

Advancing Towards Credibility of Product Review Methodology through Comparative Analysis

Avita Fuskele Jain¹, Hemendra Singh Patel², Shreya Nema³, Priya Mourya⁴

^{1, 2, 3, 4} Jabalpur Engineering College, Jabalpur, India

Abstract- In Today's era, digital market has been established thereby making Product Review- a crucial factor to make a decision. Having one honest review triggers more instead of chasing gazillion of reviews. This paper has been proposed with the motive of comparative survey to analyze various methods of product review such as sentimental analysis, maximum entropy, network techniques, natural language processing technique so as to boost the credibility and feasibility of the system.

Keywords- Blockchain, Fake Reviews, Sentimental Analysis

I. INTRODUCTION

With the spread of globalization and digitalization, E-commerce has moved a step ahead of traditional shop to shop marketing. Usually consumer's visit a ton of websites for the single product and the only distinguishable way before them is product review. The review on a product may be positive or negative. But spamming over the reviews is ruining the system. To keep this E-commerce system fair enough every consumer ought to avoid fake review. A fake review sometimes acts as a virus which fits in consumers' mind set and as a result, consumers change their mind of buying the product. Technologically advancing and building a system to exact the spamming has been our motive. Reviews should withstand the respective product rather than exaggerating or understating the product.

To achieve this, a systematic survey has been done which fills the loopholes and compare distinguished methods on different parameters. In relation with the scope these days, there has been many E-commercial sites which are working for each and every gadget or things in use. In scenario of two contradictory reviews, these parameters will be able to make difference between what is wrong and what is right. To make the system even-handed, a comparative study has been done between the various methodologies like Block chain, Decision Tree, Weighted Motifs, Sentimental analysis. The algorithm involved has different approaches which has their own pros and cons. This comparative study will provide us with the credibility of each method. Work based on Random Forest, Decision tree, Semantic and Emotional Analysis, Polarity of text, POS tagging are diagnosing the problems.

II. LITERATURE SURVEY

Spam detection over Product review is a chief issue to make E-Commerce credible. Expertise Researchers of respective fields has made commendable progress which gives rise to new heights towards the foolproof online exchange system. Related Research works has initiated the spark and evolved the methodologies to work upon. Various parameters of various approaches have been kept in mind for a constructive outcome. Starting with the collection of reviews from various websites and these reviews can be characterized on the basis of various aspects such as data can be analyzed on the account of feelings,[5]. Also, this text can be characterized as positive and negative. And to analyze this polarity of text - sentiment lexicon can be used. Sentiment lexicon gives more accuracy as it auto-generates structures and reduce the manual classification of data, [6]. Product review can be of two types Search based Review, Experience based Review. A Data-driven method has been developed in order to get the number of reviewers who give reviews without proper knowledge and even without buying of it, [13]. A Data-driven method has been developed in order to get the number of reviewers who give reviews without proper knowledge and even without buying used, [7]. Useful profiles, Templates, Ratings and Reply thickness can be advantageous in diagnosing the fake review, [1]. In order to check whether the statement is fake review or not. K mean is used to set positive review as LP and negative as LN, [8]. Involvement of machine learning has removed the silver lining of precedent methods and worked towards its betterment. There are various algorithms for machine learning such as SVM classifiers, the Naive Bayes classifier, the maximum Entropy and Random Forest, have been proposed and they are compared on various aspects, [14]. The use of unsupervised learning has uncovered the spam's which mainly analyses the natural cluster to accord with labeled and unlabeled data, [9]. Algorithms of machine learning like Decision Tree is more efficient than other algorithms like Random forest, [10]. When comparing between Lexicon based approach and machine learning, it is found out that Lexicon based approach is more effective, [11]. Further there are newer aspects for fake review detection such as review density, semantic and emotional analysis, [12].

III. METHODOLOGY

Now The four research papers of prominent authors representing different approaches to optimize the product review system have been compared. These particular methods have their own perspective towards the vision of a foolproof product review system. Combining them by extracting best points out of them will definitely result in advancement in technology.

Following are the papers studied:

- Review Chain: Untampered Product Review on BlockChain.
- Uncover Product Review pattern via weighted motifs: Lambda and delta motif mining algorithm and V- Shaped motif mining algorithm
- Sentiment Analysis of Product Review: A Review
- Detection of fake opinions on online product using decision tree and information gain.

We have discussed the main points of the research paper below:

1.1 Review Chain: Untampered Product Review On BlockChain:

This Research paper presents a Decentralized approach so as to avoid central authority using block chain technology. A mobile app ought to contain the private keys and store review hashes. A database has been made to store contents. For this, platform for computational purpose has been made and stores programming logic.

The Review chain itself has two approaches as follows-

- I) Centralized approach (authority to make important decisions is retained by top level organization).
- II) Decentralized approach (authority at levels of management and in all of the organization).

A: ERC 20 Token Decentralized Design

B: Pool Key (App) Decentralized Design`

C: ERC 20 Token Centralized Design

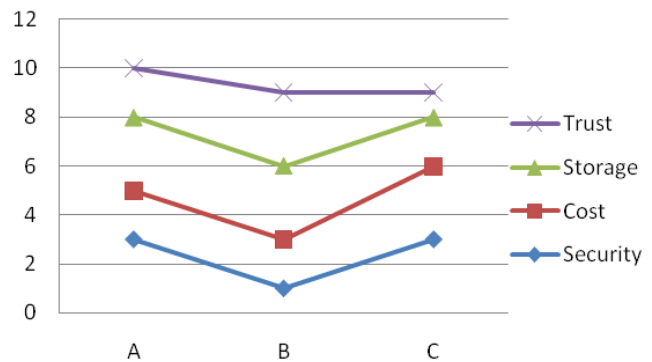


Fig.1: Comparison Of Various Design Alternatives Of Block chain.

1.2 Uncover Product Review pattern via weighted motifs:

In this paper, a network technique namely weighted motif is used to uncover underlying reviewing patterns. Pattern recognition and Network modeling has been done. Various Algorithms like: *Lambda & Delta motif mining and V-shaped motif mining* are used so that sub graphs can be enumerated to get reviewing pattern, (Li J and Ji C) [2].

```

Algorithm I: Lambda and delta motif mining
Input: PR, PP, U=∅
Output: LM, DM, U
Begin:
1 For-each reviewer R (row) in PR
2   For-each product pair <P1, P2> in R
3     If P1 ≠ P2 and RP1 ∈ U and RP2 ∈ U
4       Append <R, P1> and <R, P2> to U
5     If P1, P2 is not related in PP
6       Append <R, P1, P2 > to LM
7     Else
8       Append <R, P1, P2, L2> to DM
9     End If
10  End If
11  End For-each
12  End For-each End
    
```

```

Algorithm II: V-shaped motif mining
Input: PR, PP, U
Output: VM, U
Begin:
1 For-each product P (column) in PR
2   For-each reviewer pair <R1, R2> in R
3     If R1 ≠ R2 and PR1 ∈ U and PR 2 ∈ U
4       Append <P, R1> and <P, R2> to U
5       Append <R, P1, P2 > to VM
6     End If
7   End For-each
8 End For-each
End
    
```

1.3 Sentiment Analysis of Product Review :A Review

Extracting the opinion from the review given by consumer is the research arena of sentimental analysis. It includes Natural Language Processing, Computational Linguistics, text analytics & polarity classification of the opinion. Sentimental classification methods are usually divided into machine learning approach and lexical based approach. (T K S and Shetty J) [3].

Machine Learning Approaches:

- Supervised Approach
- Unsupervised Approach

Lexicon Based Approach:

- Dictionary Based
- Corpus Based
 - Statistical Method
 - Semantic Method

1.4 Detection of fake opinions on online product using decision tree and information gain:

Various synopses like Response, useful profile, template, star rating, reply etc are used to extract the fake review out of all the reviews given. Entropy & Information Gain are the measures to determine whether the review is spam or not

$$IG(x) = \sum_{ch=1}^n (E(x) - W * E(ch))$$

where IG = Information Gain

W = Weighted average

E(ch) = Entropy of childs

$$E(x) = - \sum p(x) \log p(x)$$

Firstly, the feature is extracted and according to defined rules and decision rule classifier/Information gain is determined and then the credibility of review found, (K.S S and Danti A) [4].

IV. RESULT



Fig.2.Process of Sentimental Analysis.

The parameters applied and studied for differentiation are:

Table 1. Comparison between different methodology based on various parameters

METHODS	ReviewChain	Detection of fake opinion on online product using Decision Tree and Information gain.	Uncover Product Review pattern via weighted motifs.	Sentiment Analysis of Product Review :A review.
POINT OF DIFFERENCE	Untampered product review on BlockChain. Author:Daniel Martens,2018, [1].	Author: Sanjay K. S, Dr Ajit Danti, 2019, [2].	Author: Jhandumli,Pin LV,Chamleji, 2018, [3].	Author: T.K Shivaprasad, Jyoti shetty,2017, [4].
ACCURACY APPROACH	High	Average	Average	Average
TECHNOLOGY USED	Decentraliz-ad Review Chain approach	Entropy Information gain	Network modelling Pattern Mining	Machine learning Lexical based approach
MACHINE LEARNING USED	Public Ethereum Blockchain platform ERC-20 Token.	WebHarvy Crawler.	Softwares-Python v.2.7.11 Jupyter,Notebook,,Elthought Canopy.	NLTK(Natural language Toolkit) and POS tagging.
PROCEDURE/FORMULATION	NO	YES	YES	YES
LIMITATIONS	App store Transaction details: Private keys are generated at the backend To make it decentralized	Entropy has been determined: E(x)=Σp(x) logp(x) where p(x) is probability of x.	Kind of networks i.e. lambda, delta and V-shaped are created and compared on the basis of: Response Useful profile Template Rating<2	A raw data is prepared using reviews: Analysis using POS, negation tagging. Sentiment classification and result is prepared
FUTURE SCOPE	Private keys can be extracted from the app. Hence affects security.	Inadequacy to achieve regression and produces discrete values.	Irregularity among binding sites.Motifs may not correspond with the accuracy of algorithms.	Cannot detect well encrypted fake reviews.
	Escalation in security of the app	To increase the accuracy of values	Text-based approaches can be done & NLB oriented review scam detection.	It can be further improved using deep learning techniques

4.1 Accuracy:

An accurate method is one, which can mark the genuine and fake review considering all the aspects. Block chain approach is high on accuracy because of the presence of nodes. Decision Tree method have different accuracy measures i.e. entropy and information gain which are considerably the best feature to find accurate fake review has 96% of success rate. Weighted motifs method has less text based approaches so it has moderate accuracy. Under sentimental analysis sometimes it becomes difficult to understand the intention of reviewer, hence it also has moderate accuracy.

4.2 Approach:

All this research papers have different approaches such as decentralized Review chain approach in block chain, a supervised approach- Entropy and Information gain in decision tree method, Enumeration based approach- Pattern Mining Network modeling. Machine learning &Lexical based approach is used in Sentimental Analysis

4.3 Technology used::

Various technologies like Public Ethereum Block chain platform, ERC 20 Token (a block chain asset), Web Harvy Crawler (a search scrapper which search keywords from data), Softwares- Python v.2.7.11, Jupyter Notebook, Elthought Canopy, NLTK(Natural language Toolkit), POS tagging (part of speech tagging).

4.4 Use of Machine Learning:

Except BlockChain, all methods make use of machine learning. Respective algorithms has been implemented so as to get the required result.for eg: Decision tree and sentimental analysis are a type of supervised machine learning. In order to maintain algorithms and machines , human involvement becomes mandatory.

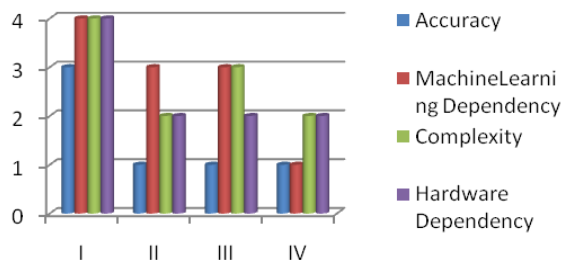


Fig. 3. Comparison between parameter of methodology of product review.

- I- Review chain: untampered product review on blockchain**
- II- Detection of fake opinion on online product using decision tree and information gain**
- III- Uncover product review pattern via weighted motifs**
- IV- Sentiment analysis of product review-A review**

4.5 Formulation/ Procedure:

All the papers which are under comparison follows different procedure like decentralized ethereal node storage, determination of entropy, building networks to find fake reviews ,classification based on sentiments are described. Some methods has used machine learning or deep learning whereas some has used mathematical formulation.

4.6 Limitations:

Every method has its own limitation which can be overlapped by the positive points of other methods. For example efficiency and security of block chain method can be further improved by reducing human involvement.. The disadvantage with the block chain method is that it depends on

human because nodes are basically humans connected to each other through computers.

4.7 Future Scope:

In this rapidly changing world technology is advancing every other moment. Hence accuracy, security, efficiency of respective methods have always the scope of improvement. Machine learning, deep learning can be put to use for the required purpose. By combining machine learning and neural system we can excel in accuracy as well as efficiency.

V. CONCLUSION

This research paper has been written to make the product review system secure, reliable and worth. Various methods has been compared on the basis of their accuracy, procedure, limitations etc . Each method has its own pros and cons. Block chain method has high accuracy but it lacks in security whereas other methods have comparatively low accuracy but has better security. Hence any suitable method can be put to use as per need.

REFERENCES

- [1] Daniel Martens, Walid Maalej “Reviewchain: Untampered product reviews on the blockchain”, 2018 in 1st International Workshop on emerging trends in Software Engineering for Blockchain in ACM, IEEE 2018. Available: <https://ieeexplore.ieee.org/document/8445057>
- [2] Jiandun Li, Pin Lv, Chunlei Ji .”Uncover Product Review Pattern via Weighted Motifs”, 5th International Conference on Systems and Informatics ICSAI, 2018. Available: <https://ieeexplore.ieee.org/document/8599334>
- [3] ShivaprasadT K,Jyoti Shetty. “Sentiment Analysis of Product Review :A Review” International Conference on Inventive Communication and Computational Technologies ICICCT, 2017. Available: <https://ieeexplore.ieee.org/document/7975207>
- [4] Sanjay K.S, Dr.Ajit Danti .” Detection of fake opinions on online products using Decision Tree and Information Gain” in 3rd International Conference on Computing Methodologies and Communication(ICCMC 2019), IEEE Xplore Part Number- CFP19K25-ART; ISBN:978-1-5386-7808-4, 2019. Available: <https://ieeexplore.ieee.org/document/8819685>
- [5] Zeenia Singla,Suckchandan Randhawa, Sushma Jain. “Statistical and Sentiment analysis of consumer product Reviews” in 8TH ICCCNT, IEEE-40222, 2017. Available: <https://ieeexplore.ieee.org/document/8203960>

- [6] Mr jayraj M.desai,Ass.prof. Swapnil R.Andhariya.” Sentiment analysis Approach to adapt a shallow parsing based sentiment Lexicon”, International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS),IEEE 2015. Available: <https://ieeexplore.ieee.org/document/7193160>
- [7] Namarata R.Bhamre,Nititn N.Patil,,”Aspect Rating analysis based product Ranking” in International Conference on Global Trends in Signal Processing, Information Computing and Communication,978-1-5090-0467-6 in IEEE 2016. Available: <https://ieeexplore.ieee.org/document/7955297>
- [8] Huaxaun Deng, Linfeng Zhao, Ning Luo, Yuan Lue, Guibing Guo, Xingwei Wang, Zhenhua Tan, Shuang Wang & Fucai Zhou.” Semi-supervised learning based fake review detection “In International Symposium on Parallel and Distributed processing with Applications and IEEE International Conference Ubiquitous Computing and Communication(ISPA/IUCC) , 2017. Available: <https://ieeexplore.ieee.org/document/8367425>
- [9] itendra Kumar Rout, Amiya Kumar Dash, Niranjan Kumar Ray. “A framework for fake Review detection: Issue and challenges” in IEEE International Conference on Information Technology(ICIT), 2018. Available: <https://ieeexplore.ieee.org/document/8724277>
- [10] Shina, Shikha Sharma, Anshu Singhal, “A study of Tree based machine learning techniques for restaurant review”, 4th International Conference on Computing Communication and Automation (ICCCA),2018. Available: <https://ieeexplore.ieee.org/document/8777649>
- [11] Afashan Ejaz,Zakia, Turabea, Maria Ratum, Shakel Khoja. “Opinion mining approach on Amazon product Review: A comparative study” International Conference on Information and Communication Technologies (ICICT),IEEE, 2017. Available: <https://ieeexplore.ieee.org/document/8320185>
- [12] Yuejun Li, Xiao Feng, Shuwu Zhang Yujen Li, “Detecting fake reviews utilizing semantic and emotional model” , 3rd International Conference on Information Science and Control Engineering.2016. Available: <https://ieeexplore.ieee.org/document/7726174>
- [13] Xiening Xu, Yu Qien, Hua Yuan.” Exploring User-Group behaviors in reviewing online product”,IEEE 2018. Available: <https://ieeexplore.ieee.org/document/8464988>
- [14] Aakanksha sharaff,Asma Soni, “Analyzing Sentiments of product reviews based on Features”, 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018. Available: <https://ieeexplore.ieee.org/document/8553794>