

Spotting Spams and Fake User Identification

Sayali Balgude¹Mohini Goykar²PrajaktaMore³Shraddha Vidhate⁴

Prof.Sarala Dabhade⁵

¹(Assistant Professor, Computer Engineering, SVPM's College of Engineering Malegaon(bk), Baramati
Email: hodcomp@engg.svpm.org.in)

²(UG Students, Computer Engineering, SVPM's College of Engineering Malegaon(bk), Baramati)
Email:sayalibalgude001@gmail.com, mohinigoykar001@gmail.com, prajaktemore148@gmail.com,
shraddhavidhate123@gmail.com)

Abstract— India is going towards digitalization. Everyone wants the things at their fingertips. So, E-Commerce websites play a vital role in it. As we all know while going through ecommerce websites or while trying to purchase any product at very first we all used to check the reviews given by people on that particular product. By analyzing top reviews we used to decide whether we should have to buy that product or not. But sometimes reviews are not genuine. They might be placed by some bots or someone else. So by using machine learning algorithms and using previous data of reviews we are going to set certain threshold limit for the reviews. If particular person is crossing that limit then his/her reviews are going into spam and respective user is fake.

Keywords: Spam Detection, Natural Language Processing, Fake User, Genuine User

I. INTRODUCTION

Nowdays, everyone wants luxury everyone wants the things to buy the products from. their home itself. and the sites which are providing it are e-commerce websites. While going through e-commerce sites, we used to check reviews of the particular. product. It means reviews play a vital role in online shopping.

Let's see how reviews matters while purchasing any product what is the impact of the reviews on customers as well as the manufacturing company. Let's take an example suppose person X wants to buy product y, then first of all he will check how many people have bought it what are the reviews on that product. If the reviews and ratings are matched and good, then person X used to buy that particular product. If someone. have put bad reviews on the product even if it is good one, then person used to revoike the idea of purchasing product .

Sometimes the quality of product genuinly not good. So people used to less rating and bad

comments on it. The company itself used to put good reviews through bot system. Due to this people used to buy that low quality product as well as .

So our task is to find the bad reviews which are not genuine one. We have done this by applying natural language processing on the dataset.

II. OBJECTIVE

In this paper a To design a real time software system that will be able detect the fake user and genuion user using NLP(Natural Language Processing).

- 1.To conduct survey of famous social networking sites for spam detection.
- 2.To identify patterns in spam attacks.
- 3.To detect spammers on social networking sites.
- 4.To analyse the solution for detecting spammers.

III. RELATEDWORK

E-Commerce Site's Fake user Review spotting and Sentiment survey using ML Technique
Every e-commerce website is on large demand just because of facility of providing reviews to the product. Any customer who is willing to purchase any product through online shopping used to see reviews. So, due to some reviews customer can be mislead. So, the main purpose of this paper is to identify the difference between reviews given in text and ratings. By finding the sentiments given through text by natural language processing we will come to know whether given reviews are actually given by any person or any bot. This paper can provide a little guidance towards the awareness of the fake and genuine reviews so that any customer who is willing to buy anything online could not be mislead through any kind of reviews.

Spotting of Fake user Identifications and Automation in Comments on E Commerce

Platform using Machine learning.

Digitalization and online shopping these are trending technologies in recent years. Everyone wants to buy the things without going to the market that is from home itself. So in this regard the ratings and reviews on any product given by customer plays very vital role. While purchasing any product through online shopping, first of all user used to see ratings and reviews on that particular product. If reviews are good then he gets a surety about the product. In this way during online shopping reviews matters a lot. Some reviews tries to put bad impact on selling and actually they not genuine one. So now this issue is becoming concerning year by year. Therefore we have implemented one system to identify that whether given reviews are fake or genuine through semi supervised learning.

IV. LITERATURE SURVEY

Decision making for buy Of Online product generally depends of reviews given by the users hence apportunisticseperate or groups. This paper introduces Some semi supervised & supervised text mining imitation to find fake online reviews also Contrast the planning of both techniques. The charge of fake review spotting has been studied in 2007, with the Scanning of review spamming in this project, the authors Scanning the case of Amazon deduce The manually labelling Fare reviews outcomes challenging as fake reviewers could exactly craft their reviews in order to make them more definitive for other users. Fake review detection analysis has mainly utilizedlexical and behavioral features. Enterprise, such as the number of reviews or the time of device where the reviews was display they were used in order to better the classification model resulting in inspire results introduced. social networking and online discourse give us the ability to express our opinions. Unfortunately, many people increasingly take social media platforms for granted and use them as a means of harassing and bullying others. This leads to cyberattacks and cyberbullying, which can in some circumstances result in horrific experiences and suicidal attempts. Such remarks must be manually identified and classified, which is a time-consuming, exhausting, and unreliable process. We have created a deep learning system that will recognise such harmful internet information in order to address this concern.

V. PROPOSED SYSTEM

In this proposed system we will be using NLP (Natural language processing) in which use the processing

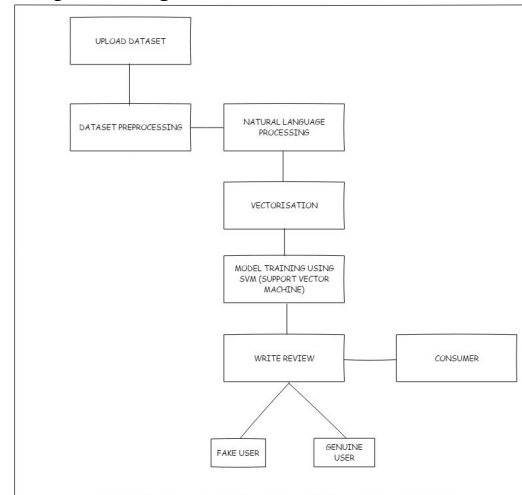


Fig 1: System Architecture

A. Dataset:-

We are Using the module that will upload the ecommerce that will upload the ecommerce reviews dataset and In those dataset the 50% dataset are fake user or 50% are Genuion userl .Recentaly we have included our newly created dataset.

B. Data Preprocessing:-

It is need to clean the data due to the existence of fake or genuion user. Fake user and genuion user text were mixed in several of the comments. We have used dealying in the fake user and genuion Datasets.

C. Model Building:-

We firstly created registration page where we like gender Birthdate, password, conform password

D. Feature Extraction:-

Firstly take an the ecommerce training dataset then perform the data preprocessing operation then we apply the NLP (Natural language processing) and SVP Support Vector Machine.

E. Expected Result:-

As an experimental output our system will generate the text output that contains the fake or real from an input of the e-commerce dataset.

VI. ADVANTAGES

- 1]in this system we have suggested this paper user can easy to use this system .
- 2]in this paper ,we have enable way for e - commerce platform out of fake reviews so the

reviews have no impact on scale of e-commerce platform.

3]through this paper ,we have to provide an automatic solution for the rude or toxic words problem people are facing when shopping on this platform as we successful remove this toxic and rude words.

4]this paper provides a technique to identify the positive and negative comments.

VII. CONCLUSION

The entire idea of project is taken on the behalf of the rise in the use of ecommerce websites. We researched a lot for the idea of the system then we came to a conclusion that machine learning is the domain from which we can built the required system. After that we have collected the dataset of reviews.

We have preprocessed it and by using natural language processing we came to a final result that whether the given review is going into spam or not and the respective user is fake or genuine.

REFERENCES

- [1].Yuejun Li, Xiao Feng, and Shuwu Zhang. Detecting fake reviews utilizing semantic and emotion model. In Information Science and Control Engineering (ICISCE), 2016 3rd International Conference on, pages 317–320. IEEE, 2016.
- [2].Nitin Jindal and Bing Liu. Review spam detection. In Proceedings of the 16th international conference on World Wide Web,pages 1189–1190. ACM, 2007.
- [3]. E. I. Elmurngi and A.Gherbi, “Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques,” Journal of Computer Science, vol. 14, no. 5, pp. 714–726, June 2018.
- [4]. Alimuddin Melleng, Anna-Jurek Loughrey, Deepak P, Text Representation Based on Sentiment and Emotion for Detection of Fake Reviews.
- [5]. Chengai Sun, Qiaolin Du and Gang Tian, Exploiting Product Related Review Features for Fake Review Detection, Mathematical Problems in Engineering, 2016.
- [6]. E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, Detecting product review spammers using rating behaviours, in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.7.
- [7]. Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. What Yelp fake review filter might be doing? In ICWSM, pages 409–418, 2013.
- [8]. Snehasish Banerjee, Alton YK Chua, and Jung-Jae Kim. Using supervised learning to classify authentic and fake online reviews. In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, page 88. ACM, 2015.
- [9]. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting spammers on Twitter,” in Proc. Collaboration, Electron. Messaging, AntiAbuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.
- [10]. S. J. Soman, “A survey on behaviors exhibited by spammers in popular social media networks,” in Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT), Mar. 2016, pp. 1–6.
- [11]. A. Gupta and R. Kaushal, “Improving spam detection in online social networks,” in Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP), Mar. 2015,pp. 1–6.
- [12]. S. Ghosh, G. Korlam, and N. Ganguly, “Spammers’ networks within online social networks: A case-study on Twitter,” in Proc. 20th Int. Conf. Companion World Wide Web, Mar. 2011, pp. 41–42.
- [13]. G. Jain, M. Sharma, and B. Agarwal, “Spam detection in social media using convolutional and long short term memory neural network,” Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21–44, Jan. 2019.
- [14]. Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-gyo Jung, Bernard J. Jansen, Creating and detecting phony online product reviews.
- [15]. J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised Learning for Fake News Detection,” IEEE Intelligent Systems, vol. 34, no. 2, pp. 76-81, May 2019.
- [16]. C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, “A machine learning approach for Twitter spammers detection,” in Proc. Int. Carnahan Conf.Secur. Technol. (ICCST), Oct. 2014, pp. 1–6.
- [17]. A. McCallum and K. Nigam, “A Comparison of Event Models for Naive Bayes

Text

Classification,” in Proceedings of AAAI-98

Workshop on Learning for Text Categorization, Pittsburgh, PA, USA, vol. 752, no. 1, pp. 41-48, July 1998.

[18]. A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, Detection of review spam: a survey, Expert Systems with Applications, vol. 42, no. 7, pp. 36343642, 2015.

[19]. A. Gupta, H. Lamba, and P. Kumaraguru, “1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter,” in Proc. eCrime Researchers Summit (eCRS), 2013, pp. 1-1.