ANNEXURE B

Detection of Bot Accounts on Social Networks Using Big Data Mining Tools
(Final Year Project - Ongoing)

Abstract:

The revolution in the history of communication came with the birth of online social media which completely transformed the way people communicate. However, every revolution brings with it some negative impacts. The ease of access of the social media platforms has made it popular amongst millions of users all around the world. As a result, these platforms have a huge volume of data including personal information about the users. Moreover, the popularity of online social media has attracted many malicious activities. The ease of access with minimal verification of new users has led to creation of several fake accounts that are used to collect private data and spread false harmful/fake content. Recently, the detection of fake accounts has become an area of research. In this project, we aim to identify the fake accounts from a leading online social media platform Twitter. Machine learning algorithms will be used for identifying the fake accounts and further verification of the authenticity of dataset will be done using Benford's Law.

Keywords: Benford's Law, Online Social Networks (OSN).

Introduction:

In this modern era of data dominance, the authenticity of data is a major issue for everyone. With the growing presence of Online Social Networks (OSNs), people have started using it as their preferred medium of communication. Everyone and anyone can use these platforms for sharing their personal information, news, opinions and even their current mood. Some of the most popular OSNs are Facebook, Twitter, Instagram, Google Plus, Reddit and LinkedIn. It is not only limited to individuals using it for personal purposes, but governments, organizations, commercial enterprises and even politicians are using these platforms to increase their reach to the masses. It also makes getting response from the audience much easier while making it convenient to convey their messages directly to a wider population. But due to popularity of OSNs, they are attractive targets for malicious entities that are trying to exploit the vulnerabilities of these platforms. OSNs have a bulk of fake accounts that are either operated by other humans or by artificial social bots. These fake accounts are generally made to take advantage of the weaknesses of the network and thus the genuine users become victims of these malicious activities. A social bot is a software to automate user activities. These activities can be generating pseudo posts which look like human generated, re-posting photos, articles etc, adding likes and comments on other posts and increasing their social network by connecting with other accounts. The level of sophistication of these bots ranges from dummy like bots that aggregate information from posts and re-post them to bots that are capable of infiltrating human conversations. Social bots have pros and cons for the users of the OSNs. Thus, it has become necessary to identify these fake accounts on the OSN platforms to preserve the security and privacy of the users. Benford's law [Gol15] represents the pattern of behavior in normal systems. It states that, in naturally occurring systems, the frequency of the number's first digits is not evenly distributed, rather they follow a pattern. Numbers beginning with '1' are more common than the numbers beginning with '9'. The equation followed by the Benford's law is given below, where P is the frequency of the digit d:

$$P(d) = \log_{10}(1 + 1/d)$$

The randomness of the OSNs makes it a normal system, hence the data generated from these networks should follow the Benford's law. This law can be used for detecting anomalies in the data and checking its authenticity.

Objectives:

The primary goal is to detect fake user accounts on social media platforms using Benford's Law and a variety of big data mining tools. The aim is to take a random set of user accounts from a social media platform and check if the set of users follow Benford's Law, if not then a variety of machine learning algorithms are used to filter out the fake users. The implementation will be done on a data set collected from twitter to check the accuracy of the proposed model in the real world.

Proposed Framework:

We are currently at a stage of utilizing the existing machine learning algorithms such as Naive Bayes Classifying algorithm, Clustering Classifier, Decision Tree, Hybrid Classifier, and various other custom algorithms based on what is required.

Once our data has been collected, we will begin to test different classifiers for the required task. The first goal is binary classification, hence deciding whether a user account is spam or non-spam. Based on the results obtained from the training data using various classifiers, we will be extracting random user accounts and testing the validity of our classification using Legitimate Index value. If it is greater than 50%, the account is legitimate, else fake.

After this process is over, we will be applying legitimate user dataset and test user's dataset to generate the respective Benford's Law Distribution graphs. If the test user's dataset follows the graph, its accuracy is verified by Benford's Law, else not.

Block Diagrams:

The two flowcharts below depict the initial and verification stages of the process. In the initial stage, we employ different predefined machine learning algorithms for training the data. We then perform the legitimacy test for checking the authenticity of user accounts. In the verification stage, we compare the distribution graphs of user data with Benford's Law Distribution graph.

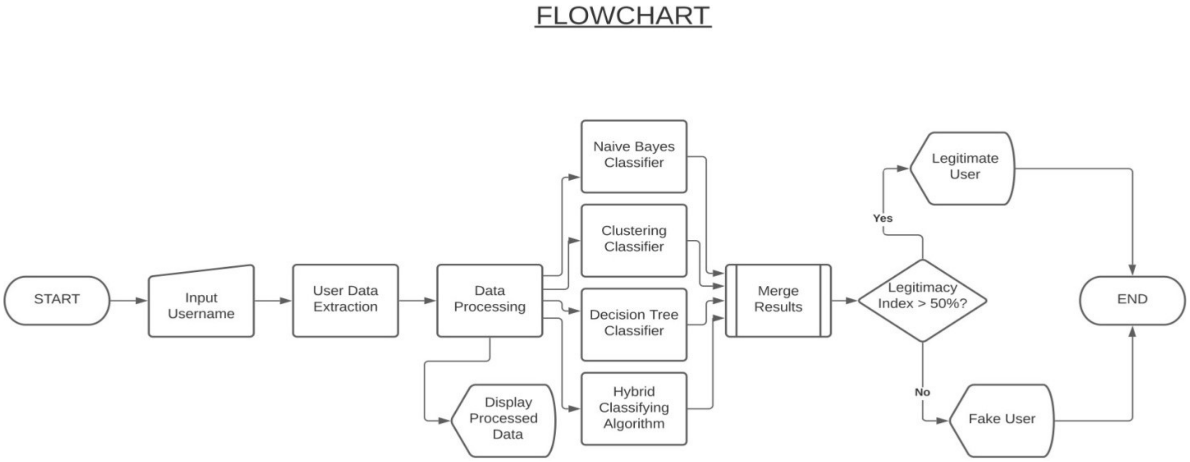


Figure 1: Classifying user accounts as fake or legitimate

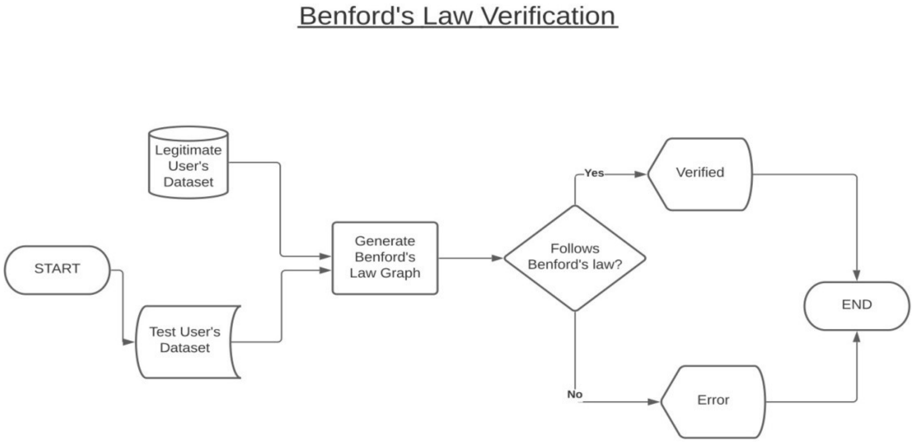
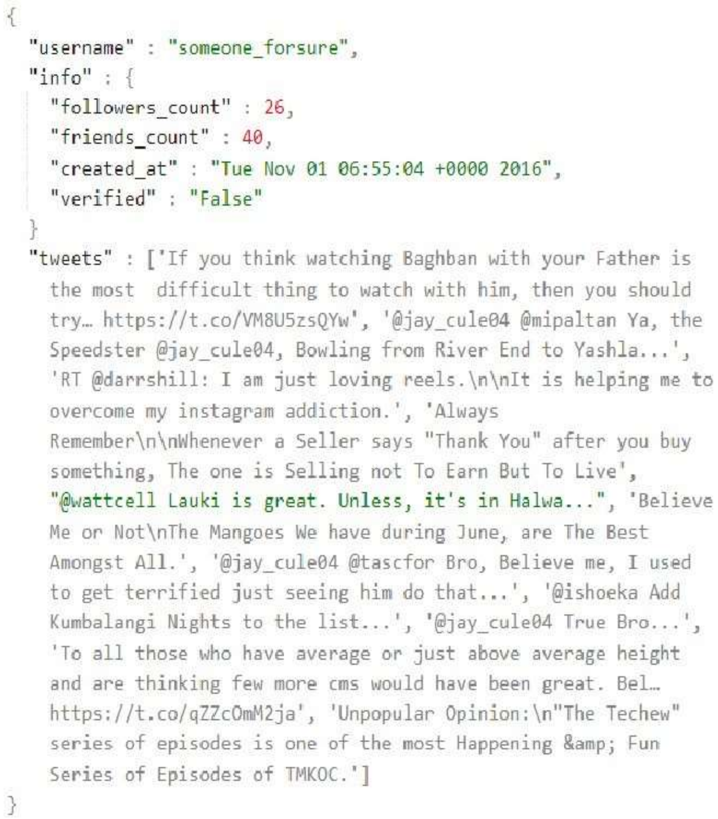

Benford's Law Verification
Legitimate User's Dataset
START
Test User's Dataset
Generate Benford's Law Graph
Follows Benford's law?
Yes
No
Verified
Error
END

Figure 2: Verification using Benford's Law

Data Preprocessing:

We collected user data from Twitter API for training our model. The get user() method of the API class in Tweepy module is used to get information of the specified user. After analyzing and removing the unnecessary fields of user data, we prepared our sample dataset. This was done using Tweepy and CSV libraries.

The following image shows one of sample user data obtained.



```
{
  "username" : "someone_forsure",
  "info" : {
    "followers_count" : 26,
    "friends_count" : 40,
    "created_at" : "Tue Nov 01 06:55:04 +0000 2016",
    "verified" : "False"
  }
  "tweets" : ['If you think watching Baghban with your Father is
    the most  difficult thing to watch with him, then you should
    try… https://t.co/VM8U5zsQYw', '@jay_cule04 @mipaltan Ya, the
    Speedster @jay_cule04, Bowling from River End to Yashla...',
    'RT @darrshill: I am just loving reels.\n\nIt is helping me to
    overcome my instagram addiction.', 'Always
    Remember\n\nWhenever a Seller says "Thank You" after you buy
    something, The one is Selling not To Earn But To Live',
    "@wattcell Lauki is great. Unless, it's in Halwa...", 'Believe
    Me or Not\nThe Mangoes We have during June, are The Best
    Amongst All.', '@jay_cule04 @tascfor Bro, Believe me, I used
    to get terrified just seeing him do that...', '@ishoeka Add
    Kumbalangi Nights to the list...', '@jay_cule04 True Bro...',
    'To all those who have average or just above average height
    and are thinking few more cms would have been great. Bel…
    https://t.co/qZZcOmM2ja', 'Unpopular Opinion:\n"The Techew"
    series of episodes is one of the most Happening &amp; Fun
    Series of Episodes of TMKOC.']
}
```

Figure 3: A sample user data from Twitter

Conclusion:

Nowadays, people on a massive scale use some or the other social media platform, but not many are aware that their privacy and security is at risk due to some fake accounts present on the very same OSNs. These fake users can interact with other genuine users, have the ability to share false information, participate in identity thefts, etc. Fake users can be operated by a bot or an individual. A lot of people are aware of dummy chat bots that aggregate information but there is not enough awareness about bots that can infiltrate human conversations. With the help of a few known machine learning algorithms these users can be identified and removed from the platform.