

## FAKE PRODUCT REVIEW MONITORING SYSTEM

Ms. Aiswarya K<sup>1</sup>, Ms. Rasna k<sup>2</sup>, Mr. Sravan Ranjith<sup>3</sup>, Mr. Sukar Anas IP<sup>4</sup>, Mr. Mahesh B.L<sup>5</sup>

<sup>1,2,3,4</sup>Students, Dept. of CSE, Yenepoya Institute of Technology, Moodbidri, India-574225

<sup>5</sup>Assistant Professor, Dept. of CSE, Yenepoya Institute of Technology, Moodbidri, India-574225

\*\*\*

**Abstract** - In the present situation, the data on the Network is growing exponentially. Social networks are generating huge amounts of data on a day to day basis, such as reviews, comments, ratings, reviews. This is a huge amount of data from the user, is of no use, unless some of the digging operations have to be applied. Because there are a lot of false opinions, and the method of analysis of an opinion, it should have been, spam detection, in order to have a real offer. At present, there are a lot of people are using the social networking platform to call for the purchase of a product or a service. Spam's opinion, there is a deep and complex issue, as there have been many false and false opinions, which have been created by organizations, or persons, who, for a variety of purposes. Write fake reviews to mislead the reader or the auto-detection systems, and the promotion of, or use of specific products, to be aware of their reputation even worse. The proposed method consists of an ontology, to track the location and IP addresses of a list of spam keywords, using naive Bayes, and a fire-detection and tracking.

### 1.INTRODUCTION

One of the fast growing areas of e-commerce. Usually, e-commerce allows customers to provide feedback in relation to the service. The presence of these recommendations can be used as a source of information. For example, companies will be able to use it for making design decisions for products or services, but, unfortunately, the importance of the evaluation will be used by some sites which have tried to create fake reviews, which are aimed at increasing the popularity of the product will be damaged. They will share their thoughts online. Before buying anything, it is natural for a person to make a review on this product. On the basis of the reviews, customers can compare different brands and their products of interest. These online reviews can be for the customer's opinion of the product. If these assessments are correct, it can help you in choosing the right product is the one that meets their needs. On the other hand, if the reviews are to be manipulated, or does not correspond with reality, the user can not be deceived. This requires us to create a system that can detect fake

reviews of the product by using the text property of the review and the rating. The value of honesty, and the measure of the false vision is to be measured with the help of data mining techniques. The algorithm can be used to keep track of a customer's review, the subject's go to the beach and the vote cast at the online reviews of the customers, it will also block the fake reviews. The detection of fake reviews that will allow you to compare the attributes of moods and attributes, reviews, accessories and tuning, with the characteristics of the different opinions and takes into account the similarity value between the two approaches. The number of video cards, for products, as a result, it is also available in the details for a user review of the product they are buying

### 2.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

- Review of the results will be graded, and the results were presented in the past

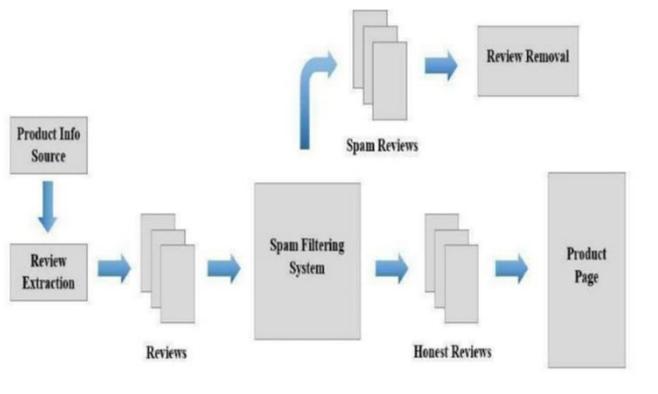


Figure-1 working flow chart

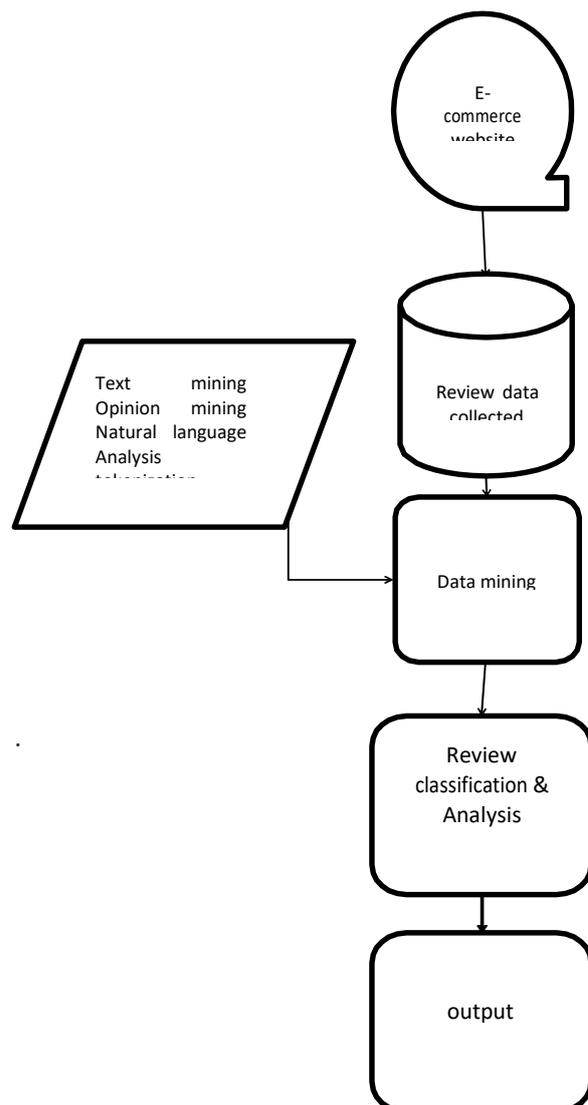
There are a variety of mining methods, which have been carried out in the field

- **Text Mining**
- **Opinion Mining**
- **Sentimental Analysis**
- **Natural language Analysis**
- **Tokenization**

Data mining is the practice of analyzing large databases to obtain that information

- **Text Mining**, has been widely used in knowledge-based organizations. Text mining is the process of examining large amounts of documents, for the purpose of identifying new information
- **Mining opinions and sentiments analysis**, and the method of analysis, which makes use of computer linguistics and natural language processing to automatically identify and extract sentiment and reviews with text (positive, negative, neutral, etc.).
- **Natural language processing or NLP**, is an acronym that stands in the broadest sense of the word, is defined as the automated processing of natural language, as a language, and text.
- **Tokenization**, is the process of replacing sensitive data with unique identification of the characters, which is to ensure that all the necessary information about the data.

Figure-2 data mining flow chart



## 2.1 Methods based on URL

We have gathered several of the layers of the details on the Amazon product, and an understanding of false beliefs. Data cleaning: the data that is received as a result of cleaning is not yet ready for analysis, so we took a number of steps to clean up the data, and the creation of new props for the analysis, there are two new attributes, sentiment\_score, and most\_rev, which were obtained by an in-depth analysis. The goal of this project is to contribute to a growing number of studies identifying fake reviews from the reviewers, by carrying out the analysis of social media, combined with the identification of abnormalities, with the help of a group of methods. The developed approach for the recognition of the fake reviewers have compared it with the previously proposed methods, it has a very wide possibilities for automation, and can dramatically reduce the search field, and the dynamic tracking of the fake users. We have instagram data on Amazon's

website with the help of the snowball method. Two additional attributes are obtained by using the help of a statistical analysis of the data and the python module `most_rev` has the highest number of reviews per day. This will help us determine whether or not a user has to a very large number of opinions, on the same day. `review_sentiment`: valuation, sentiment, and recommendations are in the range from -1 (negative), 1 (positive). This will help us determine whether or not the assessment is at odds with the user's mood. The `Field` module is used to create a sentiment score.

## 2.12 Dataset

**url:** Scraped url

**review\_bold:** Title of the review

**ratings:** star rating given by the review

**verified:** 1 = reviewer is a verified buyer, 0 = if not

**date:** date review was made

**by:** name associated with profile

**profile\_id:** ID number of profile

**most\_rev:** maximum review made in a day by profile

**by\_link:** url to profile

**helpful:** number of people who tagged review as 'helpful'

**product:** product name

**product\_link:** url to product page

**review\_sentiment:** sentiment score of the review

The networks indicate that there is a significant difference between the alleged false view (on the left) and the normal to the reviews (on the right). The nodes are the profiles to connect with each other when they are together, inspected more than 2 products to you. Place the product on the left is manually selected in order to be suspicious, as compared to other similar production. The set of data that is collected will not be marked. So, We have to work on unsupervised learning. While at the network visualization is used to identify the types of presumption, from the critics, the clusters are useful for the understanding of the common characteristics of fake reviews, and the detection of abnormalities. The model is divided with the feedback at the end of the five clusters, with the help of a simple K-Means method, with the 5 groups. Features that are included in the models, the time-tested reviews, `most_rev`, `useful`, and `review_sentiment`.

## 2.13 Summary of the Results of the Analysis of Networks and Clusters

And then we will close the list of the accused and their reviews in the clusters. We analyzed the summary of the features and functions of our list of suspects. Based on its features, we can't match up with their ratings in a single cluster. However, 77.35% of the reviews are created by users and belong to Cluster1 and Cluster3. In the field of artificial intelligence, artificial intelligence, and I think that the machines are able to mimic the actions and abilities of the human mind. The language is one of the major achievements of the human, ?????????? of human progress. It is, therefore, not surprising that a lot of the work is in progress on the integration of the language used in the field of artificial intelligence, in the form of Natural Language Processing (NLP)

## 2.14 Named Entity Recognition

The simplest and most useful techniques in NLP, and removal of entities from the text. It covers the key concepts and references in the main text. Named Entity Recognition (NER) identify the entities, such as persons, locations, organizations, dates, etc., etc.

## 2.15 Sentiment Analysis

The most commonly used method of NLP is sentiment analysis. Sentiment analysis is most useful in cases such as this, as the client, advisory, research, reviews, and comments on the social networks, where people can express their opinions and feedback. The easiest way to do it is to analyze sentiment in order to make use of the 3-point seat belt: a positive/negative/neutral). In more complex cases, as the result of a numerical score, which can be divided into as many categories as you want. In the case of our sample text, and the customer can clearly articulate the feelings in different parts of the text. Because of this, the output is not very useful. Instead, we can find the meaning of a sentence and to separate the positive and negative parts of the review. The Sentiment is from the review it can also help you choose the best of the positive and negative portions of the review

## 2.16 Text Annotation

As the name suggests, GROUPS, and has techniques that will help you to summarize large documents of text. Text is development, it is mainly used in cases, such as press releases, and research articles. Two of the

major approaches to the text is a synthesis, the extraction and abstraction. The extraction method for how to make a summary of the data portion of the text. Abstract a method for how to make a summary of the making of a fresh text that brings out the essence of the original text. There are many different algorithms that can be used to add text, such as LexRank, TextRank, and latent semantic analysis. To take the example of LexRank, this algorithm carries out the sentences, using the degree of similarity between them. Any offer will be ranked higher if it appears to be a greater number of deals, and this, in turn, is comparable with any other offer.

### 2.17 Aspect Mining

Aspect mining refers to various aspects of the text. In conjunction with the sentiment of research, all of the information from the text. One of the simplest methods for the analysis of the approaches is the use of part-of-speech tags.

### 2.18 Topic Modeling

Topic modeling is one of the most complex of the methods used for the identification of natural objects in the document. The most important advantage of special modelling is that it is unregulated. The training of the model, and the appropriate training data set is not required.

## 3. RESULTS

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 be 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.

## 4. CONCLUSIONS AND FUTURE WORK

Data mining has become an established discipline in the field of artificial intelligence (AI) and technical knowledge and (C). It has its roots in machine learning and statistics, but also in other areas of computer science. Over the past ten years, it is of the utmost importance, and because of the advancements in the information technology secure computing power, allowing for large-scale data mining. In contrast to a number of other innovations in the AI, and the community of data mining can be seen as an application, rather than the technology, so you can be assured that it is important to be in the not so distant future. This article provides a brief overview of the history and the details of the study and ideas for future directions. In our work, we have come to the conclusion that it is found to be spam reviews, with an enormous amount of unstructured data is an important task in the research. Although some of the algorithms used in the anti-spam analysis to give a satisfactory result has been achieved, no algorithm is able to solve all your problems and difficulties faced by the current generation to the next. In the analysis of any time, it is extremely important to take account of the specific indicators for the level of quality, convenience, usefulness, and ease of use. The literature describes a large number of more complex methods, which determine the sentiment analysis in relation to different aspects of the company. With our app, you will be helping to pay for the product you want, without the assumption of a scam. With our app, to analyze it, and then to publish genuine reviews of the design of the product. And you can be sure that the availability of products in development, as well as in the reviews. In the future, we will try to improve the way we calculate the reviewer sentiment score. We will try our dictionary, with feelings and emotions. We will do our best to have the word in our dictionary-and-updating the weights given to these words, to get a more accurate estimate of the score of the reviews. Sentiment analysis-or opinion mining can be used for all new applications that conform with the principles of data mining. One of the areas of future research is to implement and test the effectiveness of the application of the proposed approach for the label, and the different data-sets. The main goal of our work is to create a system that will detect spam and useless, reviews, and filtered, so that the user knows about it.

The goal of the project is to improve customer satisfaction, as well as the purchases to be reliable. It will detect fake reviews, and the development of algorithms for the extraction of the product, and create a list of words.

## 5. LIST OF REFERENCES

1. International Journal of Pure and Applied Mathematics, Vol. No. 119, No. 12 in 2018, 13203-13209
2. International Journal of Research and Development Trends (IJTSRD) Volume: 3 / Issue: 3 / March-April 2019. Online: [www.ijtsrd.com](http://www.ijtsrd.com) e-ISSN: 2456-6470
3. International Journal of Engineering Research & Technology (IJERT) <http://www.ijert.org> ISSN: 2278-0181 IJERTV8IS030040 Full. 8 Issue 03, March 2019
4. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056: 05 Издания: 06 | июнь 2018 [www.irjet.net](http://www.irjet.net) p-ISSN: 2395-0072
5. N. Jindal and B. Liu, "Opinion, спам, and analysis", International Conference on Web Search and Data Mining, 2008, str.
6. Jun Yue Li, Xian F, eng, Suzah, S, to hang, to the Detection of false opinions, and with the help of the Media, and the Emotional model IEEE2016