Statistical Analysis of Bot Classification on Twitter

Alexa Kelly

JMU College of Integrated Science and Technology

alexavkelly98@gmail.com

July 15th, 2020

Dr. Kathleen Moore
ISAT 116 Info-pocolypse Now
moore8ka@jmu.edu

Abstract

A cross-sectional case study on the influence automated agents (bots) have on Twitter, using new machine learning tools on data collected by the Cresci study. Analysis is based on experimentation of the machine learning algorithm, to propose a new way of combating the issue. Studies done in recent years using Twitter data have come up with methods to create a distinction between automated agents (bots) and real users. Machine learning algorithms have often proved their worth in the classification, creating open source classification tools such as Botometer. Based on previous research done, and our own algorithm's success, the results call for more sophisticated measures to be utilized so that accuracy maintains closest to 100% as possible, considering the constantly evolving nature of the subject.

Table of Contents

Part I
Project Overview

1 Introduction

In this information age, the spread of misinformation has never been more prevalent. Social media websites like Facebook and Twitter have replaced the daily newspaper for many of the newest generations. Such a large uptick in access to information also drives a proportional platform for propaganda and deceit. Many tweets and posts may seem harmless in nature, and may seem to be coming from a trust-worthy source. These people online look like your friends, neighbors, or even doctors and other 'experts'. But as we see through this analysis, this is not always the case.

The introduction of automated technology, or bots, has changed the game. What was once limited by human capabilities (such as time, sheer number of posts, human error, etc.) is now automated by technology. This allows for the account to increase their activity, and repetition power, exponentially.

Bot accounts are very efficient at the task they were programmed to do. They can tweet as quickly as every 5 seconds, but because of Twitter restrictions this is not possible to achieve without getting suspended. The bot software created by a human, or by a team, may autonomously perform actions such as automatically tweeting, re-tweeting, liking, following, unfollowing, or direct messaging. But, their purpose is not always positive.

1.1 Increasing threat of bot activity

The illusion of truth effect explains the phenomenon found in psychology experiments that repetition makes a fact seem more true.[5] Regardless of whether it is or not, or the persons prior knowledge, repetition has been proven to be a powerful force of persuasion. Naturally, the human instinct is to find short-cuts in deciding how plausible something is. Understanding this effect can help in avoiding falling for propaganda, according to psychologist Tom Stafford.[5] Lisa Fazio from Vanderbilt University conducted an experiment to test how the illusion of truth effect affects our prior knowledge. The results show that this powerful effect worked just as strongly for known as for unknown information.[7] This study suggests that previous, factual, knowledge is not strong enough to prevent repetition from influencing judgements of credibility.

This effect can easily be seen in the Twitter ecosystem, and it another important reason why people create bots in the first place. One of the main benefits to automated accounts is that they are able to continuously repeat information without much effort on the account owners behalf.

The intentions for creating such misinformation bots could be any of the following: political agenda, link farming, political infiltration, spamming, and spreading malicious content (also called sock-puppeting). Not all bots are malicious, though. Which is why it is important to classify bots like the way they were in the Cresci dataset. More positive bots may be created for positive

influence/ marketing (or as Cresci names it, fake_followers), broadcasting helpful information, automatically generating interesting or creative content, and automatically replying to users via direct message.

It is important as a user on social media platforms to stay accurately informed. Google and Bing have integrated social status updates directly into 'relevant' search results pages.[4] Users increasingly access Tweets via search and Twitter's trending hashtag notifications. Meanwhile, the access to information to gauge credibility has not, especially when compared to consuming content from social networks. The Tweeting is Believing study conducted an experiment with two groups of participants to measure the impact of several tweet features (message topic, user name, and user image) on perceptions of message and author credibility.[4] The results showed that tweet consumers have difficulty discerning truthfulness based on content alone.

A common motivation for searching public tweets is for seeking updates about local emergencies. Users look for what others are saying about what is going on, to gauge severity and importance. It is not unusual for users to perpetuate misinformation by sharing posts with numerous likes, shares and other seemingly credible features of the post without actually checking the linked article's contents. This creates a world where lies and truth are easier to confuse.[5] Tweets that are politically charged should not be taken at face value. It is sometimes difficult as a user to decipher between sources, and whether or not they are credible or not. One way to do this is to determine whether the account is a robot, or not.

So, RQ1: How can we better differentiate a bot from a legitimate account?

1.2 What is Twitter doing about this?

Twitter bot accounts are governed by a set of automation rules. In order to make a developer account necessary to make one, the user must write answers to a detailed questionnaire bot's purpose of creation. Acceptance to create a developer account needed to use Twitter API is not given out freely, and may take several weeks depending on time of year and the user's answers to the questionnaire.

1.2.1 Fact Checker Label

All misinformation spread on social media can have its own consequences in real life. Some rumors may be relatively harmless to the greater good, such as celebrity gossip, but others may sway people's opinions completely.[4] For very important decisions, such as voting in elections, this influence can be very damaging to democracy. On the other hand, social media has had a hand in supporting charitable causes such as donations to a GoFundMe to financially assist victims of natural disasters.

In 2020, the news of a pandemic sparked a wave of medical misinformation, and propaganda. The Fact Checker Label was introduced in March 2020 to discern information about the novel Coronavirus. These labels link to a Twitter-curated page or external trusted source containing additional information on the claims made within the Tweet. Labels go into 3 categories:

- "Misleading information" (things that haven't been confirmed to be false or misleading by experts)
- "Disputed claims" (statements where the truth or credibility is contested or unknown)

- "Unverified claims" (information that is unconfirmed at the time it is shared).

Twitter released an archive of more than 10 million tweets, from 3,841 accounts it said were affiliated with the IRA in 2017. In recent years, it has been bolstering efforts to shut down the spread of misinformation. "Since introducing these policies on March 18, [Twitter has] removed more than 1,100 tweets containing misleading and potentially harmful content" [2]. Twitter has a team that monitors malicious activity, ironically with the help of some automated tools. Not all bots are bad! "Additionally, our automated systems have challenged more than 1.5 million accounts which were targeting discussions around COVID-19 with spammy or manipulative behaviors. We will continue to use both technology and our teams to help us identify and stop spammy behavior and accounts" [2].

1.3 BotoMeter™

A tool has been created to assist users in identifying bots on Twitter; it is an OSoMe project called Botometer.[6] According to their website, Botometer is a machine learning algorithm trained to classify an account as bot or human based on tens of thousands of labeled examples. When you check an account, your browser fetches its public profile and hundreds of its public tweets and mentions using the Twitter API (Application programming interface). API's are the way computer programs communicate with each other, to request and deliver information. Twitter is unique in that it is the only social media platform with an open API. The Twitter API data that is collected, is passed to the Botometer API. Which extracts about 1,200 features to characterize the account's profile, friends, social network structure, temporal activity patterns, language, and sentiment. Finally, the features are used by various machine learning models to compute the bot scores. Bot scores are displayed on a 0-to-5 scale with zero being most human-like and five being the most bot-like.[6].A score in the middle of the scale is a signal that our classifier is uncertain about the classification.

The user also has the option to expand the bot score result, which presents several more, detailed scores. The first two sets are "feature category" scores, computed by a classification algorithm trained using only the corresponding features. It shows how the algorithm narrows down specific features. The first of these sets of scores is only relevant for accounts in English. The second set is more broad, and uses language-independent features.[6].

1.3.1 Complete Automation Probability (CAP)

Using Botometer for visualization and behavior analysis is great, but it does not provide enough information alone to make a judgement about an account. Which is why the CAP was created. It is the probability, according to the machine learning models that the account is completely automated. This range provided by CAP is the most useful aspect of the Botometer tool. It is calculated using **Bayes' theorem** to take into account an estimate of the overall prevalence of bots, to balance false positives with false negatives.[6].

1.4 Data

The features explored in the Cresci 2017 dataset, [1] that this study will measure in depth, were categorized into 4 categories. A large dataset was collected, and categorized into genuine, traditional spam bot, social spam bot, and fake followers. Each of these categories includes two comma-separated values files, with extensive information on tweets and users who match the categories description. This study's classification was also monitored by CrowdFlower, who identified the categories manually instead of relying on the machine learning algorithm.

Cresci 2017 data [1]: Genuine and spambot Twitter accounts, annotated by CrowdFlower contributors.

The Cresci 2017 data folder includes 10 subfolders containing categorized data:

- genuine_accounts: read into the algorithm as real_users and has 3474 observations and 42 variables.
- crowdflower_results
- fake_followers
- social_spambots (1-3)
- traditional_spambots (1-4).

Almost all of the subfolders within the Cresci Dataset provides two '.csv' files: one for tweets, and one for users. The crowdflower_results is an exception, this subfolder instead has 'agreggated', 'contributors', and 'detailed' .csv files.

According to the Cresci study [1], the genuine accounts are a random sample of genuine (human-operated) accounts. The social spambots 1 dataset was crawled from Twitter during the Mayoral election in Rome 2014. The Spambots 2 dataset is a group of bots who spammed hashtags with ads for paid apps on mobile devices. The Spambots 3 group advertised products on sale on Amazon.com. The deceitful activity was carried out by spamming URLs pointing to the advertised products. The traditional spambots were the Empirical evaluation and new design for fighting evolving Twitter spammers training set of spammers.

verified accounts that are human-operated retweeters of an Italian political candidate spammers of paid apps for mobile devices spammers of products on sale at Amazon.com training set of spammers used by Yang et al. in [43] spammers of scam URLs automated accounts spamming job offers another group of automated accounts spamming job offers simple accounts that inflate the number of followers of another account

A subset of the large data was created and used for our classification analysis, is called user_var. The total number of tweets, followers, friends, favorites, and lists collected and used for this analysis is:

status_tweetcount	followercount	friendscount	favoritescount	listedcount
60,476,608	7,305,633	6,603,581	16,397,745	79,795

2 Parameters, Confines, and Limitations

Any single tool doesn't always tell you the full story.

Algorithms depend on pattern recognition for results. Different machine learning techniques have their strengths and weaknesses when it comes to this task. The same goes for humans, except that we are not limited in the data we receive, or in computational time, and we have insight that a machine can not be taught. The benefit of classifying using an algorithm is the automation. We can classify more bots in a smaller amount of time by teaching the algorithm the patterns to look for. It is not to supplement or go against our own judgement, but create probabilities for accounts that fall into the grey area of 'Bot', or "Not'.

So, RQ2: Can an open-source bot identifier be improved with machine learning algorithms?

Botometer is updated on a regular basis in order to maintain accuracy, but it can realistically only be so accurate. The reason for this fact is two fold. 1. Data collection on Twitter is bounded by API limits. The API call limit is 43,200 accounts per API key, per day. Thus, giving the Botometer algorithm used to classify accounts, limited data to work with daily. 2. Twitter data is almost constantly changing. The culture of social media is very fluid, and trends change sometimes hourly. As the Twitter ecosystem naturally evolves, so does the data and patterns found within it. Because bot accounts want to seem as real as possible, they change aspects of the account to mimic real ones. This is a problem for classifying algorithms, since they are trained on a static dataset. Thus, algorithms based on such dynamic data are almost always one step behind, and gradually lose their effectiveness with time. The only way to combat this is to continuously, or as often as the API limit allows, update the algorithm with new data.

No algorithm is perfect, and the algorithms used in this analysis report (my own, and the one utilized by Botometer), will not always classify correctly. For example, Botometer sometimes categorizes accounts made for organizations as bots. Other times, the Botometer algorithm may confidently classify accounts that humans have a hard time with.

2.1 Statement of Analytical Confidence

1. In statistics, the statement of analytical confidence signifies an interval of values bounded by confidence limits within which the true value of a parameter is stated to lie with a specified probability. Our algorithm [9] has shown to have a 90.1% confidence in classification.
2. In Intelligence analysis, this confidence level is a rating analysts use to convey doubt to decision makers about an estimated statement of probability. The need for analytic confidence ratings arise from analysts' imperfect knowledge of a conceptual model. Here, it is reasonable to say that this study has Moderate Confidence. This means that the information is credibly sourced and plausible, but not of sufficient quality or corroboration to warrant a higher level of confidence. This level was chosen because of the accuracy of the algorithm, since it is above 90%, but not by much. The algorithm needs significant fine tuning to get an accuracy level above 98%. With more practice and knowledge of more efficient training models, data preparation, and more machine learning tools; the algorithm has promise of being a significant classifier.

Part II
Research

3 Summary of the Research

Misinformation runs rampant on social media sites. Unless you do your own research on every single tweet or post you see on your timeline, you are subject to internalizing misinformation. The introduction of automated agents (bots) became an inflammatory aspect to the growing issue. Bots facilitate a wider spread of the misinformation, due to the spamming nature that they tend to have. They sometimes pose as real accounts with real opinions, but are actually spam. About nine-in-ten (89%) tweeted links to popular news aggregation sites were posted by bots, not human users [3]. Typical users and the public cannot tell the difference between a bot and a real account on most occasions.

There has been some controversy, especially in light of the 2016 election, and the COVID-19 pandemic, surrounding the role social media has in the spread of misinformation. To combat this, Twitter introduced Fact check labels, among other monitoring methods such as the use of proprietary algorithms and a team to run them. The creation of tools like BotoMeter, and collections of data on behaviors of bots, helps mitigate bot activity. These monitoring methods have been received with mixed emotions. Fact checker labels used on a couple of the presidents tweets fueled outrage and discussion of limiting a social media's power to control its activity.

Efforts in recent years by the social media sites themselves, and data scientists, have assisted in controlling the spread of misinformation through automated accounts. The issue in the grand scheme of things is a new one, with undoubtedly more obstacles to be assessed. The algorithms, tools and teams involved will become more accurate as time goes on.

3.1 Pattern detection

Combining all of the datasets [1] together results in a large dataframe, that is not realistic to run on a laptop computer. Combining all three of the social spambot datasets also proved to be too taxing for the computer used to run this analysis. For example, the social_users dataset, that includes all the 'social spambot' user data collected, has 40 variables and 4,913 observations. And, the social_tweets combined dataset, which includes all the 'social spambot' tweets collected, said 25 variables and 3,457,133 observations. So, combining the other subsets of data together is not feasible. Instead, we will look at the first subset of users for each category. In other words, this report analysis will focus on:

1. fake_followers users (there is only one)
2. social_spambots_1 users
3. traditional_spambots_1 users
4. genuine_accounts users

Combining these 4 subsets of the large dataset will allow for diversification of data, and still allow for 'bot or not' classification.

The result used for classification is comprised of 8816 rows, 11 columns, and is saved as the user_var data frame. In order to proceed with classification, the data frame must have only numeric values, and a binary classifying factor (1 for bot, 0 for human).

```
'data.frame':   8816 obs. of  11 variables:
 $ id                           : num  24858289 33212890 39773427 57007623 63258466 ...
 $ statuses_count               : int  1299 18665 22987 7975 20218 15259 9551 206 93793
...
 $ followers_count              : int  22 12561 600 398 413 134 337 28 2617 1561 ...
 $ friends_count                : int  40 3442 755 350 405 401 630 105 52 2001 ...
 $ favourites_count             : int  1 16358 14 11 162 55 655 38 0 0 ...
 $ listed_count                 : int  0 110 6 2 8 1 1 0 28 0 ...
 $ profile_use_background_image: num  1 1 1 1 1 1 1 1 1 1 ...
 $ botornot.f                   : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ screenname_len               : num  9 12 10 14 11 10 13 15 13 13 ...
 $ name_len                     : num  14 14 15 18 12 7 14 16 16 14 ...
 $ desc_len                     : num  0 134 23 149 79 103 78 75 0 37 ...
```

Figure 1: Final working dataset variable overview

A larger subset of the collected data, named users here, was created and used for statistical analysis for this report, and includes variables that are not numeric.

3.2 Statistical Analysis

Let's take a look at what the mean values are for each category. Note that 'mean_statuscount' stands for the mean number of "statuses" (posts an account creates on Twitter), 'mean_followcount' stands for the mean number of followers, 'mean_followcount' stands for the mean number of followers, 'mean_friendcount' stands for the mean number of friends, and 'mean_favcount' stands for the mean number of 'favorites' (posts they have 'liked').

Account type	mean_statuscount	mean_followcount	mean_friendcount	mean_favcount
f_followers_u	72	18	370	4
social1_u	1112	1785	1854	158
social_users	1500	612	590	40
trad_spam1	221	637	1327	4
Bots (all)	726	763	1035	52
real_users	16958	1393	633	4670

The social_users row is based off of the 3 social users datasets combined. So, the means for this row is calculated on all social spambot subsets collected for this study.

Lets compare the overall mean values for bots and humans.

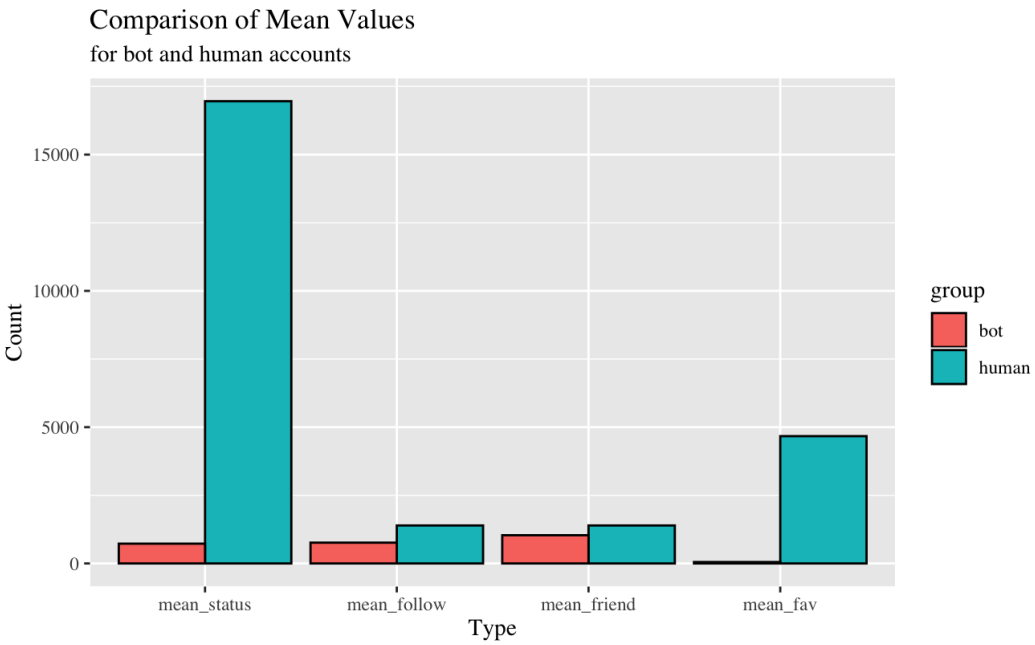

Figure 2: Mean variable values

Clearly, human accounts have way more activity across the board, compared to bots in general. This is a useful operation, because although bots are used and known largely for spamming we see here that they do not compare in overall activity. Bots typically have a shorter lifespan than human accounts, and especially so for malicious spambots. They are more likely to be taken off of Twitter for violating Twitter's policies. Bots are seen to be limitless, but we must consider the fact that they must be told what to say for the most part. They need at least some monitoring, unlike human accounts that may continuously tweet, like, retweet, direct message, and more. More complicated bots that learn on their own are more difficult to make and keep active on Twitter. Account duration is a key aspect to differentiating bot versus human accounts.

Cresci dataset also included a manual human classification of bot and genuine accounts. This proved to not be as effective as expected. The graph in Figure 3 below shows the confidence interval for the Crowdflower classifications.

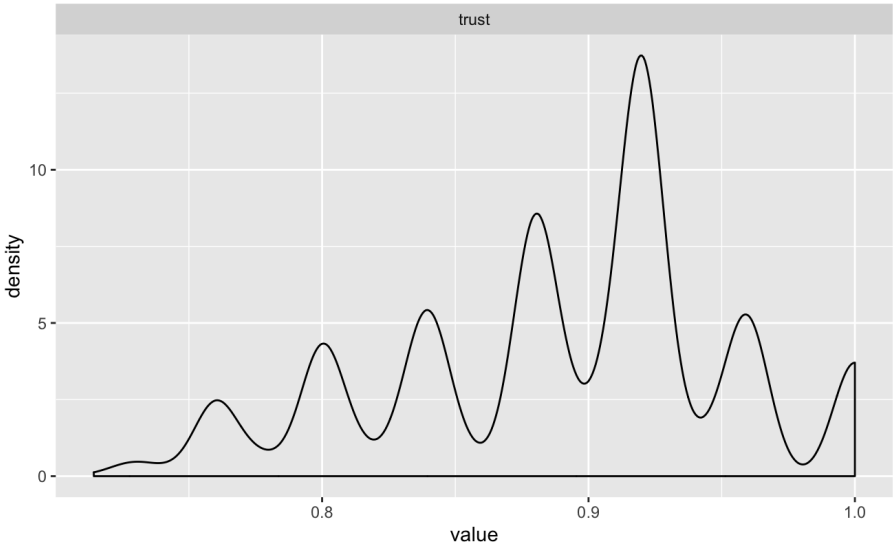

Figure 3: CrowdFlower distribution of trust

4 Methods Used

In order to solve a real life problem using machine learning, one must do extensive data preparation and cleaning to get the data frame to be in a compatible form for the methods required. The Cresci datasets were raw, and included a lot of NULL values. These NULL values must be filtered out, either by completely omitting that information, or replacing the values with the mean. In fact, any variables with over 30% NULL values should be omitted unless they are significant. Most of the columns that were comprised of NULL values could not be useful for the algorithm due tot he fact that they were not numeric or in factor form.

4.0.1 Classifying variables

Heat maps are a great tool for feature selection, and to see any association between data from different sources. It is done by visualizing clusters of samples and features. Hierarchical clustering is performed to both the rows and the columns of the data matrix. Then, the rows and columns are re-ordered according to the hierarchical clustering result, putting similar observations close to each other. The blocks of 'high' and 'low' values are adjacent in the data matrix. Color is applied accordingly, for clarity. The darker the color, the more statistically significant the feature will be. Visualizing the data matrix using a heat map can help to find the variables that appear to be related for each sample cluster.

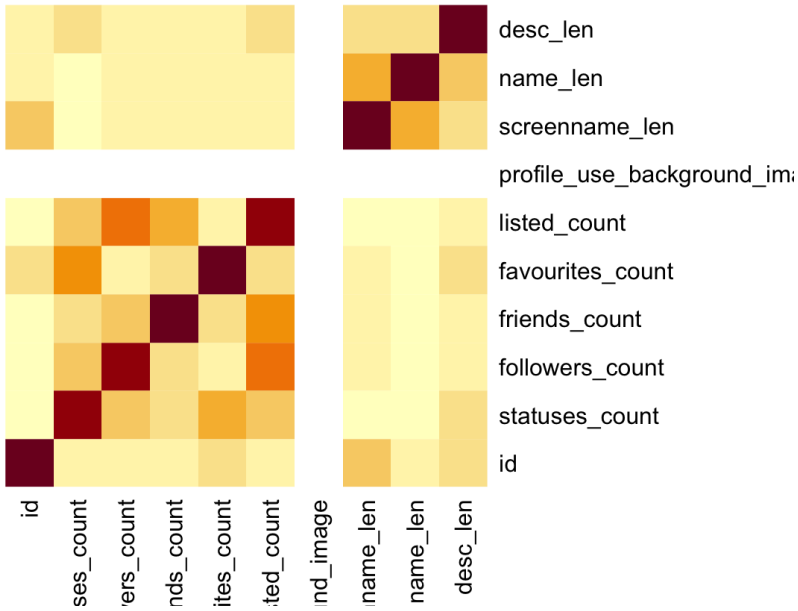

Figure 4: Heatmap result

K-means clustering a type of unsupervised machine learning algorithm that solves clustering problems. The algorithm classifies a given data set simply; by grouping alike features together. It starts by using a user defined k number of points for each cluster, called centroids. Each data point forms a cluster with the closest centroids, forming k clusters of data. It is a recursive process that only stops when converge occurs, which is when the clusters are completely homogeneously grouped. Running this piece of the larger study's algorithm, with the plot, took 10 minutes to run using Rmarkdown. This process was chosen because of its unsupervised nature. The Cresci study used a more complex supervised approach, so this case study attempted to see what the results of a different, unsupervised machine learning tool would have on the data.

Here we see that the K-means clustering created 3 clusters of sizes 5984, 878, 1954. The table below shows the mean values for each of the clusters:

```
K-means clustering with 3 clusters of sizes 5984, 878, 1954

Cluster means:
          id statuses_count followers_count friends_count favourites_count listed_count
1  326107586       7594.389        989.6984      899.5603         1644.236    11.796959
2 2548239490       7199.085        713.6355      599.8314         3639.608     4.035308
3 1193094791       4458.028        387.2600      355.1484         1721.117     2.896111
  botornot.f screenname_len name_len desc_len
1 0.64789439       11.04027 11.49883 52.74248
2 0.07744875       11.20729 10.03986 48.17426
3 0.71494371       12.99130 12.33521 64.11924
```

Figure 5: K-means clustering stats

The algorithm had a 90% accuracy in classifying, as we see in [Figure 9]. Although this percentage may seem high, accurate machine learning algorithms classify data in the high 90's, as close to 99 percent as possible. There is clearly room for improvement in both the data collected and the algorithm itself. A main way to get more accurate results is to improve the data preparation before training the dataset. Other than the heat map and clustering, more sophisticated measures could be considered to ensure that the clustering training set is as effective as possible. This can be done by selecting better features, or converting variables based on other classifiers, such as Random Forest.

4.1 Compare Inaccuracies

The Cresci study [1] used a number of methods to classify the factors as bot or human.

Their data preparation included a progressive variation of class distribution of fake followers and human followers in the dataset, from 5%–95% to 95%–5% (respectively 100 humans–1900 fake followers, 1900 humans–100 fake followers). [1]

One of the methods used was an unsupervised spambot detection via graph clustering. The approach exploits statistical features based on URLs, hashtags, mentions and retweets. Feature vectors are then compared with one another, using an Euclidean distance measure. Distances between accounts are organized in an adjacency matrix. An undirected weighted graph of the accounts using a Markov Cluster Algorithm are constructed.[1] Graph clustering and community detection algorithms, like heat maps, are applied to identify groups of similar accounts. This supervised system provides a machine learning classifier that decides whether a Twitter account is genuine or spambot by relying on the account's details; focusing on account relationships, tweeting timing and level of automation.

technique	type	detection results					
		Precision	Recall	Specificity	Accuracy	F-Measure	MCC
test set #1							
Twitter countermeasures	mixed	**1.000**	0.094	**1.000**	0.691	0.171	0.252
Human annotators	manual	0.267	0.080	0.921	0.698	0.123	0.001
BotOrNot? [14]	supervised	0.471	0.208	0.918	0.734	0.288	0.174
C. Yang *et al.* [43]	supervised	0.563	0.170	0.860	0.506	0.261	0.043
Miller *et al.* [30]	unsupervised	0.555	0.358	0.698	0.526	0.435	0.059
Ahmed *et al.* [2]♯	unsupervised	0.945	0.944	0.945	0.943	0.944	0.886
Cresci *et al.* [13]	unsupervised	0.982	**0.972**	0.981	**0.976**	**0.977**	**0.952**
test set #2							
Twitter countermeasures	mixed	**1.000**	0.004	**1.000**	0.502	0.008	0.046
Human annotators	manual	0.647	0.509	0.921	0.829	0.570	0.470
BotOrNot? [14]	supervised	0.635	**0.950**	0.981	0.922	0.761	0.738
C. Yang *et al.* [43]	supervised	0.727	0.409	0.848	0.629	0.524	0.287
Miller *et al.* [30]	unsupervised	0.467	0.306	0.654	0.481	0.370	-0.043
Ahmed *et al.* [2]♯	unsupervised	0.913	0.935	0.912	0.923	**0.923**	0.847
Cresci *et al.* [13]	unsupervised	**1.000**	0.858	**1.000**	**0.929**	**0.923**	**0.867**

♯: Modified by employing *fastgreedy* instead of *MCL* for the graph clustering step.

Figure 6: Cresci study classification results

Part III
Moving Forward

So, RQ3: What about the more sophisticated ones?

It is sometimes difficult as a user to decipher between sources, and whether or not they are credible or not. This is especially true for automated agents that are well designed to mask their intentions. They do so by mastering an alias, creating a seemingly credible online false persona. Determining whether the account is a robot, or not is much more complicated in these cases. The features that are typically seen in bots, and therefore are the features sought after by the classification algorithm, will be masked if it is a more sophisticated bot. Unfortunately, with every study that is conducted, and posted for the purpose of understanding the spread of misinformation, may prove to be counter-intuitive. Sophisticated bots and masters of misinformation will use this information to adapt and better hide the aspects that were previously held to be indicative of their intentions. The goal of studying bots and their trends is to stay one step ahead of them and be able to control their influence as much as possible.

According to Cresci[1] on Twitter, "a few novel approaches are being experimented, but the majority of detectors uses outdated techniques and approaches (we are not doing it right!)". The graph in Figure 7 below shows the impact estimation of social bots. There has been an exponential growth

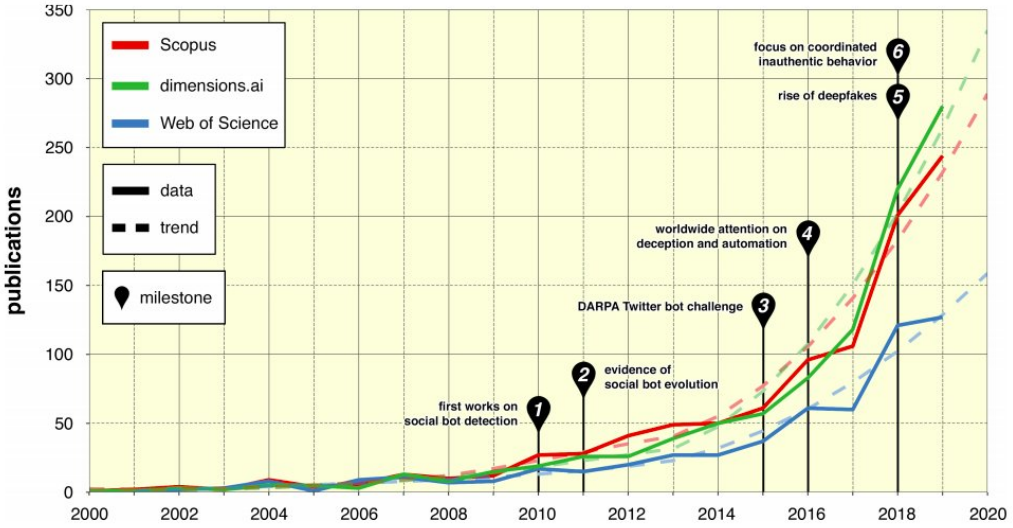

Fig. 2. Publications per year on the characterization, detection and impact estimation of social bots. Since 2014, the number of publications on the topic skyrocketed. We forecast that from 2021 there will be more than 1 new paper published per day on social bots, which poses a heavy burden on those trying to keep pace with the evolution of this thriving field. Efforts aimed at reviewing and organizing this growing body of work are needed in order to capitalize on previous results.

Figure 7: Cresci bot forecasting

Compare this trend to the steadily increasing Twitter user growth, it makes sense to see these two lines are correlated.

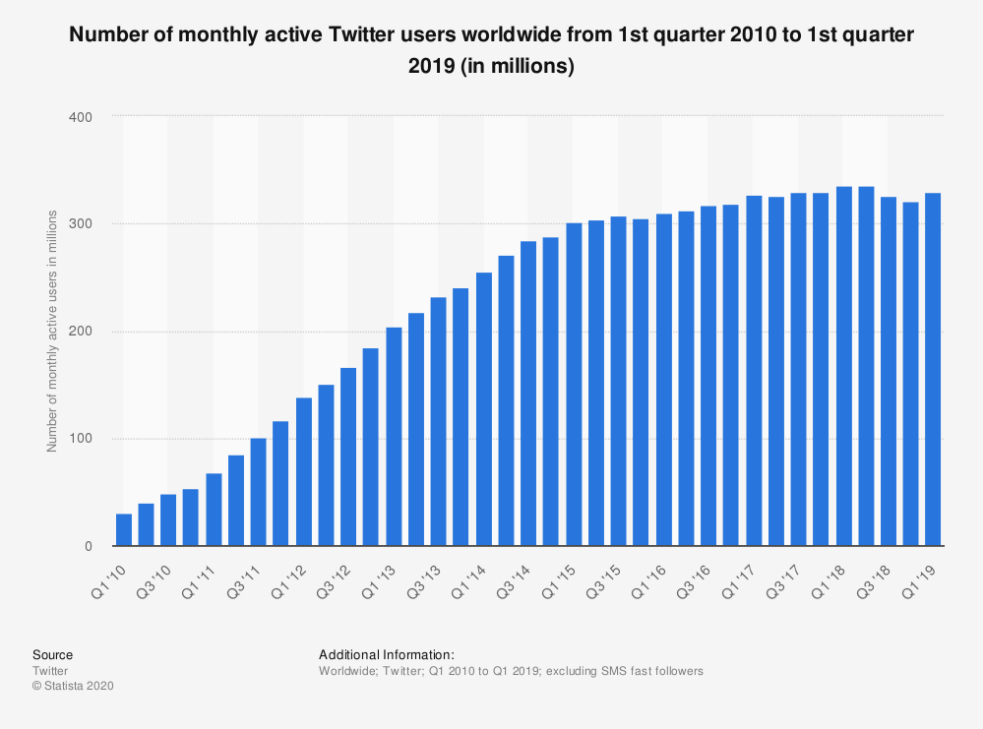

Figure 8: Twitter monthly user growth trendline [9]

Social media has proven to be an ever growing platform, and even more integrated into our lives. With this, comes an increase in the spread of information, malicious and otherwise.

5 Recommendations and Further Work

Although the fight against malicious bot accounts seems futile, it is an important process. By accurately mimicking the characteristics of genuine users, these spambots are naturally harder to detect than those studied by previous collected data.

In the Cresci study, neither the human team nor the spambot detection applications were all that successful in accurately detecting spambots. Which truly showcased that a lot the automated tools that we use to detect spambots erroneously label social spambots as genuine accounts. Clearly, there is a long ways to go until we are continuously successful in these efforts.

There is no doubt that both automated agents, and the machine learning algorithms that detect them will continue to improve to further each of their purposes. If we do not control the exponential growth that bots incur due to their automated qualities, the war on misinformation will worsen. Eventually, it is the hope that the technology will improve at a faster rate so that it may better combat this issue. So, studies like this one, and others mentioned in this report are vital. Eventually, through trial and error, and experimentation, we will figure out the most accurate way to detect

bot activity. We must not accept a 90% accuracy as good enough, especially since the data is ever evolving. But, it is a great step in the right direction.

Continuous creation and innvoations of technology and the algorthms that run them is key to creating a more accurate detection system. The original hypothesis of this project included the creation of a Twitter API to automate a bot that collects its own data on tweets to analyze. Due to the time constrictions of this project, Twitter took over 2 weeks to approve of the developer account needed in order to create the twitter bot. The next step to this analysis report is collect my own Twitter data, and rerun the algorithm suing this data. This will be used to measure the usefulness and accuracy of the algorithm.

Part IV
Appendix

```
Within cluster sum of squares by cluster:
[1] 3.160261e+20 9.793821e+19 5.703209e+19
 (between_SS / total_SS =  90.1 %)
```

Figure 9: K-means clustering result accuracy

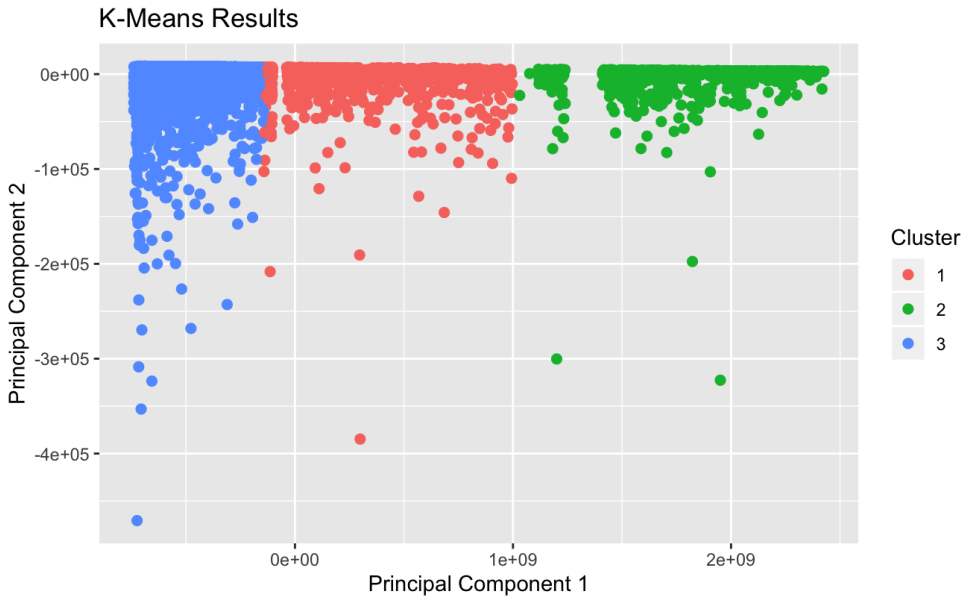

Figure 10: K-means clustering plot

Part V
Code Documentation

The GitHub repository can be found **here**. It includes this report, as well as the code for the algorithm that made the analysis, the HTML version of the Rmarkdown analysis, a link to the data, and the presentation slides.

6 References

1 Cresci 2017 data: The Paradigm-shift of Social Spambots: Evidence Theories and Tools for the Arms Race, S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi WWW '17 Proceedings of the 26th International Conference on World Wide Web Companion, 963-972, 2017

2 Gadde, V. (2020, March 16). An update on our continuity strategy during COVID-19. Retrieved July 11, 2020, from https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html

3 Wojcik, S., Messing, S., Smith, A., Rainie, L., & Hitlin, P. (2020, May 30). Twitter Bots: An Analysis of the Links Automated Accounts Share. Retrieved July 11, 2020, from https://www.pewresearch.org/internet/2018/04/09/bots-in-the-twittersphere/

4 Morris, M. R., Counts, S., Roseway, A., Hoff, A., amp; Schwarz, J. (2012). Tweeting is believing? Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12. doi:10.1145/2145204.2145274

5 Stafford, T. (2016, October). How liars create the 'illusion of truth'. Retrieved July 12, 2020, from https://www.bbc.com/future/article/20161026-how-liars-create-the-illusion-of-truth

6 Project, OSoMe. (2018). Botometer by OSoMe. Retrieved July 10, 2020, from https://botometer.iuni.iu.edu/

7 Fazio, L. K., amp; Marsh, E. J. (2009). Prior knowledge does not protect against illusory truth effects. PsycEXTRA Dataset. doi:10.1037/e520562012-049

8 Kreil, M. (2019). Lecture Notes. Retrieved July 10, 2020, from https://michaelkreil.github.io/openbots/

9 Clement, J. (2019, August 14). Twitter: Monthly active users worldwide. Retrieved July 14, 2020, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

6.1 Credit to R and the packages used

- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
- Terry Therneau and Beth Atkinson (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart
- Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2019). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.
- Alboukadel Kassambara and Fabian Mundt (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

- Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. http://www.jstatsoft.org/v21/i12/,
- Max Kuhn (2020). caret: Classification and Regression Training. R package version 6.0-85. https://CRAN.R-project.org/package=caret,