

# Enhancing NLP Techniques for Fake Review Detection

Ms. Rajshri P. Kashti<sup>1</sup>, Dr. Prakash S. Prasad<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Dept. of CSE, PIET College, Nagpur, India

<sup>2</sup>Head of Department, Dept. of CSE, PIET College, Nagpur, India

\*\*\*

**Abstract** - We are in the era of internet where people are more techno-savvy and they surf internet before buying a single item. Since buying a product online is easy and convenient these days, people also tending towards it as it save time and sometimes money. Also many branded products can be bought without thinking much about quality as name is enough for branded item. Nowadays various vendors also advertise their products through social media like facebook, whatsapp etc. Thus it is an extremely important to check their reliability before buying product. Buyer or client wants to check the opinion of other buyers regarding their purchase for that product. Most of the times review given by the user is not considered genuine as review was given without buying it. Sometimes review contains unrelated words. This makes a false impression on another customer and he or she may cancel buying it. Such as activity often referred as fake Review. Thus detecting fake reviews has become more important issue for customers to make better decision on purchase as well as for the trader to make their products reliable. This paper presents an active learning method for detecting fake and genuine reviews.

**Key Words:** Fake reviews, active learning, Rough Set Classifier, Decision tree algorithm, Random Forest.

## 1. INTRODUCTION

Online shopping is increasing day by day as every product and service is getting available easily. Vendors are getting more response to their business. More and more Mobile apps are available for online Shopping and hence also it is easier for customer to purchase any item on a click and he/she can post their reviews without much complications. People can post their views or opinions on tons of thousands of discussion groups, internet community, and forums, product/service reviews, and blogs etc. These things can be cooperatively called user-generated contents. Usually, these user-generated comments are written in natural language and people have the freedom to give their opinion as they want, as there is no monitoring system available till now. Sharing a personal view about a particular product or a service that has experienced by an individual is referred to as reviews. Online reviews can create a great impact on people across a comprehensive band of industries, but are more important in the world of e-commerce, where personal opinion and reviews on products or services are considered to be useful to make a decision whether to purchase a product or avail service. Some people usually disgruntled type of people misdirect others by posting fake reviews to

promote or harm the reputation of some particular products or services as per desire. These persons are labeled as opinion spammers and the misleading comments they provide are called fake reviews.

New buyers give importance to the feedback given by other users as do the companies that sell such products, today's individuals and older ones extensively rely on reviews available on line. People make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products. There are positive and negative reviews if the overall impression is not proper, it is doubtful that they don't buy the product. Now the customers can write any opinion text, which motivates people to give fake review of the particular product. This paper is about fake review detection techniques accuracy for amazon dataset

### 1.1 Objectives

- Develop a System to accept Reviews from Authenticated Users only.
- After accepting reviews from trustworthy clients (Genuine Users), we will be using NLP based sentiment analyzer and text mining algorithms to classify and predict positive, negative and neutral reviews.
- Fake Reviews - Unauthorized, Untrustworthy, Contents of Unrelated words.

## 2. RELATED WORK

- In 2016, M.N. Istiaq Ahsan, Tamzid Nahian, Abdullah All Kafi, Md. Ismail Hossain, Faisal Muhammad Shah proposed "Review Spam Detection using Active Learning." [1] This paper explores the opportunities of introducing active learning for detecting Review spams conducted on real life data which shows promising results. During the process, they trained model using active learning method which learns from the best samples in multiple iterations.
- In 2015, Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada, presented "Survey of review spam detection using machine learning techniques." [2] This survey paper covers machine learning techniques and approaches that have been proposed for the

detection of online spam reviews. Supervised learning is the most frequent machine learning approach for performing review spam detection; however, obtaining labeled reviews for training is difficult and manual identification of fake reviews has poor accuracy. This has led to various experiments using synthetic or small datasets. Features extracted from review text (e.g., bag of words, POS tags) are often used to train spam detection classifiers. An alternative approach is to extract features related to the metadata of the review, or features associated with the behavior of users who write the reviews.

- In 2017, SP.Rajamohana, Dr.K.Umamaheswari, M.Dharani, R.Vedackshya presented "Survey of review spam detection using machine learning techniques." [3] focused light on deceptive reviews that are available in the internet which increasingly affects businesses and customers. Hence it is important to detect and eliminate such fake reviews from online websites. This paper reveals several approaches used for review spam detection and performance measures were identified.
- In 2015, Daya L. Mevada, Prof.Viraj Daxini in their paper "An opinion spam analyzer for product Reviews using supervised machine Learning method." [4] suggested method to find opinion spam from huge amount of unstructured data has become an important research problem. This research proposes an opinion spam analyzer which automatically classifies input text data into either spam or non-spam category. The proposed system will use machine learning supervised technique.
- In 2016, Miss. Rashmi Gomatesh Adike, Prof. Vivekanand Reddy [5] presented their views in the paper "Detection of Fake Review and Brand Spam Using Data Mining Technique". This system proposes a behavioral approach to identify review spammers those who are trying to manipulate the ratings on some products. Author derive an aggregated behavior methods for rank reviewers based on the degree that they have demonstrated the spamming behaviors. They verified proposed methods by conducting user evaluation on an Amazon dataset which contains reviews of different company's products.
- In 2013, Arjun Mukherjee, Vivek Venkataraman, Bing Liu, Natalie Glance studied and presented paper on "Fake review detection: Classification and analysis of real and pseudo reviews." [6] This paper performed an in-depth investigation of supervised learning for fake review detection using Amazon Mechanical Turk (AMT) produced fake reviews and real-life fake reviews. The work in [36] showed that using AMT fake reviews and reviews (assumed non-fake) from Trip advisor achieved the classification accuracy of 89.6% with bigram features and

balanced data. This paper first performed a comparison using real-life filtered (fake) and unfiltered (non-fake) reviews in Yelp. The results showed that the real-life data is much harder to classify, with an accuracy of only 67.8%. This prompted us to propose a novel and principled method to uncover the precise difference between the two types of fake reviews using KL-divergence and its asymmetric property.

- In 2014, Jiwei Li, Myle Ott, Claire Cardie and Eduard Hovy presented their work in "Towards a General Rule for Identifying Deceptive Opinion Spam." [7] In this work, they have developed a multi-domain large-scale dataset containing gold-standard deceptive opinion spam. It includes reviews of Hotels, Restaurants and Doctors, generated through crowdsourcing and domain experts. Study of data uses SAGE to make observations about the respects in which truthful and deceptive text differs. Suggested model includes several domain-independent features that shed light on these differences, which further allow formulating some general rules for recognizing deceptive opinion spam.
- In 2004, Hu Mingqing and Liu Bing "Mining and summarizing customer reviews" [8] extract the features of the product. The customer's sentiment to individually feature of the product is shown via a summarization system, which includes- mining features of the product that have been commented upon by customers, categorize whether each view sentence in a review is positive or negative, and summarizing the results.
- In 2017, Liu, Pan, et al ventured "Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features." [9] in social networking websites they provide the user rate basis on various factor like their total consumption, activeness of user etc. base to classify whether review is spam or not.
- In 2010, Lim Ee-Peng, Nguyen Viet-An, Jindal Nitin, et al. [10] identified and demonstrated some characteristic behaviors of review spammers by proposing scoring methods for measuring the degree of spam for each reviewer. Then, a subset of extremely suspicious reviewers is selected for additional review with the help of web-based spammer evaluation software specially developed for user evaluation experiments.
- In 2012, XieSihong, WANG Guan, LINShuyang, et al. [11] in their work analyzed the irregular pattern of just the rating of a review to detect spam.

### 3. IDENTIFICATION FACTORS CONSIDERED FOR FAKE REVIEW DETECTION

- 1) **Repeated reviews and rating:** Same review comment and rating repeatedly for the same product or service.
- 2) **Username with integers:** genuine user or buyer must exhibit his/her real name and it should not contain only numbers. As only number indicates spammer since genuine buyer will have name with address in alphanumeric notation.
- 3) **Star (\*) Rating only:** Just giving star rating doesn't mean review is genuine as customer will also try to say something before giving star rating. Hence just star rating considered as fake.
- 4) **No category Details:** actual buyer will post his review by specifying the features of product. So if user post his review without specifying product category or feature, it is considered as fake.
- 5) **Rating Vs Review Sentiment:** If review found to be with highest rating but low sentiment or vice-versa i.e. if user posting high rating but his/her review comments does not match to the rating, such reviews are considered to be fake.
- 6) **Review Length:** a genuine buyer will post his review by specifying what did he like or didn't like by addressing the features of product or services. Thus review must be within a predefined length.

All of the above factors will be combined together in an active learning process. Only single factor may not give accurate identification of fake review. Thus value of all the factors will be considered in the review process.

### 4. PROPOSED WORK FLOW

1. Data Acquisition
2. Data Preprocessing
3. Active Learning
4. Classifiers /Algorithms
  - i. Rough Set Classifier
  - ii. Decision tree algorithm.
  - iii. Random forest algorithm
5. Result

### 5. PROPOSED METHODOLOGY

1. **Data Acquisition:** Data extracted from amazon dataset is used as the unlabeled data, labeled dataset is used for both training and testing purpose in this method. We use reviews extracted from

amazon.com as the unlabeled data in this learning method.

2. **Data Preprocessing:** Unstructured data in MS Excel format acquired from the source is converted into structured data i.e. in MySQL Database format. Preprocessing procedures includes- tokenization & lowercasing letters, removing stop words, removing punctuations, stemming etc.
3. **Processing unlabeled data:** It includes tagging a label to the unprocessed data. In this step we make cluster head from the structured data
4. **Active Learning:** Active learning is a special case of semi supervised machine learning which can interactively request the user to determine the class of some unknown data points to achieve the desired results.

Labeling the whole dataset manually is extremely time consuming and labor intensive. So, the algorithm actively queries the user for labeling the new, confusing data points.

In this type of learning, learner itself chooses the data point examples that's why it needs a much lower number of examples to learn a concept than it is required in typical supervised learning

Training dataset is used by the algorithm to train the model and test dataset is used later for evaluation. Certain number of samples from unlabeled dataset are selected for training and after estimation they get added into existing train dataset. The model will start training again with the new improved training set. The selection of unlabeled samples is based on a decision function which is the distance of the samples X to the separating hyper plane. Although the distance is between [-1, 1], we use absolute values because we need the confidence levels.

5. **Feature Weighing:** Here we will make use of TF-IDF ("Term Frequency-Inverse Document Frequency") vectorizer. TF-IDF is a numerical statistic used to determine the importance of a certain word in a collection or the document.

TF-IDF is generated by multiplying the term frequency with the log of the ratio of the total number of documents to the number of documents in which the term appears.

TF-IDF can be determined by-  
$$TFIDF(n, d, D) = \sum TF(n,d) * IDF(n,D) \quad (1)$$
  
Here,

TF (n,d)= the number of times the term n appears in document d.

IDF (n,D)=  $\log N$

With, N = total number of documents in a collection  
=  $|D| \mid \{d \in D : n \in d\}$  = number of documents where the term n appears.

After constructing the vectors using TF-IDF values, these sparse vectors are fed into the classifiers.

## 6. Classifiers

**a) Rough Set Classifier:** It is the machine learning method which makes use of set theory for decision making. The indiscernibility relation that produces minimal decision rules from training examples is the important notation in this method. Decision table with if-else rule is used to identify the set of feature. It is a hybridized tool that encompasses sequence Arithmetic, Rough Set Theory and Concept Lattice. The accuracy level of this classifier is 97.7%. Thus this model requires extra time and space for further classification of the output sequence into classes or subclasses

## b) Decision tree

A decision tree is a structure which consists of a root node, branches, and leaf nodes. Each internal node exhibits a test on an attribute, which results into branches emanating from it and each leaf node holds a class label. Root node is always remains the topmost node in the tree. Given an input of attributes together with its classes, a decision tree generates a sequence of rules that can be used for classifying the data.

A decision tree will be built using the whole dataset taking into consideration all features.

Basic algorithm for decision trees

- i. start with whole training set
- ii. select attribute or feature satisfying criteria that results into the "best" partition.
- iii. create child nodes based on partition
- iv. Repeat process on each child using child data until a stopping criterion is reached

## c) Random Forest

Random Forest algorithm is a supervised classification algorithm There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the result. Random Forest is the

processes of finding the root node and splitting the feature nodes will run randomly.

It is Unexpected in accuracy among current algorithms. It runs efficiently on big databases. It can handle thousands of input variables without variable removal. It gives approximations of what variables are important in the classification. It creates an inner unbiased estimate of the simplification error as the forest building grows. It has an effective method for estimating lost data and maintains accuracy when a large proportion of the data are missing. It has methods for balancing error in class people unbalanced data sets. Made forests can be saved for future use on other data. Prototypes are calculated that give data about the relation between the variables and the classification. It calculates vicinities between pairs of cases that can be used in clustering, locating outliers, or (by scaling) give interesting views of the data. The abilities of the above can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier finding. It offers an experimental method for noticing variable interactions. Remarks Random forests does not over fit.

## Algorithm of Random Forest

- i. Randomly select "K" features from total "m" features where  $k \ll m$
- ii. Among the "K" features, calculate the node "d" using the best split point
- iii. Split the node into daughter nodes using the best split
- iv. Repeat the 1 to 3 steps until "l" number of nodes has been reached
- v. Build forest by iterating steps 1 to 4 for "n" number of times thereby creating "n" number of trees

## 6. PROPOSED SYSTEM

The algorithm for the whole approach to detect review spam is given below-

### 1) Algorithm: Active Learning process

### 2) INPUT:

initialsampleTrain = The initial labelled sample training set;  
initialTestdata = The initial labelled sample test set;  
prodData = unlabeled input;

### 3) OUTPUT:

Classification result

#### 4) Load prodData

```

for all instances in prodData
create sparse vector using the tf-idf vectorizer
mat[] = sparse vector; //store s.Vector in a matrix
feed mat[] to the CLASSIFIER
accuracy = CLASSIFIER.accuracy
EVALUATE classifier measuring accuracy
unlabData = DecisionFunction(prodData instance)
Expert labeling unlabData
prodData = prodData U unlabData
END WHEN prodData = {∅}
DecisionFunction(input)
return TOP N instances consisting HIGHEST and
LOWEST average absolute confidence
    
```

#### 7. CONCLUSIONS

- Here we will try to explore the opportunities by introducing active learning for detecting fake Reviews conducted on real life data.
- This paper suggests different factors for identifying fake review & all those factors are weighted using active learning.
- During the process, we train our model using active learning method which learns from the best samples in multiple iterations.
- The feature vectors will be constructed using TF-IDF values of the review content and classifiers such as Rough Set, Decision tree and Random forest will be used for classification process.
- Our experimental approaches studies the accuracy of all sentiment classification algorithms, and determine which algorithm is more accurate. Furthermore, we were able to detect fake positive reviews and fake negative reviews through detection processes

#### REFERENCES

- 1) M. N. Istiaq Ahsan , Tamzid Nahian , Abdullah All Kafi, Md. Ismail Hossain , Faisal Muhammad Shah "Review Spam Detection using Active Learning." 978-1-5090-0996-1, pp.16, (2016).
- 2) Michael C., et al. "Survey of review spam detection using machine learning techniques." Journal of Big Data 2.1, pp.9, (2015).
- 3) Rajamohana S. P, Umamaheswari K., Dharani M., Vedackshya R., "Survey of review spam detection using machine learning techniques." ,978-1-5090-5778-8, pp.17 (2017).
- 4) Mevada D. L., Daxini V., "An opinion spam analyzer for product Reviews using supervised machine Learning method." pp.03, (2015).

- 5) Adike R. G., Reddy V., "Detection of Fake Review and Brand Spam Using Data Mining Technique.", pp.02,(2016).
- 6) Mukherjee A., Venkataraman V., , Bing Liu , Natalie Glance. "Fake review detection: Classification and analysis of real and pseudo reviews." (2013).
- 7) Jiwei Li, Myle Ott, Cardie C., Hovy E., "Towards a General Rule for Identifying Deceptive Opinion Spam."( 2014).
- 8) Minqing H., Bing L., "Mining and summarizing customer reviews", Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, pp.168-177, (2004).
- 9) Liu, Pan, et al. "Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features." Software Quality, Reliability and Security Companion (QRS-C), IEEE International Conference, IEEE, (2017).
- 10) Lim Ee-Peng, Nguyen Viet-An, Jindal Nitin, et al. "Detecting product review spammers using rating behaviors" [C]// Proceedings of the 19th ACM international conference on Information and knowledge management. New York: ACM Press, pp.939-948, (2010).
- 11) XieSihong, Guan W., Shuyang L., et al. "Review spam detection via temporal pattern discovery" ,Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining. New York: ACM Press, pp.823-831, (2012).

#### BIOGRAPHIES



Miss Rajshri P. Kashti is a MTech Student in the Department of Computer Science & Engineering, Priyadarshini Institute of Engineering and Technology, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur.



Dr. Prakash S. Prasad is working as Professor and Head of Computer Science and Engineering at Priyadarshini Institute of Engineering and Technology, Nagpur. He is having 20 Years of experience in the field of teaching to engineering students. He completed his engineering in the year 1997, master of engineering in 2006 and PhD in 2014.