International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2682- 2686
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0370

Fake Feedback Detection Using Machine Learning

MS.M Amshavalli[1], Aditya Baibhav[2], Deepak Kumar[3], MD Arif Reza[4]
[1]Assistant Professor, Department of CSE, Erode Sengunthar Engineering College, Perundurai, Erode, Tamil Nadu, India.
[2,3,4]Student, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Perundurai, Erode, Tamil Nadu, India.
Emails: *amshavalli1494@gmail.com[1], thisisadityabaibhavgmail.com[2], officialdeepurajpk@gmail.com[3], arifs01283@gmail.com[4]*

Abstract
The prevalence of false feedback increases the reliance on online media for information and interaction, which is a major challenge for businesses, consumers and reputation management. This project presents a new way to detect false positives using machine learning techniques. We propose a multi-modal model that uses natural language processing (NLP) and supervised learning algorithms to analyze text response data. Our methodology includes sentiment analysis, the extraction of language features and behavioral patterns to distinguish between genuine feedback and fake news. We evaluate our model using a comprehensive data set that includes real and synthetic feedback samples. We are Analyzing that our approach which we are going to implement in future that can achieves high accuracy and robustness and is significantly better than traditional detection methods. In addition, we discuss the implications of our findings for increasing trust in online reviews and the potential for feedback monitoring. This initiative will contribute to the growing digital presence and provide a scalable solution for stakeholders seeking to reduce the impact of false positives in various domains.

Keywords: *Reputation Management, Sentiment Analysis, Scalable Solution, Robustness, Multi-modal Mod*

1. Introduction
The detection of fraudulent feedback through machine learning (ML) has become an essential focus in the contemporary digital environment, where online reviews and feedback significantly impact businesses, consumers, and reputation management systems. The swift growth of e-commerce, social media, and various Digital platforms has facilitated the sharing of opinions and reviews by both legitimate customers and malicious actors, often influencing public perception and the success of businesses. Nevertheless, the rising occurrence of false or misleading feedback, including counterfeit reviews, deceptive ratings, and altered testimonials, poses considerable challenges in differentiating between genuine responses and misleading content.
Conventional techniques for identifying fraudulent feedback frequently prove inadequate because of the continuously changing strategies employed by individuals who create deceptive responses, which increasingly resemble authentic user interactions. To combat this escalating issue, machine learning offers a robust, data-centrist solution by utilizing sophisticated algorithms and models capable of automatically scrutinizing extensive amounts of feedback, thereby uncovering concealed patterns and nuanced inconsistencies. Methods including natural language processing (NLP), sentiment analysis, and behavioral pattern recognition have been incorporated into machine learning models to evaluate the credibility of feedback by analyzing linguistic characteristics, sentiment patterns, and user interaction behaviors. Additionally, supervised learning techniques allow for the training of systems on labeled datasets that include both genuine and artificial feedback, thereby improving their capacity to accurately classify feedback and reduce the occurrence of false positives. The implementation of multi-modal models that integrate diverse machine learning techniques enhances the reliability and accuracy of detection systems. This advancement

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2682- 2686
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0370

provides businesses and platforms with scalable, automated solutions that contribute to user trust, safeguard brand reputation, and maintain the integrity of online feedback ecosystems. As the dependence on digital feedback increases across various sectors, machine learning will be crucial in addressing the challenges posed by fraudulent feedback and promoting transparency within online communities.

2. Literature Survey

Yash Khare, Tejas Bhadane, and Khivasara are affiliated with the School of Computer Engineering and Technology at MIT World Peace University in Pune, India. [1] in his paper Fake News Detection System using Web-Extension. The methodology outlined in this paper introduces a web-based tool designed to detect fake news through the application of various machine learning models, notably Long Short-Term Memory (LSTM) networks in conjunction with GloVe word embeddings and GPT-2. During the data preprocessing phase, the titles and bodies of news articles undergo cleaning via natural language processing (NLP) methods, which involve the removal of URLs, special characters, and stop words. This is followed by the vectorization process utilizing GloVe embeddings to transform the text into numerical vectors. The LSTM model, a form of recurrent neural network (RNN), is utilized to capture long-term dependencies within the text, enabling the differentiation between genuine and false news by examining semantic patterns. When paired with GloVe embeddings, LSTM enhances the understanding of word context. Furthermore, a GPT-2 Output Detector is incorporated to recognize AI-generated content, thereby aiding in the identification of fabricated news articles. The system is implemented as a web extension, which empowers users to report suspicious content. This content is subsequently analyzed by both models (LSTM + GloVe and GPT-2), yielding a probability score that reflects the authenticity of the news. Additionally, the extension flags URLs associated with fake news for future reference. Trained on a data set comprising both real and fake news articles, the system achieved a remarkable 98.6% accuracy with the LSTM + GloVe model, surpassing traditional models such as Naïve Bayes, which recorded an accuracy of 88.91%. Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim, and Ammar Mohammed [2] in his paper, Fake Reviews Detection using Supervised Machine Learning. The proposed approach for identifying fraudulent reviews employs a blend of supervised machine learning methods and feature engineering, concentrating on the extraction of both textual and behavioral attributes from reviews and their authors to improve detection precision. During the data pre-processing stage, the Yelp data set is cleaned and prepared through techniques such as tokenization, removal of stop words, and lemmatization to ensure text uniformity. Feature extraction encompasses the collection of textual attributes, including sentiment analysis, TF-IDF, and cosine similarity, as well as behavioral characteristics such as the count of capital letters, punctuation frequency, presence of emojis, and patterns in the timing of review submissions. A variety of classifiers, such as K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forest, are utilized. The models are assessed using bi-gram and tri-gram language models in conjunction with TF-IDF for textual feature extraction. The system's effectiveness is evaluated based on accuracy, precision, recall, and f1-score, utilizing a dataset of over 5,000 reviews from Yelp. The findings indicate that KNN (K=7) surpasses other classifiers, achieving an f1-score of 82.40%, which increases to 86.20% with the inclusion of reviewer behavioral features, underscoring the importance of behavioral data in the detection of fake reviews. M.N. Istiaq Ahsan, Tamzid Nahian, Abdullah All Kafi, Md. Ismail Hossain, and Faisal Muhammad Shah from the Department of Computer Science & Engineering, Ahsanullah University of Science & Technology, Dhaka, Bangladesh, et al. [3] in his paper An Ensemble Approach to Detect Review Spam Using Hybrid Machine Learning Technique. This methodology integrates supervised learning with active learning techniques to identify fraudulent reviews. In the initial phase, the system detects and eliminates duplicate reviews by employing Kullback-Leibler divergence (KLD) and Jensen-Shannon divergence (JSD) to assess text similarity. The second phase is dedicated to creating a hybrid dataset that

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2682- 2686
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0370

combines authentic and fabricated reviews, utilizing active learning to label ambiguous instances for enhanced training efficacy. During the third phase, this dataset is utilized to train various classifiers, including Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and Maximum Entropy (Maxent), incorporating features such as unigram, bigram, and trigram word sets to improve accuracy. The model, tested on a mixture of genuine and synthetic reviews, demonstrated outstanding performance, with Naive Bayes utilizing bigram features achieving 95% precision and 88% accuracy, thereby confirming its capability in detecting review spam. Elshrif Elmurngi and Abdelouahed Gherbi, affiliated with the Department of Software and IT Engineering at École de Technologie Supérieure in Montreal, Canada., et al. [4] in his paper An Empirical Study on Detecting Fake Reviews Using Machine Learning Techniques. The research examines the identification of fraudulent reviews using Sentiment Analysis (SA) techniques, employing a data set comprising 2,000 movie reviews, evenly split between 1,000 positive and 1,000 negative reviews for the purpose of classification. The approach initiates with the gathering of movie reviews, followed by a data pre processing phase that includes the removal of stop words—such as "a," "the," and "of"—to discard terms that do not significantly aid in the classification task. This process is executed using the String To Word Vector filter available in the Weka software. Following this, four feature selection techniques are implemented to improve classification accuracy by eliminating irrelevant features. For the sentiment classification task, four machine learning algorithms are utilized: Naïve Bayes (NB), a probabilistic classifier grounded in Bayes' theorem; Support Vector Machine (SVM), a supervised learning model that discerns patterns for classification; K-Nearest Neighbor (KNN), a non-parametric approach reliant on distance metrics; and Decision Tree (DT-J48), which employs tree structures for classification purposes. The identification of fake reviews is supported by analyzing the outcomes through a confusion matrix, which evaluates true positives, false positives, true negatives, and false negatives. Ultimately, a comparative analysis of the results based on accuracy and precision indicates that SVM surpasses the other algorithms, achieving the highest accuracy in both scenarios, with and without stop word removal. Gowri Ramachandran, Daniel Nemeth, David Neville, Dimitrii Zhelezov, Ahmet Yalçin, Oliver Fohrmann, and Bhaskar Krishnamachari are the authors associated with the Viterbi School of Engineering at the University of Southern California and the Helix Foundation located in Berlin. et al. [5] proposes an WhistleBlower: Towards A Decentralized and Open Platform for Spotting Fake News. The proposed methodology presents WhistleBlower, a decentralized platform aimed at identifying fake news through the utilization of blockchain and Distributed Ledger Technology (DLT). It incorporates Artificial Intelligence (AI) and Machine Learning (ML) algorithms to enhance the accuracy of fake news detection. At the core of this platform are the detection algorithms, which assess the credibility of news articles by analyzing both their content and sources. Furthermore, the platform includes a verifiable computation framework that allows community members to run these detection algorithms on their own nodes, thereby ensuring the integrity of the results through honest computation validation. The architecture also features a Token-Curated Registry (TCR) that permits community members to contest and improve the detection algorithms. This TCR maintains a curated list of algorithms, enabling users to raise challenges if they believe an algorithm's assessment is flawed. Community engagement is encouraged, as participants earn tokens for their contributions to the cu-ration process. The system functions in a decentralized fashion, distributing computations across public nodes, which allows the community to evaluate the effectiveness of the algorithms. This innovative design not only enhances transparency but also reduces the risks linked to centralized governance, thereby fostering a collaborative and reliable environment for the detection of fake news. Claudio Marche, Ilaria Cabiddu, Christian Giovanni Castangia, Luigi Serreli, and Michele Nitti are associated with the Department of Electrical and Electronic Engineering (DIEE) at the University of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2682- 2686
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0370

Cagliari, as well as the National Telecommunication Inter University Consortium located in Cagliari, Italy. et al. [6] has proposed a Implementation of a Multi-Approach Fake News Detector and of a Trust Management Model for News Sources. The document presents a detailed two-part system aimed at identifying fake news and assessing the reliability of news sources [7]. The initial segment, known as the Fake News Detector, scrutinizes the textual content of news articles through various machine learning methodologies. It categorizes news as either authentic or fraudulent by examining several elements, including writing style—employing linguistic characteristics such as text length, informality, complexity, and variety—fact-checking by juxtaposing claims in the news against verified statements from a pre-trained network like FEVER, and sentiment analysis to evaluate the alignment between the headline and the article's content while ensuring objectivity [8]. This detector has been trained and validated using a significant dataset from Kaggle, which includes over 20,000 news articles. The second segment, the Trust Management Model, assesses the credibility of news sources based on multiple factors: expertise, which gauges the quantity and quality of news a source produces on particular subjects; relevance, which examines the frequency of requests for the source's news; and goodwill and coherence, which evaluates the historical dependability and consistency of the source over time, taking into account evolving behaviors such as misleading or deceptive news tactics [9]. This model also employs a prebunking strategy, designed to pinpoint unreliable sources before misinformation can proliferate. Both components utilize machine learning techniques and are capable of simulating real-time evaluations of news. Furthermore, the document investigates the potential use of blockchain technology for the secure storage and management of news assessments, highlighting its benefits in comparison to conventional databases [10].

3. Discussion

The initiative named "Fake Feedback Detection using Machine Learning" tackles a significant issue in the contemporary digital environment: the widespread occurrence of fraudulent reviews and misleading feedback across numerous platforms. This problem affects consumers, businesses, and online marketplaces alike, as it erodes trust and misrepresents the actual worth of products and services. Employing machine learning (ML) methodologies to identify fake feedback presents a viable solution to alleviate this concern, thereby improving consumer protection and preserving the authenticity of online reviews. In our research, we examined various machine learning algorithms, including Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and ensemble techniques like Random Forest. The findings revealed that although all models were effective in differentiating between authentic and fraudulent reviews, the Random Forest model demonstrated superior accuracy and resilience. This observation is consistent with existing research indicating that ensemble methods frequently surpass individual algorithms by combining predictions and mitigating overfitting. Additionally, the application of feature extraction methods, such as sentiment analysis and the identification of linguistic patterns, significantly improved the performance of the models. For example, reviews characterized by an abundance of emotional language or the use of unconventional phrases were more likely to be flagged as fraudulent.

Conclusion

The project named "Fake Feedback Detection using Machine Learning" signifies a notable progression in the ongoing efforts to combat fraudulent practices related to online reviews, which play a crucial role in influencing consumer choices within the rapidly expanding e-commerce landscape. By utilizing advanced machine learning methodologies, specifically the Random Forest algorithm, we have successfully established a comprehensive framework that can effectively differentiate between genuine and misleading feedback. This distinction is made possible through an in-depth examination of various attributes, such as sentiment and linguistic features, which uncover subtle indicators of fraudulent activity. Although our findings indicate the model's considerable promise for automating the detection of fake feedback, several challenges persist that must be addressed to ensure its sustained efficacy. A

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2682- 2686
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0370

significant concern is the prevalence of imbalanced datasets, where authentic reviews significantly outnumber fraudulent ones, which poses a considerable risk of skewed results that could undermine the model's precision and dependability. Additionally, as deceptive strategies continue to evolve and grow more intricate, there is an urgent necessity for continuous adaptation and retraining of the model to preserve its relevance and effectiveness in practical applications. Ethical considerations are also paramount in the implementation of such systems; the potential for false positives could result in unfair repercussions for legitimate users, thereby diminishing trust in the detection system itself.

References

[1]. Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim, Ammar Mohammed, "Fake Reviews Detection using Supervised Machine Learning," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 12, No. 1, 2021.

[2]. Claudio Marche, Ilaria Cabiddu, Christian Giovanni Castangia, Luigi Serreli, and Michele Nitti. "Implementation of a Multi-Approach Fake News Detector and of a Trust Management Model for News Sources." IEEE Transactions on Services Computing, vol. 16, no. 6, pp. 4288-4300, Nov./Dec. 2023. DOI: 10.1109/TSC.2023.3311629.

[3]. Elmurngi, E., & Gherbi, A. (2017). "An empirical study on detecting fake reviews using machine learning techniques". In Proceedings of the Seventh International Conference on Innovative Computing Technology (INTECH 2017) (pp. 107–114). IEEE.

[4]. Faiza Masood, Ghana Ammad, Ahmad Almogren, Assad Abbas, Hasan Ali Khattak, Ikram Ud Din, Mohsen Guizani, and Mansour Zuair. "Spammer Detection and Fake User Identification on Social Networks." IEEE Acces, vol. 7, pp. 68140–68150, June 2019. DOI: 10.1109/ACCESS.2019.2918196.

[5]. Mohammed Ennaouri and Ahmed Zellou, "Machine Learning Approaches for Fake Reviews Detection: A Systematic Literature Review," Journal of Web Engineering, vol. 22, no. 5, pp. 821–848, Dec. 2023. DOI: 10.13052/jwe1540-9589.2254.

[6]. M.N. Istiaq Ahsan, Tamzid Nahian, Abdullah All Kafi, Md. Ismail Hossain, and Faisal Muhammad Shah1. "An Ensemble approach to detect Review Spam using hybrid Machine Learning Technique." 19th International Conference on Computer and Information Technology, North South University, Dhaka, Bangladesh, pp. 388–394, Dec. 20162. DOI: 10.1109/CIT.2016.31.

[7]. Nidhi A. Patel, Prof .Rakesh Patel "A Survey on Fake Review Detection using Machine Learning Techniques" and it was presented in the year 2018 4th International Conference on Computing Communication and Automation (ICCCA).

[8]. Neville, Dimitrii Zhelezov, Ahmet Yalçin, Oliver Fohrmann, and Bhaskar Krishnamachari. "WhistleBlower: Towards A Decentralized and Open Platform for Spotting Fake News." 2020 IEEE International Conference onBlockchain , pp. 154-161, 2020. DOI: 10.1109/Blockchain50366.2020.00026.

[9]. Rami Mohawesh, Shuxiang Xu, Yaser Jararweh, Sumbal Maqsood "Fake Reviews Detection: A Survey" was presented at the year of 6 May 2021in IEEE.

[10]. Yash Khivasara, Yash Khare, and Tejas Bhadane. "Fake News Detection System Using Web-Extension." 2020 IEEE Pune Section International Conference (PuneCon), Vishwakarma Institute of Technology, Pune, India, pp. 119–123, Dec. 2020. DOI: 10.1109/PuneCon50868.2020.9362384.