_____

Efficient and Trustworthy Review/Opinion Spam Detection

Sanketi P. Raut
ME II (Computer),
Lokmanya Tilak College of Engineering,
New Mumbai,
Email:sanketiraut28@gmail.com

Prof. Chitra Wasnik
Computer Dept.
Lokmanya Tilak College of Engineering,
New Mumbai,
Email:chitrawasnik@yahoo.com

Abstract - The most common mode for consumers to express their level of satisfaction with their purchases is through online ratings, which we can refer as Online Review System. Network analysis has recently gained a lot of attention because of the arrival and the increasing attractiveness of social sites, such as blogs, social networking applications, micro blogging, or customer review sites. The reviews are used by potential customers to find opinions of existing users before purchasing the products. Online review systems plays an important part in affecting consumers' actions and decision making, and therefore attracting many spammers to insert fake feedback or reviews in order to manipulate review content and ratings. Malicious users misuse the review website and post untrustworthy, low quality, or sometimes fake opinions, which are referred as Spam Reviews.

   In this study, we aim at providing an efficient method to identify spam reviews and to filter out the spam content with the dataset of gsmarena.com. Experiments on the dataset collected from gsmarena.com show that the proposed system achieves higher accuracy than the standard naïve bayes.

Keywords - Review _Spam detection, Opinion, Text mining, WordNet, naïve bayes classifier, logistic Regression._

_____ ***** _____

I.   INTRODUCTION

The description of spam is not straightforward as this phenomenon is available in different forms and on different media. From all of these, one of the most widely recognized form of spam is Review spam or opinion Spam [23]. The term Spam is applied to similar abuses in other media such as Internet, Cellular Networks and VoIP platforms. While concentrating Spam related to internet there are again many forms of Spam as Email spam, also known as unsolicited bulk email (UBE), junk mail, or unsolicited commercial email (UCE), is the practice of sending unwanted email messages, frequently with commercial content, in large quantities to an indiscriminate set of recipients. Instant Messaging Spam makes use of Instant messaging (IM), it is a type of online chat which offers real-time text transmission over the Internet. Spam in blogs which is also called simply blog spam or social spam, refers to copying material from elsewhere that is not original, or promoting commercial services to blogs, wikis, guestbook etc. Video sharing sites includes Spam like posting links to sites, most likely pornographic or dealing with online dating, on the comments section of random videos or user profiles[23].

         Every consumer needs secure trustworthy, reliable and easily available information about any product while making purchase. Opinions/Reviews are middle of almost all human actions and are main influencers of our behaviours. In recent times, individuals are using web for everything, they use web to solve their queries, to find solutions of unsettled problems, to know about not so known products or services etc. They also use web, to know opinions of others before concluding their decision on purchase of a new product or service. Review system plays an important role in decision making as many People use the reviews while deciding quality of product to purchase the top quality product. Companies or merchants use opinions or feedback to take a decision to improve their sales according to intellectual things done by other challengers.

The Existing system makes use of social relationships between users of system. It is based on the identification of review spammers by incorporating social relations based on two assumptions that people are more likely to consider reviews from those connected with them as trustworthy, and review spammers are less likely to maintain a large relationship network with normal users and shows a strong correlation between social Relationships and the overall trustworthiness scores [3]. The standard collaborative filtering (CF) is not suitable for the new users who do not have much review history. The Existent method is not attack

_____

resistant, as fake reviews on system can affect the rating score and resulting in the misleading to other users. This paper proposes a method that will be identifying the fake review by first, analysing the review 1content, detecting whether the content is spam or not a spam and the providing a spam free result. The crucial goal of opinion spam detection in the review framework is to identify every fake review and fake reviewer.

II. MOTIVATION

We accept as truth that as opinions on the Web are gradually more used in practice by consumers, administrations, and companies for their decision making. These reviews or opinions are advantageous only if the reviews posted appropriately without any incorrect intention. Day by day Opinion spamming is getting worse and also more sophisticated. Detecting spam reviews or opinions have become more and more critical. The position is at present quite bad. According to our annual Local Consumer Review Survey, the importance of online reviews in the purchasing decision is growing day by day[23].

Eighty-five percent of consumers are satisfied once they have read up to 10 online reviews for local businesses and to make purchases. Many customers even say that they wouldn't buy product without reading online reviews. Reviews become even more important in the situation where customers are not able to test products practically prior to purchasing it.   But these reviews have some drawbacks which give motivation to review spam detection- bad publicity, some reviews are fake given for supporting or devaluing product.

When intention to give an opinion is not true, such opinion can be spam. Opinion spamming has become a most important issue. Apart from persons who give fake feedback in reviews and discussions, there are also business-related companies that are in the business of writing fake reviews and bogus information for their clients. There is need to detect such spamming activities to make sure that the Opinions/Reviews on the Web are a trustworthy source of important information. Therefore there is need to develop a system that will be helpful for managing reputation for business and avoiding people from getting mislead..

III. OBJECTIVE

> Our aim is to provide an Efficient system that will identify "illegal" activities (e.g., writing fake reviews/ Opinions) that try to mislead readers

> Developing a system that will filter out such spam content from reviews and provide users trustworthy opinions/feedbacks.

Opinion spam gives an untruthful opinion on a certain topic or product. It can be found among reviews and commentaries on e-commerce web sites, review Web sites, etc. The spammers try to promote or damage the reputation of people, businesses, products, or services by posting untruthful opinions [23].

The main challenge of opinion spam detection is that dissimilar to other forms of spam, it is very hard to recognize fake feedback by manually reading them. It makes the process of finding such fake data quite difficult. There is no mode to detect such fake review without considering information beyond the review text itself simply because the same opinion cannot be together truthful and fake at the same time.

IV.   LITERATURE SURVEY

The whole literature review is focused on the following literary works being done by an array of scholars and researchers from the field of Review Spam Detection. The following papers are selected for review keeping in mind the traditional and conventional approaches of Spam detection along with the emerging techniques.

4.1 LITERATURES REVIEWED
4.1.1 Spam Filtering by Semantics-based Text Classification 2016

This paper, we described a novel and efficient Chinese spam filtering approach based on semantic information delivered in the body text of emails. The fundamental step is the extracting of semantic information from texts, which will be treated as feature terms for classification later. The extraction of semantic information of text was achieved by attaching semantic annotations on the words and sentences of it. We get these feature terms through attaching annotations on text layer-by-layer, then these terms are used to build up the decision tree and selected by pruning. The method of adding annotations on text is usually applied to the pre-processing of text in natural language processing. The application of text classification in semantic extraction and feature selection is limited because of the low training speed.

4.1.2 Trust-Aware Review Spam Detection 2015

The focus is on the problem of detecting review spammers using contextual social relationships that are available in several online review systems. We first present a trust-based rating predication algorithm using local proximity derived from social relationships, such as friendships and Complements relationships, using the random walk with restart. Results show a strong correlation between social relationships and the computed trustworthiness scores.

Model works under the assumption that review spammers tend to be socially inactive. Many of them would be isolated or barely connected with other users in the system. Our prediction model only aggregates the ratings from trusted users, which potentially filters out the influence of spammers. Experiments on the collected Yelp dataset show that the proposed trust based prediction achieves a higher accuracy than standard CF method.

4.1.3 Spam Mails Filtering Using Different Classifiers with Feature Selection and Reduction Techniques 2015

Work proposes a methodology to detect an email as spam or legitimate mail on the basis of text categorization. Various techniques for pretreatment of email format are applied such as applying stop words removing, stemming, feature reduction and feature selection techniques to fetch the keywords from all the attributes and finally using different classifiers to segregate mail as spam or ham. The papers have used PCA (Principal Component Analysis) and CFS (Correlation Feature Selection) technique for feature reduction. Methodology is totally based on data mining approach for classifying ham and spam emails from large text and text embedded image datasets. Time taken to build model is less by using CFS comparatively PCA applied on different classifiers which are Naive Bayesian, SVM, Random Forest, Bayes Net. Using CFS saves a lot of time for classifiers to build than using PCA. PCA and CFS reduce the attributes without loss their value. Logistic Model Tree (LMT) classifier produce accurate results comparative to others but takes a lot of computational time. Future researches need to consideration on co-evolutionary problem of the spam filtering at server level, because while the spam filter tries to develop its prediction capacity, the spammer try to develop their spam messages in order to overreach the classifiers.

4.1.4 Opinion Spam Detection Using Feature Selection 2014

Rinki Patel and Priyank Thakkar modeled the problem as the classification problem and Naïve Bayes (NB) classifier and Least Squares Support Vector Machine (LS-SVM) are used on three different representations (Boolean, bag-of-words and term frequency–inverse document frequency (TF-IDF)) of the opinions.In this paper experiments are carried out on widely used gold-standard dataset. The paper focuses on modelling deceptive spam detection task as classification problem with deceptive and truthful as two classes. Experiments are carried out with unigram, bigram and bigram plus sequence of words approaches. Learning a classifier using appropriate number of features improves the accuracy. Detection of spam review using adjective, noun and verbs is not possible.

4.1.5 SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset 2014

The paper proposes a hybrid system of SMS classification to detect spam or ham, using Naïve Bayes classifier and Apriori algorithm. Work done here not only considered each and every word as independent and mutually exclusive but also frequent words as a single, independent and mutually exclusive. Training the system for the first time requires little bit more time than Naïve Bayes Classifier. The main contribution of this paper is better accuracy.

V. SYSTEM ANALYSIS

Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming (e.g., writing fake reviews) to promote or to demote some target products. In recent years, fake review detection has attracted significant attention from both the business and research communities [12]. It has become a common practice for people to read online opinions/reviews for different purposes. For example, if one wants to buy a product, one typically goes to a review site (e.g., amazon.com) to read some reviews of the product. If most reviews are positive, one is likely to buy the product. If most reviews are negative, one will almost certainly not buy it. Positive opinions can result in significant financial gains and/or fames for businesses, organizations and individuals. This, unfortunately, gives strong incentives for opinion spamming.

VI. PRAPOSED SYSTEM

In efficient and trustworthy Review/Opinion Spam Detection scheme we propose a method that will be identifying the fake review from gsmarena.com dataset, by first, analysing the review content, detecting whether the content is spam or not a spam and truthful or non-truthful and the providing a spam free result. The crucial goal of opinion spam detection in the review framework is to identify every fake review. Here, we use technique of Sentiment classification that determines whether an opinion is positive, negative or neural and then applies Featured base-opinion mining that discovers features of a reviewed entity with the intent of acquiring the opinion of a reviewer about that specific feature and providing a spam free content.

Advantages:

❖ Naïve bayes with logistic regression apply intelligent approach to detect data because it examines all aspects of a review message along with keyword checking that classifies a review as a spam or not spam on the basis of single word.

❖ Such result is helpful to both users and vendor application during making their respective decisions as system will be giving Spam free Results.
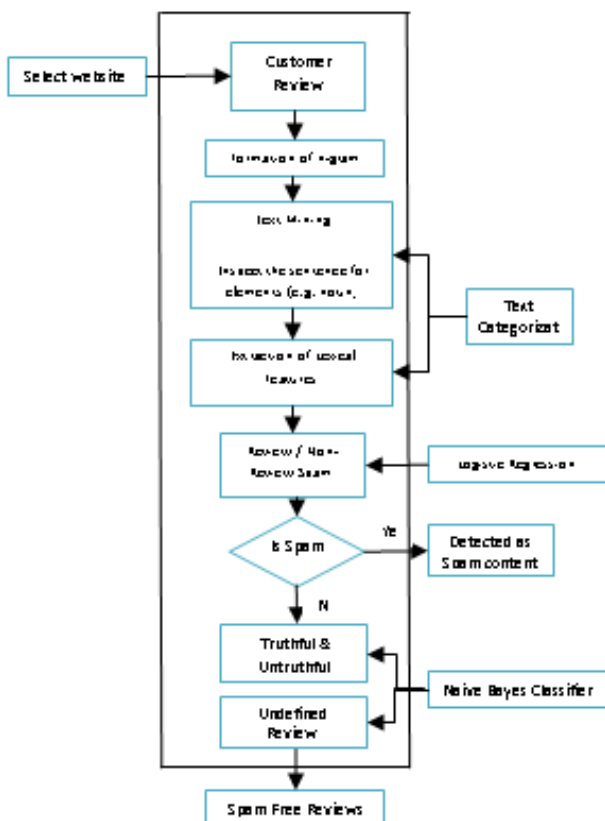
6.1 SYSTEM ARCHITECTURE

Figure 6.1: SYSTEM ARCHITECTURE

6.2 MODULE DESCRIPTION
CUSTOMER REVIEW:

Information about a product is very useful in spam detection such as, the product description and sales volume. Helps to identify whether the review is about the same product?? A feature extraction process must be performed on text documents before they can be mined. This pre-processing step transforms text documents into small units of text called features or terms.
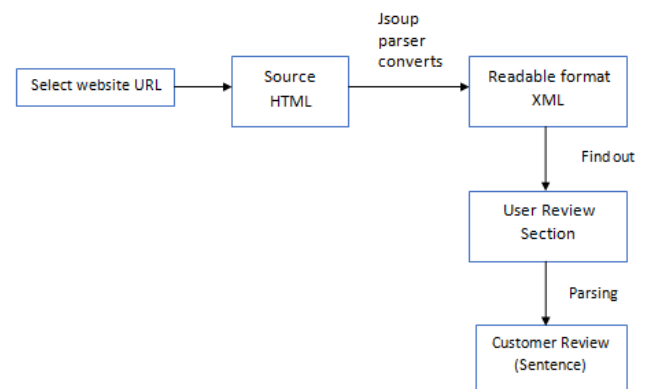

Figure 6.2: Input to System

FORMATION OF N-GRAMS:

The text of a review is called the content of the review. The content of each review is the first thing to be considered in spam detection practice. Linguistic feature such as word and POS n-grams for distinguishing malicious behaviours (for example, deceptions and lies) which can be extracted from content of a review. N-grams of texts are extensively used in text mining and natural language processing tasks. They are basically a set of co-occurring words within a given sentence and when computing the n-grams system typically moves one word forward. For example, for the sentence *"The cow jumps over the moon"*. If N=2 (known as bigrams), then the n-grams would be: the cow, cow jumps, jumps over, over the, the moon etc.

TEXT MINING:

Text mining also referred as text data mining, roughly equal to text analytics,   it is a process of deriving high-quality information from the given text. Text mining is the analysis of data contained in natural language text, mining unstructured data with natural language processing, collecting or identifying a set of textual materials. Text mining will follow text categorization (TC), which performs classification of Review text with respect to a set of one or more preexisting categories.

Text categorization is used for Finding Text duplication and conceptual similarity between reviews. This explores a method that uses WordNet concept for categorizing the Review Text. WordNet is a thesaurus for the English language based on psycholinguistics studies. WordNet ontology will capture the relations between the Review words. It refer as a data-processing resource which covers lexico-semantic categories called synsets (Synonym Sets) [22]. The synsets are sets of synonyms which gather lexical items having similar significances. I.e. certain adjectival concepts which meaning is close are gathered together. The

hyponymy is represented in WordNet is interpreted by "is-a" or "is a kind of" relationships.

REVIEW/ NON-REVIEW SPAM DETECTION:

At this stage system detects whether input sentence is actual Review or Non-Reviews. Non-Reviews include no Opinion. It may contain advertising about variety of products, sellers, other irrelevant things such as questions, answers or similes, some random text etc. To identify and filter out such spam content classifier is useful. Such content are considered as spam because they are not giving any opinion.

TRUTHFUL/UNTRUTHFUL REVIEW SPAM DETECTION:

These are reviews that are written not based on the reviewers' genuine experiences of using the products or services, but are written with some secret intentions.

In such type of Review, the reviewer often post more positive or more negative review about some product. While finding Untruthful reviews input to the system I s set of all reviews about same product, calculate the probability of word sequences of review. Set the Threshold value, and the probability is used to decide whether review is positive or negative. More Positive reviews are the opinion expressing a worthless positive feedback of a product with the intension of promoting that product. More Negative reviews are the opinion, expressing a spiteful negative feedback about a product with the intension of damaging status of product. Language model is useful while detecting and filtering out the spam content.

6.3 ALGORITHM ANALYSIS

Classification predicts categorical class labels. It classifies data (constructs a model) based on the training set and the values (class labels)in a classifying attribute and uses it in classifying new data.

6.3.1 Database creation for filtering:

In order to use Bayesian filtering method, the user needs to generate a database (fig 6.1) with words and tokens which are collected from a sample of spam review and legitimate review [2].
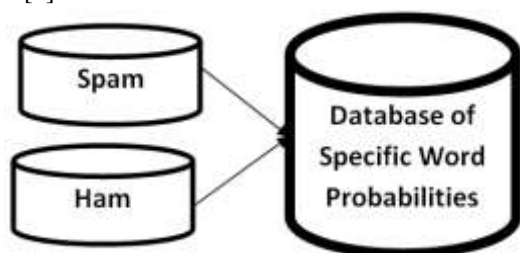

Figure 6.1: Database creation for filters [2]

A probability score value is then assigned to each word or token; the probability is based on calculations that take into account how often that word occurs in spam as opposed to legitimate review (ham). This process can be carried out by analyzing by inspecting known spam: All the words and tokens in both pools of review are analyzed to generate the probability that a particular word points to the review being spam.

Formulation of Naïve bayes and logistic regression is a combination of training and classification or testing and is described in following phases. Bayesian filtering is based on the principle that most events are dependent and that the probability of an event occurring in the future can be guessed from the previous occurrences of that event.

Classification is a two step process. First, Model construction which describes a set of predetermined classes. Second, Model usage used to classify future or unknown objects. To know the accuracy of the model, the known label of test sample is compared with the classified result from the model.

Training Phase:

Step1: Create Spam and Ham sets by collecting many reviews.

Step2: Retrieve individual tokens strings as features words.

Step3: Extract keywords-features that we want to look for in ham messages into HashSet. Extract keywords-features that we want to look for in spam messages into HashSet.

Step4: Prepare HashMap (array set), that will contain all the words that were found in the message with their frequencies for mapping relation of a feature word tokens.

Step5: Compute the class-conditional probability P for feature word w.

Classification Phase

Step1: Retrieve feature words from new review sentence.

Step2: Calculate the probabilities of review sentence for legitimate and Spam classes when it satisfies the extracted feature words from dataset.

Step3: Classify the incoming review based on the results. When the value of probability of Spamsum is greater than value of probability of hamsum or the threshold value λ, the review is treated as spam.

6.3.2 LOGISTIC REGRESSION

Spam detection is a software tool used to classify spam reviews from genuine reviews. Hence the spam detection predicts which class the review belongs to spam/no spam.

Logistic regression is a model used for prediction of the probability of occurrence of an event. It makes use of several predictor variables (features) that may be either numerical or categorical. Regularization parameter used in

logistic regression helps in preventing over fitting. Words that frequently occur in spam email are used as the feature set in the regression problem.
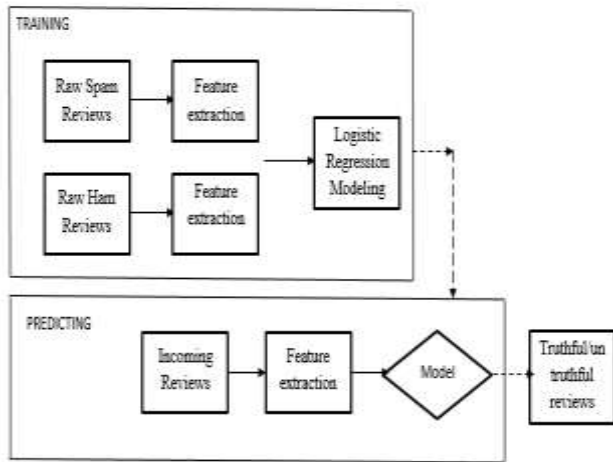

Figure 6.2: Logistic Regression Classifier

6.3.3 NAÏVE BAYES CLASSIFIER

Naive Bayes classifiers are an effective statistical technique of truthful or non-truthful Review spam filtering. They typically use bag of words features to identify spam-reviews, an approach commonly used in text classification. Naive Bayes classifiers work by correlating the use of tokens (typically words, or sometimes other things), with truthful and non-truthful-spam reviews and then using Bayes' theorem to calculate a probability that the review sentence is or is not spam.
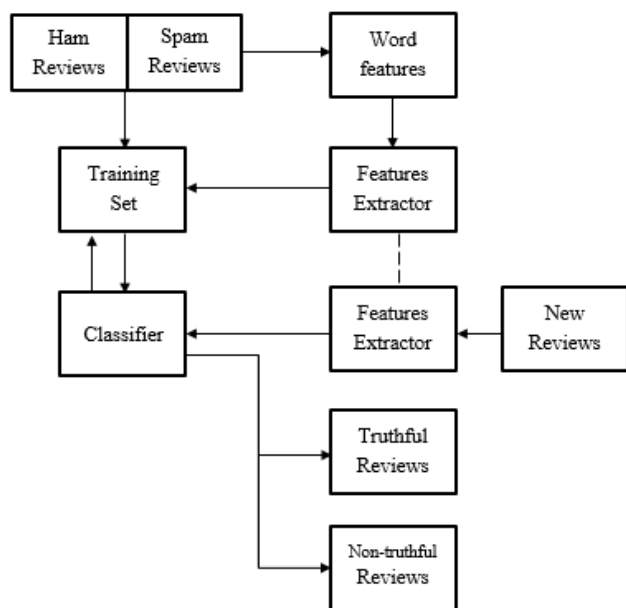

Figure 6.3: Naïve bayes classifier

Particular words have particular probabilities of occurring in non-truthful review and in legitimate truthful reviews. The

filter doesn't know these probabilities in advance, and must first be trained so it can build them up.

To train the filter, the user must manually indicate whether a new review is truthful or not truthful. For all words in each training review dataset, the filter will adjust the probabilities that each word will appear in spam or legitimate review in its database.

After training, the word probabilities (also known as likelihood functions) are used to compute the probability that review sentence with a particular set of words in it belongs to either category. Each word in the sentence contributes to the review's spam probability, or only the most interesting words. This contribution is called the posterior probability and is computed using Bayes' theorem. Then, the spam probability is computed over all words in the review, and if the total exceeds a certain threshold then the filter will mark the review as a non-truthful review. Requires a small amount of training data to estimate the parameters.

Formulation of Naïve bayes and logistic regression is a combination of training and classification or testing and is described in following phases. Bayesian filtering is based on the principle that most events are dependent and that the probability of an event occurring in the future can be guessed from the previous occurrences of that event.

Figure 6.4: home page for Spam Review Detection

Figure 6.5: Entry of selected complete URL of product

Figure 6.6: Scanning of product Reviews

Figure 6.7: Results of scanned Reviews

VII. RESULT ANALYSIS

7.1 EVALUATION OF THE ALGORITHM

In the spam filter, we need to define some indicators to evaluate the spam filter effects in addition to the public corpus. These indicators are generally form text classification and information retrieval [22]. We assume that there are $S(S=A+B+C+D)$ reviews in the test set. To facilitate the narrative, we define some variables in Table I.

	Classified as non spam by system	Classified as spam by system
Non-Spam	A	B
Real Spam	C	D

TABLE 7.1: ADJACENT TABLE

We define the evaluation indicators as follows:

a. Accuracy: Accuracy= $(A+D)/S$ * 100 %
b. Recall: Recall= $D/(D+C)$ * 100 %   This indicator reflects the ability of filtering system to found out spam.

c. Precision: Precision= $D/(D+B)$ * 100 % This indicator reflects the ability of filtering system to found spam correctly.
d. Miss Rate: Miss rate= $C/(D+C)$ * 100 %, the rate of spam which is not identified.
e. Error: Error= $(B+C)/S$ *100 %, the rate of spam which is misclassified.

7.2 EXPERIMENTAL RESULTS

We randomly selected 400 reviews as test set, including 100 spam and 300 legitimate i.e. non-spam reviews. In order to prevent the chanciness of the experimental results, we have calculated the average of the results. The experimental results of proposed system are shown in table II.

	Spam detected by system	Non-spam detected by system	total
spam	97.5	2.5	100
Non-spam	5	295	300
total	102.5	297.5	400

TABLE 7.2: EXPERIMENTAL RESULT OF PROPOSED FILTERING ALGORITHM

At end, result of spam detection is analyzed and decision will be taken on – whether each review is spam or not a spam. Such result is helpful to both users and vendor application during making their respective decisions. System will be giving spam free results.

As individual users and companies use review and opinion for decision making, it is important to detect opinion spam and opinion spammer. This approach mainly concentrates on non- review spam detection, untruthful review spam detection and truthful review spam detection and filtering. The result will give more accuracy to display a spam free Review which is helpful to both customers while buying any product and for company to improve their performance using true reviews.

We know that Naïve Bayesian algorithm is based on "Bayesian assumptions" which assume that each of the characteristics is independent by analyzing the theory. In fact, this assumption is difficult to exist. Experimental data also show that Bayesian algorithm makes important information lost, and leads to misjudging spam and

92

legitimate mail [22]. So, in our system we introduce method for review filtering field, and propose an improved filtering algorithm combined the Naïve Bayesian algorithm with logistic regression method. This method's aim is to reduce the rate of the misjudgement, and to improve the accuracy of classification. The following table III shows comparison of experimental results of the two filtering algorithms with proposed system.

	Naïve Bayesian	Improved Filtering	Proposed (Naïve bayes with logistic regression)
Accuracy	94.25	96.79	98.13
Recall	95.92	97.89	98.33
Precision	96.4	97.83	99.16
Miss Rate	4.08	2.11	1.67
Error	5.75	3.21	1.88

TABLE 7.3: COMPARISON OF EXPERIMENTAL RESULTS OFTHE TWO FILTERING ALGORITHMS WITH PROPOSED SYSTEM
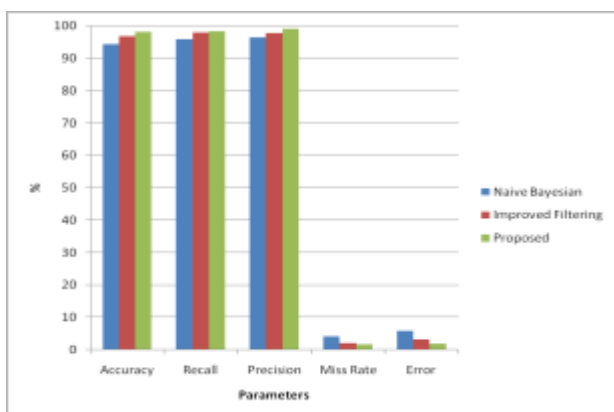

Figure 7.1: Comparison of review spam detection techniques

VIII. CONCLUSION

We have presented recent work related to spam detection is done with naïve bayes classifier, logistic regression classifier and language model, which gives more efficiency and trustworthiness while detecting and filtering the spam content. The results are promising. Supervisors, controllers, organizations can use review spam detection result as an administrative tool to supervise target e-commerce accumulation. Naïve bayes with logistic regression apply intelligent approach to detect data because it examines all aspects of a review message along with keyword checking

that classifies a review as a spam or not spam on the basis of single word. The system gives convenience to administrators, flexible settings are available. It provides efficient and trustworthy opinion and feedback. The model is tremendously appealing because of its simplicity, elegance and robustness.

REFERENCES

[1] Wei Hu, Jinglong Du, and Yongkang Xing, "Spam Filtering by Semantics-based Text Classification" 8th International Conference on Advanced Computational Intelligence, 2016 IEEE.

[2] Muhammad Iqbal1, Malik Muneeb Abid2, Mushtaq Ahmad3 and Faisal Khurshid4, "Study on the Effectiveness of Spam Detection Technologies" *I.J. Information Technology and Computer Science,* 2016.

[3] Hao Xue∗, Fengjun Li∗, Hyunjin Seo†, and Roseann Pluretti , "Trust-Aware Review Spam Detection", EECS Department, †School of Journalism and Mass Communications The University of Kansas Lawrence, KS, USA, 2015 IEEE.

[4] Amit Kumar Sharma, Amit Kumar Sharma," Spam Mails Filtering Using Different Classifiers with Feature Selection and Reduction Techniques", Fifth International Conference on Communication Systems and Network Technologies,2015 IEEE.

[5] Sonia, YMCA University of Science and Technology, "Spam Filtering by Semantic Indexing and Duplicate Detection", International Conference on Computing, Communication and Automation, 2015 IEEE.

[6] Rinki Patel, Priyank Thakkar, Department of Computer Science & Engineering Institute of Technology, "Opinion Spam Detection Using Feature Selection", Sixth International Conference on Computational Intelligence and Communication Networks, 2014 IEEE.

[7] Ishtiaq Ahmed, Donghai Guan, and Tae Choong Chung, "SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset", *International Journal of Machine Learning and Computing,* 2014 IEEE.

[8] Huayi Li∗, Zhiyuan Chen∗, Bing Liu∗, Xiaokai Wei∗ and Jidong Shao† ∗Department of Computer Science, "Spotting Fake Reviews via Collective Positive-Unlabeled Learning", International Conference on Data Mining, 2014 IEEE.

[9] Hamzah Al Najada and Xingquan Zhu, "iSRD: Spam review Detection with Imbalanced Data Distributions", 2014 IEEE.

[10] Cristina Rădulescu, Mihaela Dinsoreanu, Rodica Potolea, Technical University of Cluj-Napoca, "Identification of Spam Comments using Natural Language Processing Techniques", 2014 IEEE.

[11] M.S.Patil∗ , A.M.Bagade, " Review on Brand Spam Detection Using Feature Selection", International Journal of Advanced Research in Computer Science and Software Engineering, 2013.

[12] Arjun Mukherjee†, Abhinav Kumar†, Bing Liu†, Junhui Wang†, Meichun Hsu‡, Malu Castellanos‡, Riddhiman Ghosh‡, "Spotting Opinion Spammers using Behavioral Footprints" University of Illinois at Chicago, August 11–14, 2013,

93

[13] Sihong Xie, Guan Wang, Shuyang Lin, Philip S. Yu "Review spam detection via time series pattern discovery", ACM Proceedings of the 21st international conference companion on World Wide Web,pp.635-636,2012.

[14] *Sanjog Ray,* Information Systems Area ,Indian Institute of Management Indore, "*Identifying Influential Taggers in Trust-Aware Recommender Systems*", ACM International Conference on Advances in Social Networks Analysis and Mining, 2012 IEEE.

[15] Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu University of Illinois at Chicago, "Review Graph based Online Store Review Spammer Detection", 11th IEEE International Conference on Data Mining, 2011 IEEE

[16] Saadat Nazirova , *Institute of Information Technology of Azerbaijan National Academy of Sciences,"* Survey on Spam Filtering Techniques*" Communications and Network*, 2011.

[17] RAYMOND Y. K. LAU, S. Y. LIAO, and RON CHI-WAI KWOK, KAIQUAN XU, YUNQING XIA, YUEFENG LI, "Text Mining and Probabilistic Language Modeling for Online Review Spam Detection", ACM Transactions on Management Information Systems,2011.

[18] Guangyu Wu, Derek Greene, Pádraig Cunningham School of Computer Science and Informatics University College Dublin," Merging Multiple Criteria to Identify Suspicious Reviews", ACM 2010.

[19] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, Hady Wirawan Lauw ,"Detecting product review spammer using rating behaviors", Proceedings of the 19th ACM international conference on Information and knowledge management, pp-939-948,2010.

[20] Paulo Cortez, Clotilde Lopes, Pedro Sousa, Miguel Rocha, Miguel Rio, "Symbiotic Data Mining for Personalized Spam Filtering", WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, 2009 IEEE.

[21] Zakaria Elberrichi1 , Abdelattif Rahmoun2 , and Mohamed Amine Bentaalah1 1 EEDIS Laboratory, Department of Computer Science, University Djilali Liabès, "Using WordNet for Text Categorization" Algeria 2King Faisal University, Saudi Arabia,2008.

[22] Liu Pei-yu, Zhang Li-wei, Zhu Zhen-fang," Research on E-mail Filtering Based On Improved Bayesian", JOURNAL OF COMPUTERS, VOL. 4, NO. 3, MARCH 2009.

[23] Sanketi P. Raut, Prof. Chitra Wasnik, "*A Study on: Opinion/Review Spam Detection*", May 16 Volume 4 Issue 5 , International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC),2016.