

Fake Review Detection and Removal: A Comparative Analysis using ML and DL Models

Garima¹, Vaibhavi², Yamini Singh³, Rupika Teotia⁴, Karuna Kadian⁵, Sunita Garhwal⁶, and Vimal Dwivedi⁷

¹⁻⁵CSE Department, Indira Gandhi Delhi Technical University for Women, New Delhi, India

Email: {garimakanwaria8, jhava2201, yaminisingh.yrg, rupikateotia}@gmail.com, karunakadian@igdtuw.ac.in

⁶CSE Department, Thapar Institute of Engineering & Technology, Patiala

Email: sgarhwal@thapar.edu

⁷School of Electronics, Electricals and Computer science, Queen's University Belfast

Email: vimal.bncet@gmail.com

Abstract— Reviews are increasingly used for purchase decisions by the customers, they are important for e-commerce and social networking sites. However, not every review is necessarily authentic. Researchers have put out a variety of machine learning techniques in the past to identify false product reviews. Finding the proper machine learning algorithm to spot fake reviews for a particular type of data is crucial, though. Consequently, in this research, algorithms such as SVC (Support Vector Classifier), Decision Tree Classifier, Logistic Regression, Random Forests Classifier, Multinomial Naive Bayes and k-Nearest Neighbors are compared on different kinds of datasets like Amazon dataset, Yelp Dataset and TripAdvisor Hotel Reviews Dataset. Accuracy, Recall, Precision, and F1-Measure evaluation findings are the basis for the comparison. The findings of this study indicate that, Support Vector classifier comes out to be the best performing algorithm for detecting fake reviews when compared with the other five techniques. While the k-Nearest Neighbors algorithm has the worst performance.

Index Terms— Algorithms; Reviews; Machine Learning; Fake; Comparative Analysis; Dataset - Amazon, Yelp, TripAdvisor; Spam;

I. INTRODUCTION

Reviews are incredibly important when making decisions in today's electronic commerce. The majority of consumers check reviews of products or businesses before choosing what to purchase and where to purchase it, and whether or not to buy it. The existence of these comments can be a source of information. Businesses might use it, for instance, to choose how to produce their products or services. But regrettably, some people take advantage of the significance of reviews by fabricating reviews with the intention of either boosting the popularity of the product or undermining it.

It's customary for people to research products before making a purchase. Customers may evaluate several brands and decide on a certain product based on reviews. These internet reviews have the power to alter a consumer's perception of a product. If these reviews are accurate, it may be easier for people to choose a product that meets their needs. On the other side, if the reviews are falsified or manipulated, the user may be misled.

The suggested system provided an e-commerce website with product reviews that aid users in making decisions about what to buy. Every user has a distinctive profile on social networking and e-commerce websites in order to

submit a distinctive order. Users may search for things there and read customer reviews on them. These reviews will assist customers in determining the product's quality and assisting them in making the best choice. However, the reality is that this product review for each item cannot assist clients pick the best one.

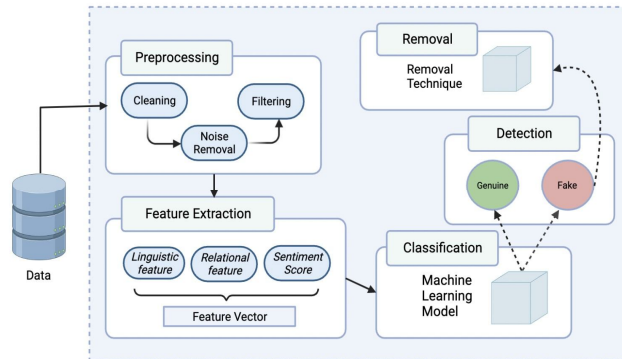


Fig. 1. Process Diagram

Customer review websites have encountered problems as a result of the way users freely express and use their comments. Platforms for social media like Twitter, Facebook, YouTube, and others let anybody express criticism of any business at any time with no restrictions or commitments. Because there are no constraints, some companies utilize social media to unjustly advertise their products, stores, or brands while disparaging those of their rivals [11].

To address this issue, many online marketplaces and retailers have implemented various techniques to detect and remove fake reviews. These techniques include the use of machine learning algorithms, natural language processing, and manual moderation by human experts.

In this paper, we focus on the problem of detecting and removing fake product reviews using a combination of machine learning algorithms and natural language processing techniques. We propose an approach that analyzes various features of product reviews, such as sentiment, grammar, and vocabulary, to identify fake reviews. Our approach can help online marketplaces and retailers improve the quality of their reviews, maintain consumer trust, and ultimately, enhance the overall user experience.

The research on review spam identification [24], [29] is extremely valuable since it can ensure the validity of reviews, decrease the cost of cleaning up fraudulent reviews on ecommerce sites and social networking sites, and give consumers a better purchasing experience. Good product evaluation will increase consumer willingness to acquire items while also protecting merchants' and consumers' rights and interests. Manufacturers can also utilize false review detection technology to gather authentic review data, assess customers' true feelings, keep updating products, and raise the standard of service by adhering to the procedure described in figure 1.

II. LITERATURE REVIEW

The most critical stage in any type of study is a review of the literature. Before constructing a model or presenting a solution, it is necessary to examine past publications in the domain. Based on the study, the author can forecast or produce the disadvantage and begin working with the mention of earlier papers.

Researchers have been utilizing supervised learning for detecting fake reviews for years, however the truth of the underlying big data sets is still inaccessible. Instead of actual false reviews, the majority of current supervised learning methods rely on pseudo-fake reviews. This give the first study on false review detection in Chinese utilizing filtered reviews from Dianping's fake review detection system in [16].

The difficulties in detecting bogus reviews are explored in [27]. The method put out in this paper functions with both labelled and unlabeled data. The study shows that supervised machine learning is 90.19% accurate, and semi-supervised machine learning is 83.70% accurate. This paper investigated and discussed all possible aspects and combines the experiment and results using the provided dataset.

In the suggested research [6], a dataset of reviews in Urdu and Roman is created, and the n-gram approach [1] is used to identify fake reviews in different languages. The research revealed that the most efficient means of identifying fake reviews is text categorization with an SVM classifier.

Several online data mining strategies are discussed in [20]. These techniques are further divided into three categories: internet structure, web content, and web usage mining. For instance, Naive Bayes, Neural Network,

and Support Vector Machine are all used in web content mining. These various kinds each employ a different methodology, tool, strategy, and algorithm to extract information from the vast amounts of data on the internet. A supervised method is applied to train on Yelp’s filtered reviews in [21]. All currently used supervised learning methods rely on pseudo-fake reviews in place of falsified reviews screened by a reputable website. Using actual Yelp data, this study assessed and evaluated performance using established research approaches. Unexpectedly, behavioral factors perform better than language features [31].

The “Amazon’s Yelp” dataset is utilized in [3], and two models, Random Forest and Naive Bayes, are created to compare model performance. As a result, the Random Forests model outperformed the Naive Bayes method by a wide margin. The problem of fake review detection is treated honestly, with the goal of selecting the best suitable algorithm to complete the work of detection and eradication. The research study examined linguistic elements in [30] such as unigram frequency, bigram frequency, unigram presence, and review length to identify fraudulent reviews on the Yelp dataset. The key issue is data scarcity, and the model requires both language and behavioral elements to function

[14]. Online reviews’ influence on businesses has significantly increased in 2018, and they are now critical in assessing business performance in a wide range of industries, from hotels and restaurants to e-commerce. Unfortunately, some consumers create bogus reviews of their businesses or competitors’ businesses in order to improve their internet reputation. Previous research has concentrated on detecting false reviews in a range of industries, such as restaurant and hotel product or company reviews. Consumer electronics companies have not yet been thoroughly studied, despite their economic importance [4].

The majority of research on fake review identification is done using supervised, binary text classification [9]. Since it involves extracting (metadata) features from review text, representing them in a machine-processable format (feature representation), and teaching a prototype that can generalize patterns based on the features and apply the patterns to previously unexplored data, it is comparable to product classification (algorithm). [28] classified this metadata as “lexical” or “non-lexical.” Words, n-grams, punctuation, and latent themes are instances of lexical features, which are textual properties. Non-lexical elements include metadata on reviews (such as ratings or stars) or associated authors.

A detailed examination of opinion-rich product reviews [12], which are frequently used by manufacturers and consumers alike, and considers how opinion spam differs from web spam and email spam, requiring various detection techniques to manage this. It analyses such spam actions in millions of Amazons’ reviews and generates models utilizing Logistic regression, feature identification such as review centric, product centric, and reviewer centric, type 1 - type 2 spam detection, and so on.

The survey assesses and summarizes current feature extraction techniques [19] in order to identify gaps according to two categories: learning algorithms and traditional statistics [7], [8]. Additionally, they conduct research on the effectiveness of other converters and neural network models that have not yet been used to identify fraudulent reviews. RoBERTa beats state-of-the-art approaches in a blended domain, with an accuracy of 91.2%, according to the experimental results on two baseline methods, and can be used as a baseline for further study.

TABLE I. DESCRIPTION OF PUBLIC DATASET SOURCES

Dataset	Dataset (Source)	Description
Amazon datasets	Jindal & Liu (2008) [12] dataset link Blitzer et al. (2007) [5] dataset link Ni et al. (2019) [22], McAuley et al. (2015) [18], He & McAuley (2016) [10] dataset link	58,38,041 reviews, 12,30,915 products 21,46,057 reviewers. Product reviews for 4 categories, including kitchen, books, dvds, and electronics. 142.8million reviews about different products.
TripAdvisor datasets	Ott et al. (2011) [25] dataset link Ott et al. (2013) [24] dataset link	800 positive ratings, 400 of which are genuine and 400 are fraudulent. 800 bad ratings, 400 of which are real and 400 of which are fraudulent
TripAdvisor and Yelp datasets	Li & Ott et al. (2014) [17] dataset link	authentic reviews: 1200 fake reviews: 1636 three domains: restaurant, hotel etc
Yelp datasets	Mukherjee et al. (2013) [21] (YelpCHI dataset) dataset link Rayana and Akoglu (2015) [26] (YelpNYC dataset) dataset link Rayana and Akoglu (2015) [26](YelpZIP dataset) dataset link Barbado et al. (2019) [4] (Write to o.arague@upm.es if you wish to acquire the dataset for analysis.)	67,395 reviews from 200 hotels and restaurants by 38,063 reviewers 3,59,052 reviews from 932 restaurants by 1,60,225 reviewers 6,08,598 reviews from 5,044 restaurants by 2,60,277 reviewers. 9456 authentic reviews and 9456 fake reviews from four US cities

The study proposes a method for identifying false reviews that takes into account the actions of reviewers as well as characteristics obtained from the reviews' content [13]. In other words, their goal is to create a white-box model that will enable people to comprehend what is happening with respect to their personal data, as opposed to Blackbox models, which are difficult to generally describe, like deep learning.

III. DATA AND METHODS

A. Datasets

The datasets for this study were obtained from Kaggle. The first dataset includes the category of the product, rating of that particular product, label (CG or OR) and review text. The dataset has 40432 occurrences and 4 ascribed characteristics. As there are 20216 fake reviews and the same number of true reviews, the dataset is balanced. For the second dataset this study considers Yelp product reviews consisting of fields like Review, Product-id, Rating, User-id, Date and Label. The third dataset include TripAdvisor hotel reviews which consists of fields like Deceptive, Hotel, Polarity, Source and Review text. To assist future researchers, relevant information has been identified and summarized for convenience and summarized in Table I.

B. Classification Algorithms

To find the false reviews, a variety of supervised learning classification algorithms are used, such as Random Forest Classifier, Decision Tree Classifier, k-Nearest Neighbors, Logistic Regression, Multinomial Naive Bayes and Support Vector Classifier.

- **Logistic Regression:** Machine learning categorization issues are solved using logistic regression. They are utilized to predict categorical variables and are comparable to linear regression. It can predict the outcome as 0 or 1, or Yes or No. Instead of the exact integers, it provides probabilistic values between 0 and 1.
- **k-Nearest Neighbors:** This approach can also solve problems related to classification and regression. This is a simple algorithm that sorts new instances by having at least k neighbors agree, then keeps all existing instances. The class with the most traits that the trade has in common wins the trade. This calculation is done using the distance function.
- **Support Vector Classifier:** Unprocessed data can be represented as points in an n-dimensional space, where n is the number of attributes that an individual has. SVM techniques allow users to classify their data. Since each feature's value is now associated with a specific coordinate, the data may then be simply categorized. Classifier lines can be used to categorize data into categories and plot them.
- **Decision Tree Classifier:** a supervised learning method for classifying situations. It may be used to efficiently categorize dependent variables that are categorical or continuous. This method divides a population into two or more homogeneous sets based on the most important trait or independent variable.
- **Random Forest Classifier:** It refers to a set of decision trees. Every tree has a class given to it, and each tree "votes" on that class to categorize new elements according to their properties. The forest's most popular category is selected (out of all trees in the forest).
- **Multinomial Naive Bayes:** The foundation of a Naive Bayes classifier is the idea that the presence of one feature in a class does not preclude the presence of any further characteristics. Even if they are connected, a Naive Bayes classifier takes into account each of these characteristics separately when calculating the likelihood of a certain result. Simple to create and effective for huge data sets are naive Bayes models. Despite being straightforward, it is known to outperform even the most sophisticated categorization methods.

IV. PROPOSED METHODOLOGY

The model consists of various phases of execution present in figure 2 which are explained as follows:

Data acquisition: Data must be acquired from various sources through web scraping then the data which is acquired needs to be processed first in order to create a rich and productive data set [15]. This dataset should contain both genuine and fake reviews. Genuine reviews are those that are written by actual customers who have purchased and used the product. Fake reviews are those that are written by individuals who have not used the product or have been paid to write a review.

1. Data pre-processing: Unstructured data is transformed into a useful and efficient format through data preparation. The data pre-processing phase includes various steps such as data cleaning. Noise removal and filtering. Noise is nothing more than a large amount of unwanted and irrelevant data that does not meet the model's data criteria, that must be removed for better performance and accurate results. Data cleaning removes data noise.

A. Data Cleaning: The data must be critical and data should be error-free and clear of unnecessary data. As a result, the material will be cleaned prior to proceeding to the next phase. Data cleansing includes checking and eliminating missing values, duplicate records, and malformations.

B. Data Transformation: A statistical transformation of a set of data is referred to as a data transformation. The data is initially transformed into a suitable form for such a data mining procedure. This enables one to organize hundreds of entries in order to better comprehend their data. Normalization, standardization, and attribute selection are all examples of transformations.

2. *Feature extraction:* A dimensionality reduction technique called feature extraction breaks large amounts of raw data into smaller, easier-to-process groups. The common feature of these huge data sets is that they have many variables that require a lot of processing power to process. The process of selecting and/or combining variables into features is called feature extraction [32]. This method considerably minimizes the amount of information that must be processed while characterizing the actual data set precisely and completely.

3. *Model Training:* Using the features that were retrieved, a machine learning model is then trained. Different machine learning algorithms can be applied. The algorithm is trained using data that has been labelled, with real reviews being classified as positive and false reviews as negative. Data can be categorized in both organized and unstructured forms. The task of classifying a given data set into groups is known as classification. The initial stage in the technique is to forecast the class of a given data point. Classes are frequently referred to by the terms target, label, and class. Classification predictive modelling refers to the issue of approximating the function of mapping specific input factors to specific output variables [2]. The key task is to find out which class or category the new data will supposedly belong to.

4. *Removal:* After classification, the data is clearly classified into two categories, genuine reviews and fake reviews, and using an appropriate machine learning model, the fake review class is removed. And to also ensure the reduction of fake reviews in a system figure 3 contains the detailed steps followed for fake reviews removal.

5. *Model Evaluation:* After the model has been trained, its effectiveness is assessed using a different test dataset. A variety of measures are used to evaluate the model's performance, including accuracy, precision, recall, and F1 score.

6. *Deployment:* The model must then be used to identify and eliminate fraudulent reviews from the online marketplace or business. The model can be used in real-time to automatically flag and remove fake reviews or provide a score indicating the likelihood of a review being fake.

V. RESULT AND DISCUSSION

After testing all of the algorithms on three separate data sets, the Amazon product review data set, the Yelp Hotel Review data set, and the TripAdvisor Review data set, this study was conducted to determine that Support Vector Machine outperformed all other classifiers. Table II shows the exact accuracy of all machine learning approaches on various data sets. To facilitate comparing the results, the output of the Deceptive (D) and Truthful (T) review types produced by the classifier are evaluated with respect to Precision (P), Recall (R), F1-Score (F1), and Accuracy (Acc). With a mean predictive accuracy of more than 88% , the Support Vector Machines Classifier predicted the fraudulent reviews the most accurately.

Logistic Regression, which had an average prediction accuracy of just over 86%, was immediately behind it. The Multinomial Naive Bayes method and the Random Forests Classifier predicted with an average precision level of about 85%. The Decision Tree Classifier, however, was able to anticipate bogus reviews with an average accuracy of just over 75%. The prediction accuracy of the k-Nearest Neighbors method, which was the worst performing algorithm, was only about 61%.

The analysis of false positive and false negative rates to determine how well various machine learning models can categorize and differentiate between fake and authentic reviews is provided with these formulas

$$Accuracy = \frac{TP+TN}{FP+FN+TP+TN} \times 100$$

$$Precision = \frac{TP}{FP+TP} \times 100$$

$$Sensitivity = \frac{TP}{FN+TP} \times 100$$

$$Specificity = \frac{TN}{FP+TN} \times 100$$

$$F1 - Score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \times 100$$

After applying the ML model (Support Vector machine algorithm in above case) to detect fake reviews, this model will remove the review and will also ensure that the same device cannot be used to create more such fake reviews shown in Fig 3. (With his or her account ID and password, the user will log into the system, see various products, and provide product reviews. To establish whether a review is legitimate or fake, the algorithm will determine the user's IP address. The system will alert the administrator and request that the review must be deleted from the system if it discovers numerous fraudulent reviews sent from the same IP address). Even though the overall model may have more time and spatial complexity, none of the current systems combine these two strategies. The Classification Report and the Confusion Matrix of the Machine Learning algorithm in the mentioned datasets (Amazon, Yelp and TripAdvisor) are presented in figures 4, 5, 6 correspondingly.

```
Report:                precision    recall  f1-score   support

      CG      0.89      0.86      0.88      7063
      OR      0.87      0.90      0.88      7088

 accuracy      0.88      0.88      0.88      14151
  macro avg      0.88      0.88      0.88      14151
 weighted avg      0.88      0.88      0.88      14151

Confusion Matrix: [[6095  968]
 [ 724 6364]]
Accuracy Score: 0.880432478270087
Model Prediction Accuracy: 88.04%
```

Fig. 4. Amazon Dataset - SVM

```
-1      0.00      0.00      0.00      39
 1      0.88      1.00      0.94      288

 accuracy      0.88      0.88      0.88      327
  macro avg      0.44      0.50      0.47      327
 weighted avg      0.78      0.88      0.82      327

Confusion Matrix: [[ 0  39]
 [ 0 288]]
Accuracy Score: 0.8807339449541285
Model Prediction Accuracy: 88.07%
```

Fig. 5. Yelp Dataset – SVM

Report:	precision	recall	f1-score	support
deceptive	0.87	0.88	0.87	281
truthful	0.88	0.87	0.87	279
accuracy			0.87	560
macro avg	0.87	0.87	0.87	560
weighted avg	0.87	0.87	0.87	560

```
Confusion Matrix: [[247  34]
 [ 37 242]]
Accuracy Score: 0.8732142857142857
Model Prediction Accuracy: 87.32%
```

Fig. 6. TripAdvisor Dataset - SVM

There are several challenges associated with detecting and removing fake product reviews:

- *Volume of reviews*: As the number of reviews for a product grows, it becomes increasingly difficult to manually review every review. This makes it challenging to detect fake reviews, particularly when there are a large number of them.
- *Diversity of language*: Fake reviews can be written in a wide range of language styles, making it difficult to develop Algorithms that can accurately detect them.
- *Mimicking genuine reviews*: As fake review detection algorithms improve, so too do the techniques used by reviewers to create fake reviews that closely mimic genuine ones. This makes it challenging to develop algorithms that can Accurately identify fake reviews.
- *Platform policies*: Platforms like Amazon have policies that prohibit fake reviews, but they are not always effective in detecting and removing them. Additionally, the policies themselves can be difficult to enforce in a consistent manner.

- *Legal considerations:* There are legal considerations associated with the removal of reviews, particularly when they are posted by individuals or companies that feel they have been wronged. This can make it challenging to remove fake reviews without causing legal issues.

TABLE II ACCURACY TABLE OF DIFFERENT ALGORITHMS ON VARIOUS DATA SETS

Algorithms		Classification Report											
Detection Techniques	Review Type	Amazon Dataset				Yelp Dataset				TripAdvisor Dataset			
		P	R	F1	Acc(%)	P	R	F1	Acc(%)	P	R	F1	Acc(%)
Multinomial Naive-Bayes	D	0.82	0.89	0.85	84.92	0.00	0.00	0.00	88.07	0.79	0.94	0.86	84.46
	T	0.88	0.81	0.84		0.88	1.00	0.94		0.93	0.75	0.83	
Logistic Regression	D	0.87	0.86	0.86	86.42	0.00	0.00	0.00	88.07	0.86	0.86	0.86	86.25
	T	0.86	0.87	0.87		0.88	1.00	0.94		0.86	0.86	0.86	
k-Nearest Neighbors	D	0.54	0.97	0.69	57.37	0.10	0.18	0.13	71.87	0.60	0.95	0.73	65.36
	T	0.87	0.18	0.30		0.88	0.79	0.83		0.88	0.35	0.51	
Decision Tree	D	0.73	0.76	0.74	73.87	0.09	0.10	0.09	76.45	0.67	0.68	0.67	67.14
	T	0.75	0.72	0.74		0.88	0.85	0.86		0.67	0.67	0.67	
Random Forest Classifier	D	0.80	0.89	0.84	83.48	0.00	0.00	0.00	88.07	0.78	0.86	0.82	81.07
	T	0.88	0.78	0.83		0.88	1.00	0.94		0.85	0.76	0.80	
Support Vector Machine	D	0.89	0.88	0.88	88.16	0.00	0.00	0.00	88.07	0.87	0.88	0.87	87.32
	T	0.88	0.88	0.89		0.88	1.00	0.94		0.88	0.87	0.87	

To improve the detection of fake reviews, future research can consider incorporating behavioral factors such as the frequency of reviews, time taken to complete a review, and the ratio of positive to negative reviews. This approach can also be used to identify spammer communities by linking reviews to groups and finding the review with the highest correlation. Additionally, researchers can investigate domain specific features that distinguish real from fake reviews, such as sentiment analysis, feature-based analysis, and aspect-based analysis.

Semi-supervised and unsupervised learning techniques that use both huge volumes of unlabeled data and small amounts of labelled data, in addition to conventional supervised machine learning techniques, may also be successful in detecting false reviews. Data augmentation and generative models can be used to generate synthetic data for training models.

As NLP techniques advance, they can play a critical role in identifying false reviews. Researchers can explore new architectures like Transformer-based models to handle the complexity of text data more accurately [23].

Ensembling, which combines the outputs of several machine-learning models, can improve the overall accuracy of fake review detection. Researchers can develop ensemble models that use different features to identify signs of fakery in reviews, such as overly positive language, a lack of detail, or suspicious patterns in timing or location.

Removing fake reviews is another challenge that researchers can address. They can develop techniques for identifying and removing fake reviews in real-time, as well as methods for removing past fake reviews.

Overall, as new technologies and methods emerge, the future of fake review detection looks promising. By developing more sophisticated and accurate methods, we can ensure that consumers can trust the online reviews they read and make informed purchasing decisions.

VI. CONCLUSION

The above study proved the relevance of fake reviews identification and how they affect the online business.

A comparative analysis of existing machine learning approaches is presented. The suggested method is tested using the Amazon Dataset, Yelp Dataset, and TripAdvisor Dataset. The developed technique uses a variety of classifiers. According to the findings, SVM classifier performs better than the other classifiers at predicting fraudulent reviews as shown in

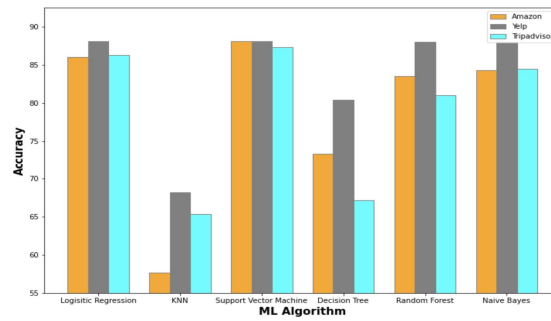


Fig. 7. Accuracy of ML Algorithms on Different Datasets

figure 7 having a mean predictive accuracy over 88%. The behavioral characteristics are not considered in the study that is being presented. Future research may take into account additional behavioral aspects, such as those that rely on how frequently reviewers execute reviews, how long it takes them to finish evaluations, and how frequently reviewers make favorable or negative ratings. It is anticipated that adding more behavioral traits would improve the performance of the approach for spotting bogus reviews.

REFERENCES

- [1] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments, pages 127–138. Springer, 2017.
- [2] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1):e9, 2018.
- [3] Syed Mohammed Anas and Santoshi Kumari. Opinion mining based fake product review monitoring and removal system. In 2021 6th International Conference on Inventive Computation Technologies (ICICT), pages 985–988. IEEE, 2021.
- [4] Rodrigo Barbado, Oscar Araque, and Carlos A Iglesias. A framework for fake review detection in online consumer electronics retailers. *Information Processing & Management*, 56(4):1234–1244, 2019.
- [5] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th annual meeting of the association of computational linguistics, pages 440–447, 2007.
- [6] Nazir M Danish, Sarfraz M Tanzeel, Nasir Usama, Aslam Muhammad, AM Martinez-Enriquez, Adrees Muhammad, et al. Intelligent interface for fake product review monitoring and removal. In 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pages 1–6. IEEE, 2019.
- [7] Elsharif Elmurngi and Abdelouahed Gherbi. Detecting fake reviews through sentiment analysis using machine learning techniques. *IARIA/data analytics*, pages 65–72, 2017.
- [8] Elsharif Elmurngi and Abdelouahed Gherbi. An empirical study on detecting fake reviews using machine learning techniques. In 2017 seventh international conference on innovative computing technology (INTECH), pages 107–114. IEEE, 2017.
- [9] Julien Fontanarava, Gabriella Pasi, and Marco Viviani. Feature analysis for fake review detection through supervised classification. In 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pages 658–666. IEEE, 2017.
- [10] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In proceedings of the 25th international conference on world wide web, pages 507–517, 2016.
- [11] Haruna Isah, Paul Trundle, and Daniel Neagu. Social media analysis for product safety using text mining and sentiment analysis. In 2014 14th UK workshop on computational intelligence (UKCI), pages 1–7. IEEE, 2014.
- [12] Nitin Jindal and Bing Liu. Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining, pages 219–230, 2008.
- [13] Nour Jnoub and Wolfgang Klas. Declarative programming approach for fake review detection. In 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA), pages 1–7. IEEE, 2020.
- [14] Parisa Kaghazgaran, James Caverlee, and Majid Alfifi. Behavioral analysis of review fraud: Linking malicious crowdsourcing to amazon and beyond. In Proceedings of the International AAAI Conference on Web and Social Media, volume 11, pages 560–563, 2017.
- [15] Raymond YK Lau, SY Liao, Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia, and Yuefeng Li. Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems (TMIS)*, 2(4):1–30, 2012.
- [16] Huayi Li, Zhiyuan Chen, Bing Liu, Xiaokai Wei, and Jidong Shao. Spotting fake reviews via collective positive-unlabeled learning. In 2014 IEEE international conference on data mining, pages 899–904. IEEE, 2014.

- [17] Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. 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), pages 1566–1576, 2014.
- [18] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, pages 43–52, 2015.
- [19] Rami Mohawesh, Shuxiang Xu, Son N Tran, Robert Ollington, Matthew Springer, Yaser Jararweh, and Sumbal Maqsood. Fake reviews detection: A survey. *IEEE Access*, 9:65771–65802, 2021.
- [20] Muhammd Jawad Hamid Mughal. Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, 9(6), 2018.
- [21] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance. What yelp fake review filter might be doing? In Proceedings of the international AAAI conference on web and social media, volume 7, 2013.
- [22] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pages 188–197, 2019.
- [23] Ray Oshikawa, Jing Qian, and William Yang Wang. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*, 2018.
- [24] Myle Ott, Claire Cardie, and Jeffrey T Hancock. Negative deceptive opinion spam. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies, pages 497–501, 2013.
- [25] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. Finding deceptive opinion spam by any stretch of the imagination. *arXiv preprint arXiv:1107.4557*, 2011.
- [26] Shebuti Rayana and Leman Akoglu. Collective opinion spam detection: Bridging review networks and metadata. In Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining, pages 985–994, 2015.
- [27] Jitendra Kumar Rout, Amiya Kumar Dash, and Niranjana Kumar Ray. A framework for fake review detection: issues and challenges. In 2018 International Conference on Information Technology (ICIT), pages 7–10. IEEE, 2018.
- [28] Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-gyo Jung, and Bernard J Jansen. Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, 64:102771, 2022.
- [29] Sunil Saumya and Jyoti Prakash Singh. Detection of spam reviews: a sentiment analysis approach. *Csi Transactions on ICT*, 6(2):137–148, 2018.
- [30] Kolli Shivagangadhar, H Sagar, Sohan Sathyan, and CH Vanipriya. Fraud detection in online reviews using machine learning techniques. *International Journal of Computational Engineering Research (IJCER)*, 5(5):52–56, 2015.
- [31] Chengai Sun, Qiaolin Du, and Gang Tian. Exploiting product related review features for fake review detection. *Mathematical Problems in Engineering*, 2016, 2016.
- [32] Eka Dyar Wahyuni and Arif Djunaidy. Fake review detection from a product review using modified method of iterative computation framework. In *MATEC web of conferences*, volume 58, page 03003. EDP Sciences, 2016.