
Shagufta Khalif¹, Kishor Mane²

¹,²D.Y.Patil College of Engineering & Technology, Kasaba Bawda, Kolhapur, Maharashtra, India.

Emails: khalifsonu084@gmail.com¹, kishortmane2022@gmail.com²

Abstract
Over the last few years, the impact of reviews on ecommerce industries and the folk who rely on the online reviews have increased. Most of the online shoppers depend on the reviews to make their purchasing decisions. Moreover, genuine reviews can assist the businesses acquire higher sales. Implementing effective systems can ensure the reliability of reviews and create trustworthiness on online platforms. Deceptive reviews can deteriorate the integrity of online feedback system. Numerous studies have explored this domain employing Machine Learning, Deep Learning, NLP methodologies etc., in the last decade. In this paper, through the survey we address the challenges which the current systems encounter where they lack adequate capabilities in detection and removal of false reviews. Diverse methods are applied on datasets like Amazon, Yelp etc., to obtain organized information and to improve the performance of the employed systems in classifying reviews as fake or genuine. Furthermore, this paper provides details of each method, their accuracies and also the future directions in this area. There is a pressing need of systems that can effectively address the issue of fake reviews as the existence of these can mislead customers leading to the decline in their preference for ecommerce.

Keywords: Deep Learning; Fake Reviews; Machine Learning Techniques.

1. Introduction
Online reviews are essential aspect of the modern consumer experience, as they provide valuable information and experiences in the execution of items and quality of items. However, not all online reviews are created as genuine, some may be fake or manipulated, potentially misleading consumers and harming businesses. In this world of Internet and E-commerce, customers can post their reviews on different platforms sharing their experiences about the product, which may help both consumers and organizations. For several years it has been seen that the quantity of people prefer to evaluate customer reviews before selecting a particular product. Referring these reviews, it makes the work of the buyer easier in making purchase decisions. While opinion mining-based systems have the potential to protect consumers and maintain the integrity of online reviews, it's important to ensure that they are accurate and reliable and that they don't infringe on individuals' freedom of expression. In this context, it's worth exploring the potential benefits and challenges of opinion mining-based fake product review checking and expulsion systems, as well as their ethical implications. Machine learning plays a pivotal role in fake product review monitoring by enabling automated and data-driven review classification. Leveraging supervised and unsupervised learning techniques, machine learning models are trained to distinguish between genuine and fake product reviews. These models analyse a wide range of textual features, including sentiment, linguistic patterns, and semantic cues, to make predictions with a high degree of accuracy. Additionally, machine learning models can adapt and evolve over time, improving their ability to detect increasingly sophisticated fake review tactics. As a result, they provide a scalable and efficient solution for
businesses and platforms to maintain the integrity of their product reviews, enhance consumer trust, and make informed decisions based on authentic customer feedback.

2. Literature Review

Opinion mining-based fake product review checking and filtering systems have emerged as a useful tool for identifying and removing fake reviews. The different authors used different techniques for fake review identification. The people who rely on e-commerce has risen, and this reliance has been totally on the base of online reviews which has significantly impacted their purchase decisions. The proposed reserves [1] involves Deep Learning methodologies such as LSTM and Convolutional Neural Network to find fake reviews. In this system two datasets are used one is gold standard dataset and another Yelp dataset. Also, it compares the efficiencies of deep learning techniques with conventional ML classifiers such as KNN, Naïve Bayes and SVM also known as state of art techniques. Ultimately it study proves that deep learning methods perform superior than the ML techniques in solving the problem caused by the fraudulent reviews in the e-commerce industries. Additionally, the work also emphasizes that by incorporating scalable dataset and using hybrid deep learning methods would provide better result in combating this issue.

In this paper [2] the author outlines a method for identification of fake news on different social media by making use of machine learning techniques. Mainly the fake news is the issue prevailing in current scenario and also the concern for trustworthiness of many people. To tackle this problem, the authors use RNN to identify fake profiles. The methods involve extraction of data from platforms like Twitter using the APIs. Stringent training rules are utilized to identify the authenticity of profiles on the social media. To eliminate machine generated entries a policy-based approach is used in pre-processing and filtration step. Features such as friends count, total followers, tweet counts and retweets are used to obtain the maximum accuracy. The RNN classifies profiles to improve accuracy using different activation functions and this system refutes to be efficient in detecting malicious accounts. The study achieves notable results, with a 0.96 accuracy on Twitter dataset and 98% accuracy on processed social media datasets, showcasing the strength of RNNs for addressing the challenges constituted by duplicate identities on social media. However, this study needs improvement, as large number of parameters for example metadata shared by the client, the dependability of shared data, and so on, are not viewed which can be useful in knowing the account’s legality.

The article [3] focuses on leveraging the importance of sentiment knowledge by identifying the limitations in the pre-trained language models like BERT, SKEP and TextCNN. This study introduces a fake review detection system in light of pre-prepared model and CNN where SKEP is a pre-trained system on enhancement on the knowledge of sentiments. A number of trials are carried on gold datasets namely Hotel, Restaurants, Doctor etc. and compares it with three baseline models and the results illustrates that BSTC surpasses the rest state of art methods in detection of fake reviews. In this system the factors considered were still not effective enough, so if it is merged with other pre-trained models, it may be possible to fortify the model’s accomplishment in finding fake reviews.

The study [4] focuses on the most serious issue faced by the customers on ecommerce websites i.e. Fake product reviews which affect their purchasing decisions as well as harm the businesses. The authors explore the cues and consequences associated with user suspicion of the authenticity of these reviews. In this article two important questions are raised initially, which are the attitudinal and physiological influences on decision making and buying when customer is doubtful of review text. The authors investigate the signs which led customer to decide a fake review, such as timing, language patterns, and sentiments. Their findings emphasize the importance of tackling fake reviews to maintain user trust and the reputation of e-commerce platforms. Understanding the consequences of consumer suspicion is crucial, as it
can lead to hesitancy in making purchases and a lack of trust in online platforms.

The authors [5] focus on the issue of tackling the fraudulent reviews available on different commercial websites. The system proposes Supervised ML algorithms such as Naïve Bayes, Support Vector Machine, etc. By using Yelp Dataset based on textual reviews the sentiment analysis classification has been included which classifies reviews into fraud and not fraud. The study also incorporated three classifiers which help in improving the efficiency. Finally, the system outcomes reveal that SVM can produce maximum accuracy when contrasted to the other methods used in the proposed system. The future work of this system suggests to deal with data from frequently varying graphs in additional fields, strongly recommended item for online shopping site for financial services to distinguish internet spamming could be taken into account.

The study [6] investigates the issue of fake reviews by using NLP, text blob and TFIDF by dividing the reviews in two classes i.e. pretend and real. Furthermore, the study divides the actual reviews in two more classes that is smart reviews and dangerous reviews after analyzing and detecting. Also, this system introduces measures to block the addresses of the users who try to post fake reviews and notify them by sending mail to the users regarding block information. By using NLP and sentiment analysis, the researchers of this paper aim to identify spammers and delete the fake reviews ultimately enhancing the trustworthiness of the review system for online shoppers. Automatic deletion/blocking of user addresses can be achieved in the future work of this system as deletion or blocking can be troublesome if done manually.

This research [7] aims at providing an efficient way of detecting false reviews through the reviewer’s profile details. The main goal of this work is to suggest high quality products to the user. SVM classifier is incorporated in this system. In this research the buyer can post the review, it is identified through UserID and Booking ID. Later Sentiment Analysis is performed and by using SVM the reviews are classified under the categories positive, negative and neutral. By examining ratings, reviews and smileys, positive and negative attitudes are determined. After this the negative reviews are dispensed with and positive reviews are added to the dataset and recommends genuine products. Customers can post the reviews only once this will help to identify fakes reviews.

The author [8] proposes an automated method for detecting fake reviews using text cleaning and supervised ML techniques. This research uses a current fake review dataset named ChatGPT-Fate-It, which are gathered from Google Play and ChatGPT. Various Data cleaning methods are used to obtained structured data; sentiment scores (positive, negative, neutral) are calculated using Sentiment Analyzer of NLTK in Python. A group of lexical and sentiment attributes are extracted to teach three classification algorithms out of which Decision tree outer performs in various evaluation parameters such as recall, accuracy and so on. This study recommends future research can be accomplished by taking into consideration additional features such as linguistic and semantic too.

The paper [9] uses a semi-supervised ML algorithm like Naïve Bayes, Random Forest for detecting fake reviews. An existing food review dataset from Amazon has been used to analyze, extract or interpret various reviewers’ behavior. Gaussian Naïve Bayes is used as it assumes features follow normal distribution. Preprocessing on the dataset has been carried out using various NLP steps. After training and testing using ML algorithms the performance is reviewed and accuracy is calculated. The model is evaluated and the results reveal that Random Forest surpasses the other classifiers in the procedure. Also, the combined accuracy after including behavioral changes produced was about 97.7%.

The author [10] portrays different ML systems in
detecting false reviews and does relative examination of every system for identifying finest of algorithms. The authors discuss about the previous works and list four most famous ways of investigating fake reviews: Naïve Bayes, Support Vector Machine, Latent Dirichlet analysis and RF. Additionally, the experimental results are computed using different evaluation parameters such as f1 score, accuracy, etc. Prominent algorithm is chosen by referring all the performance parameters. After the investigation and absolute analysis, it is proved that the Naïve Bayes outer performs the rest three methods in the paper as it is based on relative probability.

Pal, K., Poddar, S., Jayalakshmi, S.L., Choudhury, M., Saif Ahmed, S.K., Halder, S. [11] developed a model using Deep Learning techniques which integrates set of lexicons, a system for Word2vec, and AM to convey sentiment perception. In this approach Amazon Shoe dataset is used. Various preprocessing techniques are applied on the text such as Lemma, POS, stop word etc., to obtain the structured data which is the main step used in sentiment analysis. The Liu’s Lexicon and SentiWordNet were utilized to take word sentiment data and get word sentiment scores. This article incorporates a BiLSTM to join more information as semantics and sentiments. As stated in this system accuracy was 92% and in future to enhance the accuracy, abnormalities in the validation dataset can be improved.

This article [12] mainly aims on fake review detection by training a time sequential DL model to categorize reviews in the form of fake or genuine. The article uses Yelp dataset which includes 1.2million user reviews which is separated into four datasets based on different geographical locations. Feature extraction is done by mixing two classification rules of review centric and reviewer centric. To save the successive relationship of text data in review, time sequential model, LSTM is incorporated in this system. The experimental results, shown in this system depict the textual features outer perform physiological features and improve the factuality of the system.

Kamel, Ahmed & Salminen [13], addresses the prevalent issue of fake online product reviews by employing advanced language models, specifically ULMFiT and GPT-2. The research involves the creation of synthetic fake reviews using GPT-2, and the subsequent creation of a dataset for fake review detection. Notably, the author’s findings reveal that machine learning classifiers can achieve exemplary accuracy in identifying fake reviews, though human raters display significantly lower accuracy and understanding. The results obtained underline strength of machine learning algorithms in combating the rise of fraudulent online reviews, with potential implications for enhancing consumer trust, safeguarding businesses against unfair competition, and fostering greater responsibility among review platforms. A major drawback of machine detection of fake reviews is its potential for false positives, which can lead to legitimate reviews being erroneously flagged as fake.

Due to COVID-19 and its protocols like social distancing and other related reasons the number of people moving towards online shopping has increased. As most of the purchasers refer the reviews of the product before buying, the rise in the fraudulent reviews is much. So, to tackle this issue the authors [14], investigated the impact of fake reviews on the ecommerce websites has created impact on both the customer, as well as the businesses. Fake reviews can significantly affect the consumers' purchasing decisions. In response to this the researcher proposes a fake review detection system that utilizes ML methods and text classification such as SVM, KNN and SKL. Their approach leverages sentiment analysis and linguistic features to achieve enhanced accuracy on Yelp and Trip advisory dataset contrasted with different state-of-the-art techniques, emphasizing its efficiency for identifying fraudulent reviews.

The work emphasis on creating a simple and feasible website which can identify and detect genuine reviews [15]. This website may have full influence of consumers and helps them to sense about the product or service. The website was setup and designed using supervised learning model.
Product data is extracted from the dataset of the URL entered; extraction is done by python tool for web-scrapping. The data is finally examined and segregated using NLP to extract necessary essential features. Using Machine Learning models for training the dataset, the results proves that the model’s accuracy is of 89.12% on these datasets. This research of review segregation, filtering and additionally the genuineness can be further extended to include the ability to strain the real reviews whether they are good or negative.

Fake product reviews have largely impacted the opinions of the folk preferring to opt for buying the product on ecommerce websites. The study [16] lists there are three types of spam product reviews. First indicates deceitful reviews, second includes Reviews of specific brand and third no reviews. This system involves preprocessing the text data by involving steps such as removing punctuation and stop words etc. The posted reviews will go through the course of sentiment analysis, Internet Protocol address tracking, and its fluctuation from general reviews. In case of miscalculations, reviews will be analyzed and detected. At last, the fake reviews are detected using the Machine Learning algorithm and classifies them whether fake or genuine. Thereafter the fake reviews are removed from the website so that the customers may not be misled.

These days, misleading the buyer by posting fake reviews are intentionally composed to fabricate virtual influence and drawing in expected customers [17]. The authors Elmogy and Mohammed proposed a Machine Learning methods to detect fake review and applies a few Feature Engineering to take various actions(behavior) of the reviews. The work experiments using Yelp datasets of restaurants with and without behavioral features and compares the results of few classifiers like SVM, LR, KNN, Naïve Bayes and Random Forest. The bi-gram and tri-gram language models are utilized and contrast with this approach. The after effects obtained conclude that K-Nearest Neighbor provides maximum f1score incorporating behavioral features. The work also spotlights that by taking into account more behavioral features like frequency of reviews, time taken to complete the reviews, how frequent the positive and negative reviews are posted etc., may further boost the performance of the presented approach.

The research [18] reviews the literature based on various Fake Review detection systems and discusses the bounded scope of eminence the recent researches have achieved and the regulation they have had in resolving the harm caused by deceptive reviews. The study displays how different ecommerce platforms, control the Fake reviews on their websites. Due to increasing deceptive reviews on online platforms the mistrust among users is rapidly increasing which makes difficult to the new users to build a trusted network. The author reviews the work of Lazer et al an interdisciplinary work which recommends two initiatives to combat false information user empowerment and information system structure revision. Lastly this literature survey emphasis on the need of exploring the area in both theoretical and application-based systems and suggests to modify existing systems to make feasible solutions against fake reviews influence.

This approach [19] proposes a model to detect spam reviews and remove them from a website. In this research the Reviews of Chicago hotel dataset from kaggle.com is used which includes fields such as title, author, text, label. The dataset is pre-processed by applying techniques like tokenization, eradication of stop words etc. The proposed work is built using SVM, N.Bayes and TF-IDF vectorizer which provides greater accuracy in detecting spam reviews. This study is limited to smaller dataset whereas accuracy and authenticity can be improved further by expanding to scalable datasets. K.Ajay, Loganayagi.S [21-23] aims at providing an effective way in determining which algorithm is best in detecting spam and fake reviews. The models RNN and CNN are trained with 80% of the dataset and 20% of dataset is used to test [24]. By distinguishing performance of Convolutional Neural Network and Recurrent Neural Network,
the result shows that RNN has higher mean and greater accuracy [25]. In order to train the model RNN requires larger dataset of labelled reviews. Assuming the user uses smaller or inferior quality dataset it is most likely that the system may be unable to exactly identify trends associated with extremist organization (Table 1).

### Table 1 Summary of Different Techniques Used for Fake Review Detection System

<table>
<thead>
<tr>
<th>Technique</th>
<th>Ref.</th>
<th>Description</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN, CNN, LSTM</td>
<td>[1]</td>
<td>It shows that deep learning methods perform superior than ML techniques.</td>
<td>Smaller Dataset</td>
</tr>
<tr>
<td>RNN</td>
<td>[2]</td>
<td>The RNN performs profile classification to improve accuracy.</td>
<td>Limited to Twitter Dataset</td>
</tr>
<tr>
<td>CNN, BSTC</td>
<td>[3]</td>
<td>BSTC considers TestCNN, BERT and SKEP are utilized where SKEP is a pre-trained system on enhancement on the knowledge of sentiments.</td>
<td>Standard dataset can be used</td>
</tr>
<tr>
<td>Literature Survey</td>
<td>[4]</td>
<td>Critical research questions are raised.</td>
<td>Investigation from this research reveals that differences between actual fraudulent reviews and actual truthful reviews could not be viewed as false.</td>
</tr>
<tr>
<td>NB, SVM, LR, Decision Trees</td>
<td>[5]</td>
<td>It uses three classifiers which help in improving the efficiency including feature extraction, sampling and classification algorithms.</td>
<td>Strong product recommendation for online commercial site for administration to detect internet spamming is considered.</td>
</tr>
<tr>
<td>NLP, TF-IDF, Textblob</td>
<td>[6]</td>
<td>It investigates the issue of fake reviews by using NLP, text blob and TFIDF by dividing the reviews in two classes i.e. pretend and real review</td>
<td>Automatic deletion/blocking of user addresses must be incorporated.</td>
</tr>
<tr>
<td>SVM</td>
<td>[7]</td>
<td>The main goal of this work is to suggest high quality products to the customer.</td>
<td>More features can be considered in detecting the count of reviews one posts.</td>
</tr>
<tr>
<td>Random Forest and Naïve Bayes</td>
<td>[9]</td>
<td>Gaussian Naïve Bayes is used as it assumes features follow normal distribution.</td>
<td>Behavioral features</td>
</tr>
<tr>
<td>Random Forest Naïve Bayes, SVM</td>
<td>[10]</td>
<td>This work uses comparative analysis of each model for identifying the finest of algorithms</td>
<td>Naïve Bayes Provides highest accuracy</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>[11]</td>
<td>In this approach BiLSTM is used and Amazon Shoe dataset is used. Various preprocessing techniques are applied on the text such as Lemma, POS, stop word etc.,</td>
<td>Anomalies from validation dataset can be refined.</td>
</tr>
<tr>
<td>LSTN and Time sequential model</td>
<td>[12]</td>
<td>Aims on fake review detection by training a time sequential deep learning model to classify reviews as fake or genuine.</td>
<td>Less behavioral features</td>
</tr>
</tbody>
</table>
Language models- ULMFiT, GPT2 [13] | The author’s findings reveal that machine learning classifiers can achieve exemplary accuracy in identifying fake reviews | potential for false positives

SVM, KNN, LR [14] | This approach leverages sentiment analysis and linguistic features to achieve enhanced accuracy | Limited to Yelp and TripAdvisory dataset

NLP and ML models [15] | It emphasis on creating a simple and feasible website which can identify and detect genuine reviews. | Review segregation and filtration can be included.

Different Machine learning Techniques [16] | The reviews will go through the course of sentiment analysis, Internet Protocol address tracking, and its fluctuation from general reviews. | Behavioral patterns not considered

KNN, NB, LR, RF, Language models BiGram and Trigram [17] | KNN provides maximum F1score by including behavioral features | Limited to less behavioral features

Literature review [18] | Interdisciplinary work has been carried out to combat false information. | FRD techniques cannot perform in real time.

NB, SVM, TF-IDF [19] | Provides greater accuracy in detecting spam reviews. | Limited to smaller dataset

RNN, CNN [20] | RNN has higher mean and greater accuracy | RNN can work better only on large dataset

**Conclusion**

This paper provides a comprehensive analysis of the existing research on Fake review detection and removal systems. Various methods from Machine learning and deep learning, has been compared with their accuracies and datasets used, explore the key trends, gaps and areas of future improvement. This review not only contributes to the current understanding of identifying deceptive reviews but also highlights the need for further investigation to advance the field.

**References**


[18]. Paul, H., Nikolaev, A. Fake review detection on online E-commerce platforms: a systematic literature review. Data Min Knowl Disc 35,


