Machine Learning Approach for Opinion Mining Online Customer Review

B.KEJIYA[1], S.K.ALISHA[2]

[1]MCA Student, B V Raju College, Kovvada, Andhra Pradesh, India.

[2]Associate Professor, B V Raju College, Kovvada, Andhra Pradesh, India.

Abstract: In the current scenario, Internet registration is growing rapidly. Social media generates many signups daily with customer reviews, comments, and reviews. This huge amount of user-generated data is useless unless some mining is applied. Because there are so many fake reviews, review mining techniques should include spam detection to provide an authoritative review. Nowadays, some people use social media reviews to name themselves in purchasing products or services. Opinion spam is difficult to detect because many fake or fake comments have been created by groups or by humans for various purposes. They write fake reviews to mislead readers or sell automated detection devices to targeted products or to sell or downgrade them to tarnish their image. Proposed approaches include ontology, geographic region, IP vs. tracking, a dictionary of spam words using Naïve Bayes, simple brand assessment detection, and tracking account.

Keywords: Machine learning, fake reviews, online product, E-commerce, Product review monitoring.

I. INTRODUCTION

One of the fastest-growing places is e-commerce. In general, E-Commerce gives users the ability to write reviews related to your service. Such a diagnosis can be used as a source of information. For example, companies may use it to make design decisions for their products or services. Still, unfortunately, the importance of evaluation is misused by positive events that try to create misconceptions. So, both to enhance the identity and defame the product. The percentage of their views on the net.

Before buying anything, it is common human behavior to survey this product. Based on the reviews, consumers can browse different brands and finalize the products of their interest. These online reviews can exchange customer feedback about the product. If these criticisms are true, it can help consumers choose the right product that meets their needs. Conversely, if the feedback is manipulated or incorrect, it can give the user the wrong information. It prompts us to develop a machine that searches for counterfeit products using text and classification items in the review. The value of honesty and misdiagnosis can be measured using data mining strategies.

An algorithm can be used to track consumer reviews through mining themes and

online shopper reviews to target emotions and prevent misinterpretations.

Today, the use of the Internet and Internet-based marketing has become widespread. The Internet-based exhibition features many articles and editorials that generate large amounts of data. As a result, finding the best least expensive management or ideal items for the condition is more difficult. Clients only want audits or results based on what can be extracted by others based on their competition. Anyone can write something at this time, improving the number of false audits. Other companies are working on getting people to write fake extraordinary audits on their management or products or write awesome offline surveys on the management or devices of their warring parties. This process offers the wrong partnership for modern consumers who want to buy such things, and so we need such a framework so that such false audits can be isolated and eliminated. At this time, study the method of excavating unusual semi-controlled, unsupported, and guided statistics to identify fake audits according to different highlights.

II. LITERATURE REVIEW

Opinion Mining by Ontological Spam Detection Duhan and Mittal suggested an article, "Opinion Mining by Ontological Spam Detection," to help us discover. Fake reviews using Naïve Bayes as an algorithm. This device has been introduced as a "Fake Product Review Tracking System" to get fake details inside the website. This device will detect fake reviews by users and block users. To find out if the general description is incorrect or true, we can use some included classes.

If the feedback is from a spammer, then find out the person's IP address to be crossed. If some reviews are from the same IP address, the reviews are considered spam. Account usage is used to evaluate whether reviews are made using the same account. Finding the most effective brand review, i.e. Reviews are about the best brand or not, not about the product. Therefore, it is no longer useful to remember the brand rate when deciding on a product.

The review recognizes the use of negative vocabulary, i.e., faulty phrases. If there are more than 5 negative words, the diagnosis is spam.

Rajashree S et al. [2014] today, the Internet has become an important component, as it provides more convenience to its users. Many social networking sites give users a percentage of their views. People care about politics, social issues, and unique products. Today, it is not uncommon for consumers to review online reviews of this product before buying anything. Multiple sites address these reviews. They provide scores for products and show the distinction between unique products. Some companies create false reviews to influence buyers' behavior and increase their revenue. But how to detect these fake reviews is a difficult plan for consumers. In today's competitive world, any agency needs to maintain its popularity in the market. So everyone needs to understand the corporation's

opinion and the employer's manipulation. This article explores unique tactics for identifying manipulated feedback and suggests a brand new technique for selecting these manipulative assessments using the Decision Tree (DT).

Jui-Yu et al. [2013] Identifying tampering with reviews has become one of the top research issues in eCommerce as more and more consumers make their purchasing decisions based primarily on personal impressions from digital communities and e-commerce websites. However, clients should not forget that these personal analytics are more reliable than existing pure classified ads. As a result, some companies create fake personal reviews to influence customer behavior and increase their revenue. But, how to detect fraudulent reviews is a difficult task for consumers. Therefore, this study uses the Decision Tree (DT) to improve the class performance of diagnostic manipulation by introducing the eight capabilities of diagnostic manipulation. Furthermore, we attempt to explore the essential causes of manipulation in identifying criticism using communication assessments and derived technology guides. Finally, a real case of online consumer feedback on smartphones was used to testify to the effectiveness of the proposed procedure.

Benjamin et al. [2007] We deal with the problem of reading some related quotes in the text. For example, such reviews may include food, atmosphere, and service in a restaurant review. We design this project as a two-way scoring issue, which aims to develop a set of numerical scores for each item. We offer an algorithm that mutually learns the character item classification form by modeling dependencies between assigned ranks. This algorithm publishes the predictions of individual classifiers by analyzing meta-family members in all critiques, including contract and comparison. We prove that our agreement-based pairing model is more expressive than role-playing models. Our experimental effects confirm the model's strength: the algorithm provides substantial construction on each rating and a sophisticated pair rating model.

Ivan Tetovo et al. [2011] Online reviews are often viewed along with the numerical scores provided by the users for a series of services or product items. We suggest a statistical version that can find relevant themes in textual content and extract textual evidence of emotions that helps each of these item ratings, a key issue in summarizing item-based emotions. (Hu and Liu, 2004a). Our version achieves extreme accuracy, without any explicitly categorized information, except for the emotional score provided by that person. The proposed approach is well-known and can be used for distribution in other applications with relevant indicators and sequential information.

Jindal et al. [2007] Finding reviews from product reviews, forum posts, and blogs is an essential research topic with many applications. However, current studies have focused on extracting, classifying, and summarizing these resource studies. One major issue not yet been studied is the reliability of opinion spam or online reviews. In this article, we discuss this issue in product

reviews. To our knowledge, no study on this topic has been published yet, although web page spam and unsolicited email have been extensively investigated. We will see that the general definition of spam is very different from web page spam and email spam and therefore requires extraordinary detection strategies. We show that review spam is hugely based on an analysis of 5.8 million reviews and 14 million amazon.com reviewers. This document presents a classification of spam tests and then validates various techniques for detecting spam.

Jindal et al. [2008] The diagnostic test has become a valuable source of criticism about products, presentations, events, people, etc. Recently, many researchers have studied opinion assets such as product reviews, forum posts, and blogs. However, current studies have focused on classifying and summarizing emotions using herbal language processing techniques and statistical mining. One of the major issues that have been overlooked is the reliability of review spam or online reviews. In this article, we explore this challenge in the context of product reviews, which can generate reviews and be widely used by consumers and product makers. In recent years, many startups are also adding product reviews. So, it's time to look at review spam. To our knowledge, no comment has been posted on this topic yet, although web spam and email spam have been extensively investigated. We will see that opinion spam is quite specific to immovable webmail and email spam, requiring extraordinary detection techniques. Based on an analysis of five, eight million reviews and one pair, 14 million amazon.com reviewers, we show that opinion spam is important in reviews. This document discusses these spam games and offers new techniques for detecting them.

III. MACHINE LEARNING

Before we look at the data of different device control techniques, let's start by looking at which device is receiving information and which is not. Knowledge of machine acquisition is often labeled as a subfield of artificial intelligence, but I think categorization can often be misleading at first. The study of system data retrieval evolved from the study in this context. Still, in the information technology software of device data retrieval strategies, it is more beneficial to think of device data retrieval to create model files.

Getting device information involves creating mathematical models to help capture logs. "Learning" comes into play when we provide these tunable parameter models that can be tailored to the given information. In this way, it can understand that the system is learning from the facts. Once these models are healthy for pre-existing data, they can predict and understand new existing data components. I will leave the reader with a more philosophical focus that this version-based approach to "knowing" mathematics is very similar to the "learning" shown by the human brain. Effective use of these tools is important to start with some of the broadest types of processes we will talk about here.

Categories Of Machine Leaning :-

To a lesser degree, device knowledge can be categorized into main categories: supervised knowledge and non-supervised learning.

The supervised study involves modeling the relationship between somehow measured factual abilities and a label associated with statistics. Once this model is determined, it can label new and anonymous statistics. It is further divided into types of barriers and regression barriers: in type, labels are isolated classes, whereas, in regression, labels are continuous parts. We will look at examples of all kinds of supervised knowledge in the next section.

Unsupervised studies include creating unlabeled offset features of datasets and are often defined as "letting the dataset speak for itself." These models include functions such as polling and dimensional discounting. Clustering algorithms become aware of different groups of facts, even more so in the dimension of discounted algorithms looking for more brief representations of statistics. We will look at examples of both types of unsupervised knowledge in the next section.

IV. PROBLEM STATEMENT

In recent years, online reviews have been instrumental in making purchasing decisions. These reviews can provide users with useful information about products or services. However, to improperly promote or reduce the best products or services, spammers can also be deceived and bring in false reviews. Because of this behavior of spammers, customers lie and make wrong choices. So finding fake reviews (spam) is a big hassle. Review spam refers to the use of excessive and illegal techniques, including the growing number of fake reviews, to generate biased positive or negative reviews for a targeted product or service to sell or lower you, respectively. Allows Reviews created for this reason are known as fake spam or fake reviews, and authors responsible for writing such misleading material are spam or fake email reviewers.

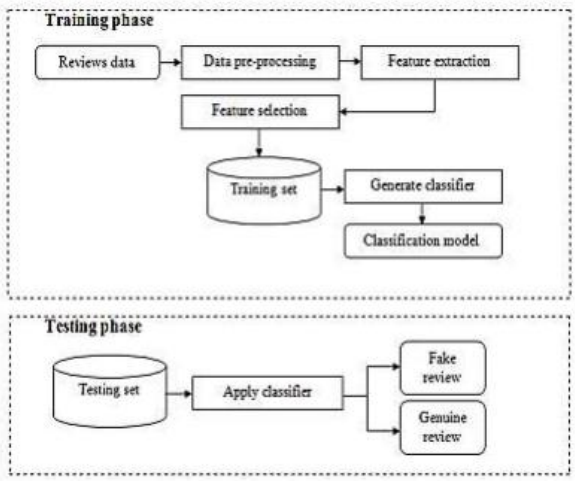
V. PROPOSED WORK

Recognized Demand Circumstances encourage introducing a technique for all of the issues described in the problem statement section above. Therefore, the proposed technique and the objectives of the work of this thesis are as follows:

Enforcing a set of rules to detect high spam, i.e., IP address, account used, dictionary of negative words centimeter use, ontology. Graphic representation of work.

- An offer with 6 specific forms of spam reviews.
- The opinion offers to mine on data filtered through spam.
- To implement ontology in detecting spam.
- Introduce a set of rules that mimics feedback with spam detection

SYSTEM ARCHITECTURE



Different ways to find spam reviews have been studied to make Opinion Mining more accurate and useful. An accurate dialogue is provided about the common techniques to determine if the diagnosis is spam. Other techniques, such as IP address tracking and ontology, have been incorporated to discover that spam reviews are a better way to get accurate results than the release of reviews.

After detecting junk criticism from existing datasets, a new dataset contains no junk emotions, and then the emotions are extracted on the new spam filtered dataset. Finally, a new set of rules is proposed that detects spam more accurately and plays with emotional mining on leaked spam logs.

WORKING METHODOLOGY:

In the past few years, online reviews have an important role to play in making a purchase decision. This is because; the reviews can provide customers with a lot of useful information about your product or service. However, in order to improve it, fictitious, or reduction in, the quality of the products or services spammers will be able to fake it and produce fake reviews. Because of this, the behavior of spammers, customers will be deceived, and they will all make bad decisions all the time. Therefore, the detection of a counterfeit (so-called "spam") opinions, it is a serious problem. Advice, spam refers to the use of excessive and unlawful methods, such as the creation of a large number of false opinions, to be positive or negative reviews about a product or service to promote, or to downgrade them. Any problems can be identified, and to motivate you to find the solutions to all the problems mentioned in the previous section of the exhibition will be a problem. Listed below are the objectives of the proposed course of action, and it does work:

- Do not install the relevant information in the application to point to the path
- The checking of the sets of data that have been installed
- Data mining that takes place

A system for monitoring a fake product reviews are based mainly in the mining industry assessments, which makes the purchase of the product is more reliable for our customers, because of the fake reviews is removed automatically from the system. The proposed system provides the user with the ability to post their own reviews."The proposed system will use your IP address to detect a fake and a real review. This is a system to detect fake reviews made by placing a fake product reviews, how to determine the IP address of the templates for the post of

customer reviews. The purpose of the results of the validation test is to check to see if they have the calculations done by the program, calculated on and off manually with the same procedure, and with the same result. The test is designed to verify that the input to the output. These are a variety of methods for detecting Spam reviews on the Internet. In order for the process to be more accurate and useful, we have been studying. There is a detailed description of the existing methods, in order to find out whether a judgment is spam or not. Other methods, such as the IP address, tracking, and Instagram will be used for the detection of spam to get a more accurate opinion, the results of the analysis.

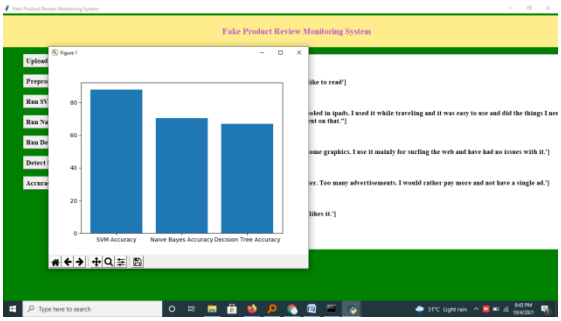


Fig.1. Tracking details.

In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and in all 3 algorithms SVM got higher accuracy.

VI. CONCLUSION

The rapid development of the Internet will increase the volume of fake and genuine surveys. As a result of these excellent reviews, no food reviews have been accepted. Some of the reasons for the fake reviews appear to be the horrible selection of merchandise, and this product lacks authenticity. Therefore, in this study, based on SVM devices, a complete misdiagnosis was designed with the help of Python software. The various techniques for detecting misdiagnosis based on fully supervised and non-supervised procedures are almost explained in this table. These current methods give less accuracy more obstacles in identifying erroneous reviews. This SVM pseudo-detection tool provides ninety-seven accuracy. 79%, and F1 rankings increase by 10%.

REFERENCES

[1] Rajashree S. Jadhav, Prof. Deipali V. Gore, "A New Approach for Identifying Manipulated Online Reviews using Decision Tree ". (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), pp 1447-1450, 2014

[2] Long- Sheng Chen, Jui-Yu Lin, “A study on Review Manipulation Classification using Decision Tree", Kuala Lumpur, Malaysia, pp 3-5, IEEE conference publication, 2013.

[3] Benjamin Snyder and Regina Brazil, “Multiple Aspect ranking using the Good Grief Algorithm “Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology2007.

[4] Ivan Tetovo, “A Joint Model of Text and Aspect Ratings for Sentiment Summarization “Ivan Department of Computer Science University of Illinois at Urbana, 2011

[5] N. Jindal and B. Liu, “Analyzing and detecting review spam,” International Conference on Web Search and Data Mining, 2007, pp. 547-552.

[6] N. Jindal and B. Liu, “Opinion spam and analysis,” International Conference on Web Search and Data Mining, 2008, pp. 219-230.