## ABSTRACT

Twitter popularity has fostered the emergence of a new spam marketplace. The services thatthis market provides include: the sale of fraudulent accounts, affiliate programs that facilitatedistributing Twitter spam, as well as a cadre of spammers who execute large scale spamcampaigns. In addition, twitter users have started to buy fake followers of their accounts. Inthis project we present machine learning algorithms we have used to detect fake followers inTwitter.We identified a number of characteristics that distinguish fake and genuine followers.We used these characteristics as attributes to machine learning algorithms to classify users as fake or genuine.

**CHAPTER 1**

**INTRODUCTION**

Twitter has become a popular media hub where people can share news, jokes and talk about their moods and discuss news events. In Twitter users can send Tweets instantly to his/her followers. Also, Tweets can be retrieved using Twitter’s real time search engine. The ranking of tweets in this search engine depends on many factors, one of which is the user’s number of followers. Twitter’s popularity has made it an attractive place for spam and spammers of all types. Spammers have various goals: spreading advertising to generate sales, phishing or simply just compromising the system’s reputation. Given that spammers are increasingly arriving on twitter, the success of real time search services and mining tools lies in the ability to distinguish valuable tweets from the spam storm .There are various ways to fight spam and spammers such as URL blacklists, passive social networking spam traps, manual classification to generate datasets used to train a classifier that later will be used to detect spam and spammers. So what is Twitter spam? As Twitter describes it in their website , Twitter spam is “a variety of prohibited behaviors that violate the Twitter Rules.” Those rules include among other things the type of behavior Twitter considers as spamming, such as:

● “Posting harmful links (including links to phishing or malware)

● Aggressive following behavior (mass following and mass unfollowing for attention), particularly by automated Means.

● Abusing the @reply or @mention function to post unwanted messages to users.

● Having a small number of followers compared to the number of people one is following

● Posting repeatedly to trending topics to try to grab attention.

Twitter actually fights spammers by suspending their accounts. But in general OSN (Online Social Networking) sites do not detect and suspend suspicious user accounts quickly. They are not willing to deploy automated methods to detect and remove spam accounts fearing that this will lead to a serious discontentment among users. Thus, they wait until a sufficient number of users report a specific account as a spam account to suspend it. However, legitimate users are unwilling to invest time to report spammers. Hence spammers are allowed more time to spread spam. Automated accounts, called bots, are common in social media. Although all bots are not bad, bots are easy means to engage in unethical and illegal activities in social media. Examples of such activities include selling accounts, spamming inappropriate content, and participating in sponsored activities. Many social metrics are calculated based on social media data.The significant presence of bots in social media will make many of these metrics useless. The exact number of bots is dynamic and unknown. The range of the estimates is between 3% to 7%. Social media sites, such as Twitter, regularly suspend abusive bots. Yet, the number of bots is growing because of almost zero-cost in creating new bots. Existing bot detection methods are not capable of fighting such evolving set of bots. There are several reasons. Current methods are mostly non-adaptive, require supervised training, and consider accounts independently. Typical features used in some of the methods need a long duration of activities (e.g. weeks) which makes the detection process useless, as the bots can initiate a fair amount of harm before being detected. Moreover, bots are becoming smarter. They mimic humans to avoid being detected and suspended and increase throughput by creating many accounts. We take a novel unsupervised approach of cross-correlating account activities, that can detect such dynamic bots as soon as two hours after starting their activities.

**CHAPTER 2**

**LITERATURE SURVEY**

There has been recent interest in the detection of malicious and/or fake users from both the online social networks and computer networking communities. For instance, Wang[4]looks at graph-based features to identify bots on Twitter, while Yang, Harkreader, and Gu [5] combine similar graph- based features with syntactic metrics to build their classifiers.

Thomas et al. [6] use a similar set of features to provide a retrospective analysis of a large set of recently suspended Twitter accounts.

Boshmaf et al. [7] instead create bots (rather than detecting them), claiming that 80% of bots are undetectable and that Facebook’s Immune system [8] was unable to detect their bots.

Lee, Caverlee, and Webb [9] create “honeypot” accounts to lure both humans and spammers into the open, then provide a statistical analysis of the malicious accounts they identified.

Java et al.[12] studied over 70,000 Twitter users and categorized their posts into four main groups: daily chatter (e.g., “going out for dinner”), conversations, sharing information or URLs, and reporting news. Their work also studied :

1) the growth of Twitter, showing a linear growth rate;

2) its network properties, showing the evidence that the network is scale-free like other social networks.

3) the geographical distribution of its users, showing that most Twitter users are from the US, Europe, and Japan.

Krishnamurthy et al. [13] studied a group of over 100,000 Twitter users and classified their roles by follower-to- following ratios into three groups:

1) broadcasters, which have a large number of followers;

2) acquaintances, which have about the same number on either followers or following; and

3) miscreants and evangelists (e.g., spammers), which follow a large number of other users but have few followers.

Wu et al. [16] studied the information diffusion on Twitter, regarding the production, flow, and consumption of information.

Kwak et al. [17] conducted a thorough quantitative study on Twitter by crawling the entire Twittersphere. Their work analyzed the follower-following topology, and found nonpowerlaw follower distribution and low reciprocity, which all mark a deviation from known characteristics of human social networks.

Kim et al. [18] analyzed Twitter lists as a potential source for discovering latent characters and interests of users. A Twitter list consists of multiple users and their tweets. Their research indicated that words extracted from each list are representative of all the members in the list even if the words are not used by the members. It is useful for targeting users with specific interests. In addition to network-related studies, several previous works focus on sociotechnological aspects of Twitter [7], [8], [19], [20], [21], such as its use in the workplace or during major disaster events. Twitter has attracted spammers to post spam content, due to its popularity and openness.

Yardi et al. [14] detected spam on Twitter. According to their observations, spammers send more messages than legitimate users, and are more likely to follow other spammers than legitimate users. Thus, a high follower-to- following ratio is a sign of spamming behavior.

Grier et al. [22] investigated spam on Twitter from the perspective of spam and click-through behaviors, and evaluated the effectiveness of using blacklists to prevent spam propagation. Their work found out that around 0.13 percent of spam tweets generate a visit, orders of magnitude higher than click-through rate of 0.003-0.006 percent reported for spam e-mail. Exploiting the social trust among users, social spammers achieve a much higher success rate than traditional spam methods.

Thomas et al. [23] studied the behaviors of spammers on Twitter by analyzing the tweets originated from suspended users in retrospect. They found that the current marketplace for.

**CHAPTER 3**

**DETAILED DESIGN**

**3.1 System Diagram**

The data set was taken from Tweepy and divided it into training set and test set ,Using machine learning algorithm and training data we’ve trained our classifier model. Test set is used on classifier model for giving prediction according to the given scenario.

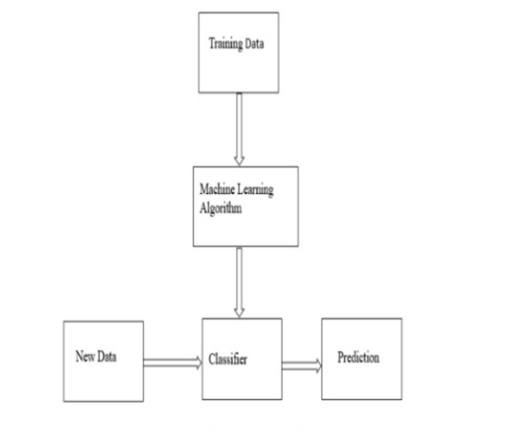


Fig 3.1 System Diagram

**3.2 Data Flow Diagram**

**3.2.1 Level-0**:

Our Dataset contains 20 attributes out of which we have selected 8 attributes based on the spearman correlation. Some of the important attributes are Follower-Friend ratio, Username, URL ratio, Number of tweets, etc. which provide information about the users. These attributes are then used by our twitter bot detection system or Machine Learning Algorithms for predicting that whether a user is a human or a bot.

**3.2.2 Level-1:**

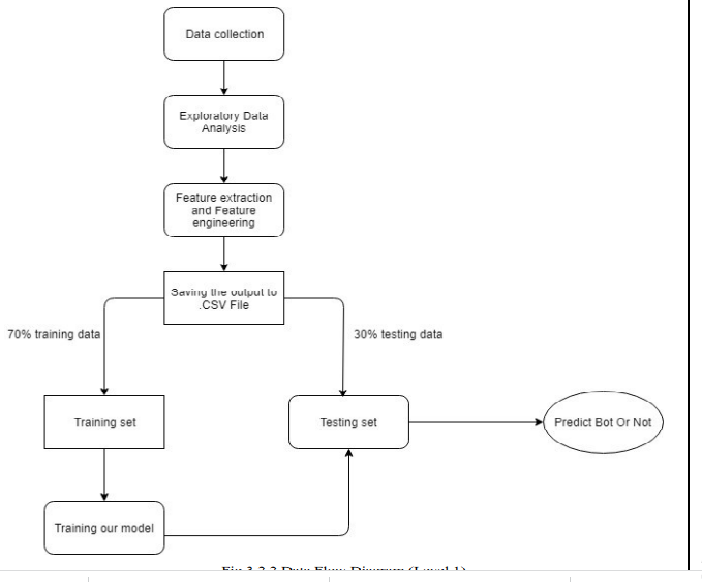


Fig 3.2 Data Flow Diagram (Level 1)

In order to use machine learning to identify fake twitter accounts, we needed a labelled collection of users, pre classified as fake or genuine. We get the real time data (dynamic data) from tweepy API which consist of 2798 training set and 578 test set. The dataset is divided into 70%(training set) and 30%(test set) on which data exploratory analysis has been done as well as it is also explored to feature extraction and feature engineering, Both training and test set are saved in .CSV format. Now next step is to train the model using training set, Attribute considered here are important which we’ve considered using spearman correlation,

Some of the important attributes are

● Listed\_count

● verified account

● Friends to follower ratio

Now test set is used on our already trained model for prediction whether the user is bot or not.

**3.3 Activity Flow Diagram**

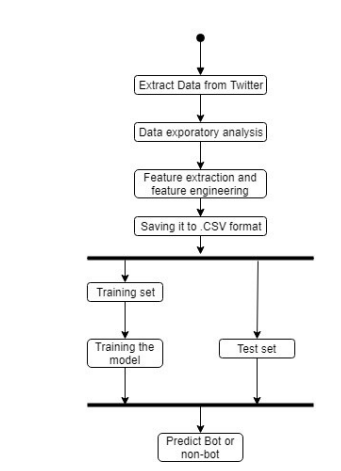


Fig 3.3 Activity Flow Diagram

**CHAPTER 4**

**PROJECT SPECIFIC REQUIREMENTS**

This project is developed using Python as development tool. Python version 3.6 is used for this purpose. The project dataset is downloaded from Kaggle and kept in the working system. For development environment , ‘Spyder’ is installed. Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package. All the required python libraries such as tweepy, pandas, numpy, matplotlib, seaborn are installed so as to provide the functionalities of python in-built functions.

**4.1 System Configuration**

**Hardware Requirements:**

System: Intel i5 2.4 GHz

RAM: 4 GB

Hard Disk: 1 TB

**Software Requirements**:

Operating System: Windows 7 or above

Coding Language: Python

Software: Anaconda

**4.2 Python-3** :

Python is one of the most common language used for machine learning because it is easy to learn, enormous packages it supports like scikit learn that contains the implementation of common machine learning algorithms, built-in library supports, pre-processed models, support for larger networks and huge toolset.

**4.2.1 Benefits of Python**

● Scikit learn- It features various classification, regression and clustering algorithms including support vector machines, random forest, gradient boosting, K-means and DBSCAN, and is designed to interoperate with the python numerical and scientific libraries NumPy and SciPy.

● Data frame object for data manipulation with integrated indexing

● Tools for reading and writing data between in memory data structures and different file formats

. ● Data alignment and integrated handling of missing data.

● Reshaping and pivoting of datasets.

● Label based slicing, fancy indexing and sub setting of large data set

. ● Data structure column insertion and deletion

● Group by engine allowing split-apply-combine operations on dataset.

● Data set merging and joining

● Hierarchical axis indexing to work with high-dimensional data in a lower-dimensional data structure.

**4.3 Anaconda Software**:

Anaconda is a python distribution, with installation and package management tools. It provides large selection of packages and commercial support. It is an environment manager, which provides the facility to create different python environments, each with their own settings. Anaconda can help with:

• Python can be installed over the multiple platforms

• Different environments can be supported separately

• Distributing with not having correct privileges and • Support and running with specific packages and libraries

**REFERENCES**

[1] “Top Trending Twitter Topics for 2011 from What the Trend,” http://blog.hootsuite.com/top-twittertrends-2011/, Dec. 2011.

[2] “Twitter Blog: Your World, More Connected,” http://blog. twitter.com/2011/08/your-world-moreconnected.html, Aug. 2011.

[3] Alexa, “The Top 500 Sites on the Web by Alexa,” http:// www.alexa.com/topsites, Dec. 2011.

[4] “Amazon Comes to Twitter ,”http://www.readwriteweb.com/amazon\_comes\_to\_twitter.

[5] “Best Buy Goes All Twitter Crazy with @Twelpforce,”http://twitter.com/in\_social\_media/ [6] “Barack Obama Uses Twitter in 2008 Presidential Campaign,”http://twitter.com/Obama/, Dec. 2009.

[7] J. Sutton, L. Palen, and I. Shlovski, “Back-Channels on the Front Lines: Emerging Use of Social Media in the 2007 Southern California Wildfires,” Proc. Int’l ISCRAM Conf., May 2008. [8] A.L. Hughes and L. Palen, “Twitter Adoption and Use in Mass Convergence and Emergency Events,” Proc. Sixth Int’l ISCRAM Conf., May 2009.

[9] S. Gianvecchio, M. Xie, Z. Wu, and H. Wang, “Measurement and Classification of Humans and Bots in Internet Chat,” Proc. 17th USENIX Security Symp., 2008.

[10] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, “Your Botnet Is My Botnet: Analysis of a Botnet Takeover,” Proc. 16th ACM Conf. Computer and Comm. Security, 2009.

[11] S. Gianvecchio, Z. Wu, M. Xie, and H. Wang, “Battle of Botcraft: Fighting Bots in Online Games with Human Observational Proofs,” Proc. 16th ACM Conf. Computer and Comm. Security, 2009.

[12] A. Java, X. Song, T. Finin, and B. Tseng, “Why We Twitter: Understanding Microblogging Usage and Communities,” Proc. Ninth WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis, 2007.

[13] B. Krishnamurthy, P. Gill, and M. Arlitt, “A Few Chirps about Twitter,” Proc. First Workshop Online Social Networks, 2008.

[14] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, “Detecting Spam in a Twitter Network,” First Monday, vol. 15, no. 1, Jan. 2010.

[15] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and Analysis of Online Social Networks,” Proc. Seventh ACM SIGCOMM Conf. Internet Measurement, 2007

[16] S. Wu, J.M. Hofman, W.A. Mason, and D.J. Watts, “Who Says What to Whom on Twitter,” Proc. 20th Int’l Conf. World Wide Web, pp. 705-714, 2011.

[17] H. Kwak, C. Lee, H. Park, and S. Moon, “What Is Twitter, a Social Network or a News Media?” Proc. 19th Int’l Conf. World Wide Web, pp. 591-600, 2010.

[18] I.-C.M. Dongwoo Kim, Y. Jo, and A. Oh, “Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users,” Proc. CHI Workshop Microblogging: What and How Can We Learn From It?, 2010.

[19] D. Zhao and M.B. Rosson, “How and Why People Twitter: The Role Micro-Blogging Plays in Informal Communication at Work,” Proc. ACM Int’l Conf. Supporting Group Work, 2009.