# Enhanced Text Mining Techniques For Spam Review Detection

Mr.Akshat Uike[1], Mr Ram Deshmukh[2], Dr.S R.Gupta[3], Dr. S.W.Ahmad[4]

[1, 2] *ME Student, CSE, Prof Ram Meghe Institution of Technology & Research, Badnera*
[3,4] *Asst. Professor, CSE, Prof Ram Meghe Institution of Technology & Research, Badnera*

## ABSTRACT

*Online reviews play a very dominant role in today's e-commerce for decision-making. Large part of the population i.e. customers read reviews of products or stores before making the decision of what or from where to buy and whether to buy or not. Review sites are more and more confronted with the spread of misinformation, i.e., opinion spam, which aims at promoting or damaging some target businesses, by misleading either human readers, or automated opinion mining and sentiment analysis systems. For this reason, in the last years, several data-driven approaches have been propose to assess the credibility of user-generated content diffusing through social media in the form of on-line reviews. As writing fake/fraudulent reviews comes with monetary gain, there has been a huge increase in deceptive opinion spam on online review websites. Basically fake review or fraudulent review or opinion spam is an untruthful review. Positive reviews of a target object may attract more customers and increase sales; negative review of a target object may lead to lesser demand and decrease in sales. These fake/fraudulent reviews are deliberately written to trick potential customers in order to promote/hype them or defame their reputations. Our work is aimed at identifying whether a review is fake or truthful one. Naïve Bayes Classifier, Logistic regression and Support Vector Machines are the classifiers using in our work.*

*This paper aims to classify movie reviews into groups of positive or negative polarity by using machine learning algorithms. In this study, we analyses online movie reviews using SA methods in order to detect fake reviews. SA and text classification methods are applied to a dataset of movie reviews. More specifically, we compare five supervised machine learning algorithms: Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK) for sentiment classification of reviews using two different datasets, including movie review dataset and movie reviews dataset. The measured results of our experiments show that the SVM algorithm outperforms other algorithms, and that it reaches the highest accuracy not only in text classification, but also in detecting fake reviews.*

*Keyword : - Amazon E-Commerce dataset, Dataset acquisition, Data preprocessing, Active Learning, Rough Set Classifier, KNN Classifier, Support Vector Machine.*

## 1. INTRODUCTION

The Internet has changed our lives since it was introduced. With rapidly expansion and usage of Internet people are now totally dependent on the web services, which also changed people's behavior of communicating and expressing their views. People post their views in their respective discussion groups, forums, social media, and blogs and in e-commerce website for a product/service. These contents are user generated which are written in natural language. Opinion sharing on a product/service is based on their personal experience which is called as reviews.

Reading about product reviews before buying the product becomes a habit, especially for potential customers. If customer wants to buy a product, they usually read reviews of other customers about the current product. If the review is positive, there is a big chance to purchase the product, otherwise if review is negative, they tend to buy other product. While, for an industry/company, the positive reviews from customers can generate decisive financial benefits for companies, the positive review can be using as input for decisions related to product design and what services are provided to customers.

In order to solve this malignant problem, we propose an interactive semi-supervised model to identify fake reviews which is evaluated later on using real life data and compared with some sophisticated prior

research work. In this paper we propose a novel approach of Active Learning to detect fake reviews. We have using original dataset of Amazon reviews. Make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products.

The fact that is if the potential customer or users gets a genuine Reading product reviews before buying the product becomes a habit, especially for potential customers. If they want to buy a product, they usually read reviews from some customers about the current product. If the review is mostly positive, there is a big chance to buy the product, otherwise if it's mostly negative, they tend to buy other products. While, for a company, the positive reviews from customers can generate significant financial benefits for businesses, it can be using as input for decisions related to product design and what services are provided to customers

Opinion mining has played a significant role in providing product recommendations to users. Efficient recommendation systems help in improving business and also enhance customer satisfaction. The credibility of purchasing a product highly depends on the online reviews. However many people wrongly promote or demote a product by buying and selling fake reviews. Many websites have become source of such opinion spam. This in turns leads to recommending undeserving products. This literature survey is done to study the various fake review detection techniques in detail and to get ourselves familiar with the works done on this subject.

Online reviews have become an important factor when people make purchase and business decisions. The increasing popularity of online reviews also stimulates the business of fake review writing, which refers to paid human writers producing deceptive reviews to influence readers' opinions. Our project tackles this problem by building a classifier that takes the review text and the basic information of its reviewer as input and outputs whether the review is reliable. The learning algorithms we experimented include logistic regression, linear discriminant analysis, multinomial Naive Bayes, support vector machines and neural networks. The results show that the neural network performs the best with a detection accuracy of 81.92%.

The rapid spread of Web technologies and the Internet has brought about a tremendous amount of user-generated content. At the same time, an increasing number of activities have turned to an online model, capitalizing on the World Wide Web's spread and advancement. E-commerce is one of the fields that have prospered in this environment, allowing consumers to engage in online transactions in an unprecedented ease of use. Buyers have the ability to contribute their opinions, through reviews, to provide criticism on the quality and value of purchased products and services. An increasing amount of prospective customers consider the general public opinion before spending their fund by reading online reviews. The effect of positive or negative opinions is therefore of paramount importance to businesses and service providers.

This led to the appearance of shady practices, involving the creation of deceptive online reviews so as to promote or demote products by manipulating consumer opinion. This opinion or review spamming, is considered a significant issue in today's online world and already, research has been undertaken to identify fraudulent content in online reviews and subsequently, protect consumer interests. So far, research has been divided among detection of spam reviews, review spammers and spammer groups, and a number of efforts have yielded promising results by capitalizing on certain clues and traits of spamming activity and developing dedicated techniques and analysis methods.

This Review Paper contributes its own approach to the field of opinion spam detection by considering spamming clues linked to both reviews and reviewers. The examined indicators of spam activity include review content, rating and timestamp but the real focus is on review author behavior, as an established spammer is directly linked to highly suspicious reviews. Review author behavior analysis is twofold, including both activity on a particular product but also past reviewing activity, as a reviewer's history of contributions is an indication of their overall reputation as an author. An additional employed technique is responsible for detecting bursts in a product's reviewing activity. And while literature focus has been shifted from review content categorization to primarily spammer behavior analysis, another contribution by this Review Paper involves an examination of review context by utilizing a deep learning classifier and approach for the first time, with the intention of ascertaining its usefulness in detecting opinion spam.

The propose methodology is evaluated on a dataset of real reviews from Amazon in accordance with the rest of the literature. Despite the absence of ground truth annotated data, which is a major issue in evaluating opinion spam detection methods, an alternative evaluation method is employed and the reported results display positive detection accuracy, attesting to the chosen spam indicators' reliability. The ultimate goal is to present an effective spam filtering system, ready to be put to practical use thanks to the straightforward approach of the propose methodology.

Fake reviews have become a pervasive problem in online review systems, wherein fraudulent users manipulate the perception of an object (e.g., a restaurant) by fabricating fake reviews. Extensive work has been devoted to identifying fake reviews via modeling different factors separately, such as user features, object characteristics, and user-object bipartite relations. However, this problem remains challenging due to the fact that more advanced camouflage strategies are utilized by malicious users. In real-world scenarios, spammers may pretend to be normal users by giving fake reviews with the similar score distribution as normal users. To address these issues, we propose to explore the temporal patterns of users' review behavior, because spammers

prefer to promote or demote the target businesses in a short period of time. In this work, we present a unified Reliable Fake Review Detection that explicitly models temporal patterns of users' review behavior into a probabilistic generative model. Moreover, the models users' underlying review credibility and objects' highly-skewed review distributions. We conduct experiments on two datasets.

Opinion Mining (OM), also known as Sentiment Analysis (SA), is the domain of study that analyzes people's opinions, evaluations, sentiments, attitudes, appraisals, and emotions towards entities such as services, individuals, issues, topics, and their attributes. "The sentiment is usually formulated as a two class classification problem, positive and negative". Sometimes, time is more precious than money, therefore instead of spending time in reading and figuring out the positivity or negativity of a review, we can use automated techniques for Sentiment Analysis.

The basis of SA is determining the polarity of a given text at the document, sentence or aspect level, whether the expressed opinion in a document, a sentence or an entity aspect is positive or negative. More specifically, the goals of SA are to find opinions from reviews and then classify these opinions based upon polarity. According to, there are three major classifications in SA, namely: document level, sentence level, and aspect level. Hence, it is important to distinguish between the document level, sentence level, and the aspect level of an analysis process that will determine the different tasks of SA. The document level considers that a document is an opinion on its aspect, and it aims to classify an opinion document as a negative or positive opinion. The sentence level using SA aims to setup opinion stated in every sentence. The aspect level is based on the idea that an opinion consists of a sentiment (positive or negative), and its SA aims to categorize the sentiment based on specific aspects of entities.

The documents using in this work are obtained from a dataset of movie reviews that have been collected. Then, an SA technique is applied to classify the documents as real positive and real negative reviews or fake positive and fake negative reviews. Fake negative and fake positive reviews by fraudsters who try to play their competitors existing systems can lead to financial gains for them. This, unfortunately, gives strong incentives to write fake reviews that attempt to intentionally mislead readers by providing unfair reviews to several products for the purpose of damaging their reputation. Detecting such fake reviews is a significant challenge. For example, fake consumer reviews in an e-commerce sector are not only affecting individual consumers but also corrupt purchaser's confidence in online shopping. Our work is mainly directed to SA at the document level, more specifically, on movie reviews dataset. Machine learning techniques and SA methods are expected to have a major positive effect, especially for the detection processes of fake reviews in movie reviews, e-commerce, social commerce environments, and other domains.

In machine learning-based techniques, algorithms such as SVM applied for the classification purposes. SVM is a type of learning algorithm that represents supervised machine learning approaches, and it is an excellent successful prediction approach. The SVM is also a robust classification approach. A recent research presented in introduces a survey on different applications and algorithms for SA, but it is only focusing on algorithms using in various languages, and the researchers did not focus on detecting fake reviews. This paper presents five supervised machine learning approaches to classify the sentiment of our dataset which is compared with two different datasets. We also detect fake positive reviews and fake negative reviews by using these methods. The main goal of our study is to classify movie reviews as a real reviews or fake reviews using SA algorithms with supervised learning techniques.

## 2. RESEARCH OBJECTIVES

Individuals and organizations increasingly use reviews from the social media for:
1. For making decisions relating to product purchases
2. For product designing and marketing
3. To make election choices
4. 31% of consumers read online reviews before actually making a purchase (rising)
5. By the end of 2014, 15% of all social media reviews will consist of company paid fake reviews

### 2.1 Our main objectives are:

1. Working on Amazon reviews data set.
2. In review data set we need to identify that review given by user was genuine or fake.
3. NLP based sentiment analyzer and text mining algorithms we will be using to classify, predict (positive, negative and neutral) reviews.
4. Any occurrence of unrelated (non referrals) words will leads to fake review.
5. Fake reviews- Unauthorized, Non trustworthy, Contents of unrelated words.

## 3. LITERATURE REVIEW

Traditional methods of data analysis have long been using to detect fake/fraudulent reviews. In effort to meet this goal researchers have turned to the fields of machine learning and artificial intelligence. A review can be classified as either fake or genuine either by using supervised and/or unsupervised learning techniques.

A number of studies have been conducted which focusing on spam detection in e-mail and on the web, however, only recently have any studies been conducted on opinion spam. [1]Jindal and Liu (2008) have worked on "Opinion Spam and Analysis" and have found that opinion spam is widespread and different in nature from either e-mail or Web spam. They have classified spam reviews into 3 types: Type 1, Type 2 and Type 3. Here Type 1 spam reviews are untruthful opinions that try to mislead readers or opinion mining systems by giving untruthful reviews to some target objects for their own gains. Type 2 spam reviews are brand only reviews, those that comment only on the brand and not on the products. Type 3 spam reviews are not actually reviews, they are mainly either advertisements or irrelevant reviews which do not contain any opinions about the target object or brand. Although humans detect this kind of opinion spam they need to be filtered, as it is a nuisance for the end user. Their investigation was based on 5.8 million reviews and 2.14 million reviewers (members who wrote at least one review) crawled from amazon.com and they have discovered that spam activities are widespread. They have regarded spam detection as a classification problem with two classes, spam and non-spam. And have built machine-learning models to classify a review as either spam or non-spam. They have detected type 2 and type 3 spam reviews by using supervised learning with manually labeled training examples and found that the highly effective model is logistic regression model. However, to detect type 1 opinion spam, they would have had to manually label training examples. Thus they had to use duplicate spam reviews as positive training examples and other reviews as negative examples to build a model.

In the paper[2] Ott, et al. 2011, they have given focus to the deceptive opinion spam i.e. the fictitious opinions which are deliberately written to sound authentic so as to deceive the user. The user cannot easily identify this kind of opinion spam. They have mined all 5-star truthful reviews for 20 most famous hotels in Chicago area from trip advisor and deceptive opinions were gathered for the same hotels using amazon mechanical trunk (AMT). They first asked human judges to evaluate the review and then they have automated the task for the same set of reviews, and they found that automated classifiers outperform humans for each metric. The task was viewed as standard text categorization task, psycholinguistic deceptive detection and genre identification. The performance from each approach was compared and they found that the psycholinguistic deceptive detection and genre identification approach was outperformed by n-gram based text categorization, but a combined classifier of n-gram and psychological deception features achieved nearly 90% cross-validated accuracy. Finally they came into a conclusion that detecting deceptive opinions is well beyond the capabilities of humans. Since then, various dimensions have been explored: detecting individual (Lim et al., 2010) and group spammers (Mukherjee et al., 2012), time-series (Xie et al., 2012) and distributional analysis (Feng et al., 2012a)

In paper [9]Yoo and Gretzel (2009) gather 40 truthful and 42 deceptive hotel reviews and, using a standard statistical test, they have manually compared the psychologically relevant linguistic differences between them. In (Mukherjee, et al., 2013), authors have briefly analyzed "What yelp filter might be doing?" by working with different combination of linguistic features like unigram, bigram, distribution of parts of speech tags and yielding detection accuracy. Authors have found that a combination of linguistic and behavioral features comparatively yielded more accuracy.

In paper [12] Prof. M. A..Pund et al(2013) presents an innovative and effective pattern discovery technique which includes the processes of pattern deploying and pattern evolving, to improve the effectiveness of using and updating discovered patterns for finding relevant and interesting information. In proposed system they can take sufficient .txt file as inputs & they apply various algorithms & generate expected results.

When making purchasing decisions, customers increasingly rely on opinions posted on the Internet (Hu, Bose, Koh & Liu, 2012). Internet users can easily and openly express their opinion about a product or brand by using social media or online product reviews and reach up to millions of potential buyers. With the assistance of opinion mining tools, businesses can retrieve valuable information with regard to product, service and marketing improvements from this kind of user-generated content (Heydari, Tavakoli, Salim & Heydari, 2016).

Online opinions thus can have great impact on brand and product reputation as well as related sales and management decisions. This gives an incentive to businesses to create, for example, positive fake reviews on their own products and negative fake reviews on their competitors' products (Akoglu & Rayana, 2015). There is a variety of ways to spam the internet with fake content. For instance, by hiring professional firms which are specialized in writing spam reviews, by using crowdsourcing platforms to employ review spammers or by using robots to create synthetic reviews. Reviews produced by someone who has not personally experienced the subjects of the reviews are called spam reviews (Heydari et al, 2015). The person who creates the spam review is called an individual review spammer. Individual review spammers working together with other review spammers are group spammers (Mukherjee, Liu & Glance, 2012).

Due to the amount of reviews posted online and the proficiency of review spammers, it is often very hard to detect spam reviews and separate them from trustworthy ones. However, it is essential not only for businesses but also for customers that review spam can be identified and removed in a reliable way. Researchers have suggested a variety of methods and tools to identify spam reviews, review spammers and spammer groups (e.g. Jindal & Liu, 2008; Mukherjee et al, 2012; Xie et al, 2012). One of these tools is reviewskeptic.com1 developed by Ott, Choi, Cardie, & Hancock (2011). The authors claim that review skeptic is able to detect spam reviews on hotels based on psychological and linguistic criteria with 90% accuracy. However, hotel reviews are only a fraction of the opinions posted on the Internet. Many reviews are related to individual products, services, brands or stores. Review skeptic claims to be a well-working yet very specialized tool for spam review detection. The aim of this research is to assess review skeptic's performance on non-hotel reviews and based on the existing literature to give recommendations on how the tool could be enhanced to detect also non-hotel review spam effectively.

The identification of spam reviews will be a relevant research topic as long as opinions will be expressed on the internet. Not only the tools for detection are improving but also the ways of producing review spam are getting more advanced. For example, sellers on Amazon 2 now have the opportunity to provide their products for free or at a discount in exchange for a review. Thereby, the review is still marked as a verified purchase and thus seems more trustworthy to potential buyers and to conventional review spam detection methods. However, the honesty of the reviews obtained in this way is highly questionable (Bishop, 2015). This example shows the importance of developing, testing and improving new methods for spam review detection which can keep up with the novel ways of producing spam reviews constantly.

In paper SP.Rajamohana, et al [18] 2017 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.

## 3. PROPOSE SYSTEM

Propose system will automatically classify user opinions into fake, genuine and neutral. This automatic system can be useful to business organization as well as to customers. Business organization can monitor their product selling by analyzing and understand what the customers are saying about products. Customers can make decision whether he/she should purchase or not purchase the products. This can helpful to people to buy valuable product and spend their money on quality products.

The posted reviews are useful only if reviews posted without any incorrect intention. Online survey shows that around 90% of customers are satisfied after reading 10 on 10 rating and good review. Many customer decide to buy product or not only by following reviews, But when intention of a person is not good, behind giving review such opinion. There is needed to detect such activities to make sure that opinions/reviews on the web are trustworthy source of information. Therefore, it is very important to know if the reviews are genuine because fake reviews lead to false reputation of the product and mislead the user.

In order to solve this malignant problem, we propose an interactive semi-supervised model to identify fake reviews which is evaluated later on using real life data and compared with some sophisticated prior research work. In this paper we propose a novel approach of Active Learning to detect fake reviews. We have using original dataset of Amazon reviews. Make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products.

The proposed methodology is evaluated on a dataset of real reviews from Amazon in accordance with the rest of the literature. Despite the absence of ground truth annotated data, which is a major issue in evaluating opinion spam detection methods, an alternative evaluation method is employed and the reported results display positive detection accuracy, attesting to the chosen spam indicators' reliability. The ultimate goal is to present an effective spam filtering system, ready to be put to practical use thanks to the straightforward approach of the proposed methodology.

This paper aims to classify movie reviews into groups of positive or negative polarity by using machine learning algorithms. In this study, we analyses online movie reviews using SA methods in order to detect fake reviews. SA and text classification methods are applied to a dataset of movie reviews. More specifically, we compare five supervised machine learning algorithms: Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK) for sentiment classification of reviews using two different datasets, including movie review dataset and movie reviews dataset.

## 4. CONCLUSIONS

In this way we proposing system will automatically classify user opinions into fake, genuine and neutral. This automatic system can be useful to business organization as well as to customers. Business organization can monitor their product selling by analyzing and understand what the customers are saying about products. Customers can make decision whether he/she should purchase or not purchase the products. This can helpful to people to buy valuable product and spend their money on quality products.

This paper proposes an ensemble methodology for identifying Fake Review by renowned learning method (Active Learning) using real life data. We using several methods to analyze a dataset of Amazon product reviews. We worked on sentiment classification algorithms to apply a supervised learning on electronic products of amazon reviews. Our experimental approaches calculated the accuracy of Rough Set Classifier, KNN Classifier and SVM i.e. sentiment classification algorithms. Additionally, we were able to classify how many given review are fake and genuine. We have using various algorithms i.e. Rough Set Classifier, KNN Classifier and SVM (Support vector machine) to identify fake and genuine reviews. We using three supervised learning algorithms to classifying Sentiment of our dataset have been compared in this paper Rough Set Classifier, KNN Classifier and SVM Classifier. Using the accuracy analysis for these three techniques, we try to find that Rough Set Classifier is the most accurate for correctly classifying count of fake and genuine reviews. Hopefully method shows very promising results. Detection processes for fake and genuine reviews depend on the best method that is using in this project.

## 5. REFERENCES

[1] Jindal, N., Liu, B.: "Opinion Spam and Analysis" in Proceedings of the International Conference on Web Search and Web Data Mining (pp. 219–230), 2008.

[2] Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding Deceptive Opinion Spam by any Stretch of the Imagination. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, pp. 309–319 (2011), 2011.

[3]Jindal, N., Liu, B. and Lim, E.-P. 2010. Finding Unusual Review Patterns Using Unexpected Rules. CIKM (2010)., 2010 92.

[4]Mukherjee, A., Liu, B., Wang, J., Glance, N. and Jindal, N. 2011. Detecting Group Review Spam. WWW (2011)., 2011

[5] Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In Proceedings of the ACM international conference on world wide web (pp. 191–200). ACM., 2012

[6] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013a). Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews, UIC-CS-03- 2013. Technical Report.

[7] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures onhuman language technologies, vol. 5, no. 1, 2012, pp. 1–167.

[8]RAYMOND Y. K. LAU, S. Y. LIAO, RON CHI WAI KWOK, KAIQUAN XU, YUNQING XIA, YUEFENG LI TextMining and Probabilistic Language Modeling for Online Review Spam Detection ACM Trans. Manag. Inform. Syst. 2, 4, Article 25 (December 2011)

[9]"Comparison of deceptive and truthful reviews" by Yoo and Gretzel (2009)

[10]Xie, S.,Wang, G., Lin, S., & Yu, P. S. (2012). Review spam detection via temporal pattern discovery. In Proceedings of the ACM international conference on knowledge discovery and data mining (pp. 823–831). ACM., 2012.

[11] Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. In Proceedings of the ICWSM. Citeseer., 2013

[12] Miss Dipti S.Charjan , Prof. Mukesh A.Pund " Pattern Discovery For Text Mining Using Pattern Taxonomy" (IJETT) Volume 4 Issue 10- October 2013.

[13]Lin, Y., Zhu, T., Wang, X., Zhang, J., & Zhou, A. (2014b). Towards online antiopinion spam: Spotting fake reviews from the review sequence. In 2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) (pp. 261–264). IEEE.

[14]Hernandez-Fusilier, D., Montes-y-Gomez, M., Rosso, P., Guzman-Cabrera, R.: Detection of Opinion Spam with Character n-grams, 2015

[15] Heydari, A., Tavakoli, M., Salim, N.: Detection of fake opinions using time series, Expert Systems with Applications, 58, 83-92, 2016

[16] Ye, J., Kumar, S., Akoglu, L., Temporal Opinion Spam Detection by Multivariate Indicative Signals, 2016.

[17] Liu, Pan, et al. "Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features." Software Quality, Reliability and Security Companion (QRS-C), 2017 IEEE International Conference on. IEEE, 2017.

[18] SP.Rajamohana, Dr.K.Umamaheswari, M.Dharani, R.Vedackshya. "Survey of review spam detection using machine learning techniques." 2017/978-1-5090-5778-8/17.