Detection of Online Spam Reviews

Suganya V[1], Shyam V[2], KalaiChandran C[3], Sakthivel S[4]
[1]Assistant Professor, Dept of CSE
[2, 3, 4]Dept of CSE
[1, 2, 3, 4] Kumaraguru College of Technology, Coimbatore.

Abstract- *Online reviews have become a most important resource for customers. It became a habit for customers to first read a review before deciding to make purchase. But it can be used fraudsters to make review spam. So, this activity can result in the wrong customer purchase decision. Product reviews play an important role in decides the sale of a particular product on the websites or applications like Flipkart, Amazon, Snapdeal, etc. In sentiment analysis, this project tries to figure out opinion of a customer through a piece of text as review. Business organizations are monitoring their product selling by analysing and understanding what the customers are saying about products. This can help the customers to purchase valuable product and spend their money on quality products.*

Keywords- Fake Reviews, Opinion Mining, Sentiment Analysis, Fake Review Detection.

I. INTRODUCTION

A. *ACONCEPTUALSTUDY*

Internet has brought up significant changes in people's lives in recent years. Due to increased amount of usage of internet, people's behaviour while expressing their views and opinions have also been changed dramatically. The use of websites to share opinions has grown rapidly, one can easily share opinions including personal experiences, emotions, feelings, not only on a product or service but has reached the political and economic realm. It has now become a common practice of the E- commerce websites to enable the customers to provide their opinion in the form of reviews of the products. Consumers who post reviews on internet are referred to as a reviewer. Such reviews not only influence social circle of the reviewer but also helps new consumers to build their conception. New consumers increasingly depend on these user-posted online reviews to make the buying decisions, thus the truthfulness of information available in the form of reviews is of almost importance.

B. OBJECTIVE ANDSCOPE

In this project the detected spam reviews are listed separately and then provides the information that either review is real or fake for, the customer to buy. This helps in understanding the quality of product to customer easily with the help of the following methods and algorithms that are used here. This project plays major role in online shopping or booking platforms to determine whether the review is a legit one or a spam. This project aims in helping customers to go for a quality purchase. People like online traders, ordinary buyers, wholesale buyers support this type of systems to get rid of fake reviews.

II. LITERATURESURVEY

A. OPINION SPAM ANDANALYSIS

In this work of past few years, sentiment analysis and opinion mining become a popular and most important task. These studies all assume that their opinion resources were real and trustful. However, they may encounter faked opinion or opinion spam problem. The paper study this issue in the context of the product review mining system. On product review site, people may write faked reviews, called review spam, to promote their products, or defame their competitors' products. It is important to identify and filter out review spam from revies.

Previous work only focuses on some heuristic rules, such as helpfulness voting, or rating deviation, which limits the performance of this task. This paper exploit machine learning methods to identify review spam. Toward the end, they manually build a spam collection from our crawled reviews. This paper first analyses the effect of the various features in spam identification. This paper also observe that the review spammer consistently writes spam. This provides another view to identify review spam: this paper can identify if the author of the review is spammer. Based on this observation, this one provides a two view. Semi-supervised method, co-training, to exploit the large amount of unlabelled data. The experiment results show that our proposed method is effective. This designed machine learning methods achieve significant improvements in comparison to the heuristic baselines.

B. *OPINIONFRAUDDETECTIONINONLINE REVIEWS BY NETWORKEFFECTS*

Opinions are big part of people relies on available content in social media in their decisions (e.g. reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies had been done recently toward this end but so far, the methodologies put forth still barely detect spam reviews, and none of them show the importance 10 of each extracted feature type. In this study, this paper proposes a novel framework, named Net Spam, which utilizes spam features for modelling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help them to obtain better results in terms of different metrics experimented on real-world review datasets from Yelp and Amazon websites. The results show that Net Spam outperforms the existing methods and among four categories of features; including review-behavioural, user-behavioural, review linguistic, user-linguistic, the first type of features performs better than the other categories.

C. DETECTION OF FAKE OPINIONSUSING TIMESERIES

In this work Online reviews play a crucial role in helping consumers evaluate and compare products and services. This critical importance of reviews also incentivizes spammers to write fake or spam reviews to secretly promote or demote some target products and services. Existing 12 approaches to detecting spam reviews and reviewers employed review contents, reviewer behaviours, star rating patterns, and reviewer-product networks for detection. This paper then extend it to the Coupled Hidden Markov Model to capture both reviewers posting behaviours and co-bursting signals. In these experiments show that the proposed model significantly outperforms state-of-theart baselines in identifying individualspammers

D. IMPACTOFONLINECONSUMERREVIEWSON SALES

The moderating role of product and consumer characteristics," in this work how product and consumer characteristics moderate the influence of online consumer reviews on product sales using data from the video game industry. The findings indicate that online reviews were more influential for less popular games and games whose players had greater Internet experience. The differential impact of consumer reviews across products in the same product category and suggests that firms' online marketing strategies should be contingent on product and consumer characteristics.

E. SPOTTING FAKE REVIEWS GROUPSIN CONSUMERREVIEWS

In this work Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming to promote or demote some target products. For reviews to reflect genuine user experiences and opinions, such spam reviews should be detected. Prior works on opinion spam focused on the detecting fake reviews and individual fake reviewers.

However, a fake reviewer group is even more damaging as this paper can take total control of the sentiment on the target product due to its size. This work studies spam detection in the collaborative setting to discover fake reviewer groups.

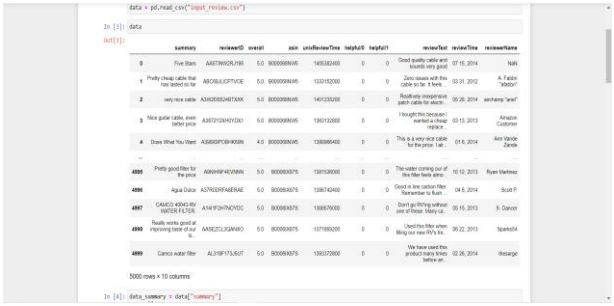
The proposed method first uses a frequent item set mining method to find a set of candidate groups. It then uses several behavioural models derived from the collusion phenomenon among fake reviewers and relation models based on the relationships among groups, individual reviewers, and products they reviewed to detect fake reviewer groups. Additionally, this paper also built a labelled dataset of fake reviewer groups.

III. PROPOSEDSYSTEM

Buying product in online is normal in today world. In this project we can detect online fake reviews by using K-means Clustering and NLP.

A. DATA PREPROCESSING:

In this module collection of datasets is taken from Kaggle website. Dataset consists of features and labels. Features and labels are extracted from dataset and as data is in string format it is converted using vectorizer technique.



B. PRODUCTS REVIEW CLASSIFICATION USING K MEANS:

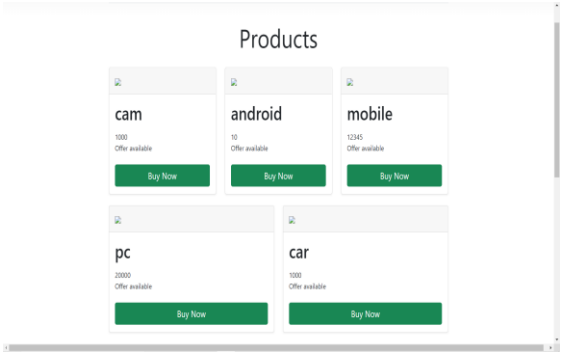
In this module all the products were listed in the page. Users can select the products and provide review for every product separately. The product review is classified into positive & negative proportions using K-means.

C.PRODUCTREVIEWANAYSISUSING NLP:

The reviews provided by the user is analyzed for its quality using NLP (Natural Language Processing). It measure sentiment and determine which parts are important in the given review. The result is then clustered using K-means Algorithm which predicts the percentage of the text quality as positive or negative.

D. INTEGRATION WITH FRONT END:

A Django administration part is created as a Web Interface with admin authentication, product creation, users & reviews seeing ability. User interaction with the project is provided for ease usage.



IV. CONCLUSION

It is important that spam reviews do not reach users as this reduces efficiency of operations. But more importantly it is necessary that no reviews reaches user those lead serious problems to the user. As the analysis and results section has revealed, looking at all the parameters collectively, it is found that SVM is the best classifier of this study. Due to these reasons, the SVM classifier should be used as it will yield positive and desirable results. It will help in classifying the spam reviews in their respective folders. On the other hand, it is also seen that ensemble of classifiers technique helps in learning capability of the classifiers. The time taken to build the model for the best classifier found for this study, which is SVM, is high. However, as already established, even though the time taken to build the model is more, the results are more effective and thus even if it takes a little longer to achieve more desirable results.

V. ACKNOWLEDGEMENT

We express our profound gratitude to the management of Kumaraguru College of Technology for providing us with the required infrastructure that enabled us to successfully complete the project.

We extend our gratitude to our Principal, **Dr. D. Saravanan**, for providing us the necessary facilities to pursue the project.

We would like to acknowledge **Dr. P. Devaki**, Professor and Head, Department of Computer Science and Engineering, for her support and encouragement throughout this project.

We thank our project coordinator **Dr. L. Latha**, Professor, Department of Computer Science and Engineering and guide **Mrs. Suganya V**, AssistantProfessor, Department of Computer Science and Engineering, for their constant and continuous effort, guidance, and valuable time.

Our sincere and hearty thanks to staff members of Department of Computer Science and Engineering of Kumaraguru College of Technology for their well wishes, timely help and support rendered to us during our project. We are greatly indebted to our family, relatives and friends, without whom life would have not been shaped to this level.

REFERENCES

[1] N. Hussain, H. Turab Mirza, I. Hussain, F. Iqbal andI. Memon, "Spam Review Detection Using the Linguistic

and Spammer Behavioral Methods," in *IEEE Access*, vol. 8, pp. 53801-53816, 2020, doi: 10.1109/ACCESS.2020.2979226.

[2] Crawford, M., Khoshgoftaar, T.M., Prusa, J.D. *et al*, *"*Survey of review spam detection using machine learning techniques", Journal *of Big Data* **2,** 23 (2015). https://doi.org/10.1186/s40537-0 15-0029-9

[3] Simran Bajaj, Niharika Garg, Sandeep KumarSingh, "A Novel User-based Spam Review Detection", Procedia Computer Science, Volume 122,2017, Pages 1009-1015, ISSN 1877- 0509,

[4] Shehnepoor. S, Salehi. M, Farahbakhsh. R, and Crespi. N, (Jul. 2017), ''Net Spam: A network-based spam detection framework for reviews in online social media,'' IEEE Trans. Inf. Forensics Security, vol. 12, no. 7, pp. 1585–1595.

[5] S. Zhao, Z. Xu, L. Liu, and M. Guo, "Towards accurate deceptive opinion spam detection based on wordorder-preserving CNN", Available CoRR, vol. abs/1711.09181,2017

[6] M.N. Istiaq Ahsan, Tamzid Nahian, Abdullah AllKafi, Md. Ismail Hossain, and Faisal Muhammad Shah. "Review spam detection using active learning", Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE 7th Annual,2016.

[7] Zhang Y. and Wallace B," A sensitivity analysis of convolutional neural networks for sentence classification", Pro-ceedings of the 8th International Joint Conference on Natural Language Processing, pages253263, Taipei, Taiwan, November 27 December1,2017.

[8] E.D. Wahyuni and A.Djunaidy, "Fake Review Detection from a Product Review using Modified Method of Iterative Computation Framework", In proceeding MATEC Web of Conferences.2016.

[9] Shivaprasad TK Jyothi Shetty "Sentiment Analysis of Product Reviews: A Review", International Conference on Inventive Communication (ICICCT2017).