An Approach to Improve the Accuracy of Detecting Spam in Online Reviews

Abida Khanam Suborna, Suman Saha, Chironjit Roy,
Md. Tojammal Haque Siddique and Shuvrodeb Sarkar

December 7, 2020

An Approach to Improve the Accuracy of Detecting Spam in Online Reviews

Abstract— **Customers or user opinion is most important and valuable information at online now-a-days especially in product reviews. Mostly customer used to make their decision for purchasing a particular product according to the other's customer reviews. Those reviews are increasing the rating of that e-commerce site. Normally reviews are considered unbiased opinion of a person whose have personal experience with a related specific product. The noticeable thing is many reviews are not real or authentic. These kinds of reviews are usually called spam and it is becoming a large problem an online and others electronic communication. For the increasing of the value of online review, the spammers are getting inspire to doing spam for them or promoting a specific e-commerce website. Also, they are demoting a specific site for payment. In this paper we have discussed about some traditional techniques for detecting spam in online public opinions. Next, we have used stacking algorithm with some traditional classifiers for the detection of spam reviews. Finally, performance of different classifiers has been evaluated through simulation experiment. From the experiment we have seen that stacking classifier provides better accuracy than other traditional classifiers.**

Keywords- ***Detecting Spam, Online Reviews, Spam in Reviews, Stacking Classifier.***

I. INTRODUCTION

A requirement for a more advanced online shopping system is increasing day by day. Many people prefer online shopping system now. Because this shopping system is an easier and better way of shopping. That's why the authorized companies offer many types of discounts for growing business in an online platform [1]. Also, they allowed to giving customer feedback on their platform of products & services. People can post his opinion about products and services on these sites with freedom [2]. Sharing a particular judgment about an appropriate product or a service based on their own experience is considered as reviews [3]. Reviews are divided into two types like (i) Given by customer or buyer his personal opinion and (ii) Authorized Company buys fake review for promotion purpose [4]. Many times, these reviews counted as fakes that are Bought to the authorized company for promotional purposes [5]. Another one of the great problems an opinion sharing websites is that spammers can quickly generate hype of the appropriate goods by writing spam reviews. These spam reviews may play an important role in raising the value of the goods or services [6]. A new person affected most of the time because when new people come on this online platform for shopping at first, he needs a judgment which good or bad product and helps to get the decision to see these product reviews [7]. For example, if a consumer wants to buy any product online, they normally go to the comment section to know about other buyer's feedback. If the reviews are positive most of them, then the user may buy, otherwise, they wouldn't buy that particular product [8]. That's why the customer gets confused about choosing his targeted product after watching these reviews and he thinking like it will be real or fake at all. Therefore, classifying the harmful reviews automatically becomes an urgent matter for the consumers as well as for the companies. For removing and avoiding these kinds of spam opinion, host or authority have lose their valuable time and energy. And this is another problem for an authority. Spamming not only causes for harmful activity. It's using for promoting a website. Especially for an e-commerce website. Spammers are hired for giving fake reviews of products for increasing product rating and attracting the customers. This is a noticeable thing for online marketplace. And this problem can be prevented. Many researchers are worked with online review spamming. They have proposed various kinds of technique for preventing spam in public opinion. In this paper we have presenting a several techniques based on some traditional machine learning algorithms. Here we also prefer sentiment analyzing technique for filtering customer's opinions and their intention.

The objective of this paper as follows:

- To analyzing the dataset of the customer's opinion with required features (spam, ham).
- To preprocessing the datasets for making easier for the classifier and selecting the required features from the datasets.
- To training the deferent individual classifier on the same preprocessed dataset.
- To analyzing the accuracy and performance of each classifier.
- Taking the best classifier according their performance and building the model for filtering spam in customers reviews.

The rest of paper is organized as below. Section II reviews some of the related research works. Section III shows the proposed diagram of the classification technique. Section IV describes proposed methodology used in the paper. Section V presents experimental results of evaluating different classifier alongside with stacking classifier and finally Section VI concludes the paper.

II. LITERATURE REVIEW

This section presents some previous works and researches done over spam filtering on public opinion. Ajay Rastogi and Monica Mehrotra discussed about the behavior of the online opinion in commercial marketing site [1]. The main purpose of their research is to detect spam based on different kinds of the online comment and opinion. Besides, they focused on the authentication of those opinions. That is, those are from individual Authentic or not. They expressed various form of opinion like comment, post, status, tweet, or review and data collection process to collect the literature on online opinion spam. And they conducted machine learning algorithms observing the behavior of the opinion of the spammers. Eshan et al. in [6] discussed the application of machine learning detect abusive in Bengali text. This paper introduced about the Bengali text in social network which is very emergent problem in our country. This research showed that how to apply the machine learning model for learning those problem and how to prevent those text, comments, vulgar image and videos. Also, they have showed the technique of this sentiment analysis on the Bengali text. They worked on machine learning algorithms and SVM linear kernel.

Rather than using a single classifier to determine the accuracy, we have used stacking classifier to improve the accuracy of prediction for detection of spam reviews. We have assembled deferent types of classifier techniques using a Meta classifier regard of stacking classifier. Normally, stacking classifier is the simplest form of stacking which can be describe as an ensemble learning technique where the predictions of multiple classifiers are used as new features to train a Meta-Classifier. And the Meta-Classifier can be any kind of classifier of our choice. And it increases the accuracy of the whole model. For that, model provides an actual prediction.

III. Proposed Model

Workflow of our classification technique shown below:

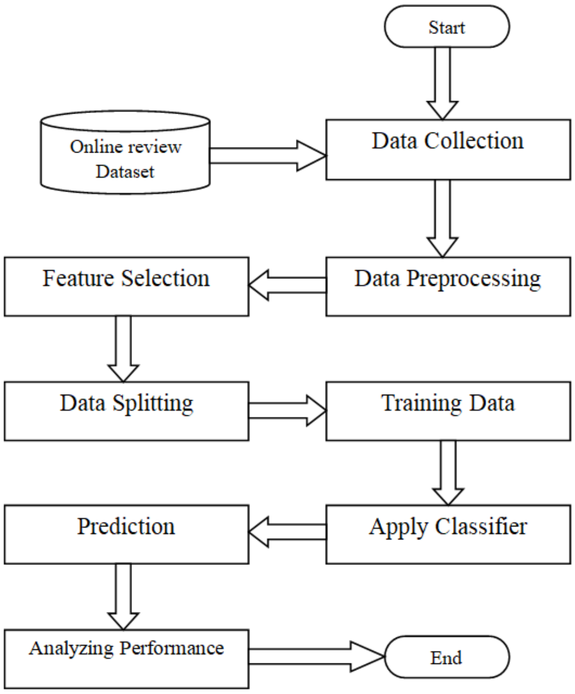


Fig. 1. Flow Diagram for Detecting Spam in Reviews

From our proposed diagram we can see, first we have collected and gathered several datasets from UCI Machine Learning Repository and Kaggle.com. We have merged those datasets according to their features. After that, we have pre-processing our dataset very efficiently for feeding our classifier without any complexity. Then we have selected our required feature from dataset which are required. Then we spitted the dataset into 2 part. 85% of the dataset for training our model, and rest of the 15% of the dataset for testing our model. After that we have feed our training dataset into the deferent types of classifier for training. Then we picked one classifier which gives the high accuracy and making an actual prediction than others.

IV. Proposed METHODOLOGY

In this part we describe the proposed methodology to filtering spam in online customer reviews.

A. About Dataset

First we take a datasets on YouTube user's review on a video and channel. In this datasets there had 5 individual datasets of different famous channels. We merged all 5 datasets. In this dataset there has 4 features and target attribute. They are CommentID, Author, Date, Content, and Class which is target attribute. Then we collected Amazon dataset on baby product reviews from kaggle.com. In this dataset there has 3 attributes. They are ProductName, Review, and Spam/Ham. In this dataset, there has 45531 values.

B. Data Pre-processing

We have pre-processed our dataset in several sections. First, we checked the missing values. If missing values exists than we fill up that at their average values of the whole dataset. Then we normalized our dataset as a method of pipeline. In the normalization section, first we converted the review text into lower case. Then removed the bigger space and bracket. After that we removed all kinds of punctuation marks from the review text. We also removed the digits and new line which is unnecessary.

C. *Feature Selection*

From the first dataset we have selected 3 independent features, they are CommentID, Date, Comment. And from the Amazon dataset, we selected all 2 features whose are ProductName and Review.

D. Proposed Classification Method

In this section we present two types of classification technique for detecting spam in customer reviews.

- Traditional Machine Learning Algorithm
- Stacking Algorithm Technique

In Traditional Machine Learning Algorithm Technique, we have used 5 classifiers. They are, Naïve Bayes Classifier, Support Vector Machine, Decision Tree, K-Nearest Neighbors and Logistic Regression. We have applied all of this classifier individually on the same pre-processed

datasets. And we have got different accuracy. Again, we have trained all the classifier with another preprocessed dataset. And compute the accuracy. After that we have compared among all of the classifier and choose the best one for our spam detecting model. As we said above that, we also work another technique which called Stacking Classifier in our proposed methodology. In the Stacking Classifier have two fundamental sections. Level 0 and level 1. In level 0, there have multiple classifiers. We provide our training dataset into them and taking an individual prediction from them. After that, those outputs feed into our Ensemble Model as input features. And this Ensemble model provides us a final prediction with high accuracy than previous classifiers.
And this Ensemble Model is our level 2 section of our Stacking Classifier. However, stacking method diagram and stacking classifier algorithm [9] are shown in figure 2 and figure 3 respectively.

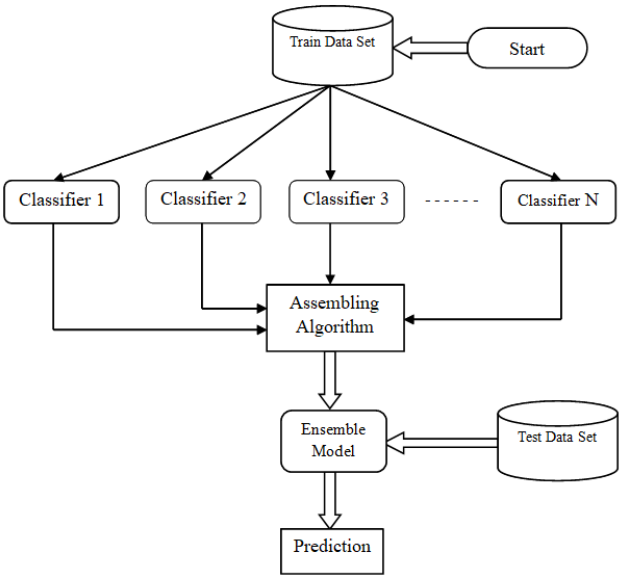
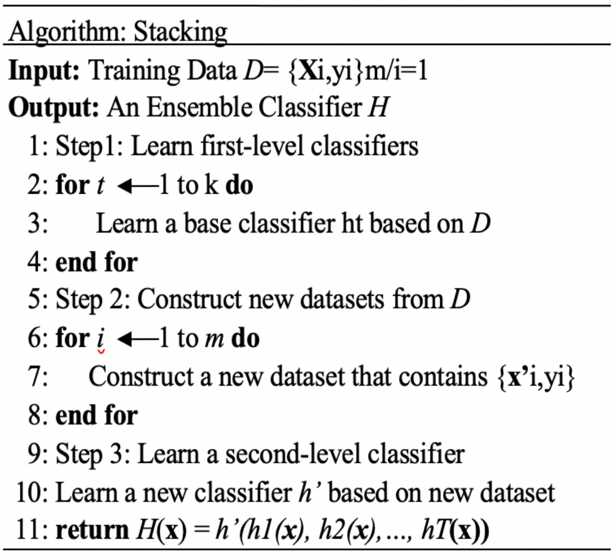


Fig. 2. Stacking Method Diagram



```
Algorithm: Stacking
Input: Training Data D= {Xi,yi}m/i=1
Output: An Ensemble Classifier H
  1: Step1: Learn first-level classifiers
  2: for t ←1 to k do
  3:     Learn a base classifier ht based on D
  4: end for
  5: Step 2: Construct new datasets from D
  6: for i ←1 to m do
  7:     Construct a new dataset that contains {x'i,yi}
  8: end for
  9: Step 3: Learn a second-level classifier
 10: Learn a new classifier h' based on new dataset
 11: return H(x) = h'(h1(x), h2(x),..., hT(x))
```

Fig. 3. Stacking Classifier Algorithm used from [9]

Initially we splitted our dataset and giving trained data into the classifier. In the first level at step 2, we are training multiple classifiers until k-times. And getting the deferent output. In step 2, we are taking a new part of the data from D dataset until m-times. And we combine all classifiers output. In step 9, we are feeding all output in second-level classifier which is ensemble model. And then we are retraining this second-level ensemble model with new constructed dataset. And that model provides us the final prediction.

E. Performance Evaluation

There are many ways to evaluating the model performance. Most importantly Confusion matrix. Confusion matrix is a representation of the above parameters in a matrix format. There are more evaluation methods are, Specificity, F1 score, Accuracy, Precision, PR curve, Recall, ROC curve. Here we have applied Precision, Accuracy, F1 score, Recall for evaluating the classifiers technique.

V. Experimental Result

In the implementation work, we applied Naïve Bayes, SVM, Decision Tree, K-Nearest Neighbor and Logistic Regression classifier using some python open-source library. We used Jupyter Notebook as an editor for implementing. Table 1 is used to show the measured performance and Table 2 shows the performance of stacking classifier alongside with others classifiers.

TABLE 1. Classifiers Performance

Classifier	Accuracy (%)	Precision (%)	F1 score (%)	Recall (%)
Naive	93	94.00	93.90	93.82
SVM	95	95.66	94.63	94.71
DT	95	96.67	95.94	95.85
KNN	85	88.39	85.41	86.50
LR	96	96.80	96.88	97.01

TABLE 2. Stacking Classifier Performance

Classifier	Accuracy (%)
K-Nearest Neighbor	91
Random Forest	93
Naïve Bayes	89
Ensemble Model	96

Now we will observe percentage of spam and non-spam reviews in dataset which predicted by stacking classifier shown in figure 4. In the figure, X axis represents the data (1 indicates spam and 0 indicates non-spam) and Y axis represents percentage of data detecting as spam or non-spam.

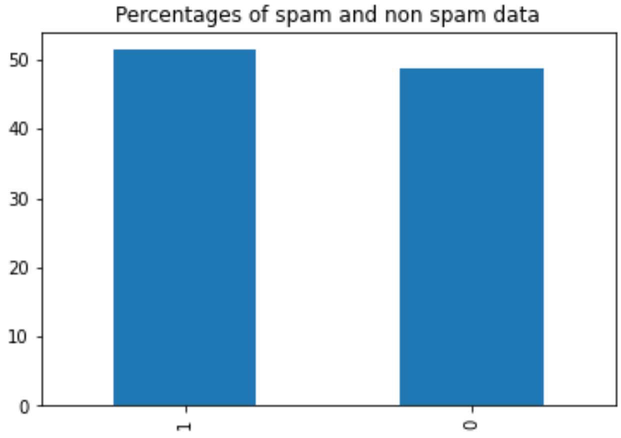


Fig. 4. Output predicted by Stacking Classifier.

VI. Conclusion

This technique of detecting spam in reviews will be very efficient and effective for e-commerce website and other's online public reviews. By training with the large number of real-life datasets, our model can be filter spam reviews with up to 97% accuracy.

REFERENCES

[1] Rastogi and M. Mehrotra. Opinion Spam Detection in Online Reviews, Journal of Information & Knowledge Management, Volume 16, Issue No. 04, 2017.

[2] S. P. Rajamohana, K. Umamaheswari, M. Dharani and R. Vedackshya. A Survey on Online Review Spam Detection Techniques, in proceeding of International Conference on Innovations in Green Energy and Healthcare Technologies, Coimbatore, India, March 15-18, 2017.

[3] L. Lota and B. Hossain. A Systematic Literature Review on SMS Spam Detection Techniques, International Journal of Information Technology and Computer Science, Volume 9, Issue No. 7, Page No. 42-50, 2017.

[4] S. Bajaj, N. Garg and S. Sing. A Novel User-based Spam Review Detection, Procedia Computer Science, Volume 122, Page No. 1009-1015, 2017.

[5] N. Jindal and B. Liu. Review Spam Detection, Proceedings of the 16th international conference on World Wide Web Conference, Banff, Alberta, Canada, May 8-12, 2007

[6] S. Eshan and M. Hasan. An application of Machine Learning to Detect Abusive Bengali Text, In proceeding of 20th International Conference of Computer and Information Technology, Dhaka, Bangladesh, December 22-24, 2017.

[7] M. Ahsan, T. Nahian, A. Kafi, M. Hossain and F. Shah. Review spam detection using active learning, In proceeding of IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference, Vancouver, BC, Canada, October 13-15, 2016.

[8] A. LI and L. SHI, Product Spam Reviews Detection based on Index Optimization, In proceeding of International Conference on Machine Learning, Big Data and Business Intelligence, Taiyuan, China, November 8-10, 2019.

[9] Tang, J., S. Alelyani, and H. Liu. Data Classification: Algorithms and Applications. Data Mining and Knowledge Discovery Series, CRC Press, page No. 498-500, 2015