Analysis and Challenges in Detecting the Fake Reviews of Products using Naïve Bayes and Random Forest Techniques

Tanveer Sajid
Capital University of Science and Technology
Wasim Jamshed (wasiktk@hotmail.com)
Capital University of Science and Technology
Navin Kumar Goyal
Suresh Gyan Vihar University
Bright Keswani
Suresh Gyan Vihar University
Gilder Cieza Altamirano
National Autonomous University of Mexico: Universidad Nacional Autonoma de Mexico
Dinesh Goyal
Poornima College of Engineering
Poonam Keswani
Dr Shyama Prasad Mukherjee University

Research Article

Keywords: Fake Reviews, Naïve Bayes, Random Forest, Review Detection, Machine Learning.

Posted Date: June 23rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2302761/v1

Analysis and Challenges in Detecting the Fake Reviews of Products using Naïve Bayes and Random Forest Techniques

Tanveer Sajid[1], Wasim Jamshed[2*], Navin Kumar Goyal[3], Bright Keswani[4], Gilder Cieza Altamirano[5], Dinesh Goyal[6] and Poonam Keswani[7]

[1]Department of Mathematics, Capital University of Science and Technology, Islamabad, Pakistan

[2]*Department of Mathematics, Capital University of Science and Technology, Islamabad, Pakistan

[3]Suresh Gyan Vihar University, Jaipur, Rajasthan, India

[4]Suresh Gyan Vihar University, Jaipur, Rajasthan, India

[5]Universidad Nacional Autonoma de Chota, Cajamarca, Peru

[6]Poornima Institute of Engineering and Technology, Jaipur, Rajasthan, India

[7]Shyam University, Dausa, Rajasthan, India

[*]Corresponding Author Email: *wasiktk@hotmail.com*

Abstract

In today's world, fake review identification and prediction is an importantarea of sentiment analysis of the E-commerce industry. The automatic fake review categorizersidentify and categorize a variety of duplicate, spam, fake and untrustworthy reviews using machine learning techniques. This paper studies various recent existing fake review detection methods using NB and RF classifiers for the Yelp and Flipkart datasets. It provides a detailed study on various fake review predictors and compares their basic and performance-based specifications. It highlights the challenges, threats, and gaps of these existing works. Further, it graphically shows the discrimination for the specifications of year-wise evolution, classifier usage, and dataset usage.

Keywords: *Fake Reviews, Naïve Bayes, Random Forest, Review Detection, Machine Learning.*

1 Introduction

Automated fake review identification, detection, analysis, and prediction becomes an important zone of sentiment analysis in the present scenario of Natural Language Processing (NLP) and the E-commerce market. Recently, the requirement of these automatic systems using Machine Learning (ML) techniques has grown greatly [1] - [8] that applies a strong base of NLP, ML, and Artificial Intelligence (AI) fields. In the industrial, corporate, and commercial realms, many predictors and analyzers have been developed, but a reliable and effective fake review-based prediction system remains a major necessity.

The top-level design of the fake review detection and identification system aims to locate and identify fake reviews. Further, it checks the system performance and efficiency using many performance-based metrics and parameters. The fake reviews can be wrong customer feedback, spam reviews, poison reviews, fabricated reviews, threatening reviews, and untruthful and deceptive reviews. Fig. 1 depicts the top-level design of this system. The system first accepts the imbalanced dataset, then pre-processes them, extracts the features or reviews, applies the ML classifiers, identifies and classifies the fake reviews, and evaluates the performance of the system using different measures.

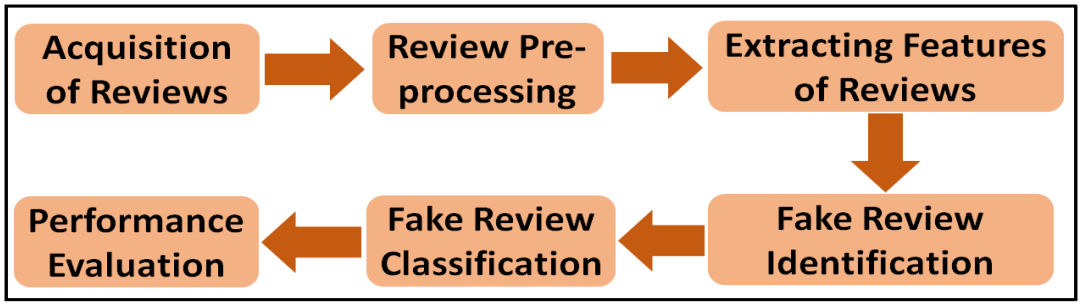

Fig. 1. Top-level design of fake review detection and identification system.

The sections are arranged in the following order. Section 2 presents the systematic and tour of various existing fake review prediction systems. Section 3 illustrates and discriminates the basic and performance specifications of these existing sentiment analyzers and systems. It provides their comparison based upon the classifiers, datasets, and performance metrics. Section 4 depicts their challenges, limitations, and threats. Section 5 analyzes these specifications graphically in terms of year-wise evolution of existing works, classifier usage, and dataset usage. Section 6 concludes the paper along with future recommendations.

2 The Journey of Existing Works

This section explains and compares the processes of the most recent automated fake identification and detection algorithms using different datasets and ML techniques. This journey describes the procedures of these existing methods to locate, identify, predict and classify the fake reviews in sentiment analysis. To demonstrate their effectiveness, these approaches are presented here one by one. The survey [1] of the prominent supervised ML techniques was presented for online spam and fake review detection and it analyzed the performance of various strategies to detect such fake reviews. It presented a thorough review of recent research on detecting fake reviews, as well as a methodology for further investigation.

The ensemble learning strategy was utilized in the fake review detection [2] by merging two active-based supervised learning algorithms. For this, it followed three different filtering phases such as KL and JS distance, Term Frequency/Inverse Document Frequency (TF/IDF) features with n-gram attributes of the review content. It experimented with the proposed approach with four types of simulations, and also labeled over 1000 unlabeled fake reviews manually

during itsActive Learning (AL) process.It was found that the Linear Support Vector Machine (LSVM) performed the best with the AL method. Further, Naïve Bayes (NB) outperformed other supervised classifiers.Another such approach [3] implemented three ensemble-based techniques along with four classifiers. It compared the findings of C4.5, Multinomial NB (MNB), Logistic Regression (LR), SVM, boosting, bagging, and Random Forest (RF).

The fake review detection model [4] implemented the ensemble approach by using the NB, SVM, and LR as base classifiers. Another model [5] proposed an ensemble approach for fake review detection of tweets. It implemented five Convolutional Neural Networks (CNN), and content-related, user-related, and n-gram attributes of the feature-based model. Another such detection model [6] classified the movie reviews of positive and negative polarity through NB, K-Star (K*), SVM, J48 Decision Tree (DT), and K-Nearest Neighbors (KNN). The fake review prediction model [7] detected the truthful and useful reviews by using deceptive and useful classifiers respectively and then provided the ranks to the reviews through a ranking model. Further, it built a repository cum dictionary to categorize the reviews as truthful reviews or deceptive reviews. It categorized the testing data as deceptive or truthful by using the deceptive classifier and also as useful or useless by using the useful classifier.

Another spam review classification method [8] implemented the ensemble of the global and local filter-related feature selection methods. It found the worst-case complexity for spam and non-spam classes as O (nm^2 + m + m (log (m))), where m and n represented the count of features and count of review instances, respectively. Further, it achieved the global feature score in $2nm^2$ iterations and sorted these features in O (m (log (m)))by using the best sorting algorithm.

The fake review detection method [9] worked upon the n-gram (unigram and bigram) features by implementing the ensemble techniques such as majority voting and stack-based. It used ten-fold cross-validation during the training and testing phases to ensure accurate classification. The performance of several deep learning approaches such as NB, SVM, and DT was compared in the study [10]. It investigated how well multiple ML approaches for detecting positive and negative false consumer reviews performed. It was discovered that such as Long Short Term Memory (LSTM) and CNN types of Deep Neural Networks (DNN), , outperformed classic ML algorithms to determine the accuracy and time performance.

The ensemble classification method [11] detected the fake reviews for Amazon Mechanical Turk (AMT) and Trip advisor datasets. Further, it evaluated the system using 5-fold cross-validation. The study [12] provided a detailed study and advanced research on several existing spam review detection strategies. It illustrated the taxonomy of ML techniques and focused on their research gaps and future recommendations. The fake review detection model [13] used several ML techniques to classify the Amazon product's reviews as positive or negative.

The ensemble ML model [14] predicted the spam reviews through KNN, Multi-Layer Perceptron (MLP), and RF, and then implemented the majority voting technique. It employed univariate, Chi-square, and information gain feature selection methods to extract 25 statistical features from the feature space. Then it filtered out the top ten optimal features. The fake review identification and classification method [15] analyzed the system's performance for the restaurant dataset by implementing ensemble learning-based approaches such as DT, RF, SVM, and Extreme Gradient-Boosting Trees (XGBT).

The sentiment prediction method [16] classified the reviews into positive, negative, or neutral reviews by implementing the Gradient Boosting Machine (GBM), RF, LR, and SVM. It analyzed the system's performance through the Gradient Boosted SVM (GBSVM) majority voting classifier. It evaluated the system results using two different datasets with TF and unigram, bigram, and trigram features of TF/IDF. Another method [17] investigated the current research by combining the findings of existing sentiment analysis studies. It gave a thorough overview of the main challenges and distinct methodologies for a variety of sentiment analysis applications. Further, it elaborated various characteristics, techniques, and datasets that are employed in sentiment analysis systems. It identified their challenges to find the points that required more research efforts. The LSTM and CNN algorithms were found to be the most popular deep learning algorithms for sentiment analysis.

The study [18] analyzed various existing fake news detection methods from 2017 to 2021. It also provided a detailed review of the recent and past false news detection using different ML algorithms. The spam review detector [19] implemented the ensemble approach with the combination of NB, LR, DT, Gradient Boosted Trees, RF, and Artificial Neural Networks (ANN). Another method [20] resampled the imbalance data, so it pruned the features to reduce the computational cost and optimized the parameters by using grid search to get the best values for the necessary parameters. Finally, to combine the optimized base classifiers, it used an ensemble classifier using majority voting and stacking techniques.

3 Specification-based Discrimination for Existing Works

Section 2 illustrated many recent systems of spam and fake review detection in sentiment analysis. This section discriminated them from each other using some basic and performances specifications. The majority of these systems were tested using standard datasets and the results were measured to evaluate the AUC, accuracy, precision, recall, F-measure, and other performance metrics.

Table 1 depicts the comparison of several existing contributions and algorithms using three basic specifications called problem-focused, classifier, and data sets, along with one performance specification of measures. All these existing methods implemented various classification techniques and their ensemble approaches. They are Gradient Boosted Trees, Boosting, Bagging, RF, NB, MNB, SVM, LSVM, KNN, K*, DT-J48, C4.5, LR, ANN, CNN, DNN, LSTM, MLP, XGBT, GBM, and maximum entropy. These systems used the standard datasets such as Gold dataset, movie review dataset, hotel dataset, restaurant dataset, HSpam balanced dataset, and 1KS10KN imbalanced dataset. These datasets were collected from various sources and domains such as Yelp and OTT, Amazon, Epinions, TripAdvisor, Twitter, social media, and Google play store. They have measured the performance of their systems by determining the accuracy, precision, AUC, f-score, and recall metrics.

Table 1. Discriminating existing fake review detection systems using basic and performance specifications.

Ref. no.	Problem-focused	Classifiers	Data sets	Performance metrics

[2]	Fake review detection using hybrid ensemble method.	Ensemble: NB, SVM, DT &maximum entropy.	3600 reviews from different domains. Real-life and pseudo reviews.	Precision: 95%, recall: 95%, f-score: 95%, & accuracy > 88%.
[3]	Spam review detection using ensemble techniques.	3 ensemble techniques: boosting, bagging & RF.	OTT and Yelp datasets.	Accuracy: 89.7%.
[4]	Ensemble classification for spam review detection.	Ensemble: NB, SVM & LR.	Amazon.	Precision: 0.882, recall: 0.881, f1-measure: 0.881 & accuracy: 88.09%.
[5]	Spam detection usingANN-based ensemble approach.	CNN with the ensemble of RF & SVM.	Twitter. HSpam balanced & 1KS10KN imbalanced datasets.	Promising results.
[6]	Fake review detection through sentiment	NB, SVM, KNN, K* & DT-J48.	Movie review dataset.	Accuracy (%) for v2.0 dataset = NB: 79.7, K*: 71.15, KNN-IBK with K as 3: 70.85, SVM: 81.35 & DT-

	analysis using ML.			J48: 71.6. %Accuracy % for v1.0dataset: NB: 70.9, K*: 69.4, KNN-IBK with K as 3: 70.5, DT-J48: 69.9& SVM: 76.
[7]	Truthful and useful review detection using opinion mining.	SVM, KNN & NB.	Balanced & imbalanced datasets. Amazon & Epinions.	Range of % Accuracy = Fake/non-fake classification for imbalanced data: 60-66. Useful/non-useful classification for balanced one: 73-79.
[8]	Ensemble of global & local feature selectors to classify the spam reviews.	MNB, LSVM, LR & C4.5.	TripAdvisor hotel review dataset & Yelp filtered review dataset.	Got best AUC score with MNB: 0.98 & 0.91 on synthetic & real datasets, respectively.
[9]	Spam review detection using ensemble ML.	Ensemble: NB, SVM & RF.	Gold standard dataset. Hotel reviews.	Majority voting ensemble = 0.86 precision for spam, 0.88 precision for non-spam, 0.89 recall for spam, 0.85 recall for non-spam&87.43%

				accuracy. Stacking ensemble with RF = 0.89 precision for spam, 0.86 precision for non-spam, 0.85 recall for spam, 0.89 recall for non-spam &87.68% accuracy.
[11]	Fake review detection using an ensemble approach.	Ensemble: SVM, NB & KNN.	Yelp & OTT datasets. Hotel reviews from AMT & Tripadvisor.	Promising results.
[13]	Analysis & detection of fake reviews using opinion mining.	SVM, NB & LR.	Online products from Amazon.	Accuracy: 90%.
[14]	Ensemble ML to classify the spam product reviews.	Ensemble: MLP, KNN & RF.	Yelp datasets: hotel & restaurant datasets. Products.	88.13% accuracy with Chi-square & 88.7% accuracy with univariate.
[15]	Fake review detection through	Ensemble: DT, RF, SVM,	Restaurant dataset. Fake reviews from	Accuracy (%) = DT (79.6), XGBT (78.3),

	ensemble learning.	XGBT &MLP.	three restaurants.	SVM (72.6), RF (71.5) & MLP (68).
[16]	Ensemble for sentiment classification of unstructured reviews.	SVM, GBM, LR & RF.	Google play store apps. Twitter dataset.	Accuracy (%) = GBSVM (93), & GBSVM with TF-IDF (90).
[19]	Spam detection using the ensemble method.	Ensemble: NB, LR, DT, gradient boosted trees, RF & ANN.	Yelp dataset. Social media.	Accuracy from 65% to 84%.
[20]	Fake review detection using an ensemble approach.	Ensemble: Bagging, boost & RF.	Yelp datasets: hotel and restaurant datasets.	Accuracy range from 80% to 90%.

Some observations are obtained here.It was found that many of them reduced their feature set and dimensions through pruning and other reduction methods. It was also discovered that some algorithms for detecting fraudulent and spam reviews included hybrid and ensemble classifiers. They prominently used NB, RF, SVM, RF, and boosting classifiers. Despite employing powerful methods, existing approaches have demonstrated very limited utilization of unbalanced datasets and

data from a variety of sources. Such limitations raised many challenges and issues in their successful implementation.

4 Challenges in Fake Review Detection

The fake review detection systems discussed in sections 2 and 3 had many challenges, issues, and threats. The challenges and threats are stated to demonstrate the effectiveness of their system implementation. So, this section presents the need for their future research to fill the gapsand limitations. They are given below.

- Need to extend with large-scale datasets [2] [16] of different languages from different domains. Feature sets need to be extended [2].
- No substantial improvement was observed when compared to MNB without the use of an ensemble approach. Used limited labeled data of real-world spam review environment. Need to build and test the new data sets [3].
- Need to extend with other classifiers [4].
- Poor performance of feature-based methods for HSpam14 data set [5]. Need to represent the features in a presentable form [5]. Need to increase the efficacy of deep learning systems by considering extra information about the tweets etc. [5].
- Extend with Amazon or eBay dataset. Extend with other programming platforms [6].Well suited for 2-class problems only [7]. Extend the proposed work with other methods for labeling online reviews [7].
- Extend the proposed work to include a different set of local and global feature selection measures. Extend with any other domain dataset of spam detection.

Extend with meta-features and textual features for performance improvement [8].

- Need to improve the accuracy results [7] [9] [11].
- Need to work upon live review data from different review websites. Extend with the content-aware classification to identify sarcasm and other human emotions [13].
- Extend spam review classification using deep learning approach and LSTM with weighted TF-IDF approach [14].
- Need of more sophisticated analysis for interactions between two parties having time dimensions such as LSTM and clustering with deep learning, evidence theory, and reinforcement learning [15].
- The proposed work needs improvement to detect spam for specific domains [19].
- Used only two Yelp datasets. Had time-consuming ablation study when there are too many features. Need to examine the robustness property of the proposed system. Using semi-supervised and unsupervised methods, extend with labeled fake review datasets. [20].

Therefore, it is observed that the data sets from Yelp and OTT have become the most popular for fake review identification and detection systems. Although these systems performed well on typical datasets, they faced numerous obstacles, including validity concerns, imbalanced data, classifier selection, feature reduction, restricted resources, low accuracy and performance, and so on. These challenges require considerable attention and effort to be mitigated to the greatest extent possible.

5 Analysis of Various Specification Categories

The basic and performance-based specifications show the results of different parameters in existing algorithms. This section graphically depicts various specification-based comparisons among existing contributions. The specifications such as year-wise system evolution, classifier type and usage, and dataset type and usage are the primary factors of such comparisons. They are depicted in Figs. 2, 3, and 4, respectively. Fig. 2 depicts the maximum research contribution of spam review detection in the years 2018 and 2020. Fig. 3 illustrates the maximum usage of various classifiers. The SVM + variants, NB + variants, and RF classifiers were found as the most prominent ones with their usage results of 20%, 16.36%, and 14.54%, respectively. The usage of other classifiers was found as 9.09% for LR and boosting + variants each, 7.27% for DT and KNN each, 3.63% for bagging, ANN + CNN, and MLP each, and 1.81% for C4.5, K*, and maximum entropy each.

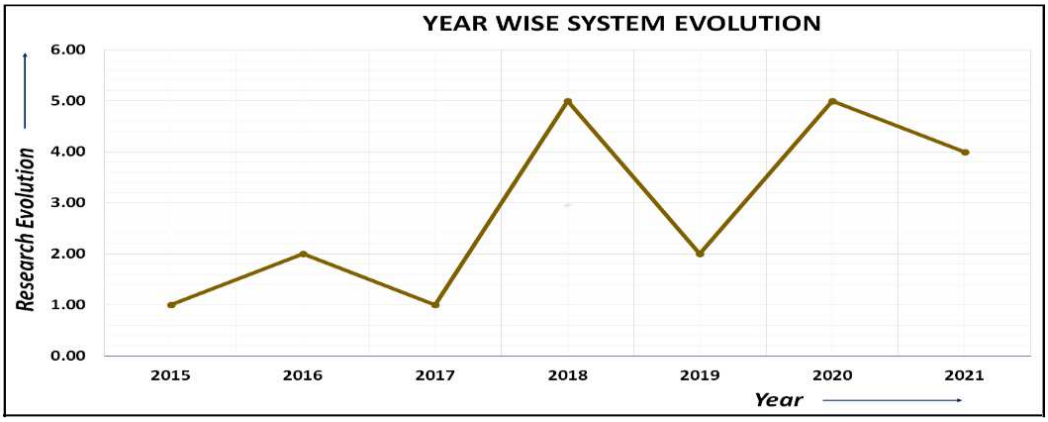

Fig. 2. Depicting the year-wise evolution.

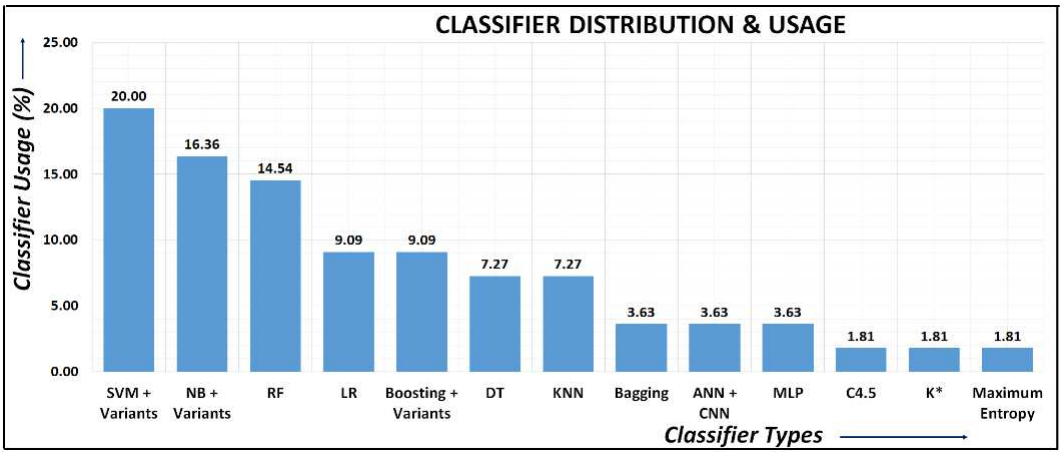

Fig. 3. Depicting the usage of various classifiers.

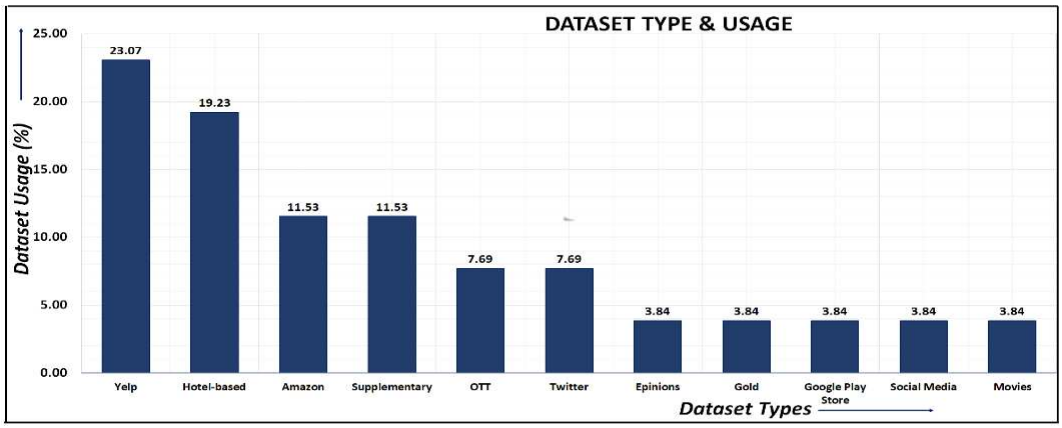

Fig. 4. Depicting various types of datasets and their usage.

Fig. 4 depicts the usage result % of various datasets of existing algorithms. The Yelp, hotel-based, Amazon and supplementary datasets were found as the most prominent ones with their usage results of 23.07%, 19.23%, 11.53%, and 11.53%, respectively. Here the hotel-based datasets included the datasets of TripAdvisor, AMT, restaurants, and other hotels. The usage of other datasets was found as

7.69% for each of OTT and Twitter, and 3.84% for each of the Epinions, gold, google play store, social media, and movies datasets.

6 Conclusions and Future Recommendations

This paper provideda thorough analysis and comparison of various specifications of existing fake review analysis and prediction systems using NB and RF. The basic specifications included the year-wise evolution, classifiers, and data sets. Along with that, the performance specification included the precision, recall, f1-score, accuracy, and AUC. The study highlighted the challenges, issues, threats, and research extensions of various existing works. This study found that the maximum research contributions were made in the years 2018 and 2020. Additionally, the usage % was found maximum for SVM and NB classifiers, and also for the Yelp and hotel-based datasets.

Although the fake review predictors and classifiers in sentiment analysis is a grown-up research area, the research techniques and challenges are also arising continuously. Their observations stated the existing research works have not worked upon Flipkart datasets, and secondly, many of them achieved low accuracy. Their limitations and risks were also explored to identify the gaps and future advancements that the fake review detection systems require. Data balanced-imbalanced issue was also found as a high concern. The future work includes accurate and efficient aspect-based sentiment analysis classification for fake review detection using an ensemble approach with real-time domains and datasets.

Ethical Approval

Not applicable

Consent to Participate

Not applicable

Consent to Publish

Not applicable

Author Contributions

TS and WJ framed the issue. NKG, BK, and GCA resolved the problem. TS, WJ, NKG, BK, GCA, DG and PK computed and analyzed the results. All the authors equally contributed to the writing and proofreading of the paper.

Fundings

None

Competing Interests

The authors announce that no conflict of interest exists.

Date Availability

All data generated or analyzed during this study are included in this published article.

References

[1] Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., Najada, H. A.: Survey of review spam detection using machine learning techniques. Journal of Big Data 2 (23), 1-24 (2015).

[2] Ahsan, M. N. I., Nahian, T., Kafi, A. A., Hossain, M. I., Shah, F. M.: An ensemble approach to detect review spam using hybrid machine learning technique. In: 19th International Conference on Computer and Information Technology (ICCIT), pp. 388-394. IEEE (2016).

[3] Heredia, B., Khoshgoftaar, T. M., Prusa, J., Crawford, M.: An Investigation of Ensemble Techniques for Detection of Spam Reviews. In: 15th IEEE

International Conference on Machine Learning and Applications, pp. 127-133. IEEE (2016).

[4] Ibrahim, A. J., Siraj, M. M., Din, M. M.: Ensemble classifiers for spam review detection. In: IEEE Conference on Application, Information and Network Security, pp. 130-134. IEEE (2017).

[5] Madisetty, S., Desarkar, M. S.: A neural network-based ensemble approach for spam detection in Twitter. IEEE Transactions on Computational Social Systems 5 (4), 973-984 (2018).

[6] Elmurngi, E., Gherbi, A.: Detecting fake reviews through sentiment analysis using machine learning techniques. In: The Sixth International Conference on Data Analytics, pp. 65-72(2018).

[7] Algotar, K., Bansal, A.: Detecting truthful and useful consumer reviews for products using opinion mining. In: CEUR Workshop Proceedings 2111, 63-72 (2018).

[8] Ansari, G., Ahmad, T., Doja, M. N.: Spam review classification using ensemble of global and local feature selectors. Cybern. Inf. Technology 18 (4), 29–42 (2018).

[9] Mani, S., Kumari, S., Jain, A., Kumar, P.: Spam review detection using ensemble machine learning. Machine Learning and Data Mining in Pattern Recognition, Springer, 198-209 (2018).

[10] Hajek, P., Barushka, A: A comparative study of machine learning methods for detection of fake online consumer reviews. In: Proceedings of the 2019 3rd International Conference on E-Business and Internet, pp. 18–22. Association for Computing Machinery, NY (2019).

[11] Baraithiya H., Pateriya, R. K.: Classifiers ensemble for fake review detection. International Journal of Innovative Technology and Exploring Engineering 8 (4), 730-736 (2019).

[12] Khan, R. A., Shoaib, F. M.: Spam review detection: a systematic literature review. TechRxiv, pp. 9-25. , IEEE (2020).

[13] Ashwini, M. C., Padma, M. C.: Efficiently analyzing and detecting fake reviews through opinion mining. International Journal of Computer Science and Mobile Computing 9 (7), 97–108 (2020).

[14] Fayaz, M., Khan, A., Rahman, J. U., Alharbi, A., Uddin, M. I., Alouffi, B.: Ensemble machine learning model for classification of spam product reviews. Complexity 2020, 1-10 (2020).

[15] Gutierrez-Espinoza, L., Abri, F., Namin, A. S., Jones, K. S., Sears, D. R. W.: Fake reviews detection through ensemble learning. Journal, 1-8 (2020).

[16] Khalid, M., Ashraf, I., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S.: GBSVM: Sentiment Classification from Unstructured Reviews Using Ensemble Classifier. Applied Sciences 10(8): 2788, 1-20 (2020).

[17] Ligthart, A., Catal, C., Tekinerdogan, B.: Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Reviews 54, 4997–5053 (2021).

[18] Kumar, S., Kumar, S., Yadav, P., Bagri, M.: A survey on analysis of fake news detection techniques. In: International Conference on Artificial Intelligence and Smart Systems, pp. 894-899. IEEE (2021).

[19] Wang, J., Xue, D., Shi, K.: An ensemble framework for spam detection on social media platforms. International Journal of Machine Learning and Computing 11(1), 77-84 (2021).

[20] Yao, J., Zheng, Y., Jiang, H.: An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization. IEEE Access 9, 16914-16927 (2021).