# A Survey Paper on Fake Review Detection System

Nikita V. Khairnar, Shruti L. Mankar, Mrunali R. Pandav, Hitesh Kotecha, Manjiri Ranjanikar

Pimpri Chinchwad College of Engineering, Nigdi, Pune, India

Corresponding author: Nikita V. Khairnar, Email: nikitak22092001@gmail.com

Nowadays, E-commerce websites has become a vital part of regular life. The pandemic has taken a quick step towards the digital world and initiated changes in online shopping behaviors. While buying products online many people read the reviews related to the product and then decide on buying it. So, reviews play an important role if a customer wants to buy any product. This leads to increased spam review volume in social networks and e-commerce websites like Flipkart, Amazon, etc. Fake reviews can be used to demote a good product or to promote a bad product, so there is a need for robust and reliable techniques to detect fake reviews which can be beneficial to the customer as well as to the vendor. The objective of this survey paper is to get an overview of different methodologies used to solve such problems. This research presents a systematic review on methods to detect spam review using different Deep Learning (DL) Approaches, Machine Learning (ML) Methods, Natural Language Processing (NLP), and Sentiment Analysis.

**Keywords**: Fake Reviews, Machine Learning, Sentimental Analysis, Deep Learning, Natural Language Processing.

2021. In Rahul Srivastava & Aditya Kumar Singh Pundir (eds.), *New Frontiers in Communication and Intelligent Systems*, 625–634. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-95502-00-4-64

# 1 Introduction

Online reviews have much importance nowadays on e-commerce sites. Before spending cash on the product, customers search for reviews available about the product. But sometimes these reviews are fake. These fake reviews can be posted by companies to increase the sale of their products. Also, some fake reviews can be posted due to competition among various companies. These fake reviews can mislead customers before purchasing a product. Fake reviews can damage the online shopping platforms as well as customers' trust in the product and the site. So to avoid this first step towards solving the problem is the detection of fake reviews. Hence a system should be designed which can distinguish fake reviews from available reviews.

Various approaches have already been proposed to develop a fake review detection system by many authors. In this paper, we have studied those proposed approaches and techniques based on machine learning, deep learning, natural language processing, and sentiment analysis.

**Materials and Methods:**

1. Fake review detection using different Machine learning techniques
2. Fake Review Detection Using Deep Learning
3. Fake review detection using Sentimental Analysis
4. Fake Review Detection Using Natural Language Processing

The next sections of the paper are as follows:

- ➢ Literature survey
- ➢ Analysis of existing fake review techniques
- ➢ Conclusion
- ➢ Future scope
- ➢ References

# 2 Literature Survey

The literature survey is based on 4 techniques and is divided into 4 sections based on it. The sections are as follows:

**2.1 Fake review detection using different Machine learning techniques:** The main purpose of this part of the literature survey is to study Machine Learning techniques that can be used for fake review detection. This section of the literature survey mainly focuses on the supervised [3][7], unsupervised[3], Semi-supervised[4][5], and ensemble machine learning approaches[6].

When it comes to machine learning, the machine learning model learns and makes predictions. Thus in [1] Patel et al. talked about the basic steps involved in creating a machine learning model for the acquisition of false reviews. The steps are collection and processing of data, extracting and selecting various features, and creating a classifier model.

In [2] Etaiwi et al. analyzed and studied how character selection methods affect the process of obtaining fraudulent reviews using various machine learning algorithms. An analysis is done using parameters such as accuracy, recall, and accuracy. Two features bag-of-words and word count are used.

The conclusion is that the feature selection method greatly affects the selection of false reviews and also affects the results of all the classifier algorithms separately.

In [3] the solution is proposed using supervised machine learning, as well as unsupervised machine learning approaches. For supervised machine learning, approaches Linguistic features, N-gram, Sentiment, POS tagging are used and for unsupervised learning approaches the features are extracted by considering reviewer information, review content, and product information. Moreover the obtained results are satisfactory but the spam analysis is incomplete if information about the reviewer such as the time at which the review was written, how many reviews a reviewer has written, IP address of the reviewer are ignored.

In [4] a semi-supervised learning-based machine learning approach is proposed. In features such as grammar, sentiment polarity, total word count, and bigram frequency count are taken into account. The sentiment polarity of an element is nothing but the orientation of the expressed sentiment. Here only minimal functionality is considered by the author during classification which impacts the expected result.

In [5] Yılmaz et al. have discussed an approach developed using semi-supervised machine learning techniques. Both textual and network-based features are used. The features generated using network-based approaches consider how many reviews a reviewer has written, IP address of the reviewer. The results obtained confirm that considering both review data features and network-based features improves the detection of spam reviews in a significant manner. But the drawback is that good results are obtained only when both the reviewed product network and the textual content of the review are taken into consideration together.

In [6] Espinoza et al. have proposed an approach based on ensemble learning. Features are generated using the Doc2vec algorithm. Boosting, Bagging, Stacking are the three main approaches that are used by the author to create an ensemble learner. As a result, it is claimed that approaches that are based on ensemble learning outperform conventional machine learning algorithms when it comes to the detection of deceiving information.

In [7] an approach that uses supervised machine learning is discussed. Here textual and behavioral features were selected by the author for the experiment. In the proposed solution supervised machine learning algorithms are used as classifier algorithms. The results obtained show that KNN(where k=7) surpasses the performance of the rest of the classifier algorithms. But the drawback is that here features that depend on how frequently the reviewer writes the review, the time is taken by reviewers to complete a review, etc were not considered.

**Table 1.** Summarizes the literature survey for fake review detection using different machine learning techniques

| RNo | Year | Description | Algorithm | Dataset | Result | Limitations |
|---|---|---|---|---|---|---|
| [3] | 2017 | The Author has proposed a solution based on both supervised machine learning and unsupervised machine learning approaches | Naive Bayes(NB), SVM, Decision tree, KNN | Obtained a dataset of fake hotel reviews by paying anonymous online workers to write fake reviews for hotels. | obtained results are satisfactory but the spam analysis is incomplete if important information about the reviewer | Were not able to gather all information about a spam reviewer due to privacy concerns hence could not reach the expected result. |

| | | | | | | |
|---|---|---|---|---|---|---|
| [4] | 2017 | The author has discussed an approach that uses semi-supervised machine learning techniques. | k-Nearest Neighbor classifier, Logistic Regression classifier, Random | Review texts on 20 hotels in the Chicago area, USA, fake reviews generated after using Amazon Mechanical Turk(AMT) more reviews collected from online sites like Yelp, TripAdvisor, Expedia, Hotels | F-score increased after applying classification based on PU learning | Here only minimal functionality is considered by the author during classification which impacts the expected result |
| [5] | 2018 | The author has discussed an approach that uses semi-supervised machine learning techniques. Textual as well as network-based features are used by the author. | Logistic Regression, k-Nearest Neighbors, Decision Trees, Support Vector Machine, Naive Bayes, Doc2vec, Node2vec | Yelp dataset | Merging representations learned from reviewer product networks and textual review data improve the detection of spam reviews in a significant manner. | The proposed approach produces good results only when both the reviewed product network and the textual content of the review are taken into consideration together |
| [6] | 2020 | The author has discussed an approach that is based on ensemble learning. | Decision Tree, Random Forest, Support Vector Machines, Extreme Gradient-Boosting Trees, Multilayer Perceptron | Developed their fake review dataset "Restaurant dataset" and also added some reviews of restaurants from google which were given by verified users | ensemble learning-based approaches outperform conventional machine learning algorithms when it comes to the detection of deceiving information. | To produce good results using this method, more parsing and slicing need to be done on the dataset. |
| [7] | 2021 | The author has discussed an approach that uses supervised machine learning techniques. | Support vector machines, Naive Bayes, k-nearest neighbor, Decision tree, Random forest, Logistic regression | Yelp dataset | KNN surpasses the performance of the rest of the classifier algorithms in terms off-score. | features that depend on how frequently a reviewer writes a review, the time is taken by reviewers to complete a review, etc were not taken into account |

**2.2 Fake Review Detection Using Deep Learning:** During the past few decades, many problems such as disease detection, image recognition, video processing, and fraud detection have been solved using the deep learning approach. Below we provided the details about different techniques of deep learning used to solve the problem of fake reviews.

In [9] explored issues on the current twitter spam detection technique and proposed a method for the classification which is based on deep learning. Also compared it with the existing syntax-based method and feature extraction method. The result is more accurate and outperformed on some datasets.

In [10] the author presents an all-embracing survey of automated Mis-Information Detection (MID) on false or disinformation, spam, and rumors, as well as fake news and it also provide a level of futuristic development review on MID where DL provides the automation for processing the data and create distinct patterns to achieve the improved better results and also helps to extract the global features.

In [12] proposed a visionary approach for the phenomenon of fake review detection. In the approach, the gathering of words that relate to each other according to the semantics is carried out by a self-organizing map commonly known as SOP. The construction of an image with a fixed size is done and the grid map is used for their representation. These images were then provided to the model for training and then classified into fake or real images with the help of CNN. Primarily, the approach uses linguistic features and extracts the lexical diversity spread with these reviews.

In [14] proposed an attention-based deep structure as a means to improve the opinion spam review detection methods and make them more effective. The model is composed of a bidirectional LSTM coupled with a Multi-Headed Self-Attention mechanism. The architecture of the model is made of four types of layers: an input embedding layer, BiLSTM, an attention layer, and a softmax layer.

In [15] investigated the above problem of misinformation. They point up that the DL is one of the leading tactics to solve such a problem. Spreading fake news and information eventually confuses so to avoid this [17] proposed a solution to 4 similar types of neural networks models which use N-gram to solve the problem.

**Table 2.** summarizes the literature survey for fake review detection using a deep learning approach

| **Ref. No** | **Year** | **Description** | **Algorithm** | **Dataset** | **Results** |
|---|---|---|---|---|---|
| [8] | 2019 | Proposed DL models for detection of fake reviews and used different approaches like CNN, LSTM, RNN, WORD2VEC along with the ML Classifiers | CNN, RNN, LSTM, ML Classifiers | Yelp (unlabeled), Ott(labeled) | The author analyzed the results then compared them with previous work. Successfully got an effective solution with high accuracy. |
| [9] | 2017 | The author proposed a model which is based on sentimental analysis along with deep learning. Here, the architectures are used for detecting patterns in the data. Patterns like question marks, unusual symbols, etc. | Sentimental analysis algorithm along with deep learning models | Politifact | Ensemble networks performance is better than that of the other types of architecture. |
| [11] | 2019 | In this proposed method, there is a novel approach that is based on Deep Learning. The author uses Word2Vec then framed binary classifiers based on the previous dataset. | CNN, RNN, Word2Vec | ---------- | Successfully proposed a new classification method |

| | | | | | |
|---|---|---|---|---|---|
| [13] | 2021 | The proposed approach uses ML and DL-based analysis. It includes the approach which will work reactively by providing the pop-up on the browser to show that there is something wrong or spam is present. | ML Classifiers along with DL algorithms | ---------- | 99.4% of accuracy |

**2.3. Fake review detection using Sentimental Analysis:** The fake review detection system can be designed using various approaches. The sentiment analysis approach is one among those. Various algorithms are included in the sentimental analysis approach. The algorithms include Naive Bayes Algorithm, Decision tree algorithm, k* algorithm, support vector machine (SVM), etc. The algorithms are based on different mathematical approaches. The sentiment analysis approach makes efficient organization of data and organizes the reviews as fake and genuine ones. The system to be developed aims to find fake reviews from available ones.

Punde et al [16] authors have explained the sentimental analysis approach using the two algorithms: Naive Bayes and Decision tree. The naive Bayes algorithm uses probability concepts in mathematics. It considers the two independent events as input and calculates the relative probabilities for the dataset. The laplacian smoothing is applied as 1.0 for training the dataset to estimate a value. This algorithm performs better with an accuracy of 96.89%.

Kauffmann et al [17], the case study of sentimental analysis is briefly explained. A large amount of data is considered in a dataset. The sentiment Analysis approach deals with the study of Natural Language Processing deals with identifying the mood or opinion of subjective elements within a text. Firstly, the classification is done based on positive and negative reviews.

Pujari et al [18], a comparison of various sentimental analysis algorithms is done. The algorithms include Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM). The performance of various algorithms is compared to find out the efficient algorithm for the detection of fake reviews. Of this, SVM achieves better efficiency than others.

Elmurngi et al [19] used sentiment analysis and a bag of words for predicting the accuracy. The classification criteria considered by the author were positive and negative emotions. In this study, the author used a hybrid textual analysis. The approach in the research paper combines machine learning approaches along with a dictionary. SVM and k-NN perform better for the lower-dimensional datasets.

Naive Bayes algorithm in sentimental analysis is applied which is based on probability distribution concept. Suppose A and B are two independent events then probability will be calculated as:

$$P(A,B) = P(A).P(B)$$

The Naive Bayes assumption can be formulated with below equation as:

$$P(y\text{-}x1 \ldots\ldots X) = P(x1\,|y)\, P(x2|y)\, ..P(xn|y)\, P(y)/P(x1)P(x2)...P(xn)$$

In this equation y is class variable and x is dependent feature of vector (size n).

Naive Bayes finds the probability of the occasion happening given the probability of event that happened.

**Table 3.** Summarizes the literature survey for fake review detection using sentimental analysis

| Ref. No | Year | Description | Algorithm | Dataset | Results | Limitations |
|---|---|---|---|---|---|---|
| 16 | 2019 | For the classification of data, data is tokenized by applying a bag of words model. The framework mainly utilizes data mining. | Naive Bayes and decision tree algorithm | ------- | Naive Bayes gives an efficiency of 96.89%. | Soft computing is not properly used in literature. |
| 17 | 2019 | The data is classified based on positive and negative reviews. The rating given by the customer is considered a parameter for classification. Polarity classification is a basic task. | Naive Bayes Probability-based algorithm, decision tree algorithm, Support vector machine (SVM). | Of the total of 1971, reviews were considered for evaluation. Out of those, 1328 were categorized as fake reviews and the remaining were predicted correctly. | From 781 reviews only 14.47% were predicted as fake reviews that were incorrect. | The literature has used online retailing platforms like Amazon. So the research is restricted to this platform only. |
| 18 | 2018 | A comparative study of different sentiment analysis algorithms is done. Various distinguishing parameters are considered while comparing. | Maximum Entropy, Decision tree algorithm, Naive Bayes Probability-based algorithm, support vector machine (SVM) | ---------- | SVM achieves better efficiency among all algorithms. | For comparison, a larger dataset was taken. |
| 19 | 2017 | The two general approaches are used-machine learning and dictionary-based. | Support Vector Machine(SVM), Naive Bayes, K-Nearest Neighbour. | ---------- | SVM performs better with large datasets. | A larger dataset was used for comparison. |

**2.4. Fake Review Detection Using NLP:** Against the changes the world is currently facing, an increasing number of daily tasks are moving online, including reading the news and being informed about relevant topics. The increase in the volume of information made available led, in turn, to fake news dissemination becoming a trending topic on the Internet. Recent events such as the COVID-19 pandemic have shown that fake news exerts a significant negative influence on societies by highlighting stories that do not necessarily report on facts. Moreover, fake news serves to misinform people and manipulate their opinions for several reasons. Research communities expressed concern about this flood of misinformation and introduced automated fake news identification solutions based on Natural Language Processing (NLP) techniques. In this study, we reference existing datasets and related work in the analysis of English fake news, discuss potential detection techniques for Romanian fake news, as well as establish future work plans for this research initiative.

**Datasets:** A challenging and crucial step in fake news identification consists of building a relevant corpus containing labeled articles. It requires a lot of effort to compile a clear objective set of articles, with large numbers of entries. To reduce the time and energy invested in such activities, certain research groups assert that the content generated by specific entities is objective and describes facts – for example, CNN and PolitiFact.

**Fake News challenge:** It is one of the most important datasets that is intensively used for testing machine learning models. The system is presented with the headline and the content of the article and is designed to determine whether the content agrees, disagrees, is discussed, or is not entirely related to the topic. Humanities analysts can use such a tool to gather different opinions on a particular subject, greatly reducing the amount of time needed to research a particular topic.

**ClaimBuster:** ClaimBuster is a tool that detects check-worthy sentences in political discourses and debates, highlighting pieces from candidates' speeches that may require additional attention. Such a solution reduces the time journalists invest in interpreting complex discourses, by spotting top priority phrases in need of a check-worthy analysis before becoming viral and misinforming information. Roughly 20,000 sentences were selected to be analyzed and categorized into "False Judgment," "Non-True Judgment" and "Improper Judgment."

**Politifact:** The 12,800 short statements were randomly selected according to the 6 categories defined by Politifact pants-fire, false, barely true, half-true, mostly-true, and true. On paper, the LIAR database described by the paper introduces the **ability** to use machine learning algorithms to analyze such statements and classify politically relevant data, by defining a structured, defined, and large enough database for such tests.

**FEND:** FEND (Fake News Detection), the architecture proposed by the authors, is centered on the idea of clustering news with similar subjects into groups, to easily compare fake articles with real ones addressing the same topic. The initial step consisted of identifying events and topics for every news article while building clusters that share a similar subset of topics in that group, whereas different clusters do not share common topics in their representative subset. The underlying assumption was that CNN and New York Times are legitimate news providers.

## 3 Results

Many different approaches and strategies to detect fake reviews are proposed and discussed by various authors. Authors have analyzed the process of feature selection, How the performance of various machine learning classifier algorithms is affected depending on the features, and how well fake review detection is done by applying different machine learning techniques. After studying these we learned about the common flow followed by various solutions to detect fake reviews which is shown in figure 1.

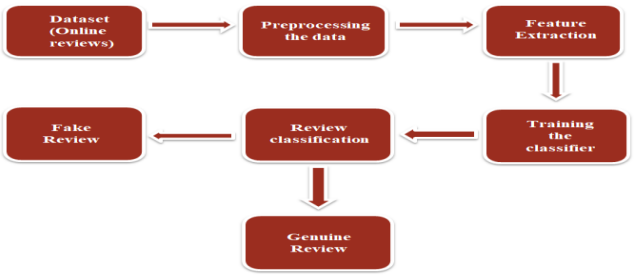


**Fig. 1.** Demonstrates the steps that are followed to detect fake reviews

| Algorithm | Accuracy |
|---|---|
| Machine Learning | 76.18% |
| Deep Learning | 81.92% |
| Sentimental analysis | 69.24% |

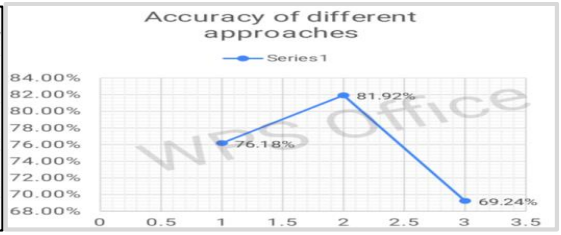


**Fig. 2.** Demonstrates comparison of different techniques

Many efforts have been made to detect the fake review. The different approaches were used for the identification of fake or genuine reviews. After studying the algorithms, we found the level of effectiveness of these approaches. As shown in the figure 2, it is clear that effectiveness of deep learning based approach is 81.92%, which is highest among the three mentioned approaches [8].

## 4 Discussion:

Recently, the review detection trap attention due to the impact on customers' behavior and purchasing decisions. We studied and analyzed different techniques that have been proposed for fake review detection so far. The survey covers the techniques which are based on deep learning, machine learning, natural language processing, and sentiment analysis. For machine learning-based proposed solutions we have discussed supervised, unsupervised, semi-supervised as well as ensemble learning-based methodologies. The best performance achieved using the combination of both techniques is Feature engineering techniques, Machine Learning, and Deep Learning Approaches. Additionally, the study discussed in the paper primarily focused on the result by applying it to the real-time dataset.

In the future, a model that detects fake reviews as well as the reviewer which continuously spams the reviews using one or more accounts can be developed, and then accordingly a system to restrict or block such review accounts can be integrated with the model. Also as the dataset increases, the efficiency of the fake review detecting system also increases. So in the future, more datasets can be used to make the system more effective. The methods that are discussed in the paper will contribute to quality management, research areas as they will help in eliminating false information.

# References

[1] Patel, N. A. and Patel, K. (2018). A survey on fake review detection using machine learning techniques. In *4th International Conference on Computing Communication and Automation (ICCCA)*, 1-6.

[2] Wael, E. and Awajan, A. (2017). The effects of features selection methods on spam review detection performance. In *International Conference on New Trends in Computing Sciences (ICTCS)*, 116-120.

[3] Rout, J. K. et al. (2017). Deceptive review detection using labeled and unlabeled data. *Multimedia Tools and Applications*, 76(3): 3187-3211.

[4] Rout, J. K. et al. (2017). Revisiting semi-supervised learning for online deceptive review detection. *IEEE Access*, 5: 1319-1327.

[5] Yilmaz, C. M. and Durahim, A. O. (2018). SPR2EP: A semi-supervised spam review detection framework. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 306-313.

[6] Espinoza, G. et al. (2020). Ensemble learning for detecting fake reviews. In *IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 1320-1325.

[7] Elmogy, A. M. et al. (2021). Fake Reviews Detection using Supervised Machine Learning. *International Journal of Advanced Computer Science Applications*, 12: 601-606.

[8] Shahariar, G. M. et al. (2019). Spam review detection using deep learning. In *IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 0027-0033.

[9] Wu, T. et al. (2017). Twitter spam detection based on deep learning. In *Proceedings of the Australasian Computer Science week Multiconference*, 1-8.

[10] Islam, Md R. et al. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, 10(1): 1-20.

[11] Kumar, S. et al. (2020). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies*, 31(2): e3767.

[12] Neisari, A., Luis, R. and Saad, S. (2021). Spam review detection using self-organizing maps and convolutional neural networks. *Computers & Security*, 106: 102274.

[13] Sahoo, Ranjan, S. and Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, 100: 106983.

[14] Sedighi, Z. et al. (2019). Opinion Spam Detection with Attention-Based Neural Networks. *Florida Artificial Intelligence Research Society Conference*, 1-5.

[15] Lee, D. H. et al. (2019). Fake News Detection using Deep Learning. *Journal of Information Processing Systems*, 15(5): 1119-1130.

[16] Punde, A. et al. (2019). Fake Product Review Monitoring & Removal and Sentiment Analysis of Genuine Reviews. *International Journal of Engineering and Management Research*, 9: 107-110.

[17] Kauffmann, E. et al. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90: 523-537.

[18] Pujari, C., Aiswarya and Shetty, N.P. (2018). Comparison of Classification Techniques for Feature-Oriented Sentiment Analysis of Product Review Data. In *Satapathy S., Bhateja V., Raju K., Janakiramaiah B. (eds) Data Engineering and Intelligent Computing. Advances in Intelligent Systems and Computing*, 542. Springer, Singapore.

[19] Elmurngi, E. and Gherbi, A. (2017). Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. In *the Sixth International Conference on Data Analytics*, 65-72.