Online Shopping Fake Reviews Detection Using Machine Learning

Afraz Moqueem[1], Fayaz Moqueem[1], Chandra Vamshi Reddy[2], Dannana Jayanth[3], and Brinta Brahma[4](✉)

[1] School of Electronics and Communication Engineering, Muffakham Jah College of Engineering and Technology, Hyderabad, India
[2] School of Civil Engineering, Gokaraju Rangaraju Institute of Technology, Hyderabad, India
[3] School of Chemical Engineering, Indian Institute of Technology Ropar, Rupnagar, India
[4] School of Electronics and Communication Engineering, Jadavpur University, Kolkata, India
brintabrahma98@gmail.com

Abstract. Online shopping has drastically reduced the tiresome job of reaching out to offline stores and selecting goods in a limited product range. Almost everything is available online, right from the basic essential goods to costlier electrical appliances in today's world. The sellers increasingly misuse these massive online platforms for increasing their product sales by posting false reviews. Consumer engagement reports suggest that around 82% of customers read online reviews before purchasing a product online. So these reviews are crucial for them to decide if the product suits them and is reliable. So, in this paper, we propose various machine learning models for detecting fake reviews and delineate and do a comparative analysis of each model to determine the best algorithm. This work plays a vital role in reducing and checking fake reviews.

Keywords: Fake reviews · Gaussian Naive Bayes · Support vector machine · Random forest · Linear discriminant analysis · Online shopping

1 Introduction

The success of any business is attributed to various factors. In this, the satisfaction of the customer holds the top place. Customer satisfaction is most important for the industry to retain the consumers and build the company's trust and brand. The reviews published by the verified customer help the company improve the product and helps other customers to know about the product. The studies show that around 84% of shoppers trust the reviews posted online as equally as a personal recommendation. Most of the shoppers who read the reviews are new to the product and are looking for the experience of the people who are familiar with it. These people need proof to trust the product for which they are paying.

It is found from the study conducted on purchase patterns of customers that around 93% read the product reviews before making a purchase. Moreover, the goods with more positive reviews stay on the top of the search result, making them more visible.

D. S. Guru et al. (Eds.): ICCR 2021, CCIS 1697, pp. 305–318, 2022.
https://doi.org/10.1007/978-3-031-22405-8_24

But, few adversaries try to gain these benefits by posting fake reviews, which intensify and significantly highlight the product even if it is not up to the mark quality. Therefore, fake reviews will falsely lure most consumers into buying the products with fake reviews. If this continues to persist, it could negatively affect the online shopping ecosystem as the consumers will lose their trust in reviews.

In the last few years, the developments in Natural Language Processing, especially its combination with machine learning, produced excellent results in classifying the text. There are many existing research works on detecting fake reviews. However, most of the works are restricted to rigid rules. So, in this paper, we propose four popular ways of detecting fake reviews: Naive Bayes, SVM, Random Forest, and Latent Dirichlet analysis.

Support Vector Machine (SVM) categories the extreme points of the dataset and form a hyperplane by drawing a decision boundary. The hyperplane is located at the extreme ends of the data. So, SVM does a great job in the segregation of two classes. In our proposed model, a Linear support vector machine is used to differentiate between fake and genuine reviews.

Naive Bayes works efficiently based on the probability theory and the Bayes Theorem to classify test tags. It functions on the concept of conditional probability, wherein multiple independent features are given as input. All the features are assumed to contribute equally to produce a classifying output.

Random forest is considered to be a powerful supervised model. It works by taking the mode of the multiple decision trees to arrive at an output. Each decision tree continuously divides the tree until it reaches the leaf node.

Linear Discriminant Analysis is a type of supervised machine learning algorithm and is widely used to overcome the problem of logistic regression like two class problems and unstability between classes.

2 Related Work

In [1], Ahmed M. Elmogy et al. Proposes a machine learning methodology for successfully detecting fake reviews posted online. The models are trained on the Yelp dataset of the restaurant's reviews. The features are extracted from the user's behavior. Different language models such as bi-gram and tri-gram are considered for the evaluation. After the appropriate data is extracted, various preprocessing techniques were performed to reduce the time complexity and to increase the accuracy of the machine learning models. This process involved Tokenization, Lemmatization followed by feature extraction and feature engineering. The dataset included approximately 5000 reviews of around 200 hotels. The dataset is then split into labeled honest reviews and fake reviews. Then the Machine learning models named SVM, Random forest, and Logistic regressions are trained. Evaluation metrics such as accuracy, recall are used to measure individual performances. In the end, the paper concludes that SVM outperformed the remaining models.

According to Wenqian Liu in [2], fake reviews lead to financial losses for the customers because of their false and deceptive information. This paper proposed a method for detecting false reviews based on the review history associated with products. They analyzed the features of the reviews using an Amazon china dataset. In the beginning, the review records of products are extracted to a temporal feature vector. The method's effectiveness is verified and compared to existing temporal outlier detection models using Amazon China dataset. The paper also scrutinized the impact caused by the parameter selection of review records.

Good feedback is crucial for any business to be successful in gaining the trust of customers. In [3], Fake reviews are detected via analysis of linguistic features. The paper used natural language processing efficiently to classify fake reviews. Fifteen linguistic features were studied and measured their importance for classification. It can be inferred from the paper that fake reviews are most likely tend to contain redundant terms and stopwords and are more often in long sentences. It is also concluded that linguistic features help to determine the fake reviews with decent accuracy.

According to Luis et al. in [4], conventional machine learning techniques need to be compared with alternatives to determine the better approach for detecting fake reviews. So, this paper compares the ensemble-based methods which are incorporated with conventional support vector machines. These techniques are compared to the traditional machine learning models. The research team in the paper created the custom-built dataset and named it "Restaurant Dataset." This dataset included 86 reviews with 43 fake and 43 genuine reviews for three restaurants. The various machine learning models used are Support vector machine, random forest, and Multilayer perceptron. The test results that Ensemble learning-based classifiers got an accuracy of 77%. It is concluded in the paper that the ensemble-based machine learning techniques outperformed the conventional machine learning models in detecting fake reviews.

In "Review Spam Detection using Machine Learning," Drasko et al. Studies spam detection approaches that are based on machine learning and put forths their overview and results. The authors present that the results yielded are different for different datasets. It is mentioned that linguistic approaches appeared in most of the research works. However, the spammer detection methods also produced efficient results. The paper concludes by saying that future research must be based on a combination of reviewer-based and content-based strategies to achieve accurate results.

The paper [6] proposed an ensemble approach using a hybrid machine learning technique for detecting spam reviews. It is mentioned that the most difficult is with the dataset since there is not sufficient large-scale real-life labeled data. Moreover, they also posed that pseudo instances cannot deliver the appropriate solution for solving a real-life problem. In this model, the duplicates were removed by using KL-JS distance measures. A hybrid dataset is constructed, which comprises both fabricated and actual data, which aids in detecting a wide range of data instances. The novelty of this work is it explores various content-based features such as tf-idf values and some linguistic features. The paper concludes that the manually created hybrid dataset works fine with supervised machine learning models to detect fake reviews.

In today's world, fake news spreads much faster than real news, which is much destructive. In order to mitigate such false news, Aravinder Pal Singh et al. in [7]. Proposed various machine learning models and Natural Language Processing Techniques for detecting Fake News. Three standard datasets were collected, and feature extraction was performed on headlines and content from each dataset. The model is evaluated for seven machine learning algorithms: Random forests SVM, Gaussian Naive Bayes, AdaBoost, KNN, MLP, and Gradient boosting. Various evaluation metrics such as accuracy and standard deviation are calculated to determine the most efficient algorithms. It is observed that the XGB classifier outperformed all other machine learning models used.

In [8], Syed Ishfaq Manzoor et al. reviewed various machine learning models in detecting fake news. It is mentioned that the ever-changing features of fake news in social media platforms make it challenging to categorize them. Still, for deep learning methods to work, one needs to compute hierarchical features—the paper elaborated on the types of data on social media which need to be focussed. The three significant forms mentioned are Text, Multimedia, and Hyperlinks. The fake news types classify as Visual-based, User-based, Knowledge-based. In visual-based, the news posts use lot more graphics which includes morphed photos, manipulated videos. In User-based news, the fake accounts are fabricated and targeted to a specific set of audiences such as age, interest, gender, etc. Knowledge-based give false scientific information on some unresolved issues.

In [9] Jane crystal, et al. points out the destructive influence of the false reviews posted online. So they figure out to solve this problem through sentiment analysis. The dataset is preprocessed, including steps such as converting characters to lower case followed by the Tokenization. In the feature selection process, the most relevant data is extracted to get an accurate output. Then the data is applied to machine learning algorithms such as LSTM, Bi-Directional LSTM, GRNN. From the literature review conducted in the paper, it is found that the Naive Bayes model is the most used method in the existing works to detect fake reviews. Various activation functions such as Relu, tanh, and sigmoid function and compared each with each machine learning model.

In [10], the authors proposed and compared various machine learning techniques on the yelps dataset. A clear explanation of each algorithm is used to provide a proper understanding of why they do better. In the end, the XGBoost classifier outperformed other models with an F1 score of 0.99 in detecting fake reviews.

In [11], Zhijie Zhang et al. analyzed the Twitter spam characteristics into user attributes, activity, and relations. A novel spam detection algorithm is developed based on the extreme machine learning named the Improved Incremental Fuzzy-kernel-regularized Extreme learning machine (I2FELM). The paper used this algorithm to detect spam efficiently.

It is clearly evident from the existing works that various techniques were used for determining fake reviews. But, most of the works lack a proper original dataset of reviews because the dataset may not be available at the time of the experiment. This also relates to the less reliable results of the works conducted with a meager inadequate dataset. So, In this paper, we have used a real-life dataset of reviews on the shopping platform Amazon. Adding to that, many previous works classified the reviews based on a rule-based approach which is dependent on just some rules to differentiate between fake and genuine reviews. This paper also addresses this problem by using machine learning algorithms for detecting fake reviews.

3 Proposed Methods

A. *Support Vector Machine (SVM)*

Support Vector Machine is one of the non-linear supervised models. Given a set of labeled training data, SVM will help us find an optimal hyperplane that categorizes new input data in one-dimensional space. In one dimensional space, the hyperplane is a point, In two-dimensional space, the hyperplane is a line, and In 3-dimensional space, the hyperplane is a surface. In general, there are many ways to form a hyperplane to separate the two classes. However, SVM constructs a perfect hyperplane that has a maximum margin from both classes. SVM consists of support vectors which are the data points that lie closest to the hyperplane. Support Vectors is an attribute informing the hyperplane. Here, the number of input features is two, so the hyperplane would be a line.

The dataset used contains genuine reviews and fake reviews. In the beginning step, the dataset is trained on the SVM classifier. The classifier plots a boundary between the fake and genuine reviews with the available dataset, and this boundary is a hyperplane. It is plotted to have the maximum possible distance from each class using the following mathematical expression.

$$\mathrm{X} = \mathrm{Z} * \mathrm{X} + \mathrm{a} \tag{1}$$

where,

X is the Classification label

Z is the Parameter of the plane

a is the Point to the position of the plane wrt to the origin (Fig. 1)

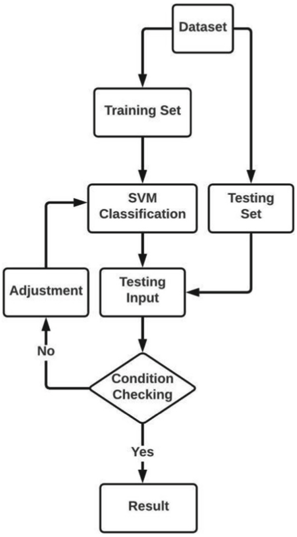

Fig. 1. Flowchart for SVM

After training the SVM classifier with the data, a hyperplane is formed between the genuine and fake reviews. Further, if we feed input to the SVM classifier, a result is popped, which indicates whether the input is a fake or genuine review. Adding to that, this input is further used to form a new hyperplane. SVM is more importantly used to find extreme data vectors since the classifier is accurate and adaptive.

Algorithm

1. Consider 2 points on a linear plane X(a1, b1), Y(a2, b2) and load their respective coordinate values. Assuming max width as Z = 0.
2. C <= 22; Assign a random value
3. Loop
4. for all {ai, bi}, {aj, bj} do
5. iterate Zi and Zj
6. end

Iterate until there is no change in the value of Z or other resource restriction requirements have been met. Ensure that only the support vectors (Zi > 0) are held.

B. Naïve Bayes (NB)

The Naive Bayes algorithm works on the principle of Bayes theorem's application and high impact judgments over the classification process through a probabilistic approach. This approach produces coherent solutions in the predictive analysis of machines and displays efficient performance in detecting fake reviews.

This approach calculates the probability of the likelihood P(xlc) by calculating the posterior probability P(clx), using the probability of class P(c) and the prior probability of predictor P(x). The below equation helps to calculate the posterior probability (Fig. 2).

$$\mathrm{P(c|x) = (P(x|c) * P(c))/p(x)}$$

Algorithm

1. Dividing by events occurring
2. Finalize all the values in the datasheet
3. Calculate mean and standard deviation, then build a separate datasheet.
4. In the end, the class is predicted by finding probabilities

Thus, the naive Bayesmodel mainly follows the Bayes theorem's application and high impact judgments over the classification process through a probabilistic approach. This model gives coherent solutions in predictive analysis of machines and shows excellence in spam filtering (such as advertisements, links, etc.).

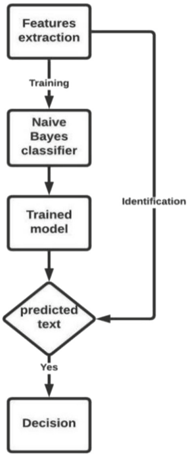

Fig. 2. Flowchart for naïve Bayes

C. Random Forest (RF)

Random forest is a method that operates by constructing multiple Decision Trees during the training phase. Although individual decision trees could predict the output efficiently for a static dataset, they lack performance when trained with the variable dataset. So, The common output of the maximum number of trees is considered as the output of the random forest algorithm. Random forests produce promising results even with variable inputs (Fig. 3).

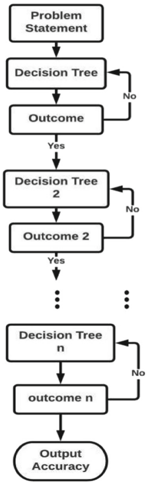

Fig. 3. Flowchart for random forest

Moreover, since we are using multiple decision trees, Random forests eliminate the data overfitting. Even a large proportion of the data is missing; Random forest could maintain the accuracy. It is very efficient in regression tasks.

Algorithm

1. From the training set, randomly select k data points
2. By using the selected data points, build a decision tree
3. Choose N number of decision trees.
4. Repeat this process
5. For new input, consider the mode of the output of the decision trees as the output of the random forest.

D. Linear Discriminant Analysis

Linear Discriminant Analysis is a type of supervised machine learning algorithm. To overcome the problem of logistic regression like two class problems and unstability between classes, linear discriminant analysis is used. LDA method uses go to linear method for multiclass classification problems. It is a dimensionality reduction technique. LDA is used as a preprocessing step for pattern classification and other application oriented scenario (Fig. 4).

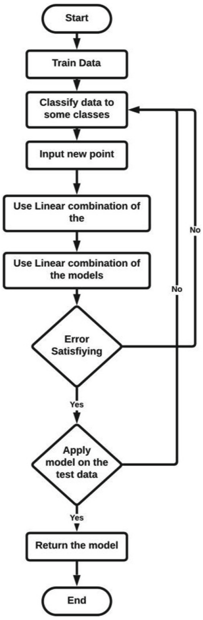
Start
Train Data
Classify data to some classes
Input new point
Use Linear combination of the
Use Linear combination of the models
Error Satisfiying
No
No
Yes
Apply model on the test data
Yes
Return the model
End

Fig. 4. Flowchart for LDA

To project input data (having N dimensions) into a smaller subspace k (where $k <= (n - 1)$) while maintaining the discriminatory information of every class.

Algorithm
Let's take a 2-D dataset

$$\begin{aligned}\text{Class C1} = \text{X1} = (\text{X1}, \text{X2}) &= \{(\text{xi}, \text{yi}), (\text{xj}, \text{yj}) \ldots\} \\ \text{Class C2} = \text{X2} = (\text{X1}, \text{X2}) &== \{(\text{xi}, \text{yi}), (\text{xj}, \text{yj}) \ldots\}\end{aligned}$$

1. Compute within class scatter matrix (captures how data is scattered within class)

Sw = S1 + S2 (To compute within class scatter matrix which is given by covariance matrix of each class)S1 is the covariant matrix of class C1 and S2 is the covariant matrix of class C2. Covariance matrix of each class is$S1 = \sum (x - \mu 1)(x - \mu 1)^T$ where $\mu 1$ is the mean of class C1. $S2 = \sum (x - \mu 2)(x - \mu 2)^T$ where $\mu 2$ is the mean of class C2. For each x, wewill have several independent matrices, here we will add all the individual.

2. Compute between class scatter matrix (SB)$SB = (\mu 1 - \mu 2)(\mu 1 - \mu 2)T$
3. Find the best LDA projection vector. Similar to PCA we find this using eigen vectors having largest Eigen value.
   $Sw^{-1}SB\ V = \lambda V$

$$\left|Sw^{-1}SB - \lambda I\right| = 0$$

Projection vectors is nothing but eigen vectors. Eigen vectors carries the best balance between all thefeatures. We need to find the highest Eigen value and corresponding Eigen vector.

$$[V1V2] = Sw^{-1}(\mu 1 - \mu 2)$$

$Y = W^T X$ where W is the projection vector and X is the input data samples. Thus dimensionality is reduced and discrimination between classes is also reduced.

4 Experimental Evaluation

4.1 Datasets and Exploratory Analysis

There are very few sufficient datasets of genuine and fake reviews are available. We have extracted the dataset from an open repository in GitHub. The dataset contains a total number of 21000 reviews, of which 50% are fake reviews and the other 50% genuine reviews.

Data Cleaning

1. All the unnecessary punctuations and symbols are removed.
2. Removed all Html tags
3. Converted all characters to lowercase
4. Expanded all the contractions
5. Lemmatization: in this stage, all word tokens are converted to root words.

Since the Machine cannot decode the text form, it needs to be converted into numerical or vectorized form. For this, we performed.

Term Frequency-Inverse Document Frequency

In this stage, each word is the statistical measure resembling the impact of the word is calculated. The weight of each word is calculated using the following equation.

$$Q_{i,j} = tf_{ij} * log(\frac{n}{df_i}) \tag{2}$$

where,

$Q_{i,j}$ is the weight for word i in the document j
n is the number of documents
tf_{ij} is the term frequency of term i in document *j*
df_i is the document frequency of term i in the collection.

Count Vectorizer

In this method, every word I'd considered a feature and a matrix is constructed, which contains the count of each word.

From these exploratory analysis, the frequency of the common words in fake reviews is displayed in the Fig. 5.

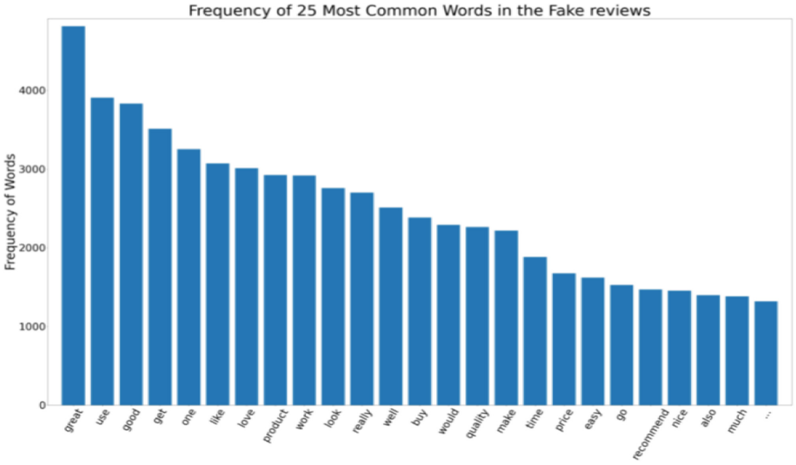

Fig. 5.

4.2 Experimental Setup

For the experimental analysis, we have used an Intel I7 10th Gen processor with a turbo speed of 4.9 GHz. This setup consisted of 2GB Nvidia Geforce MX 350 dedicated graphic memory for the fast processing speed. The model utilized TensorFlow runtime version 1.6 as a back-end engine for training and testing. The analysis used Tensorflow runtime version 1.6 backend engine for training the models as well as testing.

4.3 Evaluation Metrics

The following evaluation measures were utilized to evaluate the performance of our proposed models.

Accuracy: The measure of the number of correctly classified fake reviews to the total number of messages.

Precision: The ratio of correctly classified fake reviews to total messages classified as fake by the algorithm.

Recall: The proportion of fake reviews predicted as fake.

F1-Score: It is determined as the harmonic average of precision and recall. It is calculated with the equation.

$$\text{F-Measure} = \frac{(2 * Precision * Recall)}{Precision + Recall} \tag{3}$$

4.4 Experimental Results

To measure the performance of each machine learning model proposed, we have used various performance metrics such as accuracy, precision, recall, and f1-score. The highest value is selected and referred to a conclusion to the problem. These values for Naive Bayes, SVM, Random Forest, and Linear Discriminant Analysis are displayed in Table 1. All the algorithms produced nearly similar accuracies. However, the best algorithm is selected by considering all the performance parameters. So, it can be concluded that Naive Bayes ranks highest among the remaining algorithms in overall performance. However, Latent Discriminant analysis performs well in terms of precision. SVM uses kernel logic to solve regression and classification problems. Hence, we conclude from the sheer mathematical analysis that Gaussian Naive Bayes is the best algorithm for detecting fake reviews. The confusion matrix for the Naive Bayes model is observed in Fig. 6.

Table 1.

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	0.785800	0.760035	0.851513	0.798851
Naive Bayes	0.791991	0.759314	0.871609	0.808429
Random Forest	0.758085	0.735908	0.821700	0.772868
Linear Discriminent Analysis	0.784419	0.760078	0.847227	0.796921

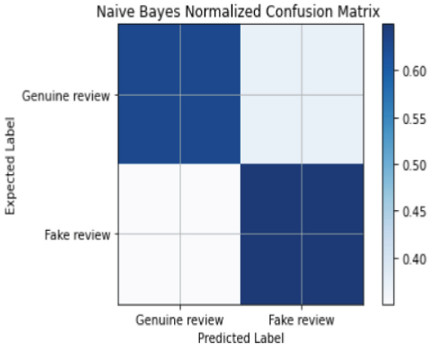

Fig. 6.

5 Conclusion

When customers shop online, the main drawback is they don't get to feel the product physically, so the reviews provided by the people who alreaady used the product makes decisive role for the buyers. Unfortunately, some of the reviews are falsly crafted and posted by few adverseries. So, it is important to detect and remove these fake reviews inorder to maintain the reliability of the reviews. In this paper, we have proposed four machine learning models namely support vector machine, Naive Bayes, Random Forest and Linear Discriminent Analysis. We have used an open Amazon review dataset from github. It consisted of 21000 reviews out of which half are real and other half are fake reviews. After detailed analysis of each algorithm and comparing four algorithms (as shown in the Table 1), it is found that Naive Bayes algorithm outperformed all other models used in the paper. As Naive Bayes classification is based on relative probablity estimates, it is able to perform very well.

References

1. Abri, F., Gutierrez, L.F., Namin, A.S., Jones, K.S., Sears, D.R.W.: Fake reviews detection through analysis of linguistic features. arXiv:2010.04260 [cs] (2020)
2. Ahsan, M.N.I., Nahian, T., Kafi, A.A., Hossain, M.d.I., Shah, F.M.: An ensemble approach to detect review spam using hybrid machine learning technique. In: 2016 19th International Conference on Computer and Information Technology (ICCIT) (2016). https://doi.org/10.1109/iccitechn.2016.7860229
3. Bali, A.P.S., Fernandes, M., Choubey, S., Goel, M.: Comparative performance of machine learning algorithms for fake news detection. In: Singh, M., Gupta, P.K., Tyagi, V., Flusser, J., Ören, T., Kashyap, R. (eds.) ICACDS 2019. CCIS, vol. 1046, pp. 420–430. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-9942-8_40
4. Elmogy, A.M., Tariq, U., Mohammed, A., Ibrahim, A.: Fake reviews detection using supervised machine learning. Int. J. Adv. Comput. Sci. Appl. **12**(1) (2021). https://doi.org/10.14569/ijacsa.2021.0120169
5. Gutierrez-Espinoza, L., Abri, F., Namin, A.S., Jones, K.S., Sears, D.R.W.: Fake reviews detection through ensemble learning. arXiv:2006.07912 [cs] (2020)
6. Liu, W., He, J., Han, S., Cai, F., Yang, Z., Zhu, N.: A method for the detection of fake reviews based on temporal features of reviews and comments. IEEE Eng. Manag. Rev. **47**(4), 67–79 (2019). https://doi.org/10.1109/EMR.2019.2928964
7. Manzoor, S.I., Singla, J., Nikita: Fake news detection using machine learning approaches: a systematic review. IEEE Xplore (2019). https://ieeexplore.ieee.org/document/8862770. Accessed 13 May 2020
8. Radovanovic, D., Krstajic, B.: Review spam detection using machine learning. In: 2018 23rd International Scientific-Professional Conference on Information Technology (IT) (2018). https://doi.org/10.1109/spit.2018.8350457
9. Rodrigues, J.C., Rodrigues, J.T., Gonsalves, V.L.K., Naik, A.U., Shetgaonkar, P., Aswale, S.: Machine & deep learning techniques for detection of fake reviews: a survey. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (2020). https://doi.org/10.1109/ic-etite47903.2020.063
10. Sihombing, A., Fong, A.C.M.: Fake review detection on yelp dataset using classification techniques in machine learning. In: 2019 International Conference on contemporary Computing and Informatics (IC3I) (2019). https://doi.org/10.1109/ic3i46837.2019.9055644
11. Zhang, Z., Hou, R., Yang, J.: Detection of social network spam based on improved extreme learning machine. IEEE Access **8**, 112003–112014 (2020). https://doi.org/10.1109/access.2020.3002940
12. Sihombing, A., Fong, A.C.M.: Fake review detection on yelp dataset using classification techniques in machine learning. In: 2019 International Conference on contemporary Computing and Informatics (IC3I) (2019).https://doi.org/10.1109/ic3i46837.2019.9055644

13. Soni, J.: Effective machine learning approach to detect groups of fake reviewers. In: ICDATA (2018)