SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673

FAKE ACCOUNT DETECTION ON TWITTER USING MACHINE LEARNING

S.Indarjeet Singh
Professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, drindar2020@gmail.com

Kondlada Shirisha
U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

Kondur Nikitha
U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

Singala Tejaswini
U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

ABSTRACT
Spam plays a major role in social media, like Twitter. Twitter is the major platform for spreading the news all over the world. So the users always choose Twitter as a platform to convert in to a target by sharing fake posts, fake news, etc. So, the spammers are used to spreading an enormous amount of false and deleterious data. Twitter is an online site where people can give their opinions, news, and everything. Furthermore, the ability to spread false information via fake identities results in the spreading of hazardous materials. To identify this spam, we are using machine learning algorithms.
Key words: Spammer's identification, Machine learning algorithms, Social media, Spam detection.

Introduction
Spammer detection and identification of fake users is detecting unwanted messages or posts on social media. There are number of methods and techniques to identify fake posts and accounts on social media. So we can detect spammers on Twitter by using fake content, URL-based spam detection, and fake users. Twitter is an online social media network for acquiring real-time information about users. So, Twitter is an online platform where everyone can share their opinions, comments, etc. Twitter is an online platform that spreads information rapidly. So it is difficult to maintain the social network's security. Therefore, it is very difficult to identify spam on OSN sites. We have to save the users from harmful attacks from spammers. Several types of research were carried out to detect the spammers on Twitter. There are a number of existing systems that presents the spammer's behaviours on Twitter. So there is a proposed methodology like machine learning to recognize the spammers on Twitter. It is estimated that 40% of social

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673
sites such as Twitter are used for spam. The spammers use most popular networking tools in order to target some specific areas, review pages, or fan pages to spread false information in the form of text. Regular highlights are shared with the users through email.
These highlights are examined so that one can improve the detection of these types of emails. By using artificial intelligence (AI), emails are classified into spam and non-spam emails. This can be possible by using feature extraction from the message headers, subject, and body. After extracting this data, we can divide them into spam or nonspam. For the detection of spam nowadays learning-based classifiers are used. In learning-based classification, spam emails contain a set of features that seperates them from actual leagal emails. Several factors increase the complexity of spam detection in learning-based models. These factors contain spam subjectivity, idea drift, language problems, overhead processing, and text latency.
One of the examples of learning-based models is an extreme learning machine (ELM). This is a modern machine learning model that has only one hidden layer. It removes slow training speed and overfitting problems when it is compared with traditional neural networks. In ELM, it requires only one cycle of iteration.

Related Work

In this section there are many existing methods that are discussed and defined based on their applications.

A. Twitter fake account detection.

B. Ercahin, O. Aktas, D. Kilinc, and C. Akyol presented a framework to detect spammers on social networks. The increase in usage of social sites led to an increase in the probability of spreading invalid information to the actual users through fake accounts, which results in the spreading of harmful content. This can result in huge damage to society. They presented a classification method for finding fake accounts on Twitter. They preprocessed the dataset using a supervised discretization technique named Entropy Minimization Discretization (EMD) on numerical features and described the results of the Naïve Bayes algorithm.

B. A survey on spammers in social media networks.

S. J. Soman presented a framework to detect the spammers by concentrating on the development of Honey pots. Spammers converted Twitter in to a target platform. The authors survey the related literature that identifies the presence of spam as well as spammers in popular social media networks.

C. Detection of spam in tweets using NLP.

S. Gharge, and M. Chavan have presented a method based on two aspects: the identification of spam tweets without knowing the previous background of the user; and the other based on the analysis of language for detecting spam on Twitter in trending topics at that time. This method tries to detect spam tweets based on the language tools. First, all the tweets are collected that are related to many trending topics, differentiating them on the basis of their content, which is either harmful or safe. Also, the performance is evaluated, and the classification of tweets as spam or not spam is processed. Thus, the above system can be used for spam detection on Twitter, focusing mainly on analyzing tweets instead of user accounts.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673

Existing System

In this section, we discuss the existing system for the detection of spam and for the identification of fake users. Social media are the simplest way for millions of people to interact with the world. Nowadays, people are spending more time on social platforms and they are sharing their opinions and their personal information on social media. Some users create unwanted, harmful links, posts, tweets, fake news, etc. Thus, the existing system does not give better results for spammer detection.

Proposed System

Here, we discuss the proposed system using algorithms to detect spammers on social media. There are several machine learning algorithms that are used in detection of spam. The proposed method is categorized into four main classes, namely (i) fake content, (ii) URL based (iii) spam on trending topics (iv) fake user identification. These include various techniques, such as regression, prediction models, malware alerting systems, and identifying fake URLs through different machine learning algorithms. So, by this method, we can compare all algorithms to get the best results and best accuracy.

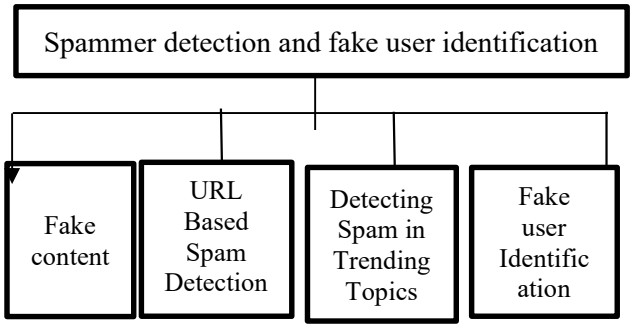


Figure.1 Spam Detection Techniques

System Architecture

The proposed system contains many techniques used for Twitter spam detection. We detect spam on social networks using different machine learning algorithms. Figure.1 shows the system architecture of spammer detection which contains the following machine algorithms.

1.Random Forests Algorithm : Random forests is a learning, flexible, and easy-to-use algorithm. Classification and regression can be done by using a random forests algorithm. A random forest algorithm contains a huge number of trees. It will be more robust If it has more trees. A random forest algorithm is used to create decision trees on some selected data samples, acquires conclusions from each tree, and selects the better solution through voting. Random forests have various applications, such as feature selection, and image classification.

2.Naive Bayes algorithm : It is most commonly used for classifying problems that are simple probabilistic classifier and it is based on Bayes' Theorem. The probability of each feature occurring in each class is determined, and it returns the outcome with the highest probability.

3.Extreme learning machines (ELM) : These are the neural networks used for classification, regression, compression, and feature learning to use a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes are not tuned. These hidden nodes are given randomly and will not get updated, or can get their features from their ancestors without any

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673

change. In most of cases, the output weights of hidden nodes are usually learned in a single step, which essentially amounts to learning a linear model.

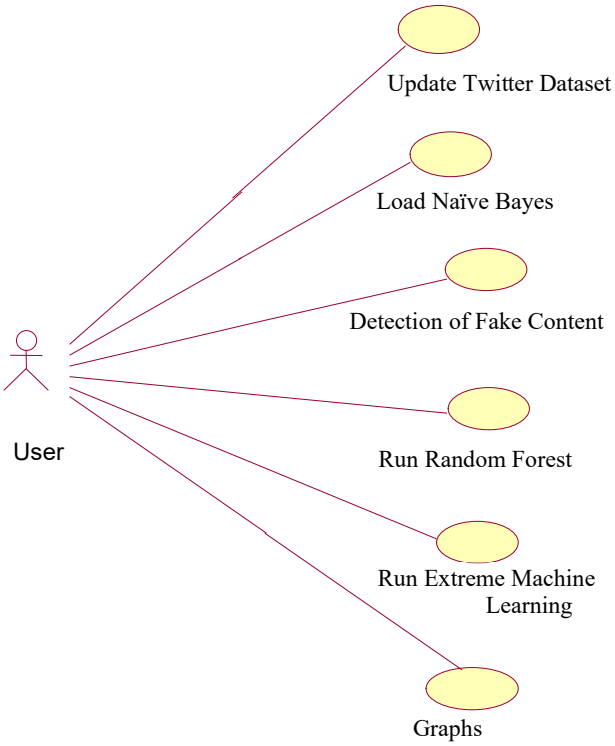


Figure.2 Architecture diagram for spammer detection

In this, the first step is to load all the tweets from all users. To analyze these Tweet texts or URL, we have to load the Naïve Bayes classifier. Then, all the user's accounts are analyzed and we identify whether the account is normal or it contains spam content using the naïve bayes algorithm. Now click on 'Run Random Forest For Fake Account' button to build a machine learning model on the above data. So we can predict whether the account is normal or fake in the future by using account details. Now click on 'Detect Fake Content, Spam URL, Trending Topic & Fake Account' to detect spam accounts and create machine learning features.

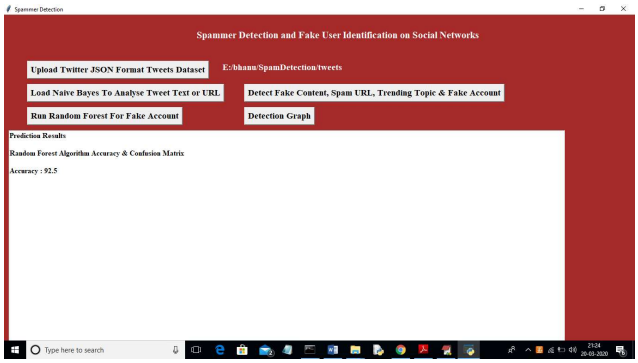
In the above picture, we can see the different buttons through which we can load different algorithms and compare them in order to find the algorithm which has high accuracy to detect spam on social networks like Twitter.

Result

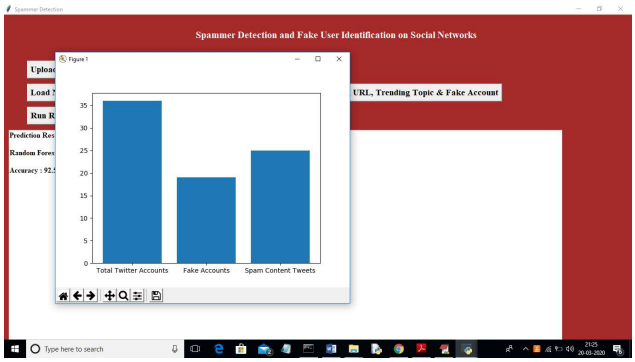
we have used different techniques for detecting spammers on Twitter. We also presented

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673

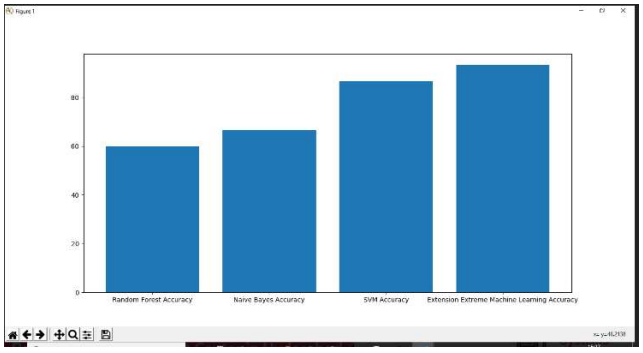
different approaches for detection of spam on Twitter and classified them into four different techniques such as detection of fake content , URL-based spam detection, detection of spam in trending topics, and fake user identification techniques.



As shown in the above figure we compared the techniques used for detection of spam and machine learning algorithms based on different features, such as user, structure, content, graph, and time features. Also, these techniques were compared with the datasets we have used. We can see that the prediction accuracy of the random forests algorithm is about 92.5 which is the highest among all the algorithms used.



In the given picture, we can see the total number of Twitter accounts we have uploaded in our dataset and it also shows the number of fake accounts and number of spam content tweets. It is expected that the presented review will help researchers to find the information on spam detection techniques used on Twitter in an integrated form.



In the above figure it is given that the accuracy of different machine learning algorithms we are using inorder to detect the spam on twitter. We can observe that the accuracy of Extension

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6875673
Extreme Machine learning algorithms is high among all of them.

Conclusion

In this, we are using machine learning algorithms to identify spammers on Twitter. There are a number of strategies presented based on the number of characteristics, such as user features, material features, graph features, structure, and time features. So we have compared all the strategies to get the best accuracy. We detect spam on Twitter based on URLs,spam detection in trend topics, and fake user detection. By using these techniques, we can detect malware in social media.

Acknowledgement

Special thanks for the guidance to our supervisor professor Dr. Indarjeet Singh.

References

[1] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ‘‘Twitter spam detection: Survey of latest methodologies and comparative study,’’ Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.

[2] C. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano, E. Scillia, and R. Surlinelli, ‘‘Spam detection of Twitter traffic: A method based on random forests and non-uniform feature sampling,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 811–817.

[3] 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. 1–12.

[4] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ‘‘Analysis for malware discovery on Twitter,’’ in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.

[5] N. Eshraqi, M. Jalali, and M. H. Moattar, ‘‘Detecting spam tweets in Twitter using a data stream clustering algorithm,’’ in Proc. Int. Congr. (ICTCK), Nov. 2015, pp. 347–351.

[6] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ‘‘Statistical features-based real-time detection of drifted Twitter spam,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.

[7] C. Buntain and J. Golbeck, ‘‘Automatic identification of fake news in Twitter threads,’’ in Proc. IEEE Int. Conf. Smart Cloud Nov. 2017, pp. 208–215.

[8] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, ‘‘A machine learning-based spam tweets detection,’’ IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.

[9] G. Stafford and L. L. Yu, ‘‘A performance evaluation of the effect of spam on Twitter trending topics,’’ in Proc. Int. Conf. Social Comput., Sep. 2013, pp. 373–378.

[10] M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, ‘‘An approach for Twitter spam detection,’’ in Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2017, pp. 466–471.

[11] R. Kaushal, and A. Gupta ‘‘Spam detection improvement in social networks’’ in Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP), Mar. 2015, pp. 1–6.

[12] M. Bouguessa, and F. Fathaliani ‘‘An approach for identifying spammers in social networks,’’ in Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2015, pp. 1–9.

DOI: 10.5281/zenodo.6875673
[13] A. Pilaniya, V. Middha, V. Chauhan, A. Gupta, U. Bana, and S. Agarwal, ‘‘Abnormal behavior detection on social networks,’’ in Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2017, pp. 1–5.
[14] G. Noh, S. Jeong H. Oh, and C.-K. Kim, ‘‘Spam detection based on social information,’’ Inf. Sci., vol. 369, pp. 481–499, Nov. 2016.
[15] M. Washha, A. Qaroush, and F. Sedes, " Controlling time for detection of spammers on Twitter,’’ in Proc. 8th Int. Conf. Manage. Digit. EcoSyst., Nov. 2016, pp. 109–116.
[16] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, ‘‘Presenting the features for spam detection on Twitter,’’ 2015, arXiv:1503.07405. [Online]. Available: https://arxiv.org/abs/1503.07405
[17] M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan, S. Din, A. Ahmad, G. Jeon, and A. G. Reddy, ‘‘Towards multilingual URL filtering: A big data problem,’’ J. Supercomput., vol. 74, no. 10, pp. 5003–5021, Oct. 2018.