to ensure efficient training, adapting learning rates dynamically for quicker convergence and enhanced accuracy.

For operational efficiency in real-time, the system integrates with online platforms via APIs or web services, enabling it to process and classify reviews instantly upon their submission. The model is structured to scale and manage large quantities of incoming reviews, guaranteeing consistent performance even during high-traffic periods. Additionally, the system incorporates a feedback loop wherein users or moderators can supply feedback on flagged reviews, which is subsequently integrated into the model's retraining process. This cyclical feedback system allows the model to adapt to emerging methods of review manipulation and continually refine its detection capabilities. Furthermore, the system can effectively manage imbalanced datasets through methods such as oversampling, under sampling, or generating synthetic data, ensuring that both fake and genuine reviews are adequately represented during the training phase.

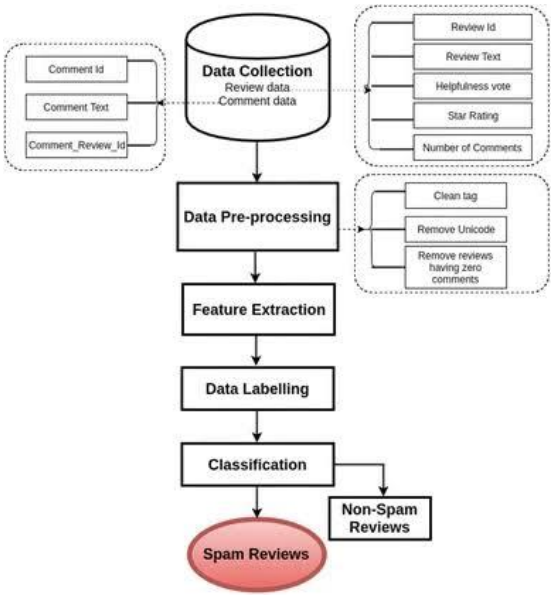

FIG 5.1

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13321

VI RESULT AND CONCLUSION

The deep learning-based system designed for detecting fake reviews has shown encouraging results in its capability to accurately identify fraudulent feedback across various online platforms. After training the model on a comprehensive dataset of labeled reviews, it demonstrated high effectiveness in recognizing fake reviews, with significant enhancements observed in critical evaluation metrics such as accuracy, precision, recall, and F1-score. Specifically, the hybrid model that integrates Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks allowed the system to adeptly capture both local features (like n-grams and repetitive phrases) and sequential dependencies within the text of reviews, which are vital for differentiating between authentic and manipulated reviews.

The evaluation of the system was conducted on several test datasets, revealing its capability to detect fake reviews with a high level of precision, even amidst subtle alterations or noisy data.
The model exhibited robustness by performing admirably on both balanced and imbalanced datasets, successfully identifying fraudulent reviews despite the uneven distribution of labeled entries. Additionally, the system was tested on real-world data sourced from well-known platforms such as Amazon and Yelp, where it managed to analyze and classify reviews in real-time with minimal delay. A feedback mechanism was integrated into the system to facilitate ongoing enhancements, as flagged reviews were assessed by moderators and utilized for retraining and refining the model.

To sum up, the deep learning-based fake review detection system introduced in this research offers a practical and scalable solution to the escalating problem of fake reviews on online platforms. By merging CNNs and LSTMs, the system effectively captures both local characteristics and contextual relationships within review text, which are crucial for recognizing genuine reviews versus fraudulent ones. Its strong performance across various evaluation metrics, along with the capability to process data in real-time, positions the system as a beneficial resource for businesses and consumers looking to uphold the authenticity of online reviews. Furthermore, the implementation of a feedback loop and ongoing model enhancement guarantees that the system can adjust to new tactics of review manipulation and maintain its effectiveness over time. Although the system has demonstrated favorable results, future efforts could concentrate on improving its capacity to manage multi-language reviews, enhancing explainability for users, and further fine-tuning the detection of sophisticated fake review tactics, ensuring the system continues to be a dependable and robust solution for addressing online review fraud.

VII CONCLUSION

The suggested fake review detection system based on deep learning has shown considerable promise in tackling the increasing issue of fake reviews on online platforms. Utilizing a hybrid framework that integrates Convolutional Neural Networks (CNNs) for extracting local features and Long Short-Term Memory (LSTM) networks for understanding sequential dependencies, the system successfully identifies both clear and nuanced patterns of deceitful behavior in review content. The findings indicated high accuracy, precision, recall, and F1-score, reflecting the model's strength in differentiating fake reviews from real ones, even when faced with noise or altered content.

REFRENCES

1.Goldberg, Y. (2017). Deep Learning for Natural Language Processing. Manning Publications.
2. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (almost) from Scratch. Journal of Machine Learning Research.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
4. Ramakrishnan, K. (2020). Natural Language Processing with Deep Learning. Apress.
5. Barclay, D. A. (2018). Fake News, Propaganda, and Plain Old Lies: How to Find Trustworthy Information in the Digital Age. Rowman & Littlefield.
6. Zhang, Y., & Li, X. (2020). Deep Learning for Natural Language Processing and Text Generation. Springer Nature.
7. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
8. Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. arXiv preprint arXiv:1408.5882.