Fake reviews detection in e-commerce using machine learning techniques: a comparative survey

Maysam Jalal Abd[1]* *and Mohsin* Hasan Hussein[2]

1,2 Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq;

Abstract. In the field of online commerce, customer reviews have great importance because they significantly influence the profits of a business. Most consumers on the internet rely on reviews to help them make decisions about what to buy because they provide a reliable way to read other people's opinions about a specific product. Since a company's reputation and profitability are directly impacted by the reliability of its online reviews, some business owners pay spammers to create fake reviews. The creation of fake reviews that influence consumers' purchase decisions is a persistent and detrimental problem. Therefore, developing techniques to help companies and customers to distinguish between genuine and fraudulent reviews are still an important but challenging task. As a result, this paper provides a survey on various machine learning techniques are proposed to deal with the problem of detecting fake reviews, as well as the performance of different techniques in spam review classification, and determine the features, strengths, and weaknesses of those methods that may require more development.

1 Introduction

In recent years, The World Wide Web has dramatically changed the way people communicate and share ideas over the last several years. These days, opinions can be expressed online through posts, tweets, reviews, comments, and other forms of interaction on various websites, such as news sites, social networking sites, and e-commerce websites. One of the methods to share opinion is write review about product or a service [1]. Reviews are statements that represent a person's suggestion, perspective, or experience with any product on the market. Users post reviews about product on e-commerce websites to share their own experiences and offer suggestions to producers, sellers, and product providers as well as to new customers. By analyzing suggestions, the user experience that is provided can assist any business in growing and improving **[2].**Fake reviews are considered a type of opinion spamming, reviewers attempt to mislead readers instead of expressing their real opinions or experiences, these reviewers either post positive reviews about products, services, or companies to help them along, or they post negative reviews about other companies to damage their reputations [3]. Therefore, detecting fake reviews has become a crucial issue for both vendors and customers to make better decisions about which products to buy [4]. There are two ways to generate fake reviews. **First:** Human-generated method involves content creators writing reviews on products that appear genuine but are not; as the writer continues writing about products but never saw them. **Second**: Fake reviews are produced by computer-generated techniques that use text generation algorithms. [5].

2 Fake review detection

Fake reviews can be classified into three types:

Untruthful opinions: describes users who post negative reviews to damage the reputation of a product or company or post positive reviews to promote a product or business.

Reviews of a brand: only provide information about the reviewers' opinions about the products' brands**.**

*Corresponding author: maysam.j@s.uokerbala.edu.iq

Reviews that aren't relevant: don't express a genuine opinion or are just advertisements [6]. It has been observed that Type 2 and 3 reviews can be easily identified by content analysis, while type 1 reviews can be challenging to identify manually because a spam review can be successfully written to mimic any genuine or real review [7]. Three basic approaches have been suggested by Jindal and Liu [8][9] for identifying fake reviews:

Review Centric Approach: The method identifies fake reviews by analyzing the content of reviews written by reviewers such as brand name, product name similarity, use of capital letters, review content similarity, and repeated use of positive and negative words.

Reviewer Centric Approach: This approach relies on reviewers' actions.it include user data and every review that they have written [1]. This method makes use of the following features: account aging, profile photo, IP address, URL length, single reviewer number, and daily maximum assessment.[10][11].

Product Centric Approach- This method focuses primarily on data related to the product such as sales rank of product, price of product etc. Also To identify fake reviews, some criteria and attributes can be defined in the following:

Maximum number of reviews: studies show 75% of people post more than five fake reviews on specific days, however, 90% of normal individuals never write more than one review in a day.

Positive review percent: If a spammer generates a large percentage of positive product reviews, it could be an indication of fake reviews**.**

Average review length: spammers often lack detailed reviews, aiming to create fake ones, while 90% of honest reviewers write longer, more than 200-word reviews [12].

Top ranked reviews ratio: posting reviews early indicates are fake reviews. A reviewer's behavior may be considered suspicious if most of their reviews are ranked highly [6].

Reviews and ratings that are reciprocated: Those that appear on the same item or service.

Inappropriate Username: genuine users or buyers must display their true names, not just numbers, as numbers alone indicate spam. Genuine buyers must have an address with alphanumeric text.

Rating Correlated to Review: Reviews can be considered fake if the user provides a high rating but the reviewer's comments don't match the rating.

No Category Details: A review must provide specific brand or product information to be considered true, as it may be considered a fake review. [13].

The majority of spammers write imaginative reviews with pronouns, adverbs, and verbs, while normal users write informative reviews with more adjectives or nouns [4].

Review gap: false reviews are typically written at specific, regular intervals to significantly increase or decrease a product's rating, whereas non-fraudulent reviewers post the reviews randomly.

Time difference: fake reviewers typically post their fake reviews at a specific time. [14]**.**

3Techniques of fake reviews detection

Detecting fake reviews usually uses machine learning techniques, it required labeled data for supervised learning, an unlabeled data set is necessary for unsupervised learning, and for semi-supervised learning, a large amount of unlabeled data is mixed with a relatively small set of labeled data. [15].

3.1. Supervised learning techniques

Supervised learning used to determine fake reviews by looking at it as the classification issue of splitting reviews for two classes: fake and non-fake reviews [16]. Elmogy et al.[17] The proposed approach to identify the fake reviews from the websites through preprocessing, textural features extraction, and training several machine learning (ML) classifiers using the computed key points to complete the classification task, in this work the KNN classifier with the value of K = 7 give a better F1 score than other classifiers. Jitendra et al. [18] The authors utilized sentiment polarity and content similarity features to differentiate between real and fake reviews. After that, they used three algorithms: decision trees, naive Bayes, and support vector machines. In practical these features together outperforms the effectiveness of each alone. And the limitation is the number of the features is restricted. Banerjee et al.[19] This work uses several supervised learning algorithms to distinguish between genuine and fake reviews based on four linguistic cues: ease of understanding, level of detail, writing style, and perception indicators. The limitation of this paper it's only used for the labeled data. Wael et al.[20] Supervised learning algorithms are utilized to detect the fake reviews through preprocessing steps and linguistic features like bag-of-words and POS, next, The classification algorithms used include Naive Bayes, gradient boosted trees, decision trees, random forests, and support vector machines. As a result, NB and SVM perform better in this case. Silpa et al. [21] used supervised learning model based on the textual information in the reviews, and sentiment classification to classify reviews as fake or genuine. They used several classifiers such as Naive Bayes, SVM, logistic regression, and decision tree; it found the SVM has the best accuracy as compared to the other classifiers.

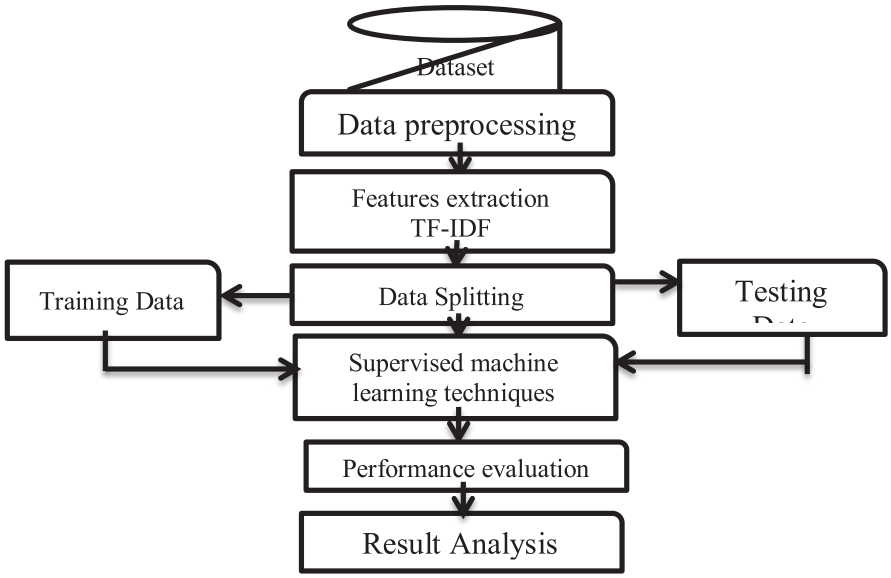

Fig1: The fake review detection model illustrates by Alsubari et al.[22].

Setievi, et al. [23] in this study compared three supervised machine learning techniques: SVM, and Random Forest , Logistic Regression. Without tuning, we start the experiment by preprocessing the data and utilizing the Term Frequency-Inverse Document Frequency (TF-IDF) feature. The experiments showed that SVM outperformed LR and RF in accuracy, with an accuracy of 88.89% before tuning and 89.77% after tuning using 5-fold cross-validation. Barbado et al.[24] Introduces a framework features for fake reviews detecting using a Yelp product review dataset. The author used the supervised machine learning techniques on the dataset, incorporating review-centric features and reviewer features. As a result, The AdaBoost algorithm achieved an accuracy of 82% in their testing. Badresiya et al.[25] Used supervised techniques for review spam detection in the domain of opinion mining on a dataset of 1600 .They use the content of reviews to detect fake reviews. Then, it presents a performance evaluation of its techniques using the Rapidminer data mining tool. The result shows that the Support Vector Machine (SVM) outperforms all other supervised techniques in detecting review spam. But use only review's content to identify fake review, in the future, to improve the accuracy of methods use other features like review ratings, review time and number of useful feedbacks etc. Siagian et al.[26] In this work machines can distinguish fake and genuine feedback by incorporating labels into the Cloud Armor dataset, limiting review counts, service counts, and collusion feedback probability. The ensemble model outperformed KNN, Logistic Regression, and SVM_rbf in terms of F1-Score, recall, accuracy, and precision, achieving high values of 97.51%, 98.19%, 93.65%, and 95.86% respectively. Wang et al.[27] Suggest a new method that based on rolling collaborative training and the fusion of multiple features to detect deceptive reviews. The method employs a multi-feature initial index system, including text, sentiment, and behavior features from reviews. To represent text using the related algorithm (Doc2vec), afterward, trains an initial sample set with seven classifiers. Hassan et al.[28] Produced an effective supervised learning method like SVM, LR, and NB to detect deceptive online reviews. They used dataset of the gold standard hotel review with features like TF-IDF, Empath, and sentiment polarity.

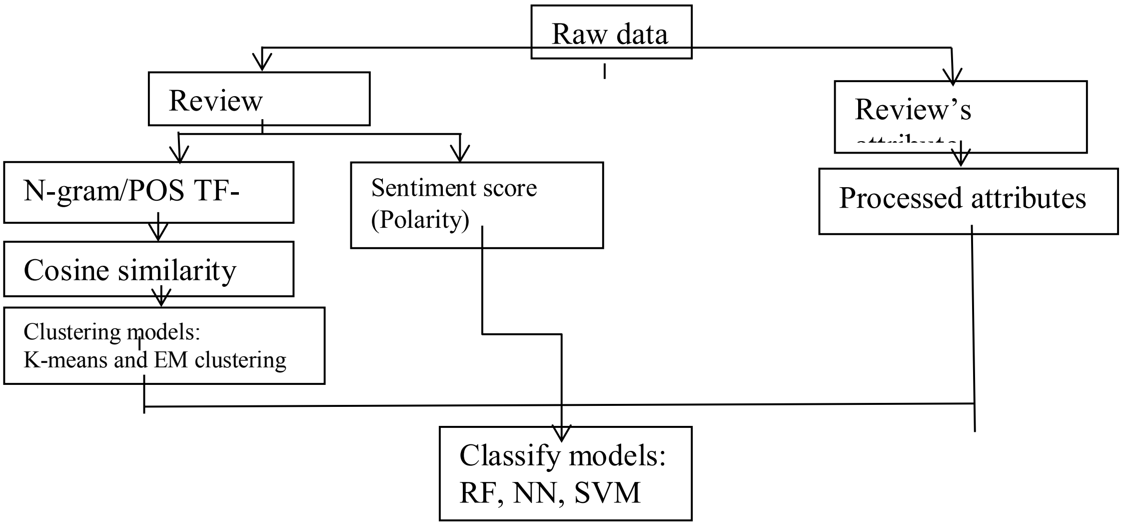

Fig 2. The fake review detection model illustrates by Le's. et al. [3].

Deep learning techniques have demonstrated their effectiveness in challenging natural language processing tasks, making them an exciting method for identifying fake reviews. Thus, Convolutional neural networks (CNNs) have been proven to perform better in natural language processing. Liu et al.[29] Suggest a network of hierarchical attention where multiple interests specifically used to capture significant, comprehensive, and multi-granularity semantic information at both layers. In particular, the initial layer of the use a CNN with an N-gram to extract the sentences' multi-granularity semantics. Next, at the second layer, the document's semantics are extracted comprehensively and significantly by combining convolution structure and Bi-LSTM (Bidirectional Long-short-term memory). Nasir et al.[30] Suggest a hybrid deep learning model that combines a recurrent neural network (RNN) and a convolutional neural network (CNN) to classify fake review, enhancing the capture of sequential text information. This analysis has employed two different datasets, FA-KES and ISOT. Also, the hybrid model was evaluated against several machine learning algorithms, including Decision Tree, CNN, KNN, RNN, and linear regression. Hajek et al.[31] It proposed a supervised learning technique that combines reviewer-centric and review features to identify fake reviews. They extracted sentiment-dependent textual characteristics using a CNN classifier, underlining the significance of combining linguistic and behavioral characteristics for best results.

Table 1: summary of using supervised machine learning techniques for fake reviews detection

Authors	Dataset	Features used	Metric	Classifier	Result
Elmogy et al.[17]	Yelp Dataset	features of reviews and behavioral features	F1_score	KNN	82.40%.
Jitendra et al. [18]	20 Chicago Hotel Review	Score of Sentiment, Linguistic and unigram features	Accuracy	Decision tree SVM Naive bayes	92.11% 88.71 91.90
Banerjee et al.[19]	Reviews of Agoda. com, Expedia.com and Hotels.com	Linguistic clues: • style of writing • details level • Word structure • Cognition Indicators	Precision	Random forest Logistic regression SVM Naive bayes	74.8% 72.8% 70 % 73.9%
Wael et al.[20]	Chicago Hotel Review	Linguistic clues N-gram	Recall	Decision tree Random forest Naïve bayes SVM	70.7% 60.3% 86.8% 86.8%
Silpa et al. [21]	Yelp dataset	review content user behaviour and	Accuracy	Naive Bayes SVM	96.04% 97.03%

		sentiment		Logistic regression Decision tree	96.04% 94.55%
Alsubari et al.[22]	Trip Advisor Website	n-grams of the text review and scores of the sentiment	Accuracy an and F1 Score	Naive Bayes SVM random forest adaptive boosting	88% 93% 95% 94%
Setievi, et al. [23]	Amazon Dataset	text features	Precision	SVM LR RF	89.78% 88.13% 87.99%
Le's.[3]	Yelp dataset	non-textual features of reviews and reviewers	Recall	ANN SVM Random Forest Decision tree	94.16% 91.84% 94.96% 88.32%
Barbado et al.[24]	Yelp dataset	review-centric and reviewer features	Accuracy	AdaBoost	82%
Badresiya et al.[25]	AMT used in ott et al.(2011) [32]	Analyzing content	Precision	Naïve Bayes SVM KNN Logistic regression Decision tree	67.50% 82.82% 66.45% 82.46% 50.52%
Siagian et al.[26]	Cloud Armor dataset	counts of reviews, count of services, and collusion feedback factor probability	Accuracy	KNN Logistic Regression SVM_rbf Ensemble Voting	97.04 91.86 93.66 97.51
Wang et al.[27]	yelp shopping website	text features, sentiment features of reviews and behavior features of reviewers	Accuracy	SVM ,RF SVM,DT RF,SVM RF,DT	84.07% 81.38% 78.89% 83.11%
Hassan et al.[28]	gold standard hotel review dataset	TF-IDF, Empath categories, and sentiment polarity	Recall	Naïve Bayes Logistic Regression SVM	73.78% 89.63% 87.19%
Liu et al.[29]	gold standard dataset consist of data from three various fields (i.e., hotel, restaurant, and doctor) released by [33]	multi-granularity semantics of the review	F1- score (in the Mix-domain)	CNN POS LSTM DSRHA Bi-GRNN	77.6% 65.8% 72.0% 89.3% 83.4%
Nasir et al.[30]	ISOT dataset, FA-KES Dataset	Sequential information in the text	Accuracy In (FA-KES dataset)	KNN CNN RNN Hybrid CNN-RNN	57% 50% 50% 60.%
			Accuracy In (ISOT dataset)	KNN CNN RNN Hybrid CNN-RNN	60% 99% 98% 99%
Hajek et al.[31]	Yelp restaurant Reviews	linguistic and behavioral features	F-score	The proposed model	81.8%

3.2. Unsupervised learning techniques:

unsupervised learning approach has the advantage of the ability to distinguish between fake and genuine reviews without the need for a labeled dataset [4]. Unsupervised methods, like Principal Component Analysis, Independent Component Analysis, and clustering techniques, use unlabeled data to learn [34]. Dong et al.[35] Developed a model called (UTSJ) unsupervised topic-sentiment joint probabilistic using the (LDA) Latent Dirichlet Allocation model, which includes four levels: document, topic, word, and sentiment. The model uses support vector machine and random forest classifiers to extract topic information from documents and sentiment from reviews. The model performed better than other baseline models on a real-world dataset from Yelp.com. Sedighi et al.[34] Using unsupervised learning representation combined with the feature selection methods to extract the appropriate features and evaluate them with using a decision tree. The RLOSD model effectively separates spam reviews from the entire corpus, as confirmed by precision, recall, and F-measure comparisons with SVM, naive Bayes, and Log Regression. Mukherjee and Venkataraman [36] The paper introduces an unsupervised method for detecting opinion spam, using a novel generative model called LSM Latent Spam Model, which exploit both spammers' linguistic and behavioral footprints, as a result the proposed model outperforms other algorithms in real-world datasets.

Kumar et al.[37] This paper presents a fully unsupervised mixture method for detecting manipulation behaviors in online reviewers, combining multivariate and univariate distributions, using four baseline techniques with equal weights. As a result, the proposed approach outperforms baseline algorithms and state-of-the-art unsupervised methods in analyzing synthetic data and real-world restaurant reviews from Yelp.com. It doesn't need the data to be explicitly labeled as spam or not. To future work, incorporate advanced linguistics with this model like sentiment analysis and models for identifying deceptive writing. Singh et al.[38] Used unsupervised machine learning: Kth Nearest Neighbor (KNN) technique to identify review outliers by analyzing reviewer, review, and product-centric features.

Li et al.[39] Proposed a novel approach (GSDNT) to identify fake review groups using aspect-oriented sentiment mining based on nominated topics. The model uses K-means algorithm, time burstiness, and content duplication, the proposed model's effectiveness was demonstrated through experimental results on a dataset from JD.com. The limitation of the GSDNT is that it relies on formally representing previously defined topics. Koggalahewa et al.[40] This paper presents unsupervised spammer detection method using user peer acceptance, which is classified as spam or genuine based on user interest distribution and peer acceptability. This method does not need labelleds dataset. It's no need to another training model when data are change. Zhong et al.[41] Unsupervised learning detects spammers using reviewer reputation scores and k-center clustering. Experiment analysis on Amazon music product reviews shows the proposed approach outperforms existing methods, combining content features and reviewer behaviors.

Y Pan and L Xu [42] Designing a new method for unsupervised fake review detection (FRD) consists of four major steps: survey research, the analysis of feature, estimation of fake index, and review selection, with a performance metric proposed for Yelp data and Dianping case study. Abutiheen et al.[43] Employed the new model with the adopted classifier "Identity classifier" to analyze the behavior of Facebook posts and comments in the Iraqi-Arabic dialect and divide them into two categories: wicked and non-wicked. The proposed classifier outperformed other classifiers with an accuracy of 85.4%.

Table 2: summary of using unsupervised machine learning techniques for fake reviews detection

Authors	Dataset	Features used	Metric	Classifier	Result
Dong et al.[35]	Yelp.com	word, sentiment , topic, and document of review	Precision	UTSJ	Restaurant: 82.29 Hotel: 87.15
Sedighi et al.[34]	Yelp.com, Dataset in fields: (Hotel, Restaurant and Doctor).	Feature selection Method	F-measure on yelp dataset	SVM Naïve Bayes LR RLOSD	72.56% 65.06% 64.64% 76.91
			F-measure On (Hotel, Restaurant and Doctor) Dataset	SVM Naïve Bayes LR RLOSD	66.9%,80.1%,73.4 % 65.9%,77.2%, 70.3% 66.4%,76.5%,71.7% 78.3%,81.8%,75.0%
Mukherjee and Venkataraman [36]	Amazon , Yelp datasets Amazon Mechanical	linguistic and behaviors	Recall On AMT Dataset	LSM-HE LSM-UP	86.1% 89.0%

	Turk (AMT)				
Kumar et al. [37]	Yelp.com	Review number Gap Review Entropy rating Rating deviation Text Length Rating Scores Review time	AUC-Score	GMM DPGMM OSVM-L STK	0.52 0.53 0.49 0.70
Singh et al.[38]	Cell Phones and Electronics products	Review Centric Features Reviewer Centric Features		KNN	Percentage of Spam= 7.68 %
Li et al. [39]	dataset collected from JD.com and TMALL.com	Sentiments (explicitly and inexplicitly)	Accuracy	GSDNT	96.42%
Koggalahewa et al.[40]	Social Honey Pot, The Fake Project and HSpam14	Content Features	Recall on HSpam14 dataset	SDPACM K-Means	93% 49%
Zhong et al.[41]	Amazon Reviews	review content based and behavior-based Characteristics	Precision F-score Recall	Proposed method	0.85, 0.73, 0.79 respectively
Y Pan and L Xu [42]	Yelp dataset, Dianping	fake-review features		FRD-R KM (*K*-means) clustering) DC(density clustering)	18.40% 10.87% 12.35%
ZA Abutiheen et al. [43]	Facebook	behavior features	Accuracy	Identity classifier	85.4%.

3.3. Semi -Supervised learning techniques:

Semi-supervised learning models include a lot of labeled data as well as a large number of unlabeled data. Hai et al.[44] The study introduced the multi-task technique called SMTL-LLR using Laplacian Logistic Regression (LLR) to identify fraudulent reviews on unlabeled data. The learning of only one task was enhanced through the model by incorporating information obtained in the three fields: (hotel, restaurant, and doctor). The results of the experiment demonstrated SMTL-LLR outperformed state-of-the-art methods. Rout et al.[45] This study demonstrates that four semi-supervised learning approaches, including of Co-training, expectation-maximization, label propagation, and PU learning, can enhance classification by incorporating new features like parts-of-speech, linguistics, count of words, and the content of sentimental. Tian et al.[46] Proposed a reliable semi-supervised approach named "Ramp One-Class SVM" to deal with a lack of labeled datasets to detect fake reviews. The experimental results that based on the Yelp dataset and AMT dataset demonstrated the proposed model obtained good results on AMT and Yelp datasets with an accuracy of 92.3%, and 74.37% respectively. Ligthart et al.[47] This study evaluated the efficacy of four semi-supervised methods for detecting fake reviews: self-training, co-training, TSVM, label propagation, and label spreading. Results showed that self-training with the naïve Bayes classifier achieved the best performance on AMT and Yelp datasets.

https://doi.org/10.1051/bioconf/20249700099

Table 3: summary of using semi-supervised machine learning techniques for fake reviews detection

Authors	Dataset	Features used	Metric	Classifier	Result
Hai et al.[44]	Doctor, Hotel, Restaurant Dataset	Text of unigram and bigram, term-frequency features	Accuracy	SMTL-LLR SVM LR	87.2% 81.8% 82.1%
Rout et al.[45]	AMT, Yelp, Trip Advisor, Expedia, and Hotels.com	Bigram frequency counts, POS features, linguistic, number of words, content and sentimental features	Precision	Co-training label propagation EM PU learning	81% 84% 85% 83%
Tian et al.[46]	AMT and Yelp Datasets	Text feature	Accuracy	Ramp-OCSVM	92.13% on AMT dataset 74.37% on yelp dataset
Ligthart et al.[47]	AMT and Yelp datasets	Unigram, Bigram,TF-IDF, and POS	Accuracy On Bigram	Self-Training + NB	93%

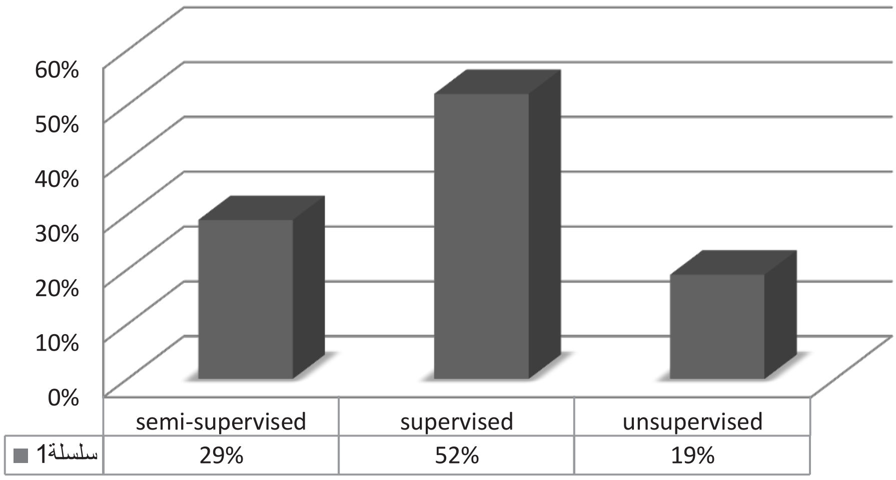

Fig 3: Percentage of techniques and approaches observed in the selected studies.

4 Data collection

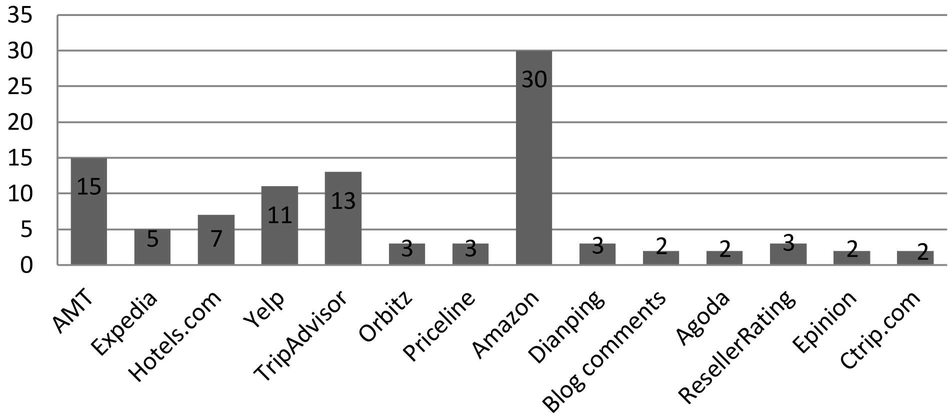

Figure 4: Distribution of the dataset used in various studies.

5 Research direction

Although the majority of existing work has made significant developments. In the field of detecting fake reviews using machine learning methods, there are still numerous unresolved issues and directions for future studies. One major challenge in that field of fake review identification is the lack of large, uniform, and various datasets that accurately represent the complexity of fake review patterns through domains and websites. Furthermore, the dataset used has a significant impact on experimental results as well as the reported effectiveness of the detection model, the dataset's unique attributes, including genuine and fake reviews, writing styles, and domain-specific language, significantly impact the effectiveness of machine learning algorithms. Another major problem is that consumer reviews are generated rapidly and diversely. Therefore, reviews are treated as big data, which requires the use of big data analysis methods to analyze and detect fraudulent reviews in real time. Therefore, future research should focus on developing algorithms that can detect fake reviews in real-time, enabling platforms to take prompt action against fraudulent content.

6 Conclusion

This paper provided a comprehensive overview of the most research that uses machine learning techniques to detect fake reviews. First, we investigated detecting fake reviews in several papers. Then we discuss the various machine learning methods used for fake review detection, including supervised, unsupervised, and semi-supervised learning, with detailed summaries provided. We have also compared the different techniques to identify the fake reviews. The result showed that Hybrid CNN-RNN, Ensemble Voting and SVM achieved the highest accuracy of 99%, 97.51% and 97.03% respectively. Thus this study aimed to provide future insights to researchers by better understanding the existing research on methodologies and machine learning techniques used so far. However, machine learning techniques have great promise for the future developments in the field of fake reviews detection, this developments help to ensure customer satisfaction, increase business revenue, and boost purchases through trustworthy and reliable reviews.

References

1. Rastogi, A., & Mehrotra, M. (2017). Opinion spam detection in online reviews. Journal of Information & Knowledge Management, 16(04), 1750036.
2. Shilpa yadav1 and K. M. Dr.Gulbakshee Dharmela2, “Fake Review Detection Using Machine Learning Techniques,” J. Emerg. Technol. Innov. Res., vol. 8, no. 4, 2021.
3. Le, H., & Kim, B. (2020). Detection of fake reviews on social media using machine learning algorithms. Issues In Information Systems, 21(1), 185-194.
4. Patel, N. A., & Patel, R. (2018, December). A survey on fake review detection using machine learning techniques. In 2018 4th international Conference on computing Communication and automation (ICCCA) (pp. 1-6). IEEE.
5. Salminen, J., Kandpal, C., Kamel, A. M., Jung, S. G., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services, 64, 102771.
6. Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., & Maqsood, S. (2021). Fake reviews detection: A survey. IEEE Access, 9, 65771-65802.
7. S. Farooq and H. A. Khanday, “Opinion Spam Detection: A Review,” Int. J. Eng. Res., vol. 12, no. 4, pp. 1–08, 2016, [Online]. Available: www.ijerd.com
8. N. Jindal and B. Liu, “Opinion spam and analysis,” WSDM’08 - Proc. 2008 Int. Conf. Web Search Data Min., pp. 219–229, 2008, doi: 10.1145/1341531.1341560.
9. N. Jindal and B. Liu, “Review spam detection,” 16th Int. World Wide Web Conf. WWW2007, pp. 1189–1190, 2007, doi: 10.1145/1242572.1242759.
10. Liu, P., Xu, Z., Ai, J., & Wang, F. (2017, July). Identifying indicators of fake reviews based on spammer's behavior features. In 2017 IEEE international conference on software quality, reliability and security companion (QRS-C) (pp. 396-403). IEEE.
11. Fontanarava, J., Pasi, G., & Viviani, M. (2017, October). Feature analysis for fake review detection through supervised classification. In 2017 IEEE international conference on data science and advanced Analytics (DSAA) (pp. 658-666). IEEE.
12. S. Rayana and L. Akoglu, “Collective opinion spam detection: Bridging review networks and metadata,” Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., vol. 2015-Augus, pp. 985–994, 2015, doi: 10.1145/2783258.2783370.
13. R. Agarwal and D. K. Sharma, “Detecting Fake Reviews using Machine learning techniques: a survey,” 2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022, pp. 1750–1756, 2022, doi: 10.1109/ICACITE53722.2022.9823633.
14. J. Luo, J. Luo, G. Nan, and D. Li, “Fake review detection system for online E-commerce platforms: A supervised general mixed probability approach,” Decis. Support Syst., no. June, p. 114045, 2023, doi: 10.1016/j.dss.2023.114045.
15. D. U. Vidanagama, T. P. Silva, and A. S. Karunananda, “Deceptive consumer review detection: a survey,” Artif. Intell. Rev., vol. 53, no. 2, pp. 1323–1352, 2020, doi: 10.1007/s10462-019-09697-5.
16. M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, “Survey of review spam detection using machine learning techniques,” J. Big Data, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0029-9.
17. A. M. Elmogy, U. Tariq, A. Ibrahim, and A. Mohammed, “Fake Reviews Detection using Supervised Machine Learning,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 1, pp. 601–606, 2021, doi: 10.14569/IJACSA.2021.0120169.
18. J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, “Deceptive review detection using labeled and unlabeled data,” Multimed. Tools Appl., vol. 76, no. 3, pp. 3187–3211, 2016, doi: 10.1007/s11042-016-3819-y.
19. S. Banerjee, A. Y. K. Chua, and J. J. Kim, “Using supervised learning to classify authentic and fake online reviews,” ACM IMCOM 2015 - Proc., 2015, doi: 10.1145/2701126.2701130.
20. Etaiwi, W., & Naymat, G. (2017). The impact of applying different preprocessing steps on review spam detection. Procedia computer science, 113, 273-279.

21. Silpa, C., Prasanth, P., Sowmya, S., Bhumika, Y., Pavan, C. S., & Naveed, M. (2023, March). Detection of Fake Online Reviews by using Machine Learning. In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 71-77). IEEE.
22. S. N. Alsubari et al., "Data analytics for the identification of fake reviews using supervised learning," Comput. Mater. Contin., vol. 70, no. 2, pp. 3189–3204, 2022, doi: 10.32604/cmc.2022.019625.
23. F. Setievi, J. Natalia, T. R. Tjhang, I. S. Edbert, and D. Suhartono, "A Comparative Study of Supervised Machine Learning Algorithms for Fake Review Detection," 2022 5th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2022, pp. 306–312, 2022, doi: 10.1109/ISRITI56927.2022.10052860.
24. Barbado, R., Araque, O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. Information Processing & Management, 56(4), 1234-1244.
25. Badresiya, M. A., & Teraiya, J. (2014). Performance analysis of supervised techniques for review spam detection. Int J Adv Netw Appl Special, (21-24).
26. H. Taneja and S. Kaur, "An ensemble classification model for fake feedback detection using proposed labeled CloudArmor dataset," Comput. Electr. Eng., vol. 93, no. July, p. 107217, 2021, doi: 10.1016/j.compeleceng.2021.107217.
27. J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi, and X. Xiao, "Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3028588.
28. Hassan, R., & Islam, M. R. (2020, December). A supervised machine learning approach to detect fake online reviews. In 2020 23rd international conference on computer and information technology (ICCIT) (pp. 1-6). IEEE.
29. Liu, Y., Wang, L., Shi, T., & Li, J. (2022). Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM. Information Systems, 103, 101865.
30. Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 1(1), 100007.
31. Sahut, J. M., & Hajek, P. (2022). Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection.
32. Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. arXiv preprint arXiv:1107.4557.
33. Li, J., Ott, M., Cardie, C., & Hovy, E. (2014, June). Towards a general rule for identifying deceptive opinion spam. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1566-1576).
34. Sedighi, Z., Ebrahimpour-Komleh, H., & Bagheri, A. (2017, December). RLOSD: Representation learning based opinion spam detection. In 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS) (pp. 74-80). IEEE.
35. Dong, L. Y., Ji, S. J., Zhang, C. J., Zhang, Q., Chiu, D. W., Qiu, L. Q., & Li, D. (2018). An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews. Expert Systems with Applications, 114, 210-223.
36. Mukherjee, A., & Venkataraman, V. (2014). Opinion spam detection: An unsupervised approach using generative models. Techincal Report, UH.
37. Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2019). Detecting anomalous online reviewers: An unsupervised approach using mixture models. Journal of Management Information Systems, 36(4), 1313-1346.
38. Singh, S. (2015). Improved techniques for online review spam detection (Doctoral dissertation).
39. Li, J., Lv, P., Xiao, W., Yang, L., & Zhang, P. (2021). Exploring groups of opinion spam using sentiment analysis guided by nominated topics. Expert Systems with Applications, 171, 114585.
40. Koggalahewa, D., Xu, Y., & Foo, E. (2022). An unsupervised method for social network spammer detection based on user information interests. Journal of Big Data, 9(1), 1-35.
41. Zhong, M., Tan, L., & Qu, X. (2019). Identification of opinion spammers using reviewer reputation and clustering analysis. International Journal of Computers Communications & Control, 14(6), 759-772.

https://doi.org/10.1051/bioconf/20249700099

42. Pan, Y., & Xu, L. (2024). Detecting Fake Online Reviews: An Unsupervised Detection Method With a Novel Performance Evaluation. International Journal of Electronic Commerce, 1-24.
43. Abutiheen, Z. A., Mohammed, E. A., & Hussein, M. H. (2022). Behavior analysis in Arabic social media. International Journal of Speech Technology, 25(3), 659-666.
44. Hai, Z., Zhao, P., Cheng, P., Yang, P., Li, X. L., & Li, G. (2016, November). Deceptive review spam detection via exploiting task relatedness and unlabeled data. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 1817-1826).
45. Rout, J. K., Dalmia, A., Choo, K. K. R., Bakshi, S., & Jena, S. K. (2017). Revisiting semi-supervised learning for online deceptive review detection. IEEE access, 5, 1319-1327.
46. Tian, Y., Mirzabagheri, M., Tirandazi, P., & Bamakan, S. M. H. (2020). A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM. Information Processing & Management, 57(6), 102381.
47. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification. Applied Soft Computing, 101, 107023.