Detection of Fake Reviews in Social Media using Machine Learning Techniques

Vaibhavalakshmi C D[1], Deepthi K[2]

[1]Student, Ramaiah Institute of Technology, Bangalore, Karnataka, India
[2]Assistant Professor, Ramaiah Institute of Technology, Bangalore, Karnataka, India

---

Abstract: E-commerce is developing all around the world at such a pace of unthinkable. Including its development, day after day the influence of online feedbacks is rising. Comments could even influence purchase choices for individuals. These days it's becoming a practice to understand customer reviews prior to actually purchasing the stuff, particularly for prospective consumers. Buyers write comments of a commodity they are purchasing that could be true or fake. The twitter data of some popular personalities and some movie reviews data can also be fake to destroy one's image. Hence it is significant to identify such fake reviews. In this proposed work, efficient techniques for detecting such fake reviews are proposed and compared. Those techniques are Naive Bayes, Support Vector Machine (SVM) and K-Nearest neighbour (KNN).

The twitter data, movie review data or any commodity review data that need to detected as a true or fake is collected and it is stored in a .csv file. The data is pre-processed and keywords are identified. The machine learning classifiers are applied for the data to detect whether it is real or fake. The three algorithms are used for this purpose. Naive Bayes algorithm is executed by giving 97% accuracy, similarly accuracy of SVM is 98% and accuracy for KNN is 90%. Thus the proposed work proves that these machine learning techniques demonstrated efficient results and SVM has better performance among them.

Keywords: E-commerce, fake reviews, twitter data, commodity, machine learning techniques, Naive Bayes, SVM, KNN, keywords, accuracy.

1. INTRODUCTION

Online purchasing is rising bit by bit since each service or product is easily accessible. Sellers are obtaining more reaction to one's corporation. Ever more Mobile applications are free for internet purchases and therefore it is even easier for consumer to order any product on even a tap and one can share their comments with no complexities. Customer comments could even generate a strong impact on society around a thorough group of sectors, but are much more essential in the e-commerce world, in which subjective thought and evaluations on services and products are regarded to be beneficial to reach a choice to either buy a product or obtain service. Several people generally frustrated kinds of persons misdirect others by sharing false comments to encourage or damage the image of any specific goods or services according to wish. Such people are known as perception spammers and the false reviews they give are considered as fake comments. Individuals form their opinions as to whether to buy the items or not by evaluating and observing the previous opinions on such goods. There are true and false comments if the general view is not appropriate, it is unlikely that they just do not purchase the stuff. Now even the consumers can post any viewpoint message that inspires customers to offer negative review of the specific item. Even with invention of vast amount of data accessible in the digital world as well as huge amount of customers, utilizing this data, retrieving the views of customers is becoming essential throughout the globalized society. Evaluating the customer comments and recommending appropriately to users premised on this comments is becoming important in facilitating the customers to have the proper knowledge regarding a worthy commodity that they would like to buy.

Although customer reviews could be beneficial, naive confidence in such comments is unsafe for either the buyers or sellers. Many consumers read research before making any online purchase. Moreover, the comments could be misleading for additional benefit or profit, so any buying decision relied on web comments should be taken carefully. Malware identification has also been tested in several fields. Internet spam and e-mail spamming are the two most commonly researched forms of spam. Sentiment spam is quite distinct from these kind of spams. Unlike all other types of spamming, it would be almost difficult, to identify false comments by individually interpreting them. The effect of online feedback on corporations has developed substantially in previous years, becoming vital to evaluate operational efficiency in a broad range of disciplines, varying across restaurants, hotels and e-commerce. Regrettably, several

people utilize dishonest ways to boost their online credibility by publishing false comments of their companies or rivals. There are different techniques to resolve this problem of fake reviews. The techniques are aids in classifying the real and fake reviews. In this proposed system three proficient techniques namely Naive Bayes, SVM and KNN are used to identify the online fake comments. The data which needs to be classified is pre-processed and is treated with each o the algorithm where SVM gives better outcomes.

2. LITERATURE SURVEY

Kolli Shivagangadhar et. al., [1] intended at determining if a review is fake or real. The classifiers used in this work are Naïve Bayes Classifier, Logistic Regression and SVM. Consumers are much more reliant on taking decisions to purchase products at either ecommerce websites or offline retail shops in the current context. As such comments are dangerous for an item's strength or weakness in profits, comments are exploited for true or false views. Distorted comments could also be called fake or dishonest comments or misleading reviews or untrue feedbacks. Phishing on misleading sentiment in recent digital world is becoming a risk to both consumers and businesses. It is a critical and challenging job to discern such fake comments. Such disappointed users also get hired to advertise such comments. As a consequence, by taking a look at every review, it's a huge challenge for a normal consumer to distinguish dishonest comments from honest ones.

Elshrif Elmurngi et. al., [2] aimed to classify film reviews through using machine learning models into classes of true or false polarisation. In this work, they analyse user reviews of films utilizing Sentiment Analysis (SA) approach to detect fake comments. Techniques of SA and message categorization are adapted to a film review dataset. Recently, due to its exciting commercial advantages, SA is now one of the main important topic for textual data. One of SA's key aspect is how to remove feelings from within the viewpoint, as well as how to identify false positive feedback and fraudulent negative comments from viewpoint comments. In addition, consumer sentiment feedback could be divided as true or false comments that a customer may use to pick a product.

Lakshmi Holla et. al., [3] discussed in detail, the multiple fake review identification methodologies, and then get to realize the jobs completed on this topic. They explains the following technologies used to classify the fake comments. Here a description of the current fake review identification approaches is specifically addressed together with their strengths and drawbacks.

The article also quickly outlines the present problems in fake review identification that needs to be discussed. Fake reviews are increasing significantly together with the huge amount of internet information accessible online. Therefore there is a clear necessity to overcome the problem of identifying fake reviews and developing appropriate techniques to enhance prediction performance. It is one of the important research topics that should be researched in modern environment.

Pankaj Chaudhary et. al., [4] purposed to provide an overview of the key review and reviewer centered applicable requirements in the writings for the detection of fake comments, in specific those methodologies using different machine learning approaches. Such methods mainly offer smarter outcomes in relation to solely unsupervised strategies that are mostly dependent on graph-based techniques that recognize conceptual linkages in forums of review.

In addition, this work suggests and tests several extra new functionality that may be appropriate to distinguish legitimate and false comments. To this end, a supervised classifier depended on Random forests was introduced, taking into account some very well-known and new attributes, and a huge-scale labelled repository from where all of these functionalities were retrieved. The positive outcomes obtained depicts the efficiency of new functionality for detecting especially singleton fake comments, and the usefulness of such a practice in general.

Rodrigo Barbado et. al., [5] recommends a functionality system to identify fake feedback measured in the consumer products sector. In the past few years, the effect of online feedback on businesses has evolved dramatically to assess operational efficiency in a range of industries, from restaurants to e-commerce. Regrettably, several consumers utilize illegal methods by publishing false feedback of their companies or rivals to boost their online credibility. Preceding study has evaluated the identification of fake comments in a variety of categories, like customer reviews or market insight in hotels and restaurants. Despite the economic significance, moreover, the area of customer electronics companies has still not been explored in detail.

Naveed Hussain et. al., [6] conducted a comprehensive analysis of current spam check research by using Systematic Literature Review (SLR) framework. The authors examined the research on basis of how

attributes are derived from the analysis datasets and the various approaches and strategies used to resolve the issue of spam identification. However, this work analyzes various measurements utilized to evaluate the phishing identification techniques that can be used for the review. This review of the literature defined two feature extraction methods for features and two separate frameworks to phishing detection.

Rajshri Kashti et. al., [7] introduces the effective way of learning to identify malicious and truthful comments. Even now multiple suppliers promote their goods via social networks such as instagram, Facebook messenger etc. Therefore, testing their accuracy prior to purchasing the product is highly necessary. Purchaser or consumer wishes to test another customers' thoughts on their purchasing of that item. Many times the user's feedback is not regarded legitimate as feedback was provided before purchasing it. Analysis often includes contradictory terms. It also makes a ridiculous perception on some other client and she or he may withdraw the purchase. So that action frequently known to as fake review. The identification of fake reviews has therefore become much more significant matter for consumers to make smarter buying decisions and also for the dealer to sell their products secure.

Alimuddin Melleng et. al., [8] explored the efficiency of emotion-based depictions for the work of establishing machine learning frameworks for identification of fake comments. The research conducted for different real-world datasets indicates that enhanced data representation could be accomplished by integrating the approaches of sensation and sentiment analysis, and also conducting portion-by-portion sentiment and expression evaluation by categorizing the feedback. Recommendations can help new clients obtain perspectives from the perspectives of several other people, especially when making decisions about buying goods and services. Around the similar time, to even get comments and sustain excellent name, businesses require comments of their goods or services.

3. PROPOSED SYSTEM

The following section gives a system outline of the proposed system. The system architecture is the conceptual model that defines the structural and behavioural representation of a proposed system. The system architecture is an outcome of the design process. It is represented by a model diagram as shown in Figure 1.

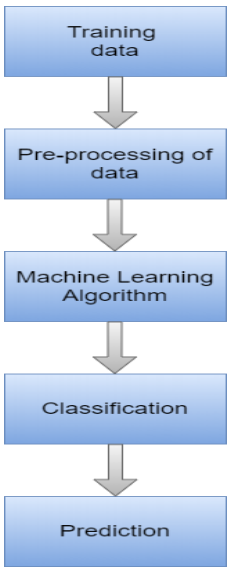

Figure 1: System Architecture

The proposed system architecture aims at the design and development of an effective system for fake review detection. The foremost step here is to collect dataset from Twitter which is need to be checked for fake or real. Data pre-processing is done for the collected data. In pre-processing unwanted words are removed and keywords are extracted. The machine learning algorithms such as Naive Bayes, SVM and KNN are applied now to predict that the review is fake or real.

- **Naive Bayes**

Naive Bayesian is an algorithm for creating classifiers, models assigning target class to problematic cases, defined as sets of extracted features, in which the classifiers are extracted from certain predefined number. For training these classifiers group of algorithms are used along with a single method on the basis of general concept, all Naive Bayes classifications presume that, target class parameter, the specific value function is regardless of the particular value of some other feature. The classification of the Naive Bayesian is dependent on Bayes' theorem with the independence assumptions among predictor variables. A Naive Bayes method is simple to construct without any difficult approximation of recursive parameters that made it especially efficient for very massive data. Bayes classifier frequently operates remarkably well and is broadly utilized since it also outperforms advanced classifier approaches. Naive Bayes works on the general principle of Bayes theorem as following:

In Naive Bayes, h is the hypothesis and d is the dataset.

$$P(h|d) = (P(d|h)*P(h))/P(d)$$

- P(h|d) is defined as the probability of hypothesis h given the dataset d.
- P(d|h) is defined as the probability of dataset d given the hypothesis h.
- P(h) is defined as the probability of hypothesis h.
- P(d) is defined as the probability of dataset d.

The maximum posterior hypothesis h can be defined as:

$$MAP(h) = max ((P(d|h)*P(h))/P(d)$$

- **SVM**

SVM is an approach for classification problems which separates information into groups. It is organized with such a latest data mobility procured into two categories, collecting the method from the beginning for whatever that is valuable. An SVM estimate is delegated to make sense from which class other data point has a position in. This classifier thus is such a semi-equal automatic algorithm. In this algorithm, the use of flat hyper-plane among two groups is far difficult to command. The query arises in any situation will this proposal incorporate the classification part to differentiate the hyper plane. So the correct answer is no, SVM does have a mechanism which is commonly referred as a kernel function to look after such issue.

Working of SVM algorithm categorised into five scenarios:

Scenario 1: Detection of the right hyper-plane

Three hyper planes are taken as A, B and C and privilege hyper planes are grouped into stars and circles. In the given figure below, hyper plane B distinguishes two classes.

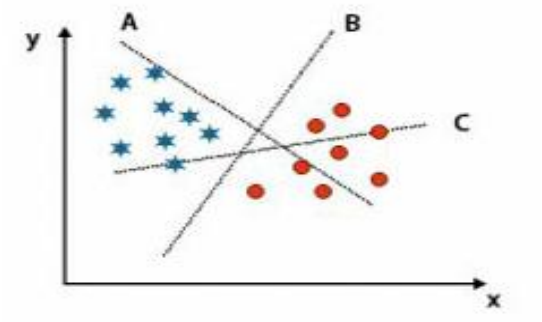

Figure 2: Three hyper plane representation

Scenario 2: Detection of the right hyper-plane

Three hyper planes A, B and C are now separated as shown in below figure 3. Margin of C is compared with both A and B.

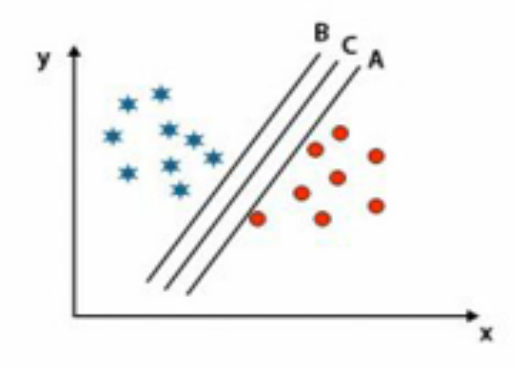

Figure 3: Three hyper plane separation

Scenario 3: Detection of the right hyper-plane

In this case, hyper-plane A is featured precisely and there is some mistake with respect to hyper-plane B. Therefore A is the privilege hyper-plane.

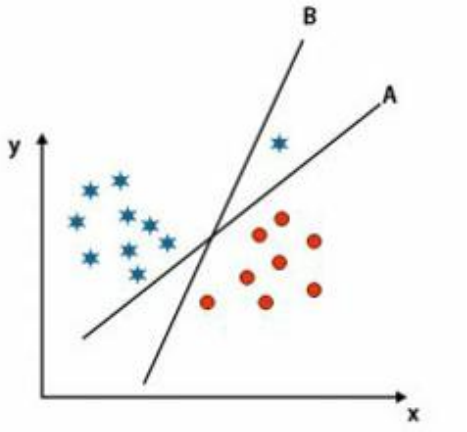

Figure 4: Privilege hyper plane A

Scenario 4: Classification of two classes

In this case, one star is represented in other class and is called anomaly. In SVM computation, the privilege hyper plane is located by ignoring anomaly.

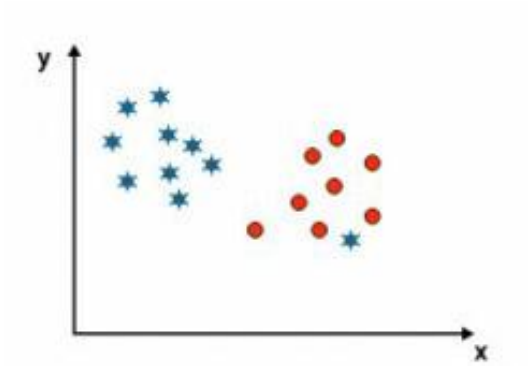

Figure 5: Star anomaly of class

Scenario 5: Distinguishing classes using fine hyper plane

To order the classes, SVM represents highlights. In this proposed work new constituent referred as z=x^2+y^2 is utilized by focusing on x and z-axis.

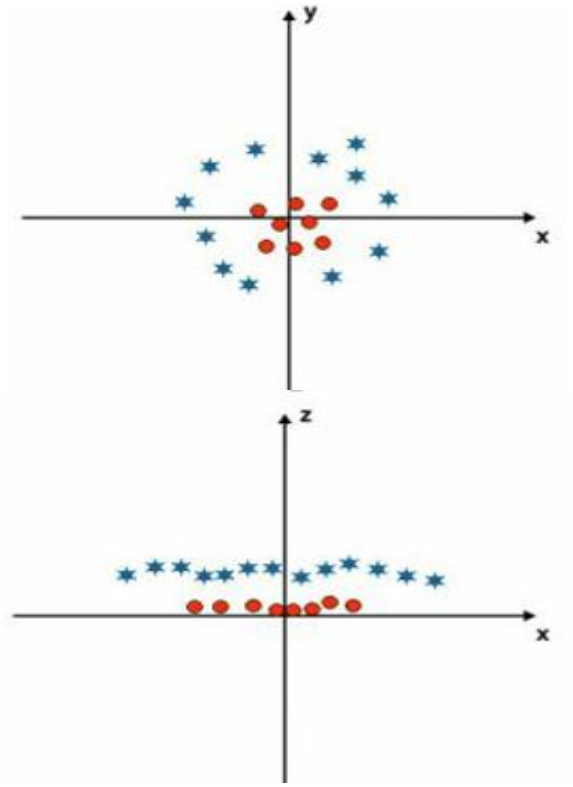

Figure 6: Fine hyper plane and focus on x and z-axis

- **KNN**

The KNN technique is a simple, convenient-to-implement machine learning technique which could be utilized to resolve both regression and classification issues. KNN approach utilizes feature identity to estimate the new data point values that also implies that a value would be allocated to the latest data point depending on how much closely it resembles the training array points. The common steps to implement the KNN algorithm are as follows:

- Loading of training and testing dataset.
- The closest data points referred to as value of K is selected.
- For every data point in testing dataset
  - The distance among testing data and every row of training data is computed using Euclidean distance.
  - They are categorised in ascending way on the basis of distance.
  - The highest K row is selected from this categorisation.
  - Class is allocated to the data point depending on most recurrent class of rows.

4. RESULTS

The results for the prediction of fake reviews using SVM algorithm are shown here. Figure 7 shows the home page of the fake review detection application. This home page incorporates the User login using user name and password. If there is new user, he needs to register before login. After registering user needs to login using proper credentials.

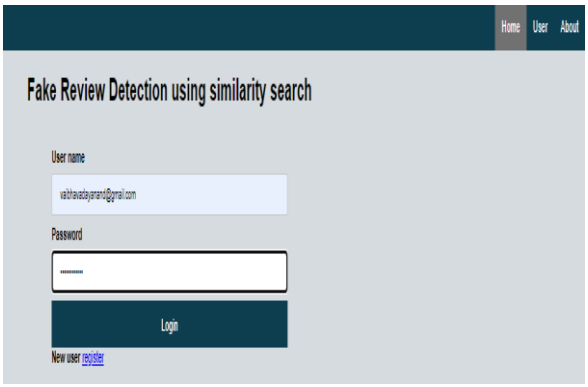

Figure 7: Home page

Once the user logs in, he need to enter the twitter details. For example as shown in figure 8, the user can enter the actor name/actress name with # or any other headlines for which he wants to detect fake reviews. Then the data is collected upon clicking on get tweets button. The data is stored in .CSV file.

Figure 8: Entering details of actor or actress

The generated .CSV file is uploaded to predict the fake news as shown in below figure 9. After uploading it will show the result as fake news or not fake news. Here in this case, it is fake news.

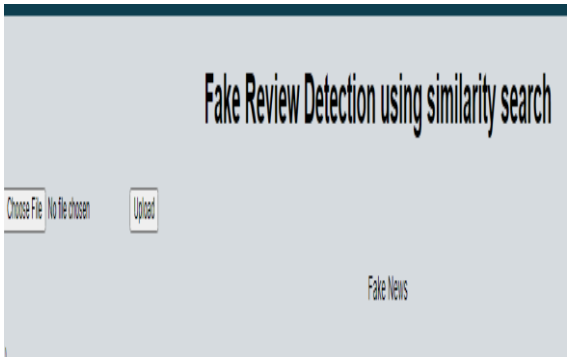

Figure 9: File uploading and detection of fake review

The below figure 10 shows that when any news is entered in the get news column it will predict whether it is real or fake. Here for example, the input is entered as "Please call immediately as there is an urgent message waiting for you".

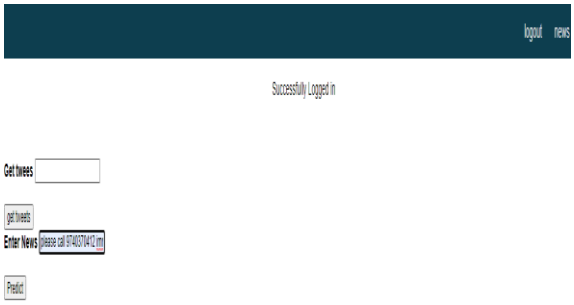

Figure 10: Entering news

The entered input news is detected as fake or not upon clicking on predict button and is displayed. The below figure 11 shows this entered input news is not fake news.

Figure 11: Not fake review detection

The fake reviews and real reviews for all the checked data are categorised and stored. Upon clicking on a button called "news" on top right of the page shown in figure 11, this stored real reviews i.e., not fake reviews for user login is displayed as shown in below figure 12.

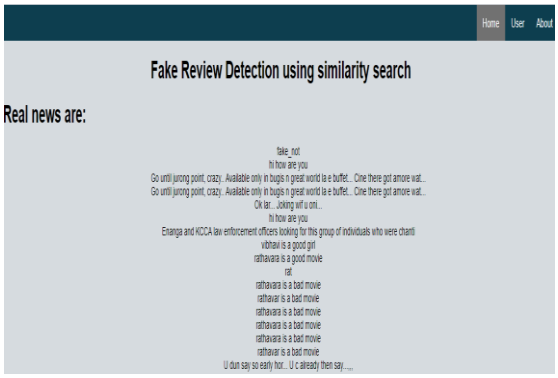

Figure 12: Display of real news

Comparison of Naive Bayes, SVM and KNN results

Comparison of the three machine learning algorithms namely, Naive Bayes, SVM and KNN is done for performance evaluation of fake review detection. Table 1 below shows the accuracy comparison. From the table it can be concluded that SVM gives better outcomes.

Table 1: Accuracy of fake review detection methods

Algorithm	Type	Accuracy
Naive Bayes	Supervised classification	97%
SVM	Supervised classification and regression	98%
KNN	Supervised classification and regression	90%

The graph is plotted to display the accuracy of all three algorithms as shown in below figure 13. The X-axis shows algorithm types and Y-axis depicts accuracy score.

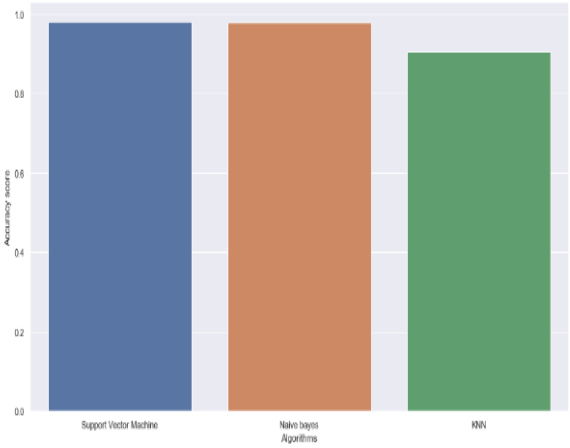

Figure 13: Graph of comparison

5. CONCLUSION AND FUTURE WORK

Evaluating and categorizing a review comment into one that is fake or true is an increasingly difficult task. Throughout the sector of fake review identification approaches there is a huge amount of research going on. Growing Process of identification does have its own benefits and drawbacks. This paper provides the overview of fake reviews and fake review detection methods. The three algorithms described here are Naive Bayes, SVM and KNN. Fake reviews from the twitter account of actor/actress are detected. Any type of the news can also be identified as fake or real by this proposed system. The results of SVM are shown. SVM is proved as better performing algorithm among all. The accuracy graph is also plotted which depicts the efficiency of SVM algorithm.

Fake review detection can be improvised by implementing some more innovations in the Machine learning algorithms. The fake review detection can be done by the advanced techniques which can detect the fake reviews quickly, appropriately and more automatically.

REFERENCES

[1] Kolli Shivagangadhar, Sagar H, Sohan Sathyan and Vanipriya C.H "Fraud Detection in Online Reviews using Machine Learning Techniques," International Journal of Computational Engineering Research, pp. 52-56, Vol. 05, 2015.

[2] Elshrif Elmurngi, Abdelouahed Gherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques," The Sixth International Conference on Data Analytics, pp. 65-72, 2017.

[3] A.Lakshmi Holla and Dr Kavitha K.S, "A Comparative Study on Fake Review Detection Technique," International Journal of Engineering Research in Computer Science and Engineering, pp. 641-645, Vol. 5, 2018.

[4] Pankaj Chaudhary, Abhimanyu Tyagi and Santosh Mishra, "Fake Review Detection through Supervised Classification," International Journal of Creative Research Thoughts, pp. 417-427, 2018.

[5] Rodrigo Barbado, Oscar Araque and Carlos A. Iglesias, "A Framework for Fake Review Detection in Online Consumer Electronics Retailers," 2019.

[6] Naveed Hussain, Hamid Turab Mirza, Ghulam Rasool, Ibrar Hussain and Mohammad Kaleem, "Spam Review Detection Techniques: A Systematic Literature Review," Applied sciences, pp.1-26, 2019.

[7] Rajshri P. Kashti1 and Prakash S. Prasad, "Enhancing NLP Techniques for Fake Review Detection," International Research Journal of Engineering and Technology, pp. 241-245, Vol. 6, 2019.

[8] Alimuddin Melleng, Anna-Jurek Loughrey and Deepak P, "Sentiment and Emotion Based Text Representation for Fake Reviews Detection," Proceedings of Recent Advances in Natural Language Processing, pp. 750-757, 2019