

## **Spam Detection on Social Network through Sentiment Analysis**

**Sunita Mahajan<sup>1</sup>, Dr. Vijay Rana**

*Research Scholar , CSE Department, Arni University, Kathgarh,  
Kangra, Himachal Pradesh, India.*

### **Abstract**

Today's we all the connected with E-Commerce websites. Some e-commerce websites are based on user reviews and some are based on certainty. The growing popularity of e-commerce and online reviews, an individual used to actively engage professionals in writing false reviews. Here, an individual may be marketing personnel, a manufacturer, a service provider, a leader, or a movie producer etc. And a false review can also be referred as a fake review, an opinion spam, a fraudulent review or a non-genuine review. The person who writes a fake review is called as a spammer or fake reviewer. If a group of people get involved in the process, then they are referred to as group spammer. A spammer used to send fake reviews either to promote a low quality product or to discredit a good quality product. User can't identify the actual quality of product.

**Keywords:** Spam detection ,Sentiment analysis, Fraud detection, Social Network.

### **INTRODUCTION**

The E-Commerce websites are commonly publicised on the basis of parameters such as certainty and user reviews. Some e-commerce websites are based on user reviews and some are based on certainty .In our proposed paper we are analyzing the e-commerce websites to check the Fraud Detection of users on basis of factors mentioned above. There are number of users who use these commercial websites to buy products online because of their comfort. Fraud Detection is the basic factors for users to use these sites for shopping. Fraud Detection in terms of the suggested parameters i.e user reviews and certainty is commonly used for analysis purpose.

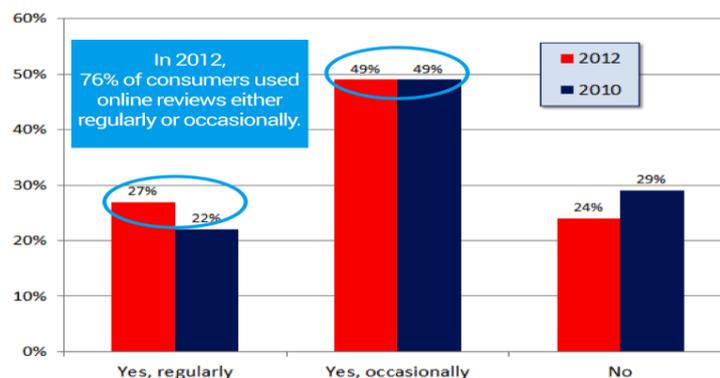
The Fraud Detection and associated parameters are critical in the success of the system. The shopping websites in order to gain profit and attract user enhance these parameters. Fraud

Detection calculation can involve hectic calculations and it is not a direct metric. Direct metric have arguments which are visible and can be calculated directly. The indirect measures such as quality are relatively difficult to calculate. The considered parameters in the proposed systems utilize both of these measures. [1]–[4] the Fraud Detection management is done in the form of search engine optimization. The search engine optimization is required in almost every area of ecommerce. Optimization will act as building block with through which Ecommerce website accomplish their offshoots. The online shopping websites increase their users commonly on the basis of user reviews and certainty. The user reviews are analysed on scale of 1 to 5. These are also known as star ratings. Higher the ratings more successful will be website. There are certain facts about the user reviews. These facts are listed as[5]

- 70% of customers consult reviews or ratings before making a final purchase
- 63% of consumers are more likely to purchase from a site if it has product ratings and reviews.
- 67% of consumers read 6 reviews or less before they feel they can Fraud Detection a business enough to make a purchase.
- As many as 79% of consumers Fraud Detection product reviews as much as a personal recommendation.
- 80% of consumers have changed their mind about purchases based on NEGATIVE information they have found online.
- On average, 75 percent of reviews posted on review websites are positive, and 71 percent of consumers agreed that reviews make them more comfortable that they are purchasing the right product or service.

The second parameter in lane for success of online shopping sites is certainty. [6]–[18]This parameter is difficult to build and consume time to build. As the sites grew older and throughout the life spam if their reviews are positive then certainty will build. Building certainty will be accomplished only if website performance is more than 90% all the times in operation.

The data about the user reviews and certainty is collected from UCI dataset. The online dataset provide the information about the year 2010-12.



**Fig 1:** Showing 76% of the users uses reviews either directly or indirectly.

The certainty parameter also taken into consideration but this parameter is difficult to calculate and hence is used less frequently as compared to user reviews.



**Fig 2:** Showing Certainty parameters values in terms of various websites

On the basis of certainty Amazon is considered best possible one. The data is collected in 2014 and represented on the basis of chart.

## RELATED WORK

There is a work which is done toward increasing the Fraud Detection within the ecommerce. The papers we have analyze are enlisted here.[13] Fraud Detection is vital for the success of online business. The case of hotel reservation is considered in this case. The hotel reservation will work properly if the booking is according to the desired standards as purposed by the consumer. Otherwise Fraud Detection breach will occur. [9]Fraud Detection in the suggested paper is accomplished with the help of analyzing the culture. Some users are not expecting change and hence cannot be forced to Fraud Detection a system which is alien to them. In such environment gathering Fraud Detection will be very difficult. Such situations are analyzed in the purposed paper. [6] Fraud Detection is enhanced by the use of this technique. The concept of user feedback is considered in this case. The feedback if positive then the Fraud Detection will be enhanced and vendor can earn profit. The Fraud Detection can also be breached if for some online trading the feedback become negative. [15] Fraud Detection concept is considered in this case. The Fraud Detection will be expressed by the use of the vendor consumer relationship. The relationship has to be strong. The concept of coupling and cohesion will be introduced in this case. [8]the review of online consumer Fraud Detection will be considered in this case. The online consumer Fraud Detection is the new concept where user buys the goods online if the online shopping portal has the Fraud Detection. The Fraud Detection will decide the profit earn by the vendor. [11]online Fraud Detection building mechanism is considered. The online word is such that some consumer afraid from it. Because of this reason Fraud Detection comes into existence. The Fraud Detection factor is important in the success of online web portal. Increase in Fraud Detection will cause more traffic toward the online ecommerce whereas breach in the Fraud Detection will cause lose. All of these situations are analyzed in this paper.

From all the papers we have analyzed that very few or nothing is done towards the increase in

Fraud Detection by the use of parameters like quality factor. A Fraud Detection web model based on a distributed search algorithm and a network of Fraud Detected broker that can authorize a Fraud Detection channel through which end to end transacting parties deal virtually directly and risk-free with each other is proposed in this paper.

From the analysed paper we get to know that parameters like certainty and user reviews are considered in mild fashion and not in functional manner. To increase the productivity of online shopping websites these parameters will play critical role.

### PROPOSED WORK

In the suggested paper we utilize dataset collected from UCI for websites like Amazon and Flipkart. The Hubspot.com is online review collecting website used in order to collect the information about website having highest rating. Flipkart is one of online shopping website which is based on online user reviews. The user reviews for Flipkart is generally between 3-5 stars.

The tabular structure of user reviews and certainty are as shown below

**Table 1:** Showing the tabular structure of data collected from UCI

Website	Origin	AVERAGE RATING(USER REVIEW)	CERTAINTY(100%)
Flipkart	India	****	62
Amazon	USA	****	92
Snapdeal	India	***	60
Ebay	India	**	56
Jabong	India	****	90

The proposed system deals with the classifiers involving K nearest neighbour on dataset to check the fraud if any within the transactions. The flowchart associated with the KNN and Euclidean distance to form hybrid approach is as follows

**Algorithm 1: kNN Algorithm for Outlier Detection**

*Require: Dataset D, Threshold M, neighbour count K*

1:  $X = \text{getOutlierScores}(D, k)$

2: for all  $p, \text{OutlierScore}[p]$  in  $X$  do

3: if  $\text{OutlierScore}[p]$  is greater than or equal to  $M$  then

4: Add  $p$  to  $L$

5: end if

6: end for

```

7: getOutlierScores(D,k)
8: if D != NULL then
9:   for all p in D do
10:     $S = \text{getKNearestNeighbours}(D,p,k)$ 
11:    for all q in S do
12:      $T = \text{getKNearestNeighbours}(D,q,k)$ 
13:     if p in T then
14:      Add q to ForwardNNk(p)
15:      $\text{ForwardNNk}(p) = \text{ForwardNNk}(p) + 1$ 
16:     end if
17:   end for
18: end for
19: end if
20: for p in D do
21:    $\text{OutlierScore}(p) = 1 - (\text{ForwardNNk}(p)/(D-1))$ 
22: return [p, OutlierScore(p)]
23: end for
24:  $\text{getKNearestNeighbours}(D,p,k)$ 
25: if D != NULL then
26:   for all q in D and p != q do
27:    Compute dist(p,q)
28:   end for
29: end if
30:  $\text{sort}(\text{dist}(p,q))$ 
31: Add k shortest distant objects from p to NNk(p)
32: return NNk(p)

```

**RESULTS: REVIEW SPAM OUTLIERS**

Total Number of reviews	78930
Number of Spam reviews detected	6064
Percentage of Spam	7.68 %

## CONCLUSION

The Fraud Detection is the factor which has many definitions. One of the definitions which we have considers is the profit factor as the Fraud Detection by the consumer toward the vendor will cause more goods will be sold online. The proposed paper will consider quality and user reviews indicating Fraud Detection of consumer toward the vendor. Breach in the Fraud Detection will cause loss of the vendor and intern's organization. Fraud Detection will be easy to lose but difficult to gain. In the future we will work toward recommender of ecommerce websites based on the security and protection. Sentiment analysis has found various capable purposes similar to market opinion, political sentiment resolve, impartiality value opinion, box office opinion etc. But, a lot work still remains to be done and it is a fertile area.

## REFERENCES

- [1] M. C. Buzzi, M. Buzzi, B. Leporini, and F. Akhter, "User Fraud Detection in eCommerce Services: Perception via Screen Reader," in *2009 International Conference on New Trends in Information and Service Science*, 2009, pp. 1166–1171.
- [2] L. H. Wei, M. A. Osman, N. Zakaria, and T. Bo, "Adoption of E-Commerce Online Shopping in Malaysia," in *2010 IEEE 7th International Conference on E-Business Engineering*, 2010, pp. 140–143.
- [3] R. Suguna and D. Sharmila, "An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms," *Int. J. Comput. Appl.*, vol. 70, no. 3, pp. 37–44, 2013.
- [4] M. Berka, T., & Plößnig, "Designing Recommender Systems for Tourism," *Proc. ENTER 2004*, 2004.
- [5] "Ecommerce Product Reviews: Why Online Store Owners Should Embrace Them." [Online]. Available: <https://www.shopify.in/blog/15359677-why-online-store-owners-should-embrace-online-reviews>. [Accessed: 28-May-2016].
- [6] S. ur Rehman and J. Coughlan, "Building Fraud Detection for online shopping and their adoption of e-commerce," pp. 456–460.
- [7] J. Jiang, G. Han, F. Wang, L. Shu, and M. Guizani, "An Efficient Distributed Fraud Detection Model for Wireless Sensor Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1228–1237, May 2015.
- [8] H. Wang, "Review of studies on online consumer Fraud Detection," in *2010 Second International Conference on Computational Intelligence and Natural Computing*, 2010, vol. 2, pp. 97–100.
- [9] F. Meskaran and Z. Ismail, "Customers' Fraud Detection in e-commerce: In collective culture setting," in *2012 International Conference on Information Retrieval & Knowledge Management*, 2012, pp. 182–186.
- [10] S. Muhammad, S. Hussain, and M. Yousaf, "Neighbor Node Fraud Detection Based

- Intrusion Detection System for WSN,” *Procedia - Procedia Comput. Sci.*, vol. 63, pp. 183–188, 2015.
- [11] Z. Zhou, Q. Ding, and B. Huang, “Online Fraud Detection building mechanism in e-commerce,” in *2012 IEEE Symposium on Robotics and Applications (ISRA)*, 2012, pp. 335–338.
- [12] S. Pradhan, V. Gay, and S. Nepal, “Improving dental care recommendation systems using Fraud Detection and social networks,” in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 4264–4269.
- [13] K. Atcharyachanvanich and O. Hitoshi, “Fraud Detection on E-Commerce Website in Thailand: A Case of Online Hotel Reservation,” in *2011 IEEE/IPSJ International Symposium on Applications and the Internet*, 2011, pp. 574–577.
- [14] Salesforce.com, “Secure , private , and Fraud Detectionworthy : enterprise cloud computing with Force . com,” 2010.
- [15] A. F. Salam, L. Iyer, P. Palvia, and R. Singh, “Fraud Detection in e-commerce,” *Commun. ACM*, vol. 48, no. 2, pp. 72–77, Feb. 2005.
- [16] C. Piao and X. Gan, “Research on Fraud Detection Management Model for E-Commerce Based on Fuzzy Clustering Method,” in *2009 IEEE International Conference on e-Business Engineering*, 2009, pp. 188–195.
- [17] F. Rashid, A. Miri, and I. Woungang, “A secure data deduplication framework for cloud environments,” in *2012 Tenth Annual International Conference on Privacy, Security and Fraud Detection*, 2012, pp. 81–87.
- [18] K. Yang and X. Jia, “An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 9, pp. 1717–1726, Sep. 2013.

