Twitter Bot or Not?

Revanth Mattapalli
Department of Computer Science
Tandon School of Engineering
rm4271@nyu.edu

Varun Elango
Department of Computer Science
Tandon School of Engineering
varunelango@nyu.edu

Vignesh Ramesh
Department of Computer Science
Tandon School of Engineering
vignesh.r@nyu.edu

***Abstract*—This paper is aimed at analyzing the effectiveness and performance of different machine learning algorithms in its ability to classify twitter accounts as bots or humans. For obvious reasons, this is treated as classification problem. The paper evaluates the generalization performance and the predictive performance of all the models on future data. Based on the results we find the best suited machine learning algorithm for the given hypothesis space**

I. Introduction

Twitter is a social networking and micro blogging service, enabling registered users to read and post short messages called tweets. There are around 300 million monthly active users in posting 500 million tweets per day. Out of all the users around 23 million is estimated to be bots(automated programs which tweets, pulls data, etc without human intervention). Our aim is to build an effective prediction model that correctly classifies bots from humans.

II. Motivation

With the rising number of bots in the twitterverse, the credibility of popularity of a user in the twitterverse seems to be diminishing and constant spam tweets have infiltrated our feeds. Moreover, with the advent of fake news spreading and influencing twitter users it is imperative that these bots are weeded out before they cause further harm. There are of course incredibly well written, good quality bots that posts useful information. Hence, It's imperative to find out if they are bots firsts, and then determine if they are useful or not. We, in this paper use machine learning techniques to classify if a twitter user account is a bot or not.

III. Related Work

Twitter has much popular recently and it has attracted spammers to post spam content, due to its popularity and openness. Fighting against spambot on Twitter has been investigated in recent works. Twitter users are categorized based on their tweets, entities and place. H. Kwak et al.[4] work also examined a quantitative study on Twitter by crawling the entire Twittersphere. Their work analyzed the follower-following ratio and found non-power law follower distribution and low reciprocity of tweets, which all mark a deviation from known characteristics of a human social network. Chaiji et al.[6] have shown how to maximize content propagation in ones own social network. In contrast, our approach aims at selecting a right set of bots on twitter to prevent information propagation.

Our goal is to support effective classification of the twitter users based on several features of a user. Sprout social[7] tool provides meaningful data about a twitter account such as tweet impressions,tweet activity using this data we can differentiate between bot and human. They [5] analyzed Twitter lists as a potential source for discovering latent characteristics and interests of the users. A User's feed in the Twitter consists of multiple followers and their following users' 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. According to their observations, spammers send more messages than legitimate users, and are more likely to follow other spammers than legitimate users.

IV. Data

Our dataset includes, equal proportion of bot users and human users. We set up five fake twitter accounts as seeds and stated more than 90 interests for each account. This resulted in each of our accounts instantly following 300+ twitter users. Our accounts were now visible to bots that follow the same users as us. These bots immediately started following our accounts and sent us messages. Moreover, there were popular hash tags such as #fiftyshadesdarker #contentmarketing #SMM #marketing #blogging #socialmedia #growthhacking which when used also attracted few bots to start following our accounts. We manually inspected these accounts as an additional check to ensure the quality of data sample.

We also searched for famous bots mentioned in blogs and websites. Most of the bots obtained using this method were usually good, non-spam bots and were relatively easier to find. Also spam bots usually have very less followers and habitually these followers turn out to be bots. We were able to unearth a few bots using this method. Of course there were exceptions to such spam bots too, which may have a lot of followers.

A. Feature Extraction

Data collected from Twitter API are distilled in 24 features which are classified in two classes. The classes and features obtained from the twitter API is discussed next.

1) User Based Features: Features extracted from user meta data have been used to classify the users(Bot Or Human). Twitter API allows us to get the meta data of the user. Features such as followers count, friendsCount,No of tweets produced by the user, screen name,description,location,languages know to him and settings.Explained in detail below.

- Screen Name: Provides us the twitter handle of user
- Location : Provides the location of user
- Description : A short information about the user. Some bots provide information in the description about there behavior.
- Followers Count : The number of followers this account currently has.
- Friends Count : The number of users this account is following.
- Created at : Date at which account is created by the user
- Verified: Tells us if the user is verified or not(Blue tick beside the name)
- Status count : No of tweets tweeted by user.
- Languages : The code for the users self-declared user interface language.
- Default Profile background : It is boolean value which provides the information if the user still has default profile background.
- Default Profile Picture: Also a boolean value which provides if the user has default profile picture.
- Name : The name of the user, as theyve defined it.Not necessary name of the user.
- Favorite Count:Number of tweets loved by user.
- Diversity : Length of name feature.
- Created Hour : Time at which account is created irrespective of date.
- URL : A URL provided by the user in association with their profile. Can be null.

2) Tweet Based Features: Features Extracted from last Tweet have been used in this class. We can get the last tweet for a particular user from Twitter API.From last tweet(Status) we can get features such as text of tweet, time of tweet tweeted, re-tweet or not, if tweet is in reply to some other user,Favorite count,Hash tags included in text,links included in text.More features are explained below.

- Truncated : If the tweet text is cut short to accommodate only 140 characters(Tweet length)
- Text : Provides tweet text tweeted by user.
- In reply to tweet id : Tweet id of main tweet
- Id : A unique id is given to each tweet.
- Favorite Count:No of users have loved this particular tweet.
- Coordinates : Provides the location from where the tweet is tweeted.
- User Mentions : Twitter handle of users mentioned in tweet.
- Hashtags : Hashtags present in the tweet.
- Retweet count : Number of times tweet is retweeted.
- Created at : Time of tweet tweeted
- Retweeted :Provides the information if the tweet is retweeted by some other user.

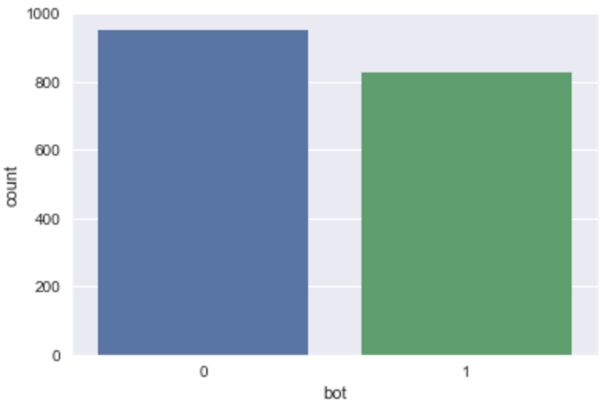

Fig. 1. No of Humans vs Bots

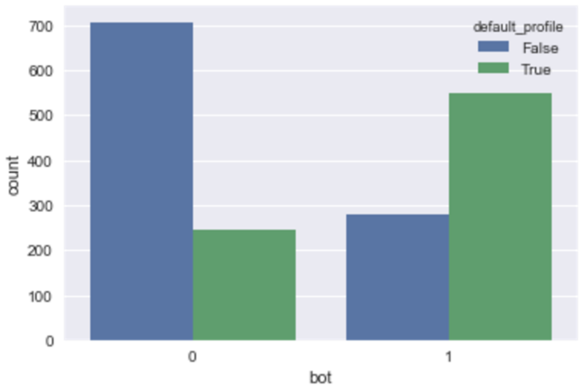

Fig. 2. No.of Default Profile on Bot and Humans

Figure 1 describes entropy of training and test data. No of humans present in the data is almost equal to that of bots. So we have high entropy.

Figure 3&4 are frequency graphs for number of followers for bot and human respectively.It can be seen that followers count for human is generally high. Followers count and probability of account being human is positively correlated. Where as for bot followers count is generally low. It can be noted that probability of account being bot and followers count is negatively correlated.

Figure 5 depicts relationship between number of users this account is following and account being verified. It can be seen that even though account has high friends count,account is human if it is verified.Which in general is behavior of bots

Figure 6 depicts the relationship between no of tweets tweeted by the user and followers count. It shows that if No of tweets for particular user is high and follower count is low then there are high chances that particular account is bot.

Figure 7 describes relationship between default picture and default background. It can be seen that probability of the user been bot is high if the user has default profile picture and default profile background. In general that probability of the profile being bot is high if the account has default profile picture can be verified from graph.

Figure 8 describes influence of account containing extended

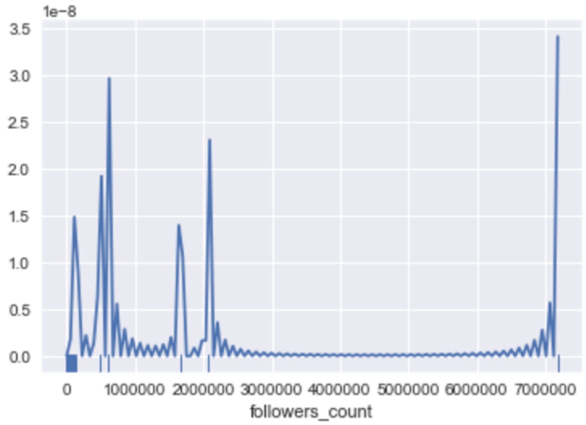

Fig. 3. No.of Followers Count on Bot

Figure 2 describes relation between user profile and default profile picture. It can be noted from the graph that most bots have not changed default profile picture. It can help in determining profile(human or bot) of the user.

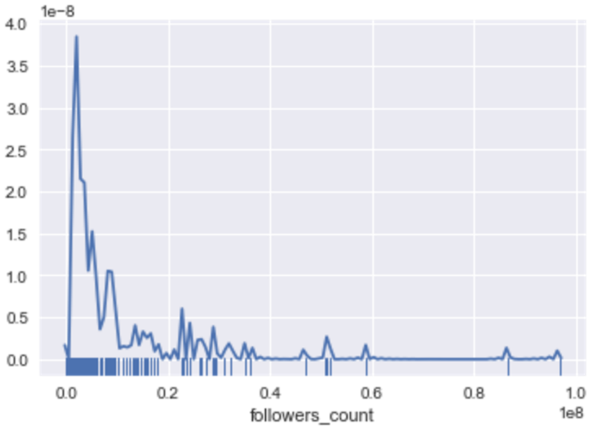

Fig. 4. No.of Followers Count on Human

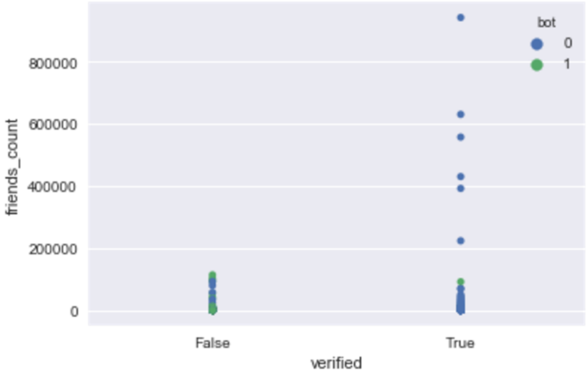

Fig. 5. Verified vs friends Count

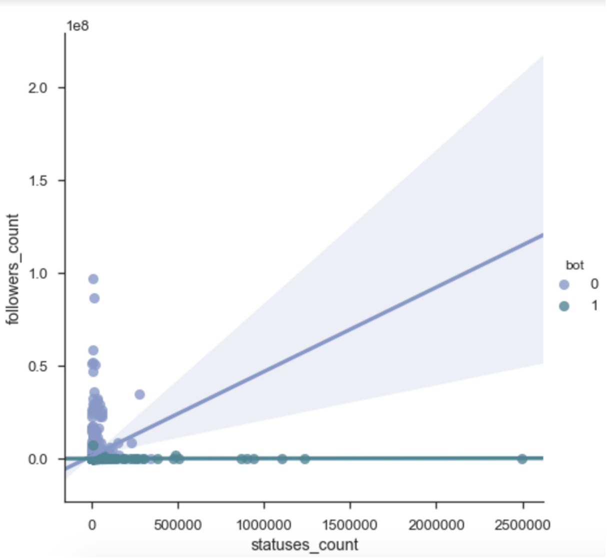

Fig. 6. Trends of bot on status Count and followers Count

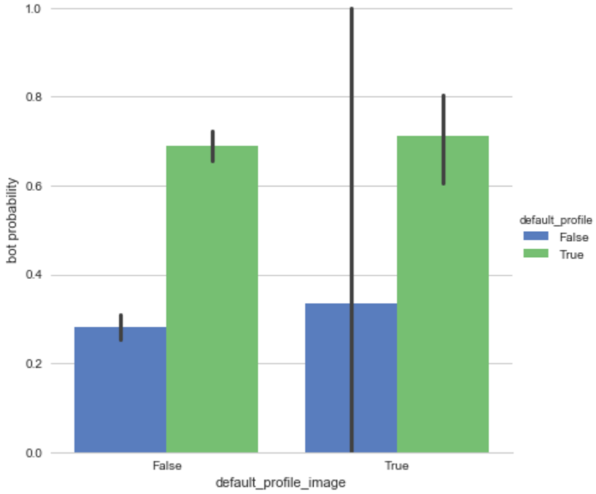

Fig. 7. Default Profile vs Default Profile Image

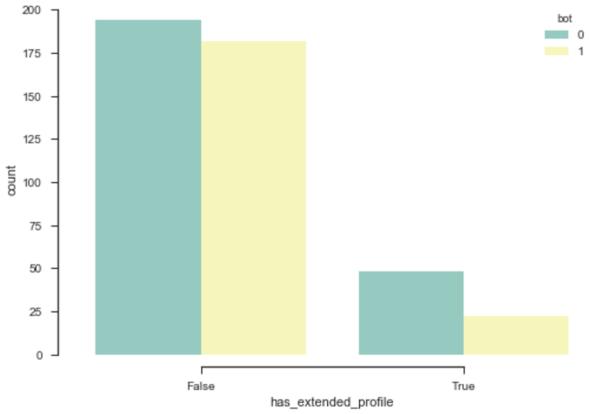

Fig. 8. No.of HasExtendProfile on bots and Humans

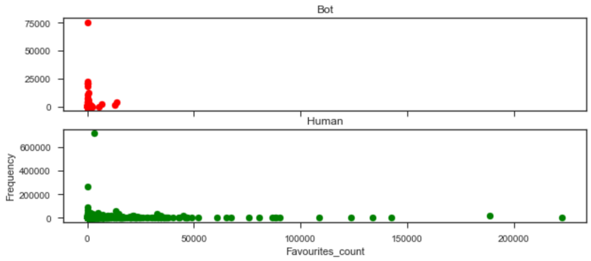

Fig. 9. Frequency on Favorites count

profile.It can be seen from graph that bot account have less probability of having extended profile.

Figure 9 depicts account having low favorite count are more probable to be bot. So favorite count and probability of account being human is positively correlated.

V. Algorithm(s) Used

The available data was split into 80% training set and 20% testing set. Dropped not important fields id, idstr, location. Converting each unique string value into a number, making out data more flexible for various algorithms, with the help of LabelEncoder in Scikit-learn. Feature selection is a process where we can automatically select those features in the data that contribute most to the prediction variable. Having too many irrelevant features in the data can decrease the accuracy of the models. By performing feature selection before modeling we can reduces Overfitting and increase Accuracy.

We have used Recursive Feature Elimination as our feature selection approach, which works by recursively removing attributes and building a model on those attributes that remain. We have chosen our base estimator as Extra Trees Classifier, which gave the optimal number of features as 6. The Figure 10 depict the distribution of the cross validation score. In which we can see that, its maximum at 6. The top 6 attributes are chosen by their feature importance score obtained from the Extra Trees Classifier model. Feature importance, estimated from the information gain given by the Extra Trees Classifier and was used to eliminate features that did not contribute to classification. The important attributes are namely followers count, friends count, listed count, favorites count, verified, tweet count. For baseline, Majority class classifier was used which had an accuracy of 0.5037.

A. Decision Tree

Decision tree classifier was trained with the training set consisting of the features considered from RFE approach namely followers count, friends count, listed count, favorites count, verified, tweet count. The decision tree classifier was able to predict with an accuracy of 0.854 using only the filtered set of features and the classifier had a cross validation accuracy of 0.841%.

The confusion matrix had a precision of 0.840 for both classes and a recall of 0.867 for both classes. The false negatives were 31 out of 461 total test records. Similarly there were 37 false positives. 10 Bot accounts that were misclassified as humans were verified bots. Remaining bot accounts had either follower counts or friends count or favorites counts or listed count similar to the mean of the respective features in human accounts. To reduce the influence of these features pruning techniques need to be used. As identifying the optimal tree in post pruning techniques is a Np complete problem we used only pre-pruning techniques by varying the max-depth of the tree and impurity threshold . Optimal depth of the tree was found to be 6 and impurity threshold 1e-8. The decision tree was now able to predict with an improved accuracy of 0.8611 and the false negatives reduced to 23 records.

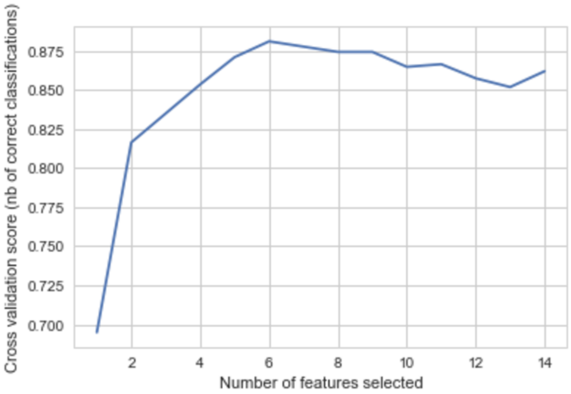

Fig. 10. Feature Selection

We also tried with the Grid Search in Scikit-learn to choose the parameters for the decision tree. The best estimator predicted by the Grid search model also has the same parameters that we predicted before.

- criterion: gini
- max depth: 6
- max features: 6
- min impurity split: 1e-08
- min samples leaf: 8

We built the decision tree with these paramaters and the tree produced is shown in the Fig. 11. With the friends count as the root of the tree.

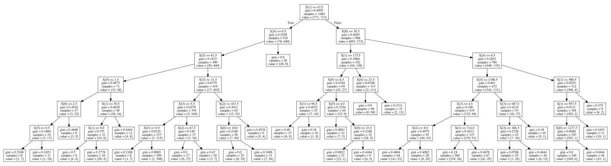

Fig. 11. Decision Tree

VI. Result

The evaluation is based on the following metrics:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

where TP, FP and FN are the numbers of true positive, false positive and false negative results respectively. After finalizing the features using feature importance values and setting parameters based on the selected features we were able to get a cross validation accuracy of 0.8611 with a precision of around 0.83 and recall of 0.891. Fig 12 shows the Receiver Operating Characteristic over all kfold cross validation by decision tree algorithms.

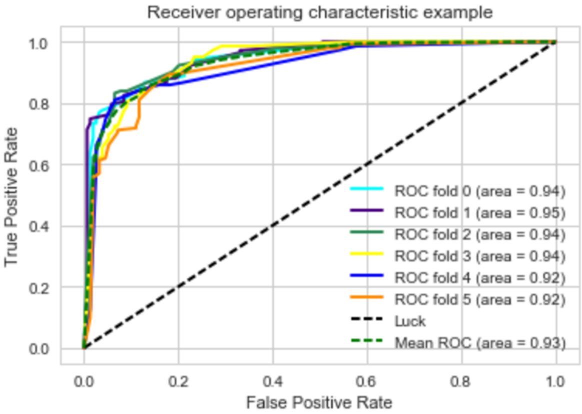

Fig. 12. Receiver Operating Characteristic

VII. Code

We have included all the codes for Data fetching, Preprocessing and ML Classification models in the following git-hub link: https://github.com/Vignesh6v/Twitter-BotorNot/

Acknowledgment

We would like to express our gratitude to, Prof. Gustavo Sandoval for his invaluable advice and guidance throughout the project.

References

[1] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer and Alessandro Flammini, *The Rise of Social Bots.* X, X, Article XX (201X), 11 pages.

[2] Kyumin Lee, Jalal Mahmud, Jilin Chen, Michelle Zhou and Jeffrey Nichols, *Who Will Retweet This? Automatically Identifying and Engaging Strangers on Twitter to Spread Information*. Harlow, England: Addison-Wesley, 1999.

[3] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

[4] H. Kwak, C. Lee, H. Park, and S. Moon, *What Is Twitter, a Social Network or a News Media?* Proc. 19th Intl Conf. World Wide Web, pp. 591-600, 2010

[5] 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.

[6] Chaoji, V., Ranu, S., Rastogi, R., and Bhatt, R. *Recommendations to boost content spread in social networks.*, In WWW, 2012.

[7] Human-Bot Interactions: Detection, Estimation, and Characterization , Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, Alessandro Flammini

[8] *http://sproutsocial.com/insights/twitter-data*