Evaluation of learning model for Fake Review Classification

Dr.Yokesh Babu S [1], Aishwarya S[2], Ketan Kolte[3], Aditya Panditrao[4]

[1]Associate Professor
Department of Internet of Things, School of Computer Science and Engineering

Vellore Institute of Technology, Tamil Nadu, India.

Department of Computer Science and Engineering

VIT

Vellore – 632014, Tamil Nadu, India

E-mail: yokeshbabu.s@vit.ac.in (aishwarya.s2022b@vitstudent.ac.in

panditrao.aditya2022@vitstudent.ac.in, ketansanjay.kolte2022@vitstudent.ac.in)

Abstract

Online reviews have grown to be an essential tool for businesses to manage their online reputation. The fake reviews circulating online are causing business to reshape their businesses. In the modern world, where the Internet has become a common convenience. Reviewing has altered how marketing is done. Reviews on e-commerce websites can be valuable to both consumers and dealers. The reviewer might enhance the trustworthiness of their brand and help other customers believe that they are behind the product. To promote or condemn a certain brand, however, dealers use customer reviews. Reviews like the ones above are referred to as phone reviews because they don't reflect the author's true views. The existence of the above mentioned false reviews makes it impossible for the buyer to evaluate sellers and amounts accurately, which makes them assume that the seller has the potential to degrade. The fraudsters' schemes alert potential customers, arrange for them to modify their trades, and prevent blind faith techniques from reaching their intended goals. Therefore, it has improved by identifying fake reviews to resurrection the true device believe. Today, evaluations that are intentionally false or misleading are produced to gain online credibility and draw in potential buyers. The ability to spot fraudulent reviews depends not just on the reviews' major characteristics but also on the reviewers' actions. In order to spot fake reviews, this research suggests a machine learning approach. This aims to study on product reviews and how to identify spam false reviews among them using supervised learning methods.

Keywords: Supervised learning Technique, e-commerce websites, Fake reviews, Machine learning, Spam reviews.

I. INTRODUCTION

In this era of the web, individuals will simply share their views regarding merchandise and services exploitation e-commerce sites, forums, and blogs. Customers tend to check with the reviews of alternative customers for creating purchase choices of merchandise from e-commerce websites. These reviews are useful for potential customers and vendors too. Vendors are capable of planning their selling ways and supporting the consumers' reviews. For instance, if numerous customers get a selected model of a portable computer and write reviews concerning problems regarding its screen resolution or processor speed, then vendors may contact the associated makers and create them tuned in to these problems and resolve them so as to extend client satisfaction towards their merchandise or services.

Targeted merchandise or services could also be promoted or degraded by mischievous users by writing either worthy reviews or uncomplimentary reviews. Therefore, the integrity of such reviews is questionable. Such reviews are referred to as fake reviews. Additionally it's been said that half a mile of customers trust on-line reviews the maximum amount as personal recommendations. Hence, detective work fake reviews seems to be a key space, and while not resolving this necessary issue, on-line sites could become an area packed with lies, virtually rendering the e-commerce business useless. Most of the previous analysis was done on building reviews and there are loads of tagged datasets out there for building reviews together with Yelp. However there's still an area for improvement within the detection of product reviews and therefore, during this report we are focusing on product reviews provided by Amazon using supervised learning techniques as supervised models tend to give more accurate results.

Reviews can be defined as positive, neutral or negative reviews that are the customer's honest opinion or that reflects the customers' genuine experience of the product,

Service or business. Online reviews have played an important role in purchasing decisions in recent years. That is because these reviews can provide your customers with a lot of useful information about your products and services. However, spammers can forge and create counterfeit goods in order to artificially advertise or degrade the quality of their products or services review. Current research shows that the reliability of the online reviews is in question. For example, 20% of the reviews on Yelp websites are faked by human writers. Such behaviour by spammers can mislead customers and lead them to make the wrong decisions. Thus, our paper aims to detect such fake reviews on Amazon platform and help the customers in making the right decisions.

Python Libraries are used to process the data. The processed dataset is trained using methods from machine learning. A machine learning approach called Random Forest combines different classifiers Tackle complex. RFC uses many decision trees. It averages the results over input dataset by using subsets to improve the accuracy of the given dataset's forecast.

Separating data in categories and regression based problems decode by using learning method supervised which uses decision tree. It is a structured classifier which is similar to tree in which columns of dataset stands as a internal nodes, decision-making process can be done with the help of fork, and classification result with the help of nodes having no children.

For categorising in positive and negative words, and also consists of algorithms based on probability this series is called as NB algorithm. Words are compared using Naive Bayes theory.

The remainder of the thesis is structured as follows Section II literature survey. Section III presents the details of the model based on learning of machine. Section IV shows results of experiment. Section V contains work that can be done in future and the conclusion of the study.

II. LITERATURE SURVEY

Fake Review Detection using Classification proposes two new features for use in reviewing products: user review frequency and term frequency. Experiments show that these features are accurate in classifying fake and trust worthy reviews. Additionally, By taking the number of occurrences of terms and also characteristics based upon reviews increases the accurateness and precision by large extent [1].

Detect unfair reviews in Amazon reviews by use of analysis which is based on sentiment analysis with the help of learning technique – supervised.

This paper study's the "Weka" tool which is implemented in machine learning algorithms to classify Amazon reviews into three categories: Positive, Negative, and Neutral. In order to limit the overall performance of the system, false reviews are recognised by taking poll, voted by the consumer. Additionally, experimental results show that the existing system uses a NB classifier for false and non-false classification, with a correctness that is quite low [4].

Machine Learning for Detection of Fake News. In this study, There are a number of possible solutions to the problem of detecting fake news, but each approach is limited in that it does not demonstrate how accurate the results are[3].This study examines various techniques for recognizing opinion spam and concludes that a self-taught

Machine learning system is the most effective [12].

A classic example of content mining is opinion mining [7]. It finds text sentiment (negative and positive) through machine learning, trains a classifier to match review characteristics, and performs sentiment analysis.

A Twitter Analysis of Sentiment Using NLTK and Machine Learning Method. This aim to study the researchers used sentiment analysis to identify positive and negative tweets from users on Twitter, and then labelled them accordingly. However, the applicability of their approach to other types of data has not been fully validated [2].

Use Opinion Mining to find true and useful consumer reviews about your products. The paper discusses a study of how to find reviews that are genuine and useful for products. This was done using a supervised learning method [5].The writer used NB, SVM, Random forest and highest entropy classifiers in detecting false reviews on the dataset that they have collected. The collected dataset is around 10,000 negative tweets related to Samsung products and their services [8].Mani, [10] developed a supervised learning algorithm to identify fake reviews using unigram and bigram features by utilising the ensemble model. The model has two phases. In the first, the classification methods Random Forest, Naive Bayes, and Support Vector Machine were employed. Correctness of the classification model was improved during the second phase using the stacking and voting ensemble techniques.

Reviews are seen as real means of exchanging input on either good or bad services; any attempt to skew such reviews by including false or misleading information is seen as dishonest behaviour and is referred to as a fake review [6].Huaxun deng and the other researcher's proposes fake reviews make it impossible for customers to make wise decisions, which might harm the platform's trustworthiness. So, an algorithm has been created to detect fraudulent reviews on commercial platforms in order to combat this[11].The methods currently in use are examined in this research along with approaches to spot phoney reviews. The study concludes that more research is needed to increase the reliability of fake review

Detection and makes recommendations for filling this knowledge gap [13].Ahmed M. Elmogy and other researcher's describes that how to identify fake reviews.

It looks at different features that can be extracted from reviews to identify fake reviews, and compares the performance of different classifiers[14].Aishwarya, Kashid and the remaining researcher's describes that the

Review sites are facing increasing criticism for spreading misinformation and promoting certain businesses through spammy opinion pieces [15]. Recently, researchers have proposed various ways to assess the credibility of user-generated content, using methods such as supervised machine learning. By analysing review and reviewer data

III. PROPOSED SOLUTION

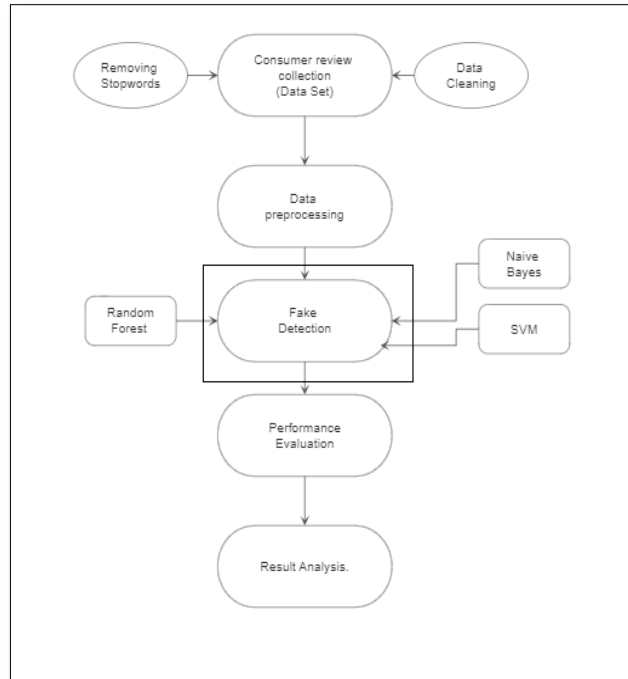The proposed system for detecting fake review is depicted in the Figure 1.

Figure 1. Proposed system for detecting fake review

An algorithm is designed for this purpose as shown below.

Detection Algorithm
Input: Data set.csv file containing reviews.
Step 1: Importing necessary libraries
Step 2: Pre-processing of Data
Step 3: Data cleaning
Step 4: Removing Stop words
Step 5: Removing Punctuation marks
Step 6: Splitting dataset into train and test dataset
Step 7: Model Training
Step 8: Model Evaluation
Step 9: Output (Accuracy)

The proposed system for detecting fake review is depicted in the Figure 1.

It consists of:

- Pre-processing: The dataset is first cleaned and all the stop words and symbols such as '#','@' are removed from the end or middle of text. Emoticons if present in the dataset are removed and data is tokenized. Duplicates if present in text after cleaning is removed. Word length of each text is calculated and texts whose length is Less than 3 and greater than 100 are removed. Word Cloud is generated for this pre-processed.
- Train and Test Data: The pre-processed data is then separated into test and train dataset for further processing and obtaining information through the machine learning algorithms.
- Analysis through different machine learning method. Different machine learning algorithms such as tree based decision, RF and GNB are trained using the train dataset and testing is done through test dataset to analyse the which model provides us with most accurate and precise results.

METHODOLOGY

Methodology shown is in below steps.

Step 1: Data set collection

Our project is based on utilizing machine learning techniques to assess the sentiment value of the

Standard dataset.

Step 2: Data Pre-processing

In addition to being a crucial step in the text mining process, data preparation is also crucial to many supervised

Learning approaches. Data preparation has been divided into the following categories:

(i) **Stop words Tokenization and Removal** – Stop words are that word that are necessary to filter before model is trained. Some of these words, which are frequent terms (such as "the," "a," "I," "of," "you," "and," and "it"), confuse our classifier rather than adding any useful information to our labelling scheme or adding value to the meaning of a phrase.

(ii) **Attribute Selection** – Selection of columns from the dataset that can be used to train the model and after choosing the selection of the right attributes can greatly improve categorization accuracy.

Naïve Bayes (**Multinomial**)

The NB method is based on the Bayesian rule of conditional probability in independent machine learning methods. Assumptions between functions.

Random Forest

A random forest is a classifier that uses tree. In trees there are subsets are given with the data of subset random forest check the data and gives the prediction accuracy. In random forest there are multiple trees for prediction it uses data from each tree and predict outcomes.

Support Vector Machine (SVM)

SVMs are processed with supervised learning method relevant algorithms that can learn and analyse the datasets use categorization.

IV. EXPERIMENTATION AND RESULTS

Dataset is obtained from kaggle after analysing variety check the length of each review get calculated and according to that length visual is depicted in Figure-2.
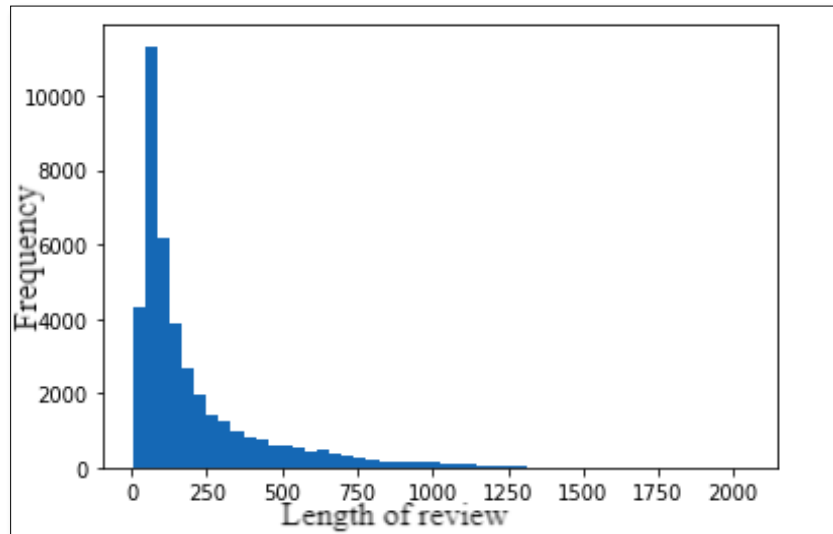

Fig-2. Represents the length of review from dataset.

After, pre-processing and cleaning is depicted in figure 3.

	category	rating	label	text_
0	Home_and_Kitchen_5	5.0	CG	love well made sturdy comfortable love itvery ...
1	Home_and_Kitchen_5	5.0	CG	love great upgrade original ive mine couple year
2	Home_and_Kitchen_5	5.0	CG	pillow saved back love look feel pillow
3	Home_and_Kitchen_5	1.0	CG	missing information use great product price
4	Home_and_Kitchen_5	5.0	CG	nice set good quality set two month

Figure 3- Pre-processed Dataset

On the pre-processed data, various machine learning algorithms are trained, including multinomial NB, RF and SVM. For comparative analysis, evaluation criteria like accuracy, evaluation criteria like accuracy, the macro average, and the weighted average are used.

The accuracy of several models is shown in Table 1, 2 and 3, and its visual depiction is shown in Figure 3.
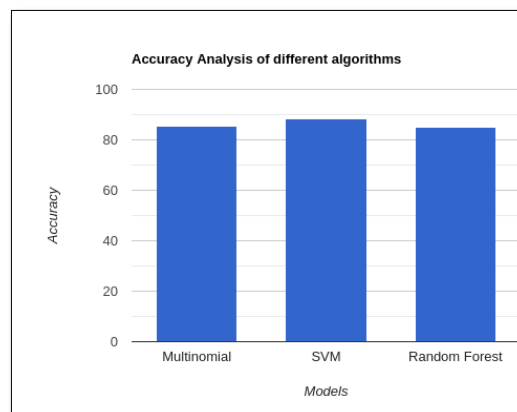

Fig-3 Accuracy Analysis of different algorithms

Classification report, Confusion matrix and Accuracy Score for MNB algorithm is represented by Table-I

```
Classification Report:              precision   recall  f1-score   support

              CG      0.83      0.90      0.86      7134
              OR      0.88      0.81      0.85      7017

        accuracy                          0.85     14151
       macro avg      0.86      0.85      0.85     14151
    weighted avg      0.86      0.85      0.85     14151

Confusion Matrix: [[6389  745]
 [1332 5685]]
Accuracy Score: 0.8532259204296516
```

Table-1 .MNB Bayes Algorithm.

For MB algorithm, after training and testing we got precision for computer generated review as 83% and for original reviews as 88%. In terms of recall, the value for CG is 90% and for OR is 81%. In terms of f1-score, the value for 86% and for OR is 85%. The overall accuracy for the model by considering CG and OR is 85%, further if we calculate the macro average then the values of recall and f1-score, precision are 85%, 85% and 86% respectively. Also for weighted average we calculate then the values of recall and f1-score, precision are 85%, 85% and 86% respectively.

Classification report, Confusion matrix and Accuracy Score for Random Forest Classifier is represented with the help of Table-II

```
Classification Report:              precision   recall  f1-score   support

              CG      0.82      0.90      0.86      7134
              OR      0.88      0.80      0.84      7017

        accuracy                          0.85     14151
       macro avg      0.85      0.85      0.85     14151
    weighted avg      0.85      0.85      0.85     14151

Confusion Matrix: [[6388  746]
 [1384 5633]]
Accuracy Score: 0.8494806020775917
Model Prediction Accuracy: 84.95%
```

Table-2 RFC

For RFC, after training and testing we got precision for computer generated review as 82% and for original reviews as 88%. In terms of recall, the value for CG is 90% and for OR is 80%. In terms of f1-score, the value for 86% and for OR is 84%. The overall accuracy for the model by considering CG and OR is 85%, further if we calculate the macro average then the values of recall and f1-score, precision are 85%, 85% and 86% respectively. Also for weighted average we calculate then the values of recall and f1-score, precision are 85%, 85% and 86% respectively.

Classification report, Confusion matrix and Accuracy Score for Support vector Machine is represented with the help of Table-III

```
Classification Report:              precision   recall  f1-score   support

              CG      0.90      0.86      0.88      7134
              OR      0.86      0.90      0.88      7017

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

Confusion Matrix: [[6134 1000]
 [ 679 6338]]
Accuracy Score: 0.8813511412621016
Model Prediction Accuracy: 88.14%
```

Table-3 Support Vector Machine.

For SVM, after training and testing we got precision for computer generated review as 90% and for original reviews as 86%. In terms of recall, the value for CG is 86% and for OR is 90%. In terms of f1-score, the value for 88% and for OR is 88%. The overall accuracy for the model by considering CG and OR is 88%, further if we calculate the macro average then the values of recall and f1-score, precision are 88%, 88% and 88% respectively. Also for weighted average we calculate then the values of recall and f1-score, precision are 88%, 88% and 88% respectively.
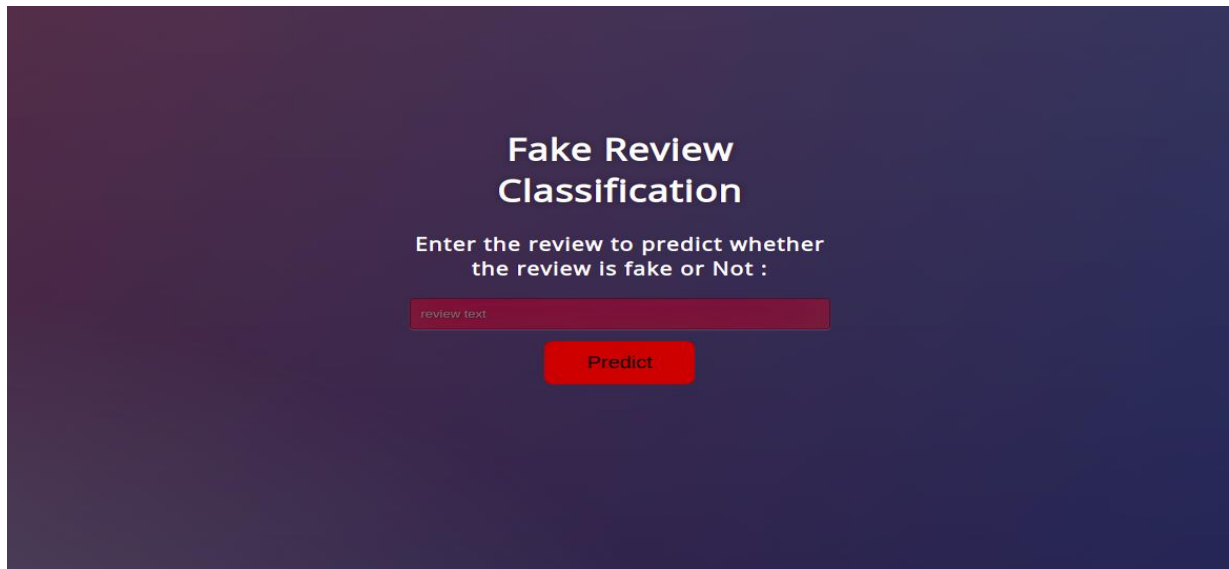
Fig. 4 Website for real time interaction.

Figure-4 shows the UI for entering the review for prediction purpose. It consists of textbox which will take input for review text and the predict button which will give prediction output after get clicked.
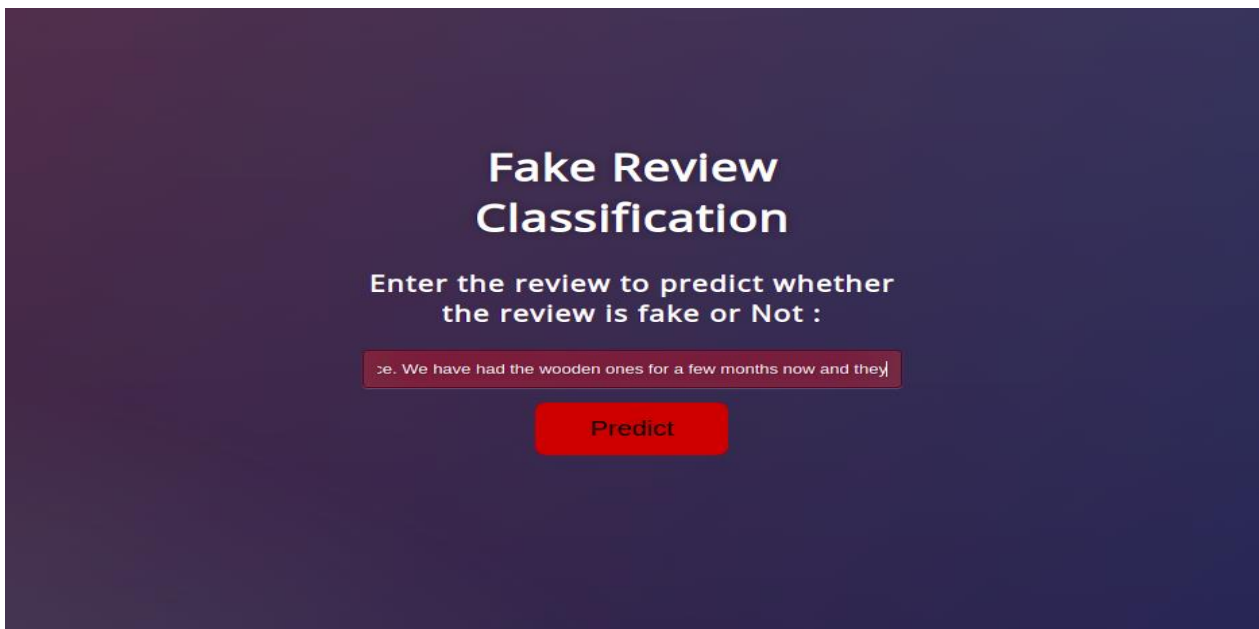

Fig.5 User Input

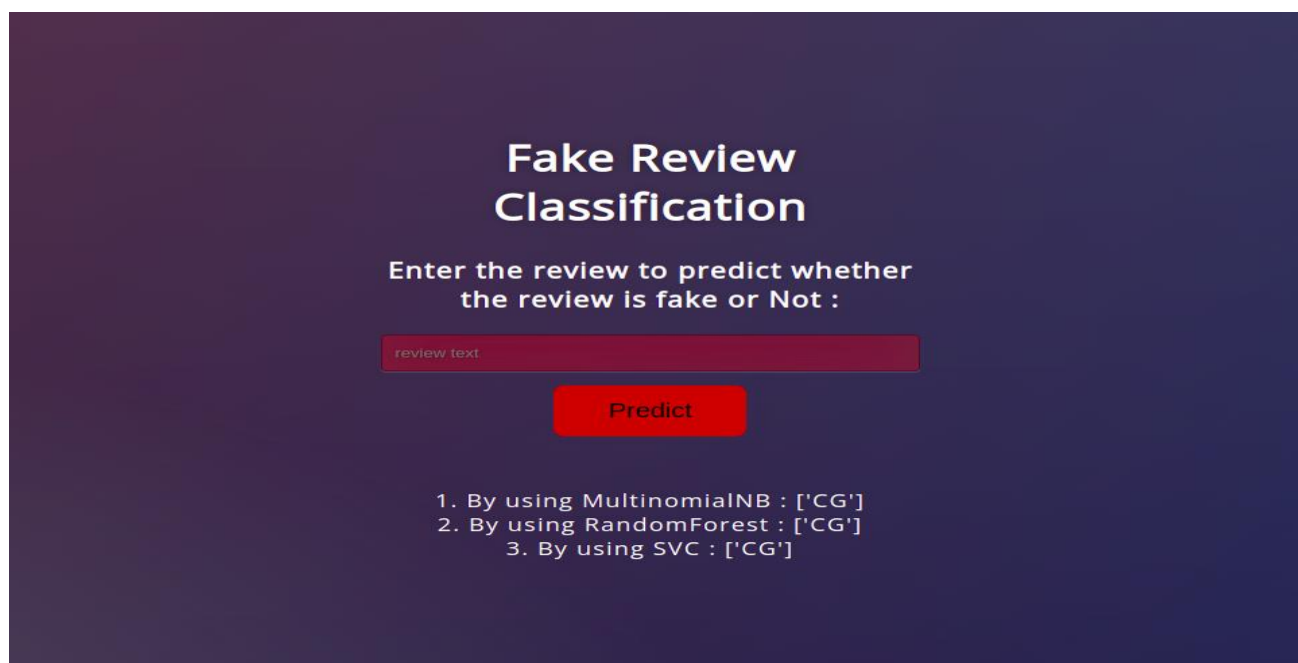Figure-5, Represents the example review entered by user on UI page.

Fig. 6 Output by 3 algorithms on website.

Figure-6, this figure shows the prediction output, i.e, the output generated by the three algorithms for the input of the example review text. As we can see from the figure-6 for the entered example review test we got output as, By using Multinomial NB we got output as 'computer generated review' , By using Random Forest classifier we got output as 'Computer generated review' and By using Support vector machine we got output as 'Computer generated review'

According to our findings, Support Vector Machine offers us the highest level of accuracy, hence it was taken into account when developing our model for Fake review classification.

V. CONCLUSION AND FUTURE WORK

To Manual label the reviews is time consuming, our study aims to create a system that can effectively identify and classify the reviews in less time we have analysed three algorithms namely MNB, RFC and SVM. In our comparative study we found that MNB algorithm is giving accuracy of 85%, RFC algorithm is giving 84.95% and SVM is giving output accuracy of 88.14%. So, according to the accuracy of this three algorithms we claim that support vector machine is giving output more efficiently and accurately to classify fake review. In future work, we will try to replace pre-processing techniques to improve the accuracy of the model and also try to add more attributes so that overall accuracy will get increased.

REFERENCES

1. Neha S. Chowdhary and Anala A. Pandit, PhD, "Fake Review Detection using Classification", International Journal of Computer Applications (0975 – 8887) Volume 180 – No.50, June 2018

2. B. Wagh, J. V. Shinde and P. A. Kale, "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques," International Journal of Emerging Research in Management and Technology, vol. 6, no. 12, pp. 37-44, December 2017.

3. N. O'Brien, "Machine Learning for Detection of Fake News," [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/119727/1078649610-MIT.pdf [Accessed: November 2018].

4. E. I. Elmurngi and A.Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science, vol. 14, no. 5, pp. 714–726, June 2018.

5. Kalpana Algotar and Ajay Bansal, "Detecting Truthful and Useful Consumer Reviews for Products using Opinion mining", [Online].Available: https://ceur-ws.org/Vol-2111/paper8.pdf [Accessed: November 2021].

6. R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Information Processing & Management, vol. 56, no. 4, pp. 1234 – 1244, 2019.

7. C. C. Aggarwal, "Opinion mining and sentiment analysis," in Machine Learning for Text. Springer, 2018, pp. 413–434.

8. A. Molla, Y. Biadgie, and K.-A. Sohn, "Detecting Negative Deceptive Opinion from Tweets." in International Conference on Mobile and Wireless Technology. Singapore: Springer, 2017.

9. G. Shan, L. Zhou, and D. Zhang, ''From conflicts and confusion to doubts:Examining review inconsistency for fake review detection,'' Decis. Support Syst., vol. 144, May 2021, Art. no. 113513.

10. S. Mani, S. Kumari, A. Jain, and P. Kumar, ''Spam review detection using ensemble machine learning,'' in Proc. Int. Conf. Mach. Learn.Data Mining Pattern Recognit. New York, NY, USA: Springer, 2018,pp. 198–209.

11. H. Deng, L. Zhao, N. Luo, Y. Liu, G. Guo, X. Wang, Z. Tan, S. Wang and F. Zhou, ''Semi-supervised learning based fake review detection,'' in Proc. IEEE Int. Symp. Parallel Distribution Process. Appl. IEEE Int. Conf.Ubiquitous Comput. Commun. (ISPA/IUCC), Dec.2017, pp. 1278–1280.

12. A. Ligthart, C. Catal, and B. Tekinerdogan, ''Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification,''Appl. Soft Comput., vol. 101, Mar. 2021, Art. no.107023.

13. Rami Mohawesh, shuxiang xu, son n. tran, "Fake Reviews Detection: A Survey"april 26, 2021, IEEE Access.