

# Fake Review Detection

Aaditya Shete  
Department of  
Computer Engineering  
Ramrao Adik Institute of Technology  
Navi Mumbai, Maharashtra 400706  
Email: aadityashete@gmail.com

Lalima Sharda  
Department of  
Computer Engineering  
Ramrao Adik Institute of Technology  
Navi Mumbai, Maharashtra 400706  
Email: lalima.sharda@gmail.com

Shubham Wankhede  
Department of  
Computer Engineering  
Ramrao Adik Institute of Technology  
Navi Mumbai, Maharashtra 400706  
Email: shub.09.sw@gmail.com

Tanvi Chavan  
Department of  
Computer Engineering  
Ramrao Adik Institute of Technology  
Navi Mumbai, Maharashtra 400706  
Email: chavantanvi96@gmail.com

Rashmi Dhumal  
Department of  
Computer Engineering  
Ramrao Adik Institute of Technology  
Navi Mumbai, Maharashtra 400706  
Email: rashmi.dhumal@rait.ac.in

**Abstract**—E-commerce is getting enormous day by day if not in every minute. Online Reviews play a very important role in this field and are proving itself to be auspicious in terms of decision making from a customers point of view. That is why spamming, fake reviews and fabricated opinions are on the rise. Materially, it has become a profitable business which hampers the ingenuosness of the real fact. This project spans over the Data Mining and Analysis and Text Mining domains. Data Analytics is increasingly becoming a trending practice with the purpose of constructing valuable information from data. This has lead to new challenges faced by experts everyday, which includes integration of data from multiple data sources, reliability of the data sources, high-dimensional data, scalability of models, distributed computing etc. Lately, opinion mining has attracted a great deal of research attention, but very few have been able to implement sophisticated techniques in this field. Recent works have been primarily focussed on identifying deceptive opinion spams that have been deliberately written in sound authentic language. Other techniques include Natural Language Processing(NLP) that extracts meaningful features from text and then applying various machine learning techniques to conduct review apm detection. This project uses text features and models to calculate the probability of a review being fake using the Naive Bayes algorithm.

## I. INTRODUCTION

Expressions can take many forms. The number of people expressing themselves is increasing day by day. In case of internet and the web they take the form of blogs, comments, reviews, forums, journals, etc. One of the most widespread forms of expression is reviews. Many sites have made it compulsory for their users to post reviews. Reputation of restaurants, movies and products depend a lot upon the reviews they get from the customers. Today, the world is flocking with E-Commerce stores. Nearly every business venture has an e-commerce platform owing to easy access to the Internet everywhere and knowledge about the procedure. There are a number of parameters which contribute to define the success and credibility of an Ecommerce store. However, one very important factor in elevating the reputation, standard and

evaluation of an Ecommerce store is Product Reviews. Not taken much into serious account, Product Reviews provide an Ecommerce store with one of the most valuable resources available: Customer Feedback.

## II. RELATED WORK

A study showing efficacy in classifying product review by semantic meaning in [2] proposes an approach that incorporates spelling correction in review text and then employing a hybrid of Decision Trees and Naive Bayes algorithm. First, a web crawler is used to fetch comments from a particular web page. Then spelling corrections are carried out to make the comment more sensible and for knowing the polarity of words using Word Net dictionary. Then stemming is performed to remove the stop words. After classifying the positive and negative words using Nave Bayes algorithm, the overall polarity is calculated using decision tree.

Shathi, Hossain and Nadim in [3] introduces a Weight Matrix during training text documents which is a combination of term frequency (TF) and inverse class frequency (ICF) and later this weighted term is powered by a significant number and added with the posteriori value during the prediction time of Naive Bayes algorithm to establish a better and efficient performance of classification task. The precedence base element TF results an additional weight for each word whereas, ICF gives each common word a low score. Thus, the combined term Weighted Matrix balances weight where necessary. The result shows an improvement in performance accuracy of the Naive Bayes classifier on a small size data set.

The study proposed in paper [4] improves the accuracy of Naive Bayes classification using improved information gain, one of methods of feature extraction, by reducing the impact of low-frequency words. In this paper, a wide corpus of NLTK is used. According to the test results, the accuracy of the classification improved significantly. Experimental result showed that the information gain method reduces the

TABLE I  
A STUNNING TABLE

Research Paper	Data set used	Methodology Used	Advantages	Disadvantages
Sentiment Analysis of Flipkart reviews using Naive Bayes and Decision Tree Algorithm	Flipkart data set	Naive Bayes, Decision Tree algorithm	Calculates polarity of customer reviews and also compares products	Spelling mistakes and slang words affect the accuracy of Sentiment Analysis Algorithms
Enhancing Performance of Naive Bayes in Text Classification	A collection of 2225 English text documents with predefined categories	Weight matrix to enhance efficiency of Naive Bayes	Improvement in performance of a smaller size of training dataset, where standard NB performs poorly	Experiments show rare demotion of accuracy instead of dramatic enhancement in performance owing to a few errors
A text classifier of English Movie Reviews Based on Information Gain	NLTK Movie Corpus	Naive Bayes(using Information Gain)	Removes redundancy between features, selecting a subset of features better to improve the performance	Time complexity and space complexity needs further research

influence of low frequency words. The method brought about improvement in the accuracy of text classification, but the time complexity and space complexity of algorithm, requires further research. Movie Review Corpus were classified by the degree of preference of movies for audience, and divided into positive and negative, the count of each of positive and negative is 1000. Currently, the corpus is used more emotion classification field corpus.

### III. PROPOSED METHOD

The proposed system will analyze the reviews and classify them as whether they are spam or not. In this project, we introduce a more proactive approach that allows us to differentiate valid and spam review. The main aim is to design and develop a fake review detection system based on the classification algorithm Nave Bayes. Following is modus operandi of our project:

Figure 1 shows the Basic flow of the system.

#### A. Data Acquisition

Using a set of labeled data sets from Amazon, the Naive Bayes algorithm is trained. For un-labeled data, the Amazon live web page is scraped for new reviews and then preprocessed for the learning method.

#### B. Preprocessing

We apply a number of preprocessing techniques to deal with noisy, missing, and inconsistent data which might disturb decision-making process. Low-quality data will produce poor quality mining results and classification results. We ensure that the quality of the data we use in this experiment is up to the mark.

#### C. Stop Word Removal

Any word that may not contribute to the classification of the review as fake or real is removed. Stop words are basically a set of commonly used words in any language, not just English. The reason why stop words are critical to many applications is that, if we remove the words that are very commonly used in a given language, we can focus on the important words instead.

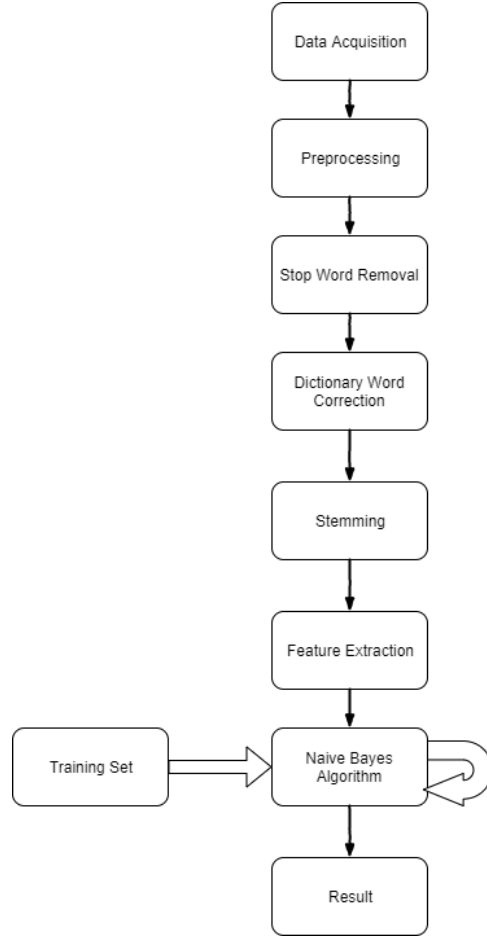


Fig. 1. Basic flow of the system.

#### D. Dictionary Word Correction

The remaining words are then checked for common typing or spelling mistakes and based on a Standard English dictionary, it is corrected accordingly.

#### E. Stemming

It is a process of reducing words to their root form (generally a written word form). The purpose is to discern the nature of the word and understand its context. From the words deducted from the above steps, root words are extracted based on the Stanford Core NLP Library.

#### F. Feature Extraction

This step starts with an initial set of labelled parts-of-speech data, i.e., the training set. Based on the derived values, it facilitates subsequent learning from new data. The words thus generated are labelled with the category of parts-of-speech they belong to base on the context.

#### G. Naive Bayes Classifier

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity,

Naive Bayes can often outperform more sophisticated classification methods. The Nave Bayes Text Classifier is first trained using the labeled Amazon Review Dataset. After that, it is used for processing of unlabeled data. Following sections elucidates the process step wise.

---

**Algorithm 1** Pseudocode

---

1. Given training data set D which consists of documents belonging to different class say class A and B.
  2. Calculate the prior probability of class A=number of objects of class A / total number of objects  
Calculate the prior probability of class B=number of objects of class B / total number of objects
  3. Find  $n_i$ , the total number of word frequency of each class.  
 $na$ = the total number of word frequency of class A.  
 $nb$ = the total number of word frequency of class B.
  4. Find conditional probability of keyword occurrence given a class.  
 $P(\text{word1} / \text{class A}) = \text{wordcount} / n_i(A)$   
 $P(\text{word1} / \text{class B}) = \text{wordcount} / n_i(B)$   
 $P(\text{word2} / \text{class A}) = \text{wordcount} / n_i(A)$   
 $P(\text{word2} / \text{class B}) = \text{wordcount} / n_i(B)$   
.....  
.....  
 $P(\text{wordn} / \text{class B}) = \text{wordcount} / n_i(B)$
  5. Avoid zero frequency problems by applying uniform distribution.
  6. Classify a new document C based on the probability  $P(C / W)$ .  
a) Find  $P(A / W) = P(A) * P(\text{word1} / \text{class A}) * P(\text{word2} / \text{class A}) * \dots * P(\text{wordn} / \text{class A})$ .  
b) Find  $P(B / W) = P(B) * P(\text{word1} / \text{class B}) * P(\text{word2} / \text{class B}) * \dots * P(\text{wordn} / \text{class B})$ .
  7. Assign document to class that has higher probability.
- 

Fig. 2. Fake review detection using Naive Bayes

#### H. Evaluation Metrics

The efficiency parameters that will be used to tally the results are as follows:

True Positives (TP) - These are the correctly predicted positive values which means that the review class is genuine and the result predicted is also genuine.

True Negatives (TN) - These are the correctly predicted fake reviews which means that the review class is fake and value of predicted class is also fake.

False Positives (FP) When actual review class is fake and predicted class is genuine.

False Negatives (FN) When actual review class is genuine but predicted class in fake.

Once these four parameters have been defined, we can calculate Accuracy, Precision, Recall and F1 score, which are the evaluation metrics. Performance measure is done in the testing phase and can be calculated by how accurately and how frequently the reviews are classified as real or fake.

## IV. EXPERIMENT

### A. Dataset

One of the main challenges encountered by fake review detection is the absence of ground-truth, because fake reviews are difficult to find, even for human readers. Thus, the project treats the duplicates/near-duplicates of reviews as the fake ones.

*Labeled Data:* The project will use the Amazon review data set. This data set contains product reviews and meta data from Amazon, including 142.8 million reviews spanning May 1996 - July 2014. This data set includes reviews (ratings, text, helpfulness votes), product meta data (descriptions, category information, price, brand, and image features), and links (also viewed/also bought graphs).

*Unlabeled Data:* The project will work on unlabeled data from the Amazon review data set and determine whether the review is real or fake. New reviews will be scraped from Amazon live web page and then preprocessed.

## V. CONCLUSION & FUTURE WORKS

It has been observed from the analysis that customers buy various products from various e-commerce websites by reading the reviews the product has so far. But the customer has no idea whether those reviews are authentic or not. The main objective of the project is to remove those fraudulent reviews and give the customer a good online shopping experience as well as to save the customers precious time and money. The main features of the software includes flexibility, quickness, convenience, reliability and efficiency as well as it provides an easy way to detect fake reviews. The software could very well be enhanced further as per the requirements. There exist other kinds of review or reviewer related features that are likely to make a contribution to the prediction task. In the future, we could further investigate different kinds of features to make more accurate predictions. The following are some modifications or possible up gradation that can be pursued in the future-

1. *Duplicate review detection:* Many time reviews are just simply copied, and are used to inflate the positive and/or negative opinion. We would like to be able to detect those reviews.

2. *Meat of the review:* In some cases, the whole review is a spam which turns out to be a sale ad or something completely unrelated. It will be good for the system to tag the part of the review that is most likely related to the product (actual review) and those who are unlikely to be related. Furthermore, the less domain knowledge required the better.

*3. Evaluating the likelihood of fake review:* So far our system flags potential problems to the user without quantifying the likelihood of the review is actually a fake. While we see that as a non-trivial problem, maybe more elaborate machine learning techniques can help us generate such likelihoods.

#### REFERENCES

- [1] M.N. Istiaq Ahsan, Tamzid Nahian, Abdullah Ali Kafi, Md. Ismail Hossain, Faisal Muhammad Shah Review Spam Detection using Active Learning. Department of Computer Science Engineering Ahsanullah University of Science Technology 2016 IEEE.
- [2] Sentimental Analysis of Flipkart reviews using Nave Bayes and Decision Tree algorithm - Gurneet Kaur, Abhinash Singla, International Journal of Advanced Research in Computer Engineering Technology (IARCET) Volume 5 Issue 1, January 2016
- [3] Enhancing Performance of Naive Bayes in text Classification of Introducing an Extra Weight using less number of training samples - Shahnaj Parvin Shathi, Md. Delowar Hossain, Md. Nadim, Sayed Golam Rasul Riyadh, Tangina Sultana, Department of Computer Science and Engineering, Department of Electronics and Communication Engineering, Hajee Mohammad Danesh Science and Technology University, Bangladesh
- [4] A text classifier of English Movie Reviews Based on Information Gain - Lianjing Jin, Wei Gong, Wenlong Fu, Hongbin Wu, Information Engineering School, Communication University of China, CUC Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture Beijing Key Laboratory of Modern Entertainment Technology Beijing, China
- [5] Lim, Ee-Peng, et al. "Detecting product review spammers using rating behaviors." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.
- [6] Performance Analysis and Improvement of Nave Bayes in Text Classification Wei Zhang, Feng Gao, IEEE
- [7] Some Effective Techniques for Nave Bayes Text Classification Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng