As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 9(02): 64-68, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0902011
Available online at: http://www.ijmtst.com/vol9issue02.html

Spammer Detection and Fake user Identification on Social Networks

Dr.D.Suneetha[1] | K. Mani Chandana[2] | M. Darahaasa [2] | P. Darshini Ratna Chowdary [2] | M. Yeshwanth[2]

[1]Professor & HOD, Department of CSE, NRI Institute of Technology, India
[2]B.Tech Student, Department of CSE, NRI Institute of Technology, India

To Cite this Article
Dr.D.Suneetha, K. Mani Chandana, M. Darahaasa, P. Darshini Ratna Chowdary, M. Yeshwanth. Research on Spammer Detection and Fake user Identification on Social Networks. International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 64-68. https://doi.org/10.46501/IJMTST0902011

Article Info
Received: 28 December 2022; Accepted: 23 January 2023; Published: 31 January 2023.

ABSTRACT

Modern social media dominates many areas. Social media use has skyrocketed. Social media helps us connect with people and express ourselves. This enabled identity fraud, fact falsification, and other assaults. Recent evidence reveals that social media accounts outnumber active users. This suggests a recent spike in fake accounts. Social media networks struggle to identify fake accounts. Due to the rise of fake accounts, advertisements, etc. on social media, detecting them is crucial. Traditional methods fail to distinguish real from fake accounts. The aforementioned articles are outdated due to advances in fake profile creation. The new models employed automated posting and commenting, false facts, and promotional material in spam to identify fake accounts. Multiple algorithms, each with their unique traits, are needed to combat fake accounts. Naive bayes, support vector machine, and random forest no longer identify fake accounts. This work provided an innovative technique to address this problem. Three-attribute decision trees and gradient boosting were used. Engagement, phony activity, and spam comments are examples. Machine Learning and Data Science helped us uncover bogus profiles.

KEY WORDS: Classification, fake user detection, online social network, spammer's identification.

1. INTRODUCTION

The use of social media is an essential component of everyone's life in the modern culture of today. Keeping in contact with friends, disseminating information, and other activities are the primary goals of using social media. The number of people using various forms of social media is skyrocketing at an alarming rate. Instagram has lately seen a meteoric rise in popularity among users of several social networking platforms. Instagram has become one of the most popular social media platforms due to the fact that it has more than one billion active users. People who have a significant number of followers on Instagram have been given the title of social media influencers since the platform's introduction to the social media landscape. The corporate organization is now turning to these social media influencers as a go-to destination to sell their goods and services.

The proliferation of people using social media platforms has turned out to be both a blessing and a curse for modern civilization. The use of social media platforms to commit online fraud or to disseminate false information is rapidly growing in prevalence. On social media platforms, the majority of misleading information comes from fake accounts. Businesses and other organizations who spend a significant amount of money on social media influencers need to determine whether or not the following that an account has earned is the result of organic growth. As a result, there is a significant need for a tool that can identify bogus accounts and determine with high levels of precision whether or not an account is fake. In this research, we make use of classification methods from the field of machine learning in order to identify bogus accounts. Finding a fake account is mostly dependent on a number of different characteristics, such as the interaction rate and the amount of bogus activity.

2. LITERATURE SURVEY

Benevenuto et al. [2] conducted research on the challenge of identifying spammers on Twitter. In order to do this, a substantial dataset from Twitter is compiled; this dataset includes more than 5400 million users, 1.8 billion tweets, and 1.9 billion connections. Following this step, the number of variables that are related with the content of tweets as well as the characteristics of users are identified for the purpose of identifying spammers. These parameters are taken into consideration as aspects of the machine learning process for the purpose of classifying users, namely to determine whether or not they are spammers. The tagged collection in pre-classification of spammers and non-spammers has been done in order to identify the strategy for detecting spammers on Twitter. This was done in order to recognize the approach. A crawling operation has been initiated on Twitter in order to collect the IDs of its users, of whom there are around 80 million. Every user on Twitter has a unique numeric ID assigned to their account, which may be used to quickly and easily identify their profile. After that, the processes that are necessary for the building of a labelled collection and the acquisition of different desirable attributes are carried out. In other words, the procedures that are absolutely necessary to be investigated in order to build up a collection of users that may be classified as spammers or as users that do not spam. At the very end, user characteristics are determined on the basis of the user's behavior, such as who the user interacts with and how often they connect with that person. In order to give credence to this hunch, characteristics of people who have used the tagged collection have been investigated. To separate one user from another, two distinct sets of characteristics—namely, the content attributes and the user behavior attributes—are taken into consideration. The wordings of tweets that are posted by users are the property of content attributes, which collect aspects that are related to the way users compose tweets and have the property of the wordings of tweets that are posted by users. On the other hand, user behavior characteristics collect certain elements of the behavior of users in the context of the posting frequency, engagement, and influence on Twitter. These attributes are gathered together. The total number of followers and following, the age of the account, the number of tags, the fraction of followers per followings, the number of times users have replied, the number of tweets received, the average, maximum, minimum, and median time among user tweets, and daily and weekly tweets are all considered to be user characteristics. In all, 23 aspects of the user's behavior were taken into consideration for this study. The results of the technique that was presented demonstrate that even with the differentiated set of features, the framework is able to identify spammers with a high frequency. This was determined by looking at the results. Jeong et al. [17] investigated the use of follow spam on Twitter as a method for the dissemination of provocative public messages. According to their findings, spammers follow authorized users and are followed by authorized users. Methods of categorization have been suggested, and they are now used for the identification of follow spammers. The focus of the social relation is cascaded and formulated into two mechanisms, namely, social status filtering and trade significance profile filtering; each of these mechanisms uses two-hop subnetworks that are centered at each other. These mechanisms are referred to as social status filtering and trade significance profile filtering, respectively. In order to combine the qualities of both the social status and the trade importance profile, assemble approaches and cascade filtering have been presented as potential solutions. A two-hop social network is focused on each user in order to obtain social information from social networks. This is done in order to determine whether or not a user is a phony. Positive findings were obtained from the experiment that used data taken from the actual world and were designed to test the credibility and reliability of the Twitter system. Using partial data allowed for real-time and lightweight spammer identification using TSP and SS filtering, respectively, which were both suggested. Both algorithms produce some false positives, but the real positives they produce are not superior to one another in terms of the collusion rank. A hybrid strategy that combines aspects of both filtering methods is proposed as an option. The experiment was carried out on one thousand legitimate users as well as one thousand spammer identities that had social status and TSP capabilities. The outcome of the strategy that was provided demonstrates that the schemes are scalable due to the fact that they examine user-centered twohops in the social network rather than investigating the whole network. This research considerably increases the performance of false and true positives than the prior approach. To detect spammer insiders, Meda et al. [21] suggested a method that makes use of a sample of non-uniform data inside a machine learning system via the application of random forest algorithm. The random forest and non-uniform feature sampling methods are the primary foci of the framework that has been developed. The random forest is a learning method that may be used for classification and regression. It accomplishes its tasks by first creating a number of decision trees during the preparation phase, and then allowing those trees to vote on which tree they believe should be the winner. The bootstrap aggregating approach as well as the unplanned selection of features are both included into the system. A strategy that uses non-uniform sampling of features is used in order to achieve an upper limit on the error generalization of the random forest. The authors prepared the dataset with the purpose of gathering users with undefined behaviors for the purpose of testing the performance of the random forest algorithm in the reference where the user categorization is undetermined. This was done in order to gather the data needed for the purpose of testing the algorithm. The selection of features may be broken down into two distinct sub-categories: the random selection of features and the domain expert selection of features. In order to demonstrate how effective the non-feature sampling approach is, we selected two different datasets. The first dataset consists of 1,065 people, of whom 355 are classified as spammers and 710 are classified as nonspammers according to 62 criteria. The dataset was produced. The author is responsible for the construction of the second dataset. The purpose of the experiments is to recreate two scenarios that are diametrically opposed to one another at the time of the feature selection phase. The selection of features for the first experimental group is made by domain experts, whereas the features used for the second experimental group are chosen at random. The effectiveness of the enhanced feature sampling approach was shown by the findings of the studies. David et al. [24] provided a method for determining the identity of a false user based on the information obtained from the Twitter site. A feature set consisting of 71 low cost variables was generated with the help of user profiles and timelines. These variables distinguish between aspects of a timeline that are based on its content and features that are based on its metadata. Features that are based on metadata relate to all of the information that supports or defines the primary content. The process of feature engineering entails a number of processes, all of which are designed to investigate whether or not the data provides a concise recollection of certain alterations that were identified when supporting decision trees with the aggregated features. The significance variable is used in order to locate the most effective and productive combination of features drawn from the available set of features. All of the feature sets were rated using the aforementioned four distinct metrics. In order to improve the accuracy of five supervised classifiers—decision trees, support vector machines, Naive Bayes, random forests, and single hidden layer feed forward artificial neutral networks—the results of validated classification are used to make selections. These selections are then used to validate the results of the classifiers. According to the findings of the methodology that was provided, the random forest algorithm, when applied to 19 different feature sets, achieved the greatest accuracy on average. Additionally, it demonstrates that the devices that are practicable and the detection methods that are generally successful may be devised to address the issue at hand. In addition, a research that was conducted by Keretna and his colleagues [26] focuses on authenticating real accounts as opposed to phony accounts with the use of Whiteprint, which is the biometric writing style. Text-mining strategies were used to segregate the feature sets, and supervised machine learning methods were used throughout the training process for the knowledge-based system. After initiating the recognition approach for segmenting the characteristics, the similarity of the characteristics vector was evaluated in relation to all of the characterized vectors already existing in the knowledge base. After that, the account with the vector that is the most similar to the confirmed one is identified. The sets of characteristics that are comparable to the issue at hand are chosen for examination. The Stanford POS is used in the process of feature extraction. Taking the messages on Twitter as a point of reference, the features have been segmented according to the characteristics of Twitter, and photographs and videos are only allowed when they are linked to from other sources. After that, the approach is used on a variety of different accounts so that the robustness and effectiveness of the suggested methodology may be evaluated. A method for locating spammers on Twitter was developed by Meda et al. [27]. [Citation needed] The training portion of the proposed framework is carried out offline, and its primary emphasis is on the construction of a random forest-based classifier that is derived from the collection of initial training sets. After performing feature extraction, the resulting data is then "parsed" to provide a user profile for a Twitter user. Because machine learning strategies need to be applied to numerical features, there is a need to convert profiles into vectors so that they may be used in conjunction with the ML-Module. The process of separating and converting certain traits into actual numbers is known as the feature extraction process, and it is used to differentiate spammers from other users. In this stage, the sample that was obtained in the previous step is used to teach the classifier. After the classifier has been trained, its parameters are locked down, and the system begins to participate in the categorization of Twitter posts during run time. The following fundamental stages are included in the run time phase: (a) The Twitter streaming API is used to collect Twitter traffic that reoccurs as Twitter reports in the format of JSON; (b) the profiles of Twitter users are constructed based on the features extracted from Twitter reports; and (c) the classifier assigns a category as spammer or non-spammer to the trial sample. The findings of the research demonstrate that the strategy that was presented is superior to a variety of other models in terms of its efficacy.

3. PROPOSED SYSTEM

The present method identifies bogus accounts via the use of an algorithm called random forest. When the appropriate inputs are provided, as well as when all of the inputs are provided, it functions at an optimal level. When part of the inputs are missing, it makes the algorithm's job of producing the output more difficult. A gradient boosting approach was used by our team in order to conquer such challenges that were presented by the offered systems. The gradient boosting technique is quite similar to the random forest algorithm, which relies heavily on decision trees as its primary building block. In addition to this, we altered the process by which we locate the phony accounts; more specifically, we developed new strategies for locating the accounts. The ways that are used include fake activity, engagement rate, and spam comments. These inputs are used in the construction of decision trees, which are then utilized in the gradient boosting process. This method provides us with an output even if part of the inputs are not there. This is the primary reason for selecting this algorithm. Because of the use of this approach, we were able to get findings that were quite precise.

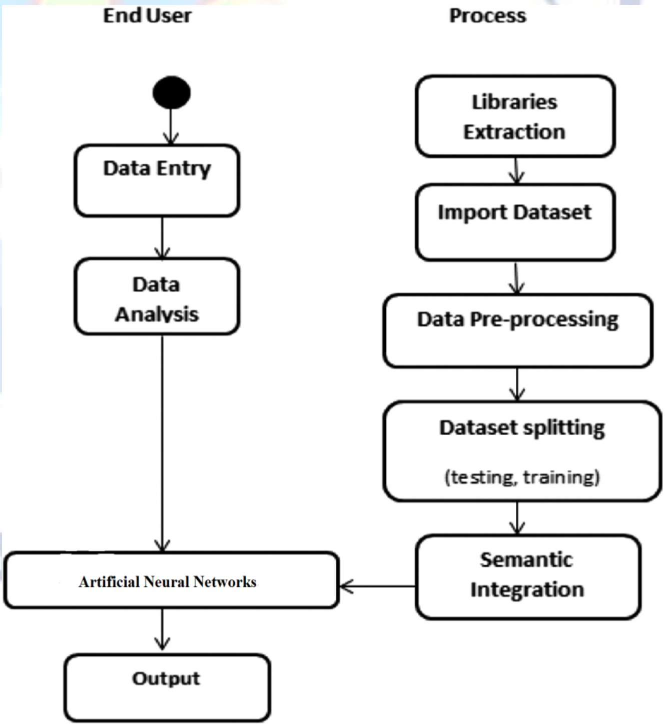
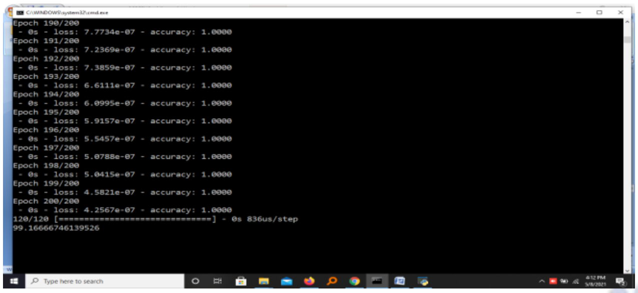


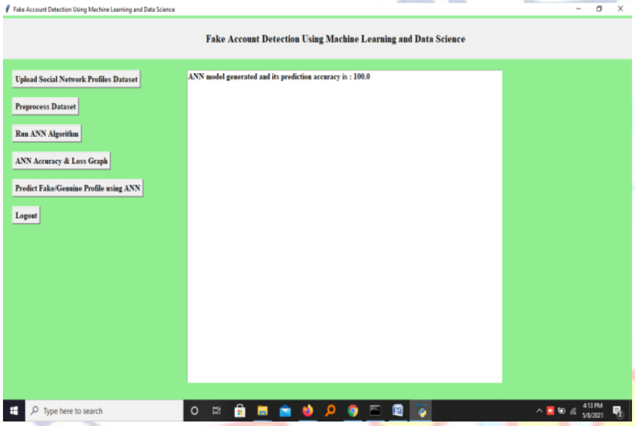
Figure 1: System Architecture for Proposed System

4. RESULTS



```
C:\WINDOWS\system32\cmd.exe
Epoch 190/200
 - 0s - loss: 7.7734e-07 - accuracy: 1.0000
Epoch 191/200
 - 0s - loss: 7.2369e-07 - accuracy: 1.0000
Epoch 192/200
 - 0s - loss: 7.3859e-07 - accuracy: 1.0000
Epoch 193/200
 - 0s - loss: 6.6111e-07 - accuracy: 1.0000
Epoch 194/200
 - 0s - loss: 6.0995e-07 - accuracy: 1.0000
Epoch 195/200
 - 0s - loss: 5.9157e-07 - accuracy: 1.0000
Epoch 196/200
 - 0s - loss: 5.5457e-07 - accuracy: 1.0000
Epoch 197/200
 - 0s - loss: 5.0788e-07 - accuracy: 1.0000
Epoch 198/200
 - 0s - loss: 5.0415e-07 - accuracy: 1.0000
Epoch 199/200
 - 0s - loss: 4.5821e-07 - accuracy: 1.0000
Epoch 200/200
 - 0s - loss: 4.2567e-07 - accuracy: 1.0000
120/120 [==============================] - 0s 836us/step
99.16666746139526
```

Figure2: Accuracy Results



Fake Account Detection Using Machine Learning and Data Science

ANN model generated and its prediction accuracy is : 100.0

Upload Social Network Profiles Dataset
Preprocess Dataset
Run ANN Algorithm
ANN Accuracy & Loss Graph
Predict Fake/Genuine Profile using ANN
Logout
Type here to search

Figure 3 : ANN Accuracy Results

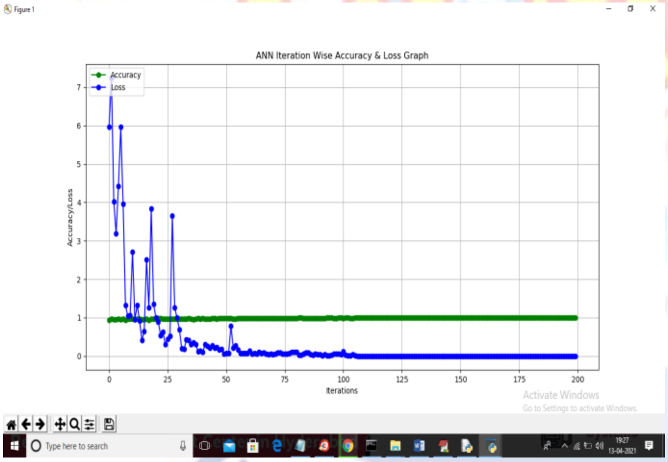


Figure 4 : ANN Accuracy & Loss Graph

5. CONCLUSION

As a result of this investigation, we have devised an inventive method for identifying phony accounts on OSNs. We were able to remove the requirement for manual prediction of a phony account by applying machine learning algorithms to their maximum potential. Manual prediction requires a significant amount of human resources and is also a procedure that takes a lot of time. The development of more sophisticated methods for creating false accounts has rendered the currently used techniques outdated. The current system relied on unreliable variables, which led to its instability. In the course of our investigation, we made use of constant parameters like engagement rate and fake activity in order to improve the precision of our predictions.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol, ''Twitter fake account detection,'' in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.

[2] 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.

[3] S. Gharge, and M. Chavan, ''An integrated approach for malicious tweets detection using NLP,'' in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 435–438.

[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ''Twitter spam detection: Survey of new approaches and comparative study,'' Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.

[5] 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.

[6] 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.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ''Twitter analysis for real-time malware discovery,'' in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.

[8] N. Eshraqi, M. Jalali, and M. H. Moattar, ''Detecting spam tweets in Twitter using a data stream clustering algorithm,'' in Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK), Nov. 2015, pp. 347–351. [9] 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.

[9] C. Buntain and J. Golbeck, ''Automatically identifying fake news in popular Twitter threads,'' in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.