International Journal of Scientific Research in Engineering and Management (IJSREM)

Volume: 04 Issue: 04 | April -2020 ISSN: 2582-3930

An approach to detect fake reviews based on Extreme Gradient Boosting using Review-centric features

Kaushik Daiv[1], Mrunal Lachake[2], Prathamesh Jagtap[3] ,

Srishti Dhariwal[4], Dr. Siddhivinayak Kulkarni[5]

[1]Kaushik Daiv Computer Science & MIT College of Engineering, Pune
[2]Mrunal Lachake Computer Science & MIT College of Engineering, Pune
[3]Prathamesh Jagtap Information Technology & MIT College of Engineering, Pune
[4]Srishti Dhariwal Computer Science & MIT College of Engineering, Pune
[5]Dr. Siddhivinayak Kulkarni Computer Science & MIT College of Engineering, Pune

---

Abstract - The impact of reviews on any e-commerce site is of great importance, as it can be the base for a buyer's decision to buy any product. Buyer tries to evaluate the authenticity and quality of the product using the feedback given by the previous purchasers in the form of review. But sellers, taking advantage of this, are posting reviews in an attempt to promote or defame a product in particular. Such reviews which are not a genuine opinion of an individual are termed as fake reviews. The existence of such fake reviews makes the buyer unable to make the right judgments on sellers as well as products, causing the credibility of the platform to downgrade. Thus, in this paper, we propose a method to detect fake reviews employing a very popular method called eXtreme Gradient Boosting (XGBoost) using review-centric features. XGBoost with CountVectorizer feature extraction technique achieves 83% accuracy whereas XGBoost with Term frequency-inverse document frequency (Tf-idf) technique achieves up to 82% accuracy. This study will provide an insight on the XGBoost classifier using review-centric features for identifying fake reviews to the future researchers of this domain.

Key Words: CountVectorizer, Ensemble learning, Fake review, Feature Extraction, XGBoost.

1.INTRODUCTION

In this era of the Internet, people can easily share their views about products and services using e-commerce sites, forums, and blogs. Consumers tend to refer to the reviews of other consumers for making purchase decisions of products from e-commerce websites. These reviews are helpful for potential customers and vendors too. Vendors are capable of designing their marketing strategies based on the consumers' reviews. For example, if various consumers buy a specific model of a laptop and write reviews regarding issues concerning its screen resolution or processor speed, then vendors might contact the associated manufacturers and make them aware of these issues and resolve them in order to increase customer satisfaction towards their products or services.

Targeted products or services may be promoted or degraded by mischievous users by writing either applaudable reviews or derogatory reviews. Therefore, the integrity of such reviews is questionable [3]. Such reviews are known as fake reviews. Recently the media news from the New York Times and BBC has stated that "Nowadays, fake reviews are very frequent on websites, and recently a photography company was exposed to thousands of customer fake reviews". Also it has been reported that 88% of consumers trust online reviews as much as personal recommendations [8]. Hence, detecting fake reviews appears to be a key area, and without solving this important issue, online sites may become a place full of lies, almost rendering the e-commerce business useless. To counter this issue, some researchers have already made some progress in detecting fake reviews that we discuss in Sec. 2. However, most of the previous research was done on hotel reviews and there are a lot of labeled datasets available for hotel reviews including Yelp. But there is still a room for improvement in the detection of product reviews and hence, in this paper we are focussing on product reviews provided by Amazon using supervised learning techniques as supervised models tend to give more accurate results than unsupervised or semi-supervised models [10].

This paper proposes a method for detecting fake reviews for Amazon products using an ensemble learning technique called XGboost. An ensemble, in the context of machine learning, can be broadly defined as a machine learning system that is constructed with a set of individual models working in parallel and whose outputs are combined with a decision fusion strategy to produce a single answer for a given problem [6]. The idea of ensemble learning is principally based on the theory foundation stone that the generalization ability of an ensemble is usually much stronger than that of a single learner [6]. We elected to employ such a technique of ensemble learning for the detection of fake reviews for its capacity to boost the weak learners into a strong one. There are three types of ensemble techniques viz. bagging, boosting, and stacking. XGBoost is an end to end tree boosting algorithm which was proposed by the authors of [2]. They proposed it as a novel sparsity-aware algorithm for sparse data and weighted quantile like a sketch for approximate tree learning. In data scientist's words, XGBoost is a decision-tree-based ensemble

Machine Learning algorithm that uses a gradient boosting framework. Recently, XGBoost has become more popular because of winning Kaggle competitions and outperforming classifiers and neural network models in some cases. Thus, studies by Chen et al.[2] and Ahmed et al.[1], and Kaggle winning solutions give us an overview that XGBoost can be explored for fake review classification. In this paper, we have not discussed decision trees and random forest as our main focus is on XGBoost, if readers want to explore those concepts, they can refer to ref.[17].

XGBoost and ensemble learning lay a foundation for our method for detection of genuine reviews with the capability to adapt the versatility of the review content using review-centric features. To the best of our knowledge, there is only one previous research proposed by Ahmed et al.[1] employing XGBoost in the field of fake review detection which considers only review text for training the model. Our study is unique as we have extended their research by taking into account other review-centric features which we will discuss in Sec. 5.1. The two main contributions of this study are as follows: (i) Provide researchers and practitioners with insight and further improvement prospects on the fake review detection problem based on XGBoost using review-centric features. (ii) Present the effect of using the "verified purchase" feature (description in detail in Sec. 5.1) for identifying the fake reviews.

The remaining paper is divided into the following sections: Section 2 discusses the related work to fake review detection, XGBoost, ensemble learning, and feature selection. Section 3 states the problem definition. Section 4 gives an insight of the dataset used. Section 5 describes the detailed methodology and feature extraction techniques used in our study. Section 6 discusses the experiments. Section 7 illustrates the results of the model. Section 8 involves discussion of the results and performance of the model. Finally, in Section 9 we present our final conclusions and Section 10 suggests possible areas for future work.

2. Related Work

In recent years, researchers have studied a lot about fake reviews and their detection techniques involving machine learning paradigms. The research shows that detecting fake reviews is not only limited to content of review, but it is also evident that reviewer's behaviour also plays a big role in detecting the fakeness of the review.

While there are various machine learning techniques viz. supervised learning, semi-supervised learning, unsupervised learning techniques; in [10], researchers used supervised machine learning algorithms - Logistic Regression, K-Nearest Neighbor and Naive Bayes classifiers on Yelp dataset. Their research proposed topic features (latent dirichlet allocation, average topic probabilistic, big topic probabilistic), readability features (automated readability index, Coleman-Liau index), review content and n-gram features along with some behavioural features achieving an overall accuracy of 97.2% for logistic regression.

A study by Faliang Huang et al.[6] stated Ensemble learning is a powerful machine learning paradigm that has exhibited applications. An ensemble in the context of machine learning can be broadly defined as a machine learning system that is constructed with a set of individual models working in parallel and whose outputs are combined with a decision fusion strategy to produce a single answer for a given problem [6]. Their study states that the idea of ensemble learning is principally based on the theory foundation stone that the generalization ability of an ensemble is usually much stronger than that of a single learner. This application proves helpful to boost weak learners into a strong one.

While the ensemble learning did show a new way to approach this detection problem, Brian Heredia et al. [4] suggested a way to improve ensemble learning by employing ensemble learning with feature selection techniques. They employed three of these techniques: Select-Boost, Select-Bagging, and Random Forest. The study shows that applying a combination of ensemble and feature selection (Select-Boost) shows significant improvement in performance when compared to using solely Multinomial Naive Bayes Classifier (MNB). The results indicate the combination of Select-Boost, MNB, and Chi-Squared (or signal-to-noise) to be the best performing model, significantly outperforming all other methods, with the exception of RF500.

Furthermore, A study by Sifat Ahmed et al. [1] stated that traditional machine learning techniques do not have a significant impact when it comes down to accurately detecting the real-life fake reviews. So the study employs boosting algorithms [6] such as XGBoost [2], Adaptive Boosting (AdaBoost) along with some feature selection techniques like Tf-idf and Chi2 achieving the highest accuracy of 95% with XGBoost and Chi2. In their research, they labelled their unlabelled dataset by manually labelling few reviews and then using active learning, which somehow loosened the reliability. As, the study by Sandifer et al. [11] reported the accuracy of human judgement was about 61.9%, which favoured the experiments of our predecessors. Taking this into consideration, we went through a technical approach and used a labelled dataset for our classification. We tried to pick up the best ideas out of these above-mentioned works and get them together to create an even more efficient and capable detection technique.

3. Problem Definition

To design and build an efficient and robust fake review classification model for e-commerce product reviews using the eXtreme Gradient Boosting algorithm (XGBoost) which follows the idea of gradient boosting for classifying fake and genuine reviews. Also provide a consensus strategy for feature extraction and text preprocessing.

4. Dataset

The dataset used in this paper is "amazon_reviews" which is openly available on Kaggle. The dataset contains Amazon product reviews which are labelled as __label1__ and __label2__ as fake review and genuine review respectively. The dataset contains a total of 21,000 uniformly distributed reviews in which 50% are fake reviews and 50% are genuine reviews. The dataset contains different columns viz. rating, verified purchase, product category, product id, product title along with review content and its label. You can find the full dataset by the link provided in ref. [18].

5. Proposed Method

This section elaborates the proposed methodology for the above-stated problem definition in Sec. 3. In this paper, we propose a method to detect fake reviews by employing XGboost using Tf-idf and CountVectorizer feature extraction techniques by using review-centric features on a labelled Amazon product reviews dataset available on Kaggle. Also, our research states the importance of the "verified purchase" feature for fake review classification. We divide our methodology in six steps as follows:

1. Collect the data of labelled reviews.
2. Preprocessing the dataset.
3. Feature Extraction.
4. Training the XGBoost random forest model.
5. Tuning the hyperparameters of the model.
6. Evaluating the model on test data.

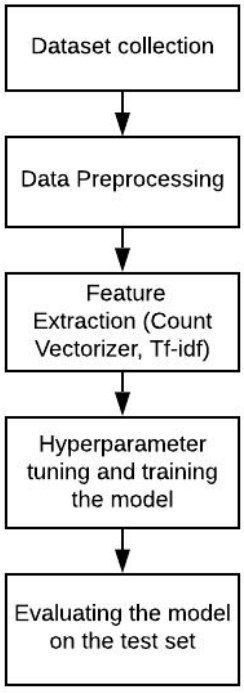

Fig. 1: Block Diagram of Proposed Model

First step is to collect the labelled dataset of reviews. There are multiple datasets available online which are used in previous research, however finding a labelled dataset for reviews was a difficult task. Fortunately, we were able to find a labelled dataset on Kaggle provided by Amazon. The dataset contains a total of 21,000 reviews in which 50% are fake reviews and 50% are genuine reviews.

After finding the dataset, the next step involves preprocessing the data. The dataset cannot be directly used to train the classifier model as the model cannot handle the text data. Preprocessing includes removing stop words and punctuations, stemming, lemmatization, etc. Preprocessing is discussed in the Sec. 5.1 in detail.

In the feature extraction phase, Tf-idf and CountVectorizer were used. Tf-idf increases the weight of uncommon words and decreases the weight of common words and results in creating a vector of features [1]. CountVectorizer creates a vector of features considering the frequency of the words in the review. Our research shows that CountVectorizer outperforms Term frequency-Inverse Document Frequency (Tf-idf) which may be because the dataset has reviews on several distinct product categories resulting in many uncommon words.

Followed by the feature extraction phase, we trained our XGBoost Random Forest classifier model on the set of features extracted in the feature extraction phase. Random Forest is an ensemble learning method which constructs a number of decision trees and ensembles them for classification. In our research, we used an ensemble of 150 decision trees.

Training the model is just not sufficient to classify the reviews successfully. For better accuracy we need to tune the hyperparameters of the model viz. n_estimators, criterion, max_depth, learning_rate, etc. Sec. 4 describes the process of hyperparameter tuning in detail.

After the model is trained and tuned, the model needs to be evaluated to understand its performance. Evaluation of the model is done by testing it on unlabelled test data and calculating the accuracy, precision and recall. Results shall be discussed in Sec. 7.

5.1 Feature Extraction and Data Preprocessing

Variety of features that have been proposed and used separately by supervised approaches to identify fake reviews in previous research are review - centric and reviewer - centric features. In some cases review - centric features are considered separately. In other cases, reviewer - centric features are taken into account. Our study employs review - centric features with XGBoost. To the best of our knowledge, this is the first study to propose the use of review - centric features with XGBoost classifier.

From the previous research proposed in [10] and [16], we tried to pick the best features which would help identify the fake reviews. We have considered the following features:

1. **Rating**

   Users rate the product from 1 to 5 representing satisfaction/dissatisfaction about the product. This feature can be used to validate that the review written and the ratings given by the reviewer are intended only in one direction and do not contradict. Also, ratings to the fake reviews usually deviate from the average rating of the product [10]. Thus, helping in classifying the fake reviews.

2. **Verified Purchase**

The verified purchase feature means Amazon has verified that the person writing the review has purchased the product at Amazon and didn't receive the product at deep discount. This feature helps to consider those reviews which are genuine as we get to know which purchaser has actually bought the product and used it. This is the first study on fake review classification which states the effect of using the "verified purchase" feature for the classification.

3. **Review length**

The length of the review is also considered for training the model as the previous research by Xinyue Wang et al. [10] suggests that reviews written by spammers are very short and are intended to defame / promote the product.

Along with the above three review-centric features, review content is also considered for classification. The review text needs to be processed before passing to the model. The first step of preprocessing includes removal of all the characters and expressions other than letters as the XGBoost model cannot make any sense of the punctuations and expressions. The review text is tokenized into a list of words and then each word is converted to its base form by stemming; followed by removal of stopwords. Stopwords are the frequently occurring words useful syntactically and grammatically that do not add any value to the model. A corpus of these words is generated. The CountVectorizer class provided by the "sklearn" library in Python language is used to represent the corpus of words using a sparse matrix where each word acts as a column and the review as a row having the most frequent 750 words from the corpus. This sparse matrix of 750 most frequent words are used as a feature vector to the model along with the verified purchase, rating and review length of the product. Similarly, we also create a feature vector of 750 words using TfidfVectorizer class.

6. Experiments

This section describes the experiments in our research. Dataset cannot be passed directly for training. At first, relevant feature columns were extracted (rating, verified purchase, review length). Then, categorical features were converted to numerical features using "LabelEncoder". Text data was extracted using CountVectorizer and Tf-idf (max_features = 750). This was followed by calculating length for each review. Finally, our data having a feature vector of 753 (750 features of review content, rating, verified purchase, review length) was generated. Train test split of 80-20% was carried out on the data which resulted in 16800 samples as training data and 4200 samples as test data.

The XGBoost model of Random Forest classification (XGBRFClassifier) provided by "xgboost" library was trained on the train data for both CountVectorizer and Tf-idf. Followed by this, hyperparameters of the model were tuned using GridSearch and ValidationCurve. Following is the list of optimal hyperparameters that result in best accuracy: n_estimators = 150, criterion = 'gini', min_samples_leaf = 3, min_samples_split = 5, max_depth = 15, learning_rate = 0.01, gamma = 0.4, min_child_weight = 0.5, colsample_bytree = 0.3, colsample_bylevel = 0.2.

Also, to examine the impact of using "verified purchase" as a feature for fake review classification, the XGBoost with CountVectorizer with the same hyperparameters was trained without using "verified purchase" as a feature and the model was evaluated for the performance.

The last phase of the experiment involves evaluation of the model to understand the performance. Binary classification involves classifying the data in groups of two - yes / no, true / false, fake / genuine, etc. Target variables in such problems are not continuous but they predict the probabilities to be yes / no. Such models are evaluated using a metric called Confusion Matrix. Using confusion matrix, we have calculated accuracy, precision and recall for both XGBoost with CountVectorizer and XGBoost with Tf-idf and XGBoost with "verified purchase" and without "verified purchase".

7. Results

Classifiers	Accuracy	Precision	Recall
XGBoost with Tf-idf	0.823	0.775	0.856
XGBoost with CountVectorizer	**0.830**	**0.768**	**0.868**

Table 1: Performance of XGBoost with Tf-idf and CountVectorizer.

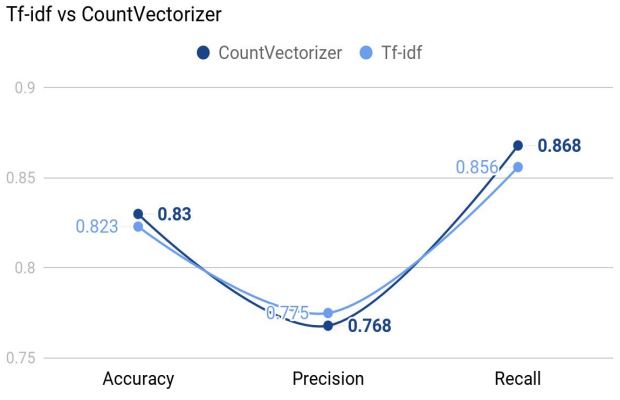

Fig. 2: Accuracy, precision, recall comparison - Tf-idf vs CountVectorizer.

Classifiers	Accuracy	Precision	Recall
XGBoost without "verified purchase"	0.653	0.698	0.639
XGBoost with "verified purchase"	**0.830**	**0.768**	**0.868**

Table 2: Performance of XGBoost using CountVectorizer with and without "verified purchase" feature.

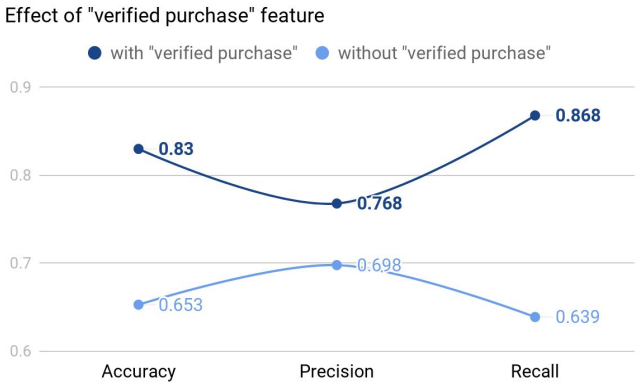

Fig. 3: Accuracy, precision and recall comparison - with "verified purchase" vs without "verified purchase"

8. Discussion

Our results from Table 1 show that XGBoost classifier with Tf-idf and CountVectorizer using review-centric features performs extremely well on our dataset by achieving an accuracy of 82% and 83% respectively. 1% of test data is around 400 reviews. Thus, it is evident that CountVectorizer classifies around 400 reviews more correctly than Tf-idf. Though, accuracy of CountVectorizer feature extraction technique outperforms Tf-idf by a margin of 1%; precision of Tf-idf feature exceeds CountVectorizer by the same margin. Table 1 and Fig. 3 show the accuracy, precision and recall of Tf-idf and CountVectorizer.

We also observe that, "verified purchase" feature has a significant impact on fake reviews classification. Fig. 4 shows the accuracy of XGBoost with CountVectorizer using "verified purchase" feature is 83% whereas when the feature is not used for classification, the same model achieves an accuracy of 65%. This result suggests that the "verified purchase" feature is very effective for identifying fake reviews. Table 2 and Fig. 4 show the accuracy, precision and recall of XGBoost using CountVectorizer with "verified purchase" and without "verified purchase".

As in this field many researchers have already tried out several algorithms which proved to be noteworthy, since the lack of a golden standard dataset there is no model yet to give a solid foundation to this field. This issue was handled by Ahmed et al. [1] by labelling an unlabelled dataset by manual labelling and active learning approach which achieved an overall accuracy of 95% with XGBoost and Chi2. However, manual labelling questions the authenticity of the achieved results as accuracy of human judgement is about 61.9%, which may have favoured the experiments of their results [11]. Thus, we extend their research using a labelled dataset for our classification and apply purely algorithmic computation and processing. Finally, from the above discussion, we recommend using the XGBoost algorithm with CountVectorizer in the future for classification of fake reviews, since it is able to achieve a good accuracy. Also, it is interesting to note the impact of the "verified purchase" feature in the domain of fake review classification.

9. CONCLUSIONS

In this paper, we have discussed the impact of XGBoost model for identifying the fake reviews using review-centric features. Using XGBoost algorithm for fake review detection is a new and very popular approach and there is a lot to explore in this branch. Along with review content, we have provided a set of review-centric features for classification of the fake reviews. One of the review-centric features we propose in this paper is "verified purchase". Our research shows that using "verified purchase" as a feature for classifying fake reviews has an outstanding effect. In addition to this, we have proposed two feature extraction techniques viz. Tf-idf and CountVectorizer and therefore, conclude that implementing XGBoost with CountVectorizer on the used dataset has achieved an accuracy of 83% and an accuracy of 82% with Tf-idf.

The work proposed in this paper acts as a platform for further research in XGBoost in fake review detection. This study might be helpful for future researchers who want to improve the fake review detection system using XGBoost. The use of "verified purchase" feature for the classification is a prominent contribution of our research to the domain of fake review classification.

10. FUTURE SCOPE

XGBoost algorithm being new and popular for fake review detection, there is a lot of scope to explore in this branch. This paper has presented the effect of feature "verified purchase" on the result, so further research can consider this feature for classification. This paper focuses only on review-centric features, future researchers can work on reviewer-centric features with Xgboost. Additionally, future work may involve testing this process on other data sets to see if results generalize. In future, we try to increase the accuracy of XGBoost by exploring other effective features and parameters and try to build a better model.

REFERENCES

1. Ahmed, Sifat, and Faisal Muhammad. "Using Boosting Approaches to Detect Spam Reviews." In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1-6. IEEE, 2019.
2. Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd ACM SIGKDD

International Conference on Knowledge Discovery and Data Mining, pp. 785-794. 2016.
3. Khurshid, Faisal, Yan Zhu, Zhuang Xu, Mushtaq Ahmad, and Muqeet Ahmad. "Enactment of Ensemble Learning for Review Spam Detection on Selected Features." International Journal of Computational Intelligence Systems 12, no. 1 (2018): pp. 387-394.
4. Heredia, Brian, Taghi M. Khoshgoftaar, Joseph D. Prusa, and Michael Crawford. "Improving detection of untrustworthy online reviews using ensemble learners combined with feature selection." Social Network Analysis and Mining 7, no. 1 (2017): pp. 37.
5. Ahsan, MN Istiaq, Tamzid Nahian, Abdullah All Kafi, Md Ismail Hossain, and Faisal Muhammad Shah. "An ensemble approach to detect review spam using hybrid machine learning technique." In 2016 19th International Conference on Computer and Information Technology (ICCIT), pp. 388-394. IEEE, 2016.
6. Huang, Faliang, Guoqing Xie, and Ruliang Xiao. "Research on ensemble learning." In 2009 International Conference on Artificial Intelligence and Computational Intelligence, vol. 3, pp. 249-252. IEEE, 2009.
7. Vidanagama, Dushyanthi U., Thushari P. Silva, and Asoka S. Karunananda. "Deceptive consumer review detection: a survey." Artificial Intelligence Review 53.2 (2020): pp. 1323-1352.
8. Dematis, Ioannis, Eirini Karapistoli, and Athena Vakali. "Fake review detection via exploitation of spam indicators and reviewer behavior characteristics." In International Conference on Current Trends in Theory and Practice of Informatics, pp. 581-595. Edizioni della Normale, Cham, 2018.
9. Fontanarava, Julien, 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), pp. 658-666. IEEE, 2017.
10. Wang, Xinyue, Xianguo Zhang, Chengzhi Jiang, and Haihang Liu. "Identification of fake reviews using semantic and behavioral features." In 2018 4th International Conference on Information Management (ICIM), pp. 92-97. IEEE, 2018.
11. Sandifer, Anna V., Casey Wilson, and Aspen Olmsted. "Detection of fake online hotel reviews." 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST). IEEE, 2017.
12. Jadhav, Amitkumar B., Vijay U. Rathod, and Hemantkumar B. Jadhav. "Improving Performance of Fake Reviews Detection in Online Review's using Semi-Supervised Learning." (2019).
13. Day, Min-Yuh, et al. "Exploring Review Spammers by Review Similarity: A Case of Fake Review in Taiwan." Proceedings of the third international conference on electronics and software science (ICESS2017), 2017, pp. 166.
14. X. Wu, Y. Dong, J. Tao, C. Huang and N. V. Chawla, "Reliable fake review detection via modeling temporal and behavioral patterns," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 494-499.
15. S. P. Rajamohana, K. Umamaheswari, M. Dharani and R. Vedackshya, "A survey on online review SPAM detection techniques," 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), Coimbatore, 2017, pp. 1-5.
16. H. Deng et al., "Semi-Supervised Learning Based Fake Review Detection," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), Guangzhou, 2017, pp. 1278-1280.
17. Ali, Jehad, et al. "Random forests and decision trees." International Journal of Computer Science Issues (IJCSI) Vol 9 Issue 5 No 3 (2012): 272
18. Amazon labelled dataset for fake product reviews used in this paper is available on: https://www.kaggle.com/lievgarcia/amazon-reviews.