

# Monitoring of Fake Reviews using Support Vector Machine and IP Tracking

Azal Fatima<sup>1</sup>, Suneeta Singh<sup>2</sup>, Shreya Khare<sup>3</sup>, Vineeta<sup>4</sup>

<sup>1, 2, 3, 4</sup>Dr. A.I.T.H., Kanpur

**Abstract:** *In the modern era of digitization, online shopping has become an emerging trend. The buyers usually check the supposition of others before buying a product .*

*The reviews given to a solitary item decides if the item holds credibility. Maker can pick up knowledge into its items quality and shortcomings dependent on the reviews of the clients. In this way, client reviews play an important role in ecommerce, since numerous choices are made dependent on them.*

*Nowadays it is seen that the organizations will in general give good reviews to its very own item all together to build its sales ,and on other hand there are a few customers who intend to either fame of defame a product by giving fake reviews. Hence ,a framework which can recognize such kinds of fraud revivis, be it positive or negative is required .*

*The proposed framework will use the support vector machine and ip tracking techniques for achieving the desired goal.*

**Keywords:** *digitization ,review, ecommerce ,fake, support vector machine, ip tracking*

## I. INTRODUCTION

As online shopping is getting prevalent day by day, the number of reviews got from clients about the item grows rapidly. Thus, a buyer goes over different reviews in the site, yet regardless of whether those reviews are certifiable isn't recognized by the customer purchasing the item on the web.

For a well-known item, the surveys can go up to thousands. This makes trouble for the potential client to peruse them furthermore, to choose whether to purchase or not purchase the item. In some sites, some great audits are included by the item organization individuals itself to make item popular.

Also there are certain reviewers who in order to either fame of defame a product post fake reviews about the product. In this way, clients purchasing items online is misguided.

To discover and evacuate those fake reviews, this framework is presented. The fundamental thought is to follow the IP address of the client presenting the review also, on the off chance that a similar sort of review is originating from indistinguishable IP address, at that point those reviews are evacuated.

The framework uses combined approach of machine learning and support vector machine to rate an item dependent on the surveys put together by the clients.

## II. EXISTING SYSTEM

When doing any type of online shopping, most users will spend a good amount of time reading user reviews if they are available. A survey performed by Inc.com has shown that:

- A. 91 percent of people regularly or occasionally read online reviews.
- B. 84 percent trust online reviews as much as a personal recommendation.
- C. 68 percent form an opinion after reading between one and six online reviews.

Obviously shoppers esteem the criticism given by other customers as do the organizations that sell such items. Blogs, websites, discussion boards and so forth are a store of client remarks which are profitable also, rich wellsprings of literary information. Thus people depend broadly on the reviews accessible on the web.

It implies that they settle on their choice of whether to purchase items or not by reading down the existing assessments on those items. In fact if a potential client gets a positive impression of an item by considering the present opinions for that item, it is exceedingly likely that he will really buy the item.

Normally if the level of positive suppositions is impressive, then the overall impression will be very positive. Similarly, on the other side if the general impression is negative, it is possible that they don't purchase the item. Presently any individuals can compose any feeling content, this can inspire the people, and associations to give undeserving spam assessments to elevate or to ruin some objective items. These spammed opinion information is called opinion spam.

### III. LITERATURE SURVEY

It is extremely hard to distinguish fake or spam reviews. The fake reviewers give undeserving reviews to the items to increase or decrease the sales in less amount of time. There are two types of reviews seen :

#### A. *Opinionated False Reviews*

These reviews contain various assessments which could be harmful for the sale of the item. These kind of reviews give false undeserving audits and evaluations so as to increment or decrement the sale of the items..

#### B. *Reviews On Brands*

Such reviews are not regarding a particular product but the company which provides it. These types of reviews do not help the quality sales of the product.

### IV. PROPOSED SYSTEM

The framework proposed will incorporate a few techniques like:

- 1) *IP Address Discovery*: Discovery of ip address which will help the administrator to check regardless of whether the given reviews are from substantial clients or are from a spammer.
- 2) *Sentiment analysis of the Review*: It is essential for framework to comprehend whether the review is positive or negative, which will additionally get the deviation from both of the positivity or cynicism in the reviews. The analysis will comprehend generally speaking perspective of the items with the goal that few spam reviews won't influence the in general statics of items.

These approaches are described as follows:

The framework tracks the IP address of the client presenting the audit of the item. At that point the admin will experience the database containing the reviews.

The database will contain the reviews, client subtleties i.e., the client who presented the audit alongside the followed IP address. The admin will look for the occasions where numerous audits suggesting a similar significance have been put together by a similar IP address.

These reviews are distinguished as fake reviews and the overseer will evacuate those reviews. The framework likewise rates an item dependent on the client reviews of the items.

This is finished utilizing the AI and opinion mining procedures. The grouping will comprise of three classes, positive, negative and impartial.

A classifier is fabricated utilizing the "sklearn machine library". The LinearSVC (a direct help vector machine) classifier imported from sklearn library is utilized as a classifier.

The classifier is prepared utilizing a database containing the reviews and its appraisals. This is accomplished using the TfidfVectorizer imported from sklearn feature extraction library. Likewise, pre-handling of the audit content is finished by expelling the stopwords (which don't pass on any nostalgic importance). This undertaking of preprocessing is cultivated utilizing nltk library which contains English stopwords.

#### STEPS

- a) The ADMIN/USER can login to the system and perform their respective tasks in the system.
- b) The posted reviews will pass through the process of ip address tracking .
- c) The rating of a product will be done using sentiment analysis and machine learning.
- d) If any spam is found in the reviews which are analysed then the spam is handled by the system admin only.

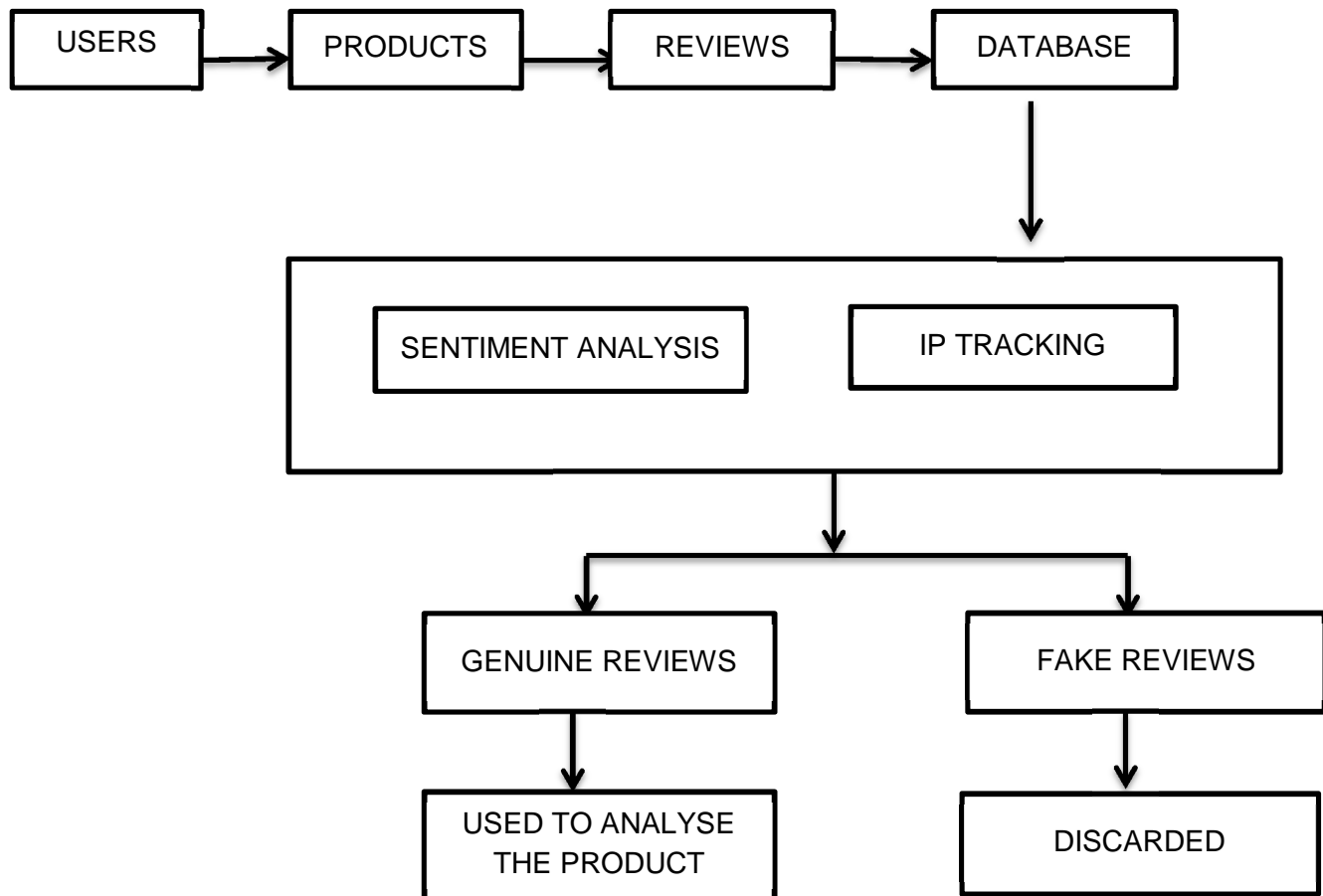


Figure 1. Product review analysis process

### V. SUPPORT VECTOR MACHINE WITH SCIKIT-LEARN

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. SVM constructs a hyperplane in multidimensional space to separate different classes. SVM generates optimal hyperplane in an iterative manner, which is used to minimize an error. The core idea of SVM is to find a maximum marginal hyperplane (MMH) that best divides the dataset into classes.

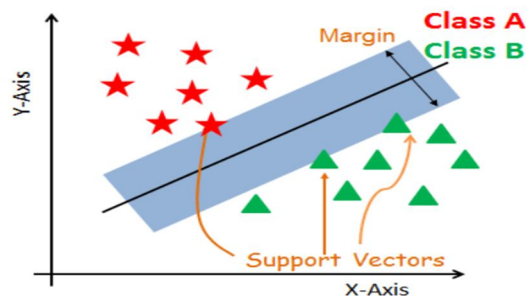


Figure 2. SVM method for classification



#### A. Support Vectors

Support vectors are the data points, which are closest to the hyperplane. These points will define the separating line better by calculating margins. These points are more relevant to the construction of the classifier.

#### B. Hyperplane

A hyperplane is a decision plane which separates between a set of objects having different class memberships.

#### C. Margin

A margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vectors or closest points. If the margin is larger in between the classes, then it is considered a good margin, a smaller margin is a bad margin.

### VI. CONCLUSION

Finding the sentiment spam from gigantic amount of information information has turned into a significant research issue. Presently business associations and scholastics are advancing their endeavors to locate the best framework for sentiment spam analysis. More future work is required on further improving the presentation of the supposition spam examination. There is a colossal need in the business for such applications on the grounds that each organization needs to know how shoppers truly feel about their products. The proposed system is capable of easily rating a product using sentiment analysis. The system is easy to use and implement. It can also track the user posting fake reviews from the same ip address and therefore preventing spread of fake information regarding the product

### REFERENCES

- [1] Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013)."New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems. 28 (2): 15–21. doi:10.1109/MIS.2013.30.
- [2] Manleen Kaur Kohli, Shaheen Jamil Khan, Tanvi Mirashi, Suraj Gupta, "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining"
- [3] Analyzing and detecting review spam, Bing liu of department of computer science university of illinois at chicago.