A Novel Approach for Fake Comments and Reviews Detection on the Online Social Networks

Akshat Gaurav*[1], B.B. Gupta[2], Kwok Tai Chui*[3], Dragan Peraković[4], Priyanka Chaurasia[5] and Ching-Hsien Hsu*[6]

[1]*Ronin Institute, Montclair, New Jersey 07043, U.S.*

[2]*National Institute of Technology Kurukshetra, Kurukshetra-136119, Haryana, India & Asia University, Taichung 413, Taiwan & Staffordshire University, Stoke-on-Trent ST4 2DE, UK*

[3]*Hong Kong Metropolitan University (HKMU) , Hong Kong, China,*

[4] *University of Zagreb, Croatia,*

[5]*Ulster University, Magee campus, Londonderry, UK*

[6]*Department of Computer Science and Information Engineering, Asia University, Taiwan*
& Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan
& Department of Medical Research, China Medical University Hospital, China Medical University, Taiwan
**Corresponding Author*

Abstract

As the primary source of information dissemination, social media networks have surpassed traditional news organisations for the first time. Nonetheless, as the number of people who use social media websites grows, they become more susceptible to the spread of misinformation, making it increasingly difficult to distinguish between real news and false news in real time. In this paper, we proposed a machine learning technique for the detection of fake comments in social networks. According to the results of the experiment, it is clear that the machine learning technique efficiently detects the fake comments.

Keywords

Deep learning, Fake comments, Machine learning

1. Introduction

For many reasons, social media has overtaken email as the most important distribution and consumption medium for news and information. For starters, getting news via social media is usually faster and less expensive than getting news from conventional sources[1]. Further, engaging and interacting with other readers by commenting, debating, and fighting with them is a great method to get one's points through while also encouraging participation and participation. Despite these developments, spreading real-time information with the use of social media has contributed to the spread of disinformation, often referred to as fake comments and reviews.

International Conference on Smart Systems and Advanced Computing (Syscom-2021), December 25–26, 2021

akshat.gaurav@roninstitute.org (A. Gaurav*); gupta.brij@gmail.com (B.B. Gupta); jktchui@ouhk.edu.hk (K. T. Chui*); dperakovic@fpz.unizg.hr (D. Peraković); p.chaurasia@ulster.ac.uk (P. Chaurasia); robertchh@gmail.com (C. Hsu*)

Recommendations and feedback are becoming more important as internet communication technology advances. In today's world, people are increasingly relying on internet reviews to assist them make a purchasing choice. For company owners, internet comments are a way to develop and better their enterprises. Product enhancements may be made based on user input via online comments. However, internet remarks are not always honest, and false online comments are common [2]. There are company owners that will pay for nice or poor evaluations regarding their rivals' items to be written. Consumers are misled by these phoney reviews and end up purchasing inferior goods because of it. Both consumers and sellers depend on honest evaluations to guide their purchasing choices, and false testimonials may have a significant impact on both. Innocent clients may suffer financial losses as a result of this. As a result, many people are interested in learning how to spot fake comments. Most shopping websites, on the other hand, have solely addressed the issue of negative ratings and comments. As a result, the detection of false reviews is critical in both corporate and academic settings[3].

2. Related Work

Several academics have offered a number of methods for identifying various cyber threats [4, 5, 6, 7, 8] as well as phoney reviews and comments[9, 10]. In this part, we discuss some of the most extensively used fraudulent review and comment detection techniques.

There are a number of popular social networking sites, like Facebook and Twitter, that are used by internet users to obtain information on the World Wide Web[11]. Spammers and hackers who transmit harmful depicted objects in the form of spam through social networking platforms are examples of content protection[12, 13].

Online social networking has developed as one of the most popular methods to exchange information and communicate with others in the course of everyday life. Online social networking is a fun and convenient way to meet new people and keep in touch with old friends. There are, however, worries about user privacy and account security. During events, people use social media sites like Twitter and Facebook to disseminate false information. Author in [14] presented a method to identify phoney accounts by studying the features of dangerous information that spreads in real-time. Fake profiles are created by stealing the personal information of a real user and utilising that information to establish a new profile. As time goes on, the profile gets hacked such that it may send friend requests to a friend of the original account holder. The suggested method outlines our chrome extension-based architecture for detecting bogus Twitter accounts by examining several attributes.

Web applications are automatically scanned for XSS attack vectors using XSS-explorer, a universal and automated server-side flexible framework proposed by the author in [8]. Extensive XSS attack investigations are generated for each web application's injection locations, which may be explored using the built-in XSS-explorer tool. This strategy relies on approaches that allow for the accurate filling of injection locations in forms with relevant information. Identification of these points allows us to search for every possible web page of the application, allowing us to look for more attack vectors and speeding up the process of finding them.

3. Setup

The suggested method uses six core machine learning techniques to identify bogus news. Tokenization and stop phrasing are the first steps in the recommended technique. Stastical approaches are used to assess the performance of each ML methodology.

It might take a long time and be difficult to determine if a remark is real or not. Because of this, a previously gathered and recognised dataset of bogus news was deployed. We drew inspiration for our project from the Kaggle database. The dataset includes a header, title and text columns, as well as a fake/real comment item flag.

It is necessary to remove stop words from text before adding it into machine learning models. Our models will perform better if we can use these methods to find and optimise the most relevant terms. There are a lot of useless words and strange characters in our datasets since we utilise real-world news items. Our data collection was simplified as a consequence of the removal of these extraneous characters. Stop words are removed as the last step in preprocessing. They were thus omitted from all of our testing due to the potential for excessive noise they would have generated.

In the last section, we used machine learning models to pre-process the data. In order to get over this limitation, we employ a counter vectorized to translate the text data into vector form. Finally, we evaluate the performance of multiple machine learning models using statistical techniques.

4. Results and Disscusion

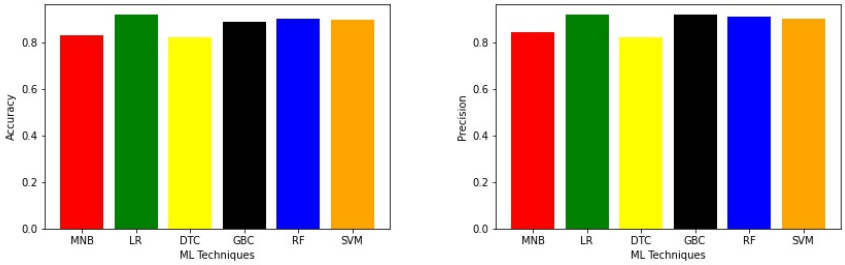
It is possible to improve the classification accuracy of various techniques by varying the quantity of labelled data used in training . When just a tenth of the labelled data is utilised, the suggested detection methods demonstrate an improvement in performance of up to three percent, while simultaneously improving the accuracy and lowering the false postive. After extracting and transforming the undesired content using tokenization algorithms. Following that, we trained distinct models using machine learning techniques. The performance of these machine learning models is evaluated using following eqations, and result is represented in figure 1.

$$Accuracy = \frac{\delta_P + \delta_N}{\delta_P + \delta_N + \delta'_P + \delta'_N} \tag{1}$$

$$Racall = \frac{\delta_P}{\delta_P + \delta'_N} \tag{2}$$

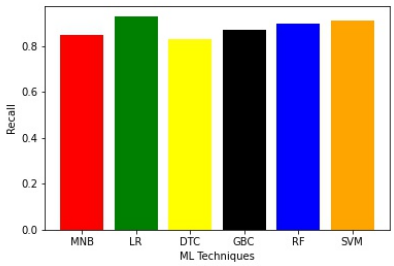
$$Precision = \frac{\delta_P}{\delta_P + \delta'_P} \tag{3}$$

$$F1 - Score = 2 \times \frac{\delta P \times \delta R}{\delta P + \delta R} \tag{4}$$

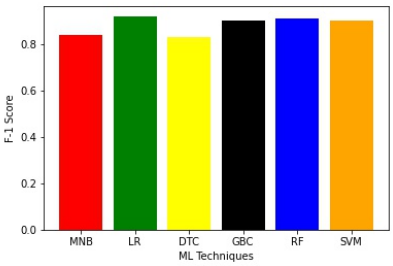


(a) Accuracy

(b) Precision



(c) Recall



(d) F-1 Score

Figure 1: Statistical Parameters Calculation

5. Conclusion

We examined social media reviews and comments in this research and created a novel approach for detecting fraudulent reviews and comments. We used machine learning technologies to identify fake reviews and comments. To show our system's usefulness and efficiency, we compared it to a variety of other temporal outlier detection approaches. Detecting fraudulent reviews using review data involves various challenges. Our study could not conclusively establish when a product is most likely to be the subject of fake reviews and comments, which is an exciting topic for more research.

References

[1] K. Kumari, Online social media threat and it's solution, Insights2Techinfo (2021) 1.

[2] H. Deng, L. Zhao, N. Luo, Y. Liu, G. Guo, X. Wang, Z. Tan, S. Wang, F. Zhou, Semi-supervised learning based fake review detection, in: 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), IEEE, 2017, pp. 1278–1280.

[3] W. Liu, J. He, S. Han, F. Cai, Z. Yang, N. Zhu, A method for the detection of fake reviews based on temporal features of reviews and comments, IEEE Engineering Management Review 47 (2019) 67–79.

[4] A. Gaurav, A. K. Singh, Light weight approach for secure backbone construction for manets, Journal of King Saud University-Computer and Information Sciences (2018).
[5] D. P. Ivan Cvitić, G. Praneeth, Digital forensics techniques for social media networking, Insights2Techinfo (2021) 1.
[6] Z. Zhou, A. Gaurav, B. Gupta, H. Hamdi, N. Nedjah, A statistical approach to secure health care services from ddos attacks during covid-19 pandemic, Neural Computing and Applications (2021) 1–14.
[7] Z. Zhou, A. Gaurav, B. B. Gupta, M. D. Lytras, I. Razzak, A fine-grained access control and security approach for intelligent vehicular transport in 6g communication system, IEEE Transactions on Intelligent Transportation Systems (2021).
[8] S. Gupta, B. B. Gupta, Robust injection point-based framework for modern applications against xss vulnerabilities in online social networks, International Journal of Information and Computer Security 10 (2018) 170–200.
[9] J. Zhao, H. Wang, Detecting fake reviews via dynamic multimode network, International Journal of High Performance Computing and Networking 13 (2019) 408–416.
[10] A. Gaurav, B. Gupta, A. Castiglione, K. Psannis, C. Choi, A novel approach for fake news detection in vehicular ad-hoc network (VANET), in: International conference on computational data and social networks, 2020, pp. 386–397. Tex.organization: Springer.
[11] S. R. Sahoo, B. B. Gupta, Classification of various attacks and their defence mechanism in online social networks: a survey, Enterprise Information Systems 13 (2019) 832–864.
[12] S. R. Sahoo, B. B. Gupta, Classification of spammer and nonspammer content in online social network using genetic algorithm-based feature selection, Enterprise Information Systems 14 (2020) 710–736.
[13] S. R. Sahoo, B. B. Gupta, Hybrid approach for detection of malicious profiles in twitter, Computers & Electrical Engineering 76 (2019) 65–81.
[14] S. R. Sahoo, B. Gupta, Real-time detection of fake account in twitter using machine-learning approach, in: Advances in computational intelligence and communication technology, Springer, 2021, pp. 149–159.