International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

Identification of Spam Profile and Fake Users for Twitter

Gunjan Gupta[1] , Ranjit More [2], Sayali Sudhakar Tembe [3] Shantanu Durgadas Takalgavankar[4] , Jayanti Kamalasekaran [5]

[1,2,3,4] B.E. Student, Dept. of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune- 411041, Maharashtra, India

---

Abstract - Millions of people around the world invest and spends lot of time in social networking sites. The connections between users and such web networks, such as Twitter and Facebook, have an immense effect on everyday life and are sometimes undesirable. Spammers have become a priority forum for popular social networking platforms, dispersing a massive quantity of useless and malicious content. For starters, Twitter is one of the best-used websites of all time and therefore facilitates irrational spamming. Fake accounts are giving users undesirable tweets to support websites that not only impact legitimate users but even compete with the consumption of resources. Nowadays many rely on social media material, such as reviews and suggestions of a subject or product, in their decisions. It gives Spammers the ability to write spam reviews of goods and services for various reasons. The chance that anyone will leave a comment. The detection of such spammers and spam material is a highly research-intensive subject and while a large amount of studies has been performed recently, spam reviews still have little to no methodology and neither demonstrate the value of any form of feature extracted. Though this project we are implementing a methodology to detect Fake social network spammers and accounts. The methodologies detect users who interact spontaneously with others by recognizing Network topology abnormalities

Key Words: Random forest algorithm,K means algorithm,Minimum weight algorithm

1.INTRODUCTION

With the increasing use of Social media sites like Facebook and twitter, there is possibilities of fake account usage and ability of spammers to write fake reviews. This problem is increasing day by day in Online Social Media(OSN) users. Fake user's and spam detection has become critical problem in modern generation because of its highly increasing usage. Thus there is a need to obtain a solution to this problem. Therefore, Machine learning has been used to solve this problem. Our proposed model can identify the fake users and spam detection due to huge usage of social media. Through this project, we are applying methodologies through which we can identify the fake user identification and spam detection

2. Literature Survey

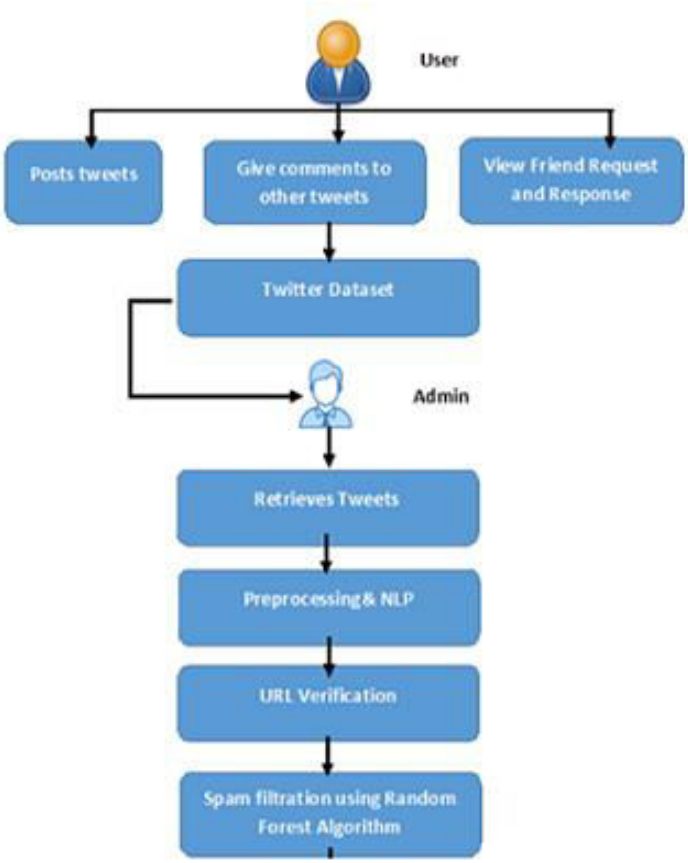
This section discusses previous studies related to the detection of spam profiles in social networks (OSN). The following are some of the most remarkable studies. Aswani, et.al, in 2018 suggested a novel hybrid methodology by incorporating analytics from social media and bio-inspired computation intended with the view of recognizing spam profiles in twitter. The K-Means algorithm was combined with Levy Flight Firefly Algorithm (LFA) along with chaotic maps. A total of 13 statistically relevant features of OSN research were used for this research, which included 18.44,701 tweets of 14,235 Twitter accounts. By analysing the tweets based on these 13 properties, the findings revealed accuracy of approximately 97.98%. The lack of consideration of substance and grammar was one of the great drawbacks of the proposed scheme. Alongside the use of non-English language terminology, the most recent tweets contain satire (Aswani, et.al, 2018). Dutta et.al has developed a technique of attribute collection to boost the specificity of classification for spam, producing improved classification outcomes by choosing a smaller sub set of attributes. This algorithm for the assortment of attributes was developed using rough concepts. Experiments were performed on five separate data sets for spam detection and validated with the proposed algorithm's efficiency. The technology suggested picked a much smaller subset of characteristics than traditional techniques with positive outcomes of the classification of literature than other techniques (Dutta, et.al, 2018). In 2015, Eshraqi et.al analysed multiple research on social media spam identification. The majority of checked web sites such as Twitter and Facebook were based in this report.

Table -1: Sample Table format

Preparation of Manuscript			
Margins : Top	0.5”	Bottom	0.5”
Left	0.5”	Right	0.5”
Margin : Narrow	Font	Cambria / 10 pt	
Title of paper : 16 Point	Heading	13 Point	
Sub Heading :12 Point	Spacing	Single line spacing	

3. Analysis and Desi

3.1 System Architect



3.2 Models and Methodologies

3.2.1User Module

User module In this module, there are n number of users present. Users should register before doing any operations. Once a user registers, their details will be stored to the database. After registration is successful, he has to login by using authorized user name and password. Once login is successful user will do some operations like viewing their profile account details like spam or normal, search users and send friend requests, viewing friend requests, searching posts and recommendations to friends and viewing review comments on the posts. User can post the tweet, comments on tweets, retweets etc. The user can view the friend requests which are sent by other users. Which includes sending user details with their tags such as user name, user image, date of birth, Email ID, phone number and Address and user can accept the request by clicking on the "waiting" link. If use wish to send a friend request to a particular user then click on the "request" button, then request will be sent to that particular user

3.2.2 Admin Module

In this module, the admin has to login by using valid user name and password. After login successful he can do some operations Viewing and authorizing users, View Spam accounts details, viewing friend request response, all recommended posts, all posts with all reviews, all positive and negative reviews, removing users etc. The admin can then perform the operations on identifying the fake users and spammer detection. For that first he will extract all the tweets of the users, will do the pre-processinglike removing missing values, null values and extra spaces etc. After pre-processing he will use Natural Language processing (NLP) tool to remove the @, # from the tweets, removes the stop words, remove urls, etc. URL Verification will be done after that. By applying Random Forest algorithm of Machine Learning we are going to identify the spammer through the classification process of categorization of the spammer. Decision Tree algorithm can be used for solving regression and classification problem. The goal of using a Decision Tree is to create a training model. Naïve Bayes is a supervised learning algorithm used for solving classification problem. This algorithm is used to identify the fake users in twitter

3.3 Algorithms Used

3.3.1 Random Forest Algorithm

In this research work we are using random forest which is comes under supervised learning in machine learning. Random forest algorithm which is used to classification, through which we are going to identify the spammer. We had to categorize the spammer after that we are going to identify the spammer. Steps for Random Forest algorithm Step 1: Gather the different training data from the training dataset. Step 2: In each data which we are gathered we have to take the particular information. Step 3: Finally, we have to predict the data

3.3.2 Naïve Bayes Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text Classification that includes a high-dimensional training dataset. Naïve Bayes classifier is one of the simple and most effective classification algorithms which help in building the fast machine learning models that can make quick predictions. Steps of Naïve Bayes Algorithm 1: Calculate the prior probability for given class labels 2: Find Likelihood probability with each attribute for each class 3: Put these value in Bayes Formula and calculate posterior probability. 4: See which class has a higher probability, given the input belongs to the higher probability class.

3.3.3 Decision Tree Algorithm

Decision Tree algorithm belongs to the family of supervised learning algorithms. It can be used for solving regression and classification problem. The goal of using a decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules referred from training data. In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next Node.

4. Working Module

Categories. For classification, we have identified four means of reporting spammers that can be helpful in identifying fake identities of users. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification.
Moreover, the analysis also shows that several machine learning-based techniques can be effective for identifying spams on Twitter. However, the selection of the most feasible techniques and methods is highly dependent on the available data.

5. CONCLUSIONS

In this research, we are working on a solution for detecting spammers and fake reviews on Online Social Networks (OSNs). This study introduces machine learning approach Addressing the problem of spam detection and fake users. We will be using NLP process to extract the tweets with meaningful information and based on that unauthorized users and fake review comments, post can be identified using Random Forest and KMeans algorithm

REFERENCES

[1] FAIZA MASOOD1, GHANA AMMAD1, AHMAD ALMOGREN 2, (Senior Member, IEEE), ASSAD ABBAS 1, HASAN ALI KHATTAK 1, (Senior Member, IEEE), IKRAM UD DIN 3, (Senior Member, IEEE), MOHSEN GUIZANI 4, (Fellow, IEEE), AND MANSOUR ZUAIR5, "Spammer Detection and Fake User Identification on Social Networks", IEEE Access, 2019.