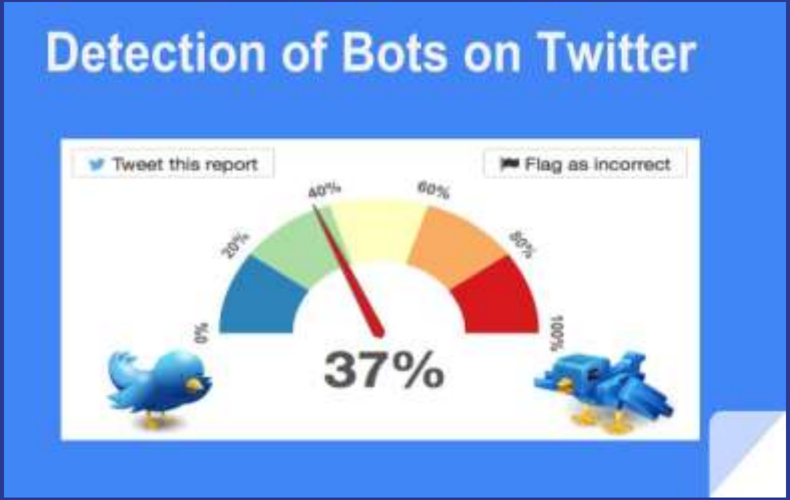
Detection of bots on Twitter



Rahul's Twitter celebrity fuelled by bots, says report

22nd October, 2017

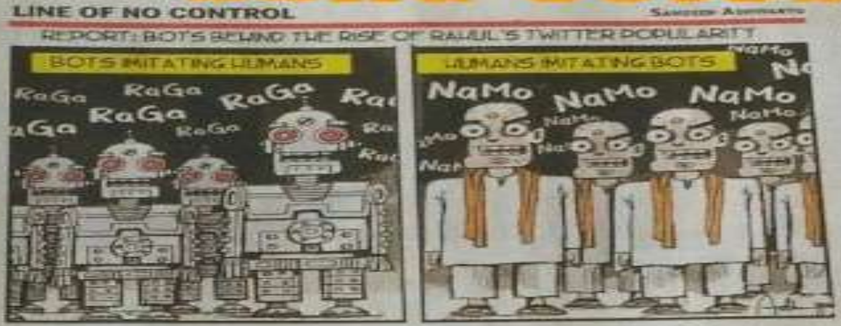
BJP Says Oppn Manipulating Social Media, Cong Alleges Sabotage

Times News Network

New Delhi: A media report suggesting the rise in Congress vice-president Rahul Gandhi's Twitter profile might be driven by "bots" that do not post original tweets but only re-tweet sparked a bitter war of words, with BJP claiming a bid to manipulate social media and Congress alleging sabotage.

The trigger for the skirmish was an agency report referring to a post by Rahul on October 15 sarcastically suggesting PM Narendra Modi needed "to hug Donald Trump once more" after the US President praised Pakistan's leaders. "Modi ji quick, looks like President Trump needs another hug (sic)," Rahul tweeted.

While the tweet gained 30,000 retweets, an agency report said a "close analysis" of "alleged bots" showed that they had been routinely retweeting Rahul's tweets. The profiles of these handles, it said, were from countries like Kazakhstan, Russia or Indonesia, had a low follower count, and did not feature original tweets, but only retweets from other accounts. It listed the profile links of 10 alleged "bot" accounts that had retweeted Rahul.



Information and Broadcasting minister Smriti Irani tagged the agency report with the barb, "Perhaps @OfficeofRG planning to sweep polls in Russia, Indonesia and Kazakhstan?" Sports minister Rajyavardhan Rathore wrote, "In sports, this would come under 'doping'.... hey wait! does 'dope' remind you of someone." The controversy trended under #RahulWaveInKazakh.

BJP leaders seized on the report to allege Rahul's social media profile was being managed but Congress quickly cited its vice-president's latest tweet about BJP opposing a Tamil movie, and asked saffron leaders to respond how had it touched 17,000 retweets in no time.

Rebuffing charges, Congress's social media in-charge Divya Spandana dubbed the agency report as "factually incorrect".

"The agency article says 10 handles have been retweeting Rahul repeatedly. It appears that these handles have retweeted Rahul's Trump tweet. How that happened and who did it, only Twitter can clarify because they claim there are no bots in their system. So, who created these handles and who retweeted, only Twitter can reply," she told TOI.

Divya said, "BJP is perhaps trying to sabotage our network with bots because they can't handle Rahul's growth."

The social media has been replete with claims of political parties that rivals were using bots and buying followers and retweets. There have also been suggestions that Rahul was using "names like Trump" in his tweets to gain traction on social media since they have greater global following.

Divya asked, "Even Rahul's tweet about Tamil movie Mersal today has notched 17,000 retweets, 35000 likes. How has it happened then?"

As the twitter fire over Rahul Gandhi spread, the controversy took a new turn. Many took to Twitter Audit, a tool that claims to analyse how many of an account's followers are fake.

This is done by taking a random sample of up to 5,000 followers and studying them for "number of tweets, date of the last tweet, and ratio of followers to friends" to tell if they are fake. This tool gave Gandhi an audit score of 51% — over 1.8 million of his 3.8 million followers were fake, while the score for PM Narendra Modi's account was 37% — over 22 million of his 35.6 million followers were fake.

Interestingly, "research" by two handles — cited by Congress managers — said the alleged bot handles quoted by the agency report were activated just four days ago, well after Rahul's tweet about Trump posted 30,000 retweets.

Why bots, not humans, make news go viral

Automated Accounts Behind 66% Of Tweets Linking To Sites: Study

Automated accounts or 'bots' play a big role in disseminating information on Twitter, accounting for two-thirds of tweets linking to popular websites, a study showed.

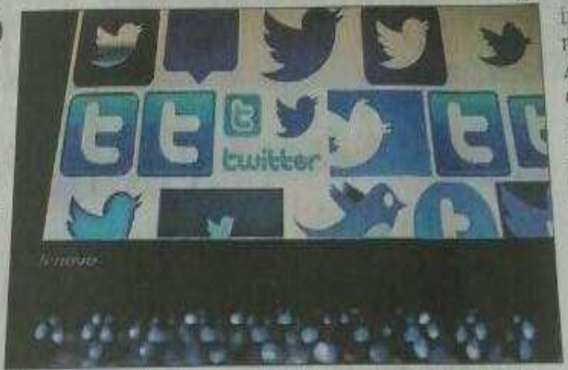
The Pew Research Center report found bots were a major source for diffusing information on news, sports, entertainment and other topics.

The researchers found that of all tweeted links to popular websites, 66% were shared by accounts that appeared to be automated rather than human users.

While bots have gained attention due to concerns over Russian-sponsored manipulation of social media during the 2016 political campaign and for other hot-button topics, the researchers said they made no effort to distinguish between 'good' or 'bad' bots.

"The study does not find evidence that automated accounts currently have a liberal or conservative 'political bias' in their overall link-sharing behaviour," the researchers wrote.

Twitter's policy on automated accounts, last updated in November, allows bots to operate but with limitations.



AFP

The Pew Research Center report found bots were a major source for diffusing information on news, sports, entertainment and other topics

The policy allows for bots to "automatically broadcast helpful information" or "run creative campaigns that auto-reply to users."

But Twitter's rules forbid automatic posts about trending topics or using automation "to attempt to influence or manipulate trending topics." It also bans the use of multiple accounts to generate more activity.

"These findings illustrate the extent to which bots play a prominent and pervasive role in the social media environment," says Pew researcher Aaron Smith. "Automated accounts are far from a niche phenomenon: They share a significant portion of tweeted links to even the most prominent and mainstream publications and online outlets. Since these accounts can impact the information people see on social media, it is important to have a sense their overall prevalence on social media."

Pew researchers said some examples of automated accounts included a Netflix bot which tweets when new shows are added to the online streaming service, and another which posts random images from the Metropolitan Museum of Art.

CNN operates a "breaking news" bot and the New York Times has one that provides live analysis of NFL games.

The researchers said a small number — some 500 — of highly active bots are responsible for a disproportionate share of the tweeted links, or 22%.

Pew found that an unusually large share of 'adult content' is retweeted by bots, accounting for 90% of links to popular adult sites. That coincides with findings from other researchers on campaigns of "pornbots" which advertise sex and pornographic websites. The Pew report examined some 1.2 million English language tweets linking to 2,315 of the most popular websites in a six-week period in [illegible] 2017. [illegible]

Apr 11, 2018

Literature Review

Online Human-Bot Interactions: Detection, Estimation, and Characterization

Using clustering analysis to characterize bots as

- Spammers
- Self-Promoter
- Accounts that post content from connected applications.

The best classification performance of 0.95 AUC was obtained by the Random Forest algorithm.

- In some cases, the boundary separating the bot and human is not sharp.
- Too many feature sets used.

A New Approach to Bot Detection: Striking the Balance Between Precision and Recall

Bot detection which considers recall in its formulation along with precision using AdaBoost and algorithm.

- Increased calculation complexity.
- Decreased precision.
- Need of a boosting algorithm.

TABLE V: The Precision, Recall and F_1 measure of different models on Honeypot dataset.

Method	Precision	Recall	F_1
$Heuristic_{URL}$	49.69%	96.39%	65.58%
$Heuristic_{Retweet\%}$	50.05%	99.33%	66.56%
$Heuristic_{Length}$	50.00%	99.82%	66.63%
$Heuristic_{Time}$	49.99%	99.96%	66.65%
SVM	62.41%	62.52%	62.47%
$AdaBoost$	79.76%	72.41%	75.91%
$BoostOR$	71.42%	82.48%	**76.55%**

Who is Tweeting on Twitter: Human, Bot, or Cyborg?

To classify a Twitter user as a human or a bot by observing the difference among human and bot in terms of tweeting behavior, tweet content, and account properties.

		Classified			Total	True Pos.%
		Human	Cyborg	Bot		
Actual	Human	949	51	0	1000	94.90%
	Cyborg	98	828	74	1000	82.80%
	Bot	0	63	937	1000	93.70%

The problem

Objective

Classify a Twitter user as a human or a bot by observing the difference among human and bot in terms of tweeting behavior, tweet content, and account properties.

Context

- Around 9 percent to 15 percent of Twitter's active monthly users are bots
- So, out of 319 million active monthly users, that translates into 28.7 million to 47.9 million bots.

Problem statement

To assist human users in identifying who they are interacting with,our project focuses on the classification of human and bot accounts on Twitter,by using the combination of features extracted from user's’ account to determine the likelihood of being a human or bot.

Process

Features

Identifying Features

We will use various features, like followers to friends ratio, URL ratio, etc. for identification.

Classification

Classify bots from humans

The decision maker uses the features identified to determine whether is a human or bot and classify the account user accordingly.

Analytics

Visualization of Analysis

Visualise each feature showing the differentiating characteristics of a bot.

And displaying the final analysis/classification.

Features

To Identify a Bot

1. Huge amount of following, small amount of followers (Followers to Friend Ratio)
2. Coming from an API (Tweeting device)
3. URL Ratio
4. Reciprocity
5. Entropy

Implementation

1. Followers to Friend Ratio

- We have the individual columns containing the followers and friends count of every user (both humans and bots) in the dataset.
- We will create a new column having the ratio of followers and friends count in the final dataset

2. Recognizing the Tweeting Device

- Identify the user-id or tweet-id and using the Twitter API, recognize the device from which the tweets were tweeted from and classify them as mobile or laptop, API, third-party component, etc.
- Store this information in a column and use for classification.

3. URL Ratio

- We have the feature column that contains the URL count that is extracted from the account user's tweets.
- We will create a new column having the ratio of URL count to total tweets count in the final dataset.

$$Heuristic_{URL}(u) = \frac{|\{x | x \in tweets^u, x \text{ contains URL}\}|}{|tweets^u|}$$

4. Reciprocity

1. For every user, randomly choose 20 users from the list of users whom he/she follows.
2. Check whether the follower follows back the user or not
3. Find the ratio of ‘number of users who follow back’ to 20
4. Store this ratio in the dataset

5. Entropy component

- The entropy component detects periodic or regular timing of the messages posted by a Twitter user.
- If the entropy or corrected conditional entropy is low for the inter-tweet delays, it indicates periodic or regular behavior, a sign of automation.
- High entropy indicates irregularity, a sign of human participation.

$$Heuristic_{Time}(u) = \frac{1}{|tweets^u| - 1} \sum_{i=2}^{N} (t_i - t_{i-1})$$

Logistic Regression - The Classifier

Logistic regression does not try to predict the value of a numeric variable given a set of inputs. Instead, the output is a ***probability*** that the given input point belongs to a certain class(human or bot).

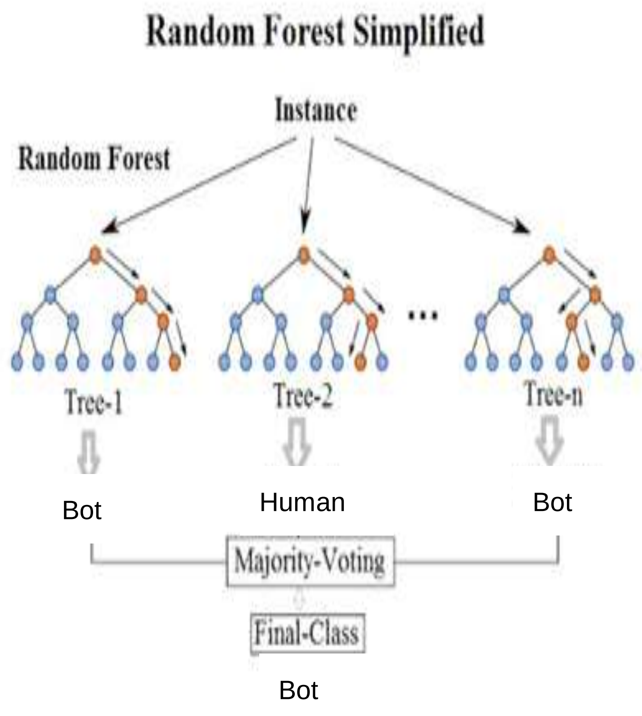
- 0 = you are absolutely sure that the user is a human.
- 1 = you are absolutely sure that the user is a bot.
- Any value above 0.5 = you are pretty sure about that user being a bot. Say you predict 0.8, then you are 80% confident that the user is bot. Likewise, any value below 0.5 you can say with a corresponding degree of confidence that the user is not a bot.

It is clear that the data points MUST be separable into the two aforementioned regions by a linear boundary.

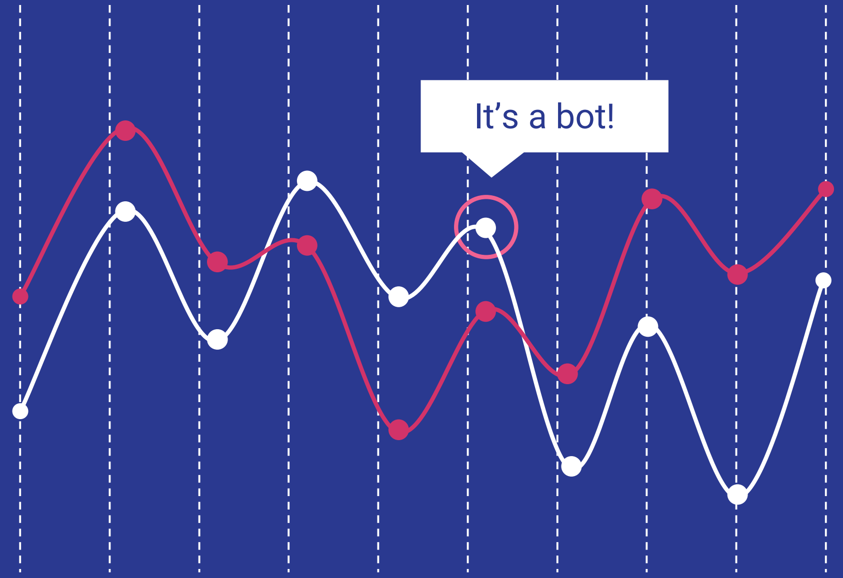
Random Forest - The Classifier

Random forest classification algorithm is used for training the dataset.It is the collection of decision trees. The random forest learning is also robust when training with imbalanced data set. It provides high accuracy rate when training with large dataset.

In random forest classification algorithm, a random instance of data is chosen from the training dataset. With the selected data, a random set of attributes from the original dataset is chosen.In a dataset, where M is the total number of input attributes in the dataset, only m attributes are chosen at random for each tree where m<M.



Visualization



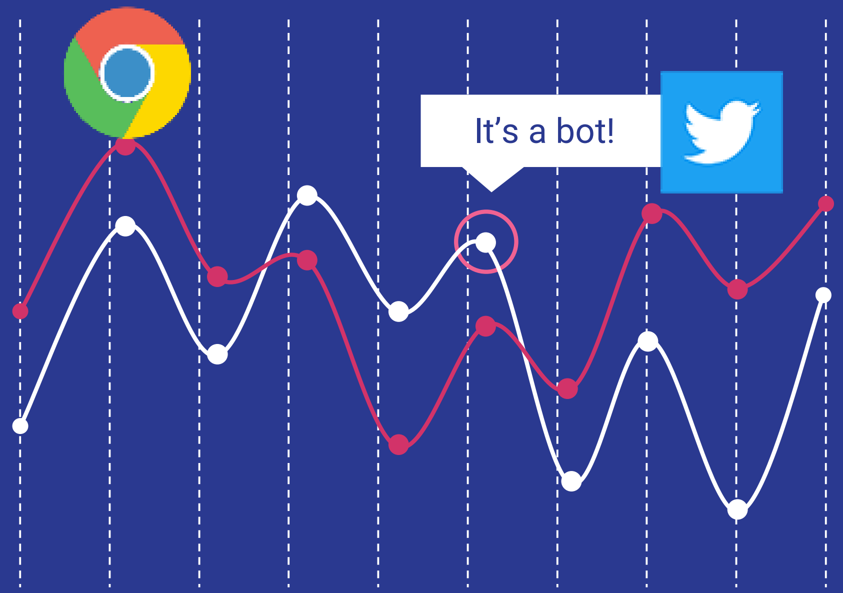
Analytics

Seaborn harnesses the power of matplotlib to create beautiful charts in a few lines of code. The key difference is Seaborn's default styles and color palettes, which are designed to be more aesthetically pleasing and modern.

Seaborn offers various features such as built in themes, color palettes, functions and tools to visualize univariate, bivariate, linear regression, matrices of data, statistical time series etc which lets us to build complex visualizations.

Installation - pip install seaborn

Real-time detection



Algorithm

1. User visits the web application.
2. User signs in using his/her Twitter account and allows the web application to read the user's twitter feed.
3. The web application, uses the access given by the user to fetch the user's Twitter feed using the Twitter API
4. For every tweet in the user's Twitter feed, the web application sends the userId of the tweeter's account to the server.
5. The server, on receiving an user-id, check the Redis cache for an existing classification output.

    a. If the cache has the classification output, then return this output to the web application
    b. Else,

        i. The server uses the Twitter API to fetch details necessary for the classification model to predict the output.
        ii. The server, on successful fetch of user details, sends them to the AWS Lambda function for classification.
        iii. The Lambda function, on receiving the user details, fetches the trained machine learning model from Amazon S3 and uses it to classify the user. It returns this output back to the server.
        iv. The server, on receiving the output, formats it accordingly and sends the output to the web application.

6. The web application, on receiving the classification output, displays it along with the tweet to the user.

Block Diagram

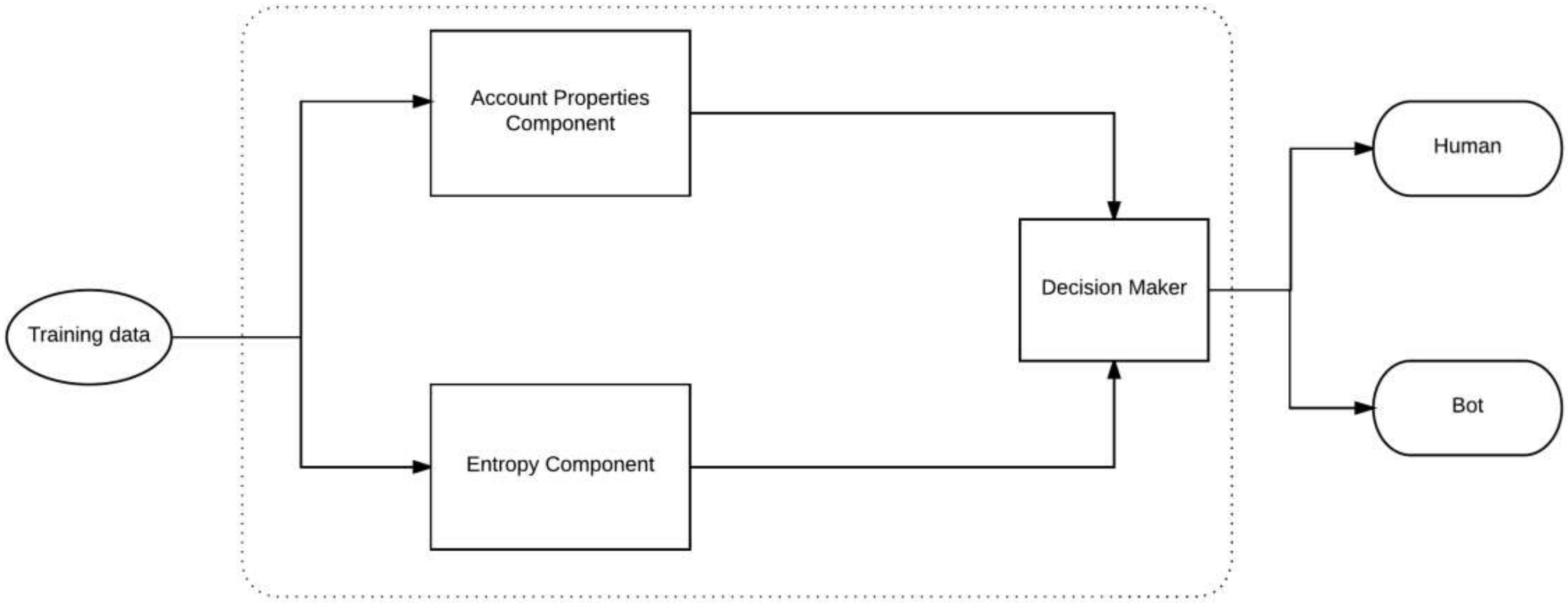


Fig: Block Diagram of Training Phase

Block Diagram

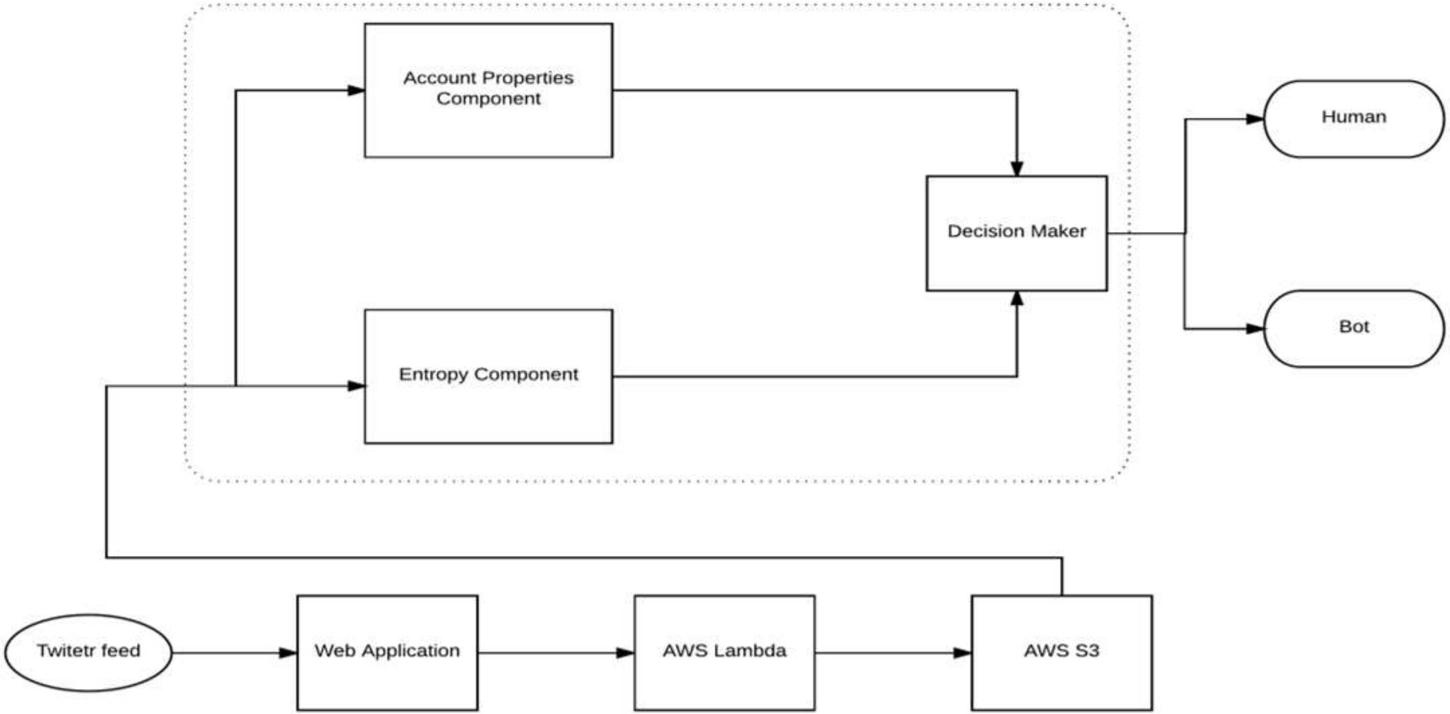
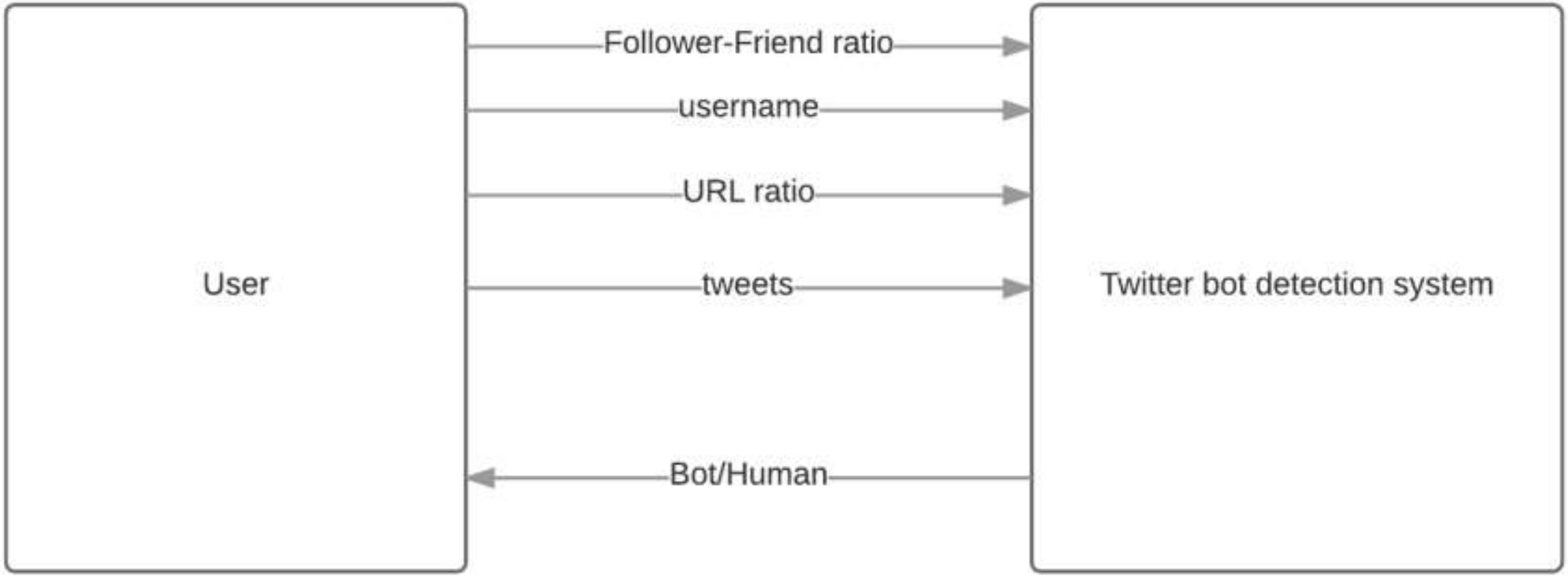
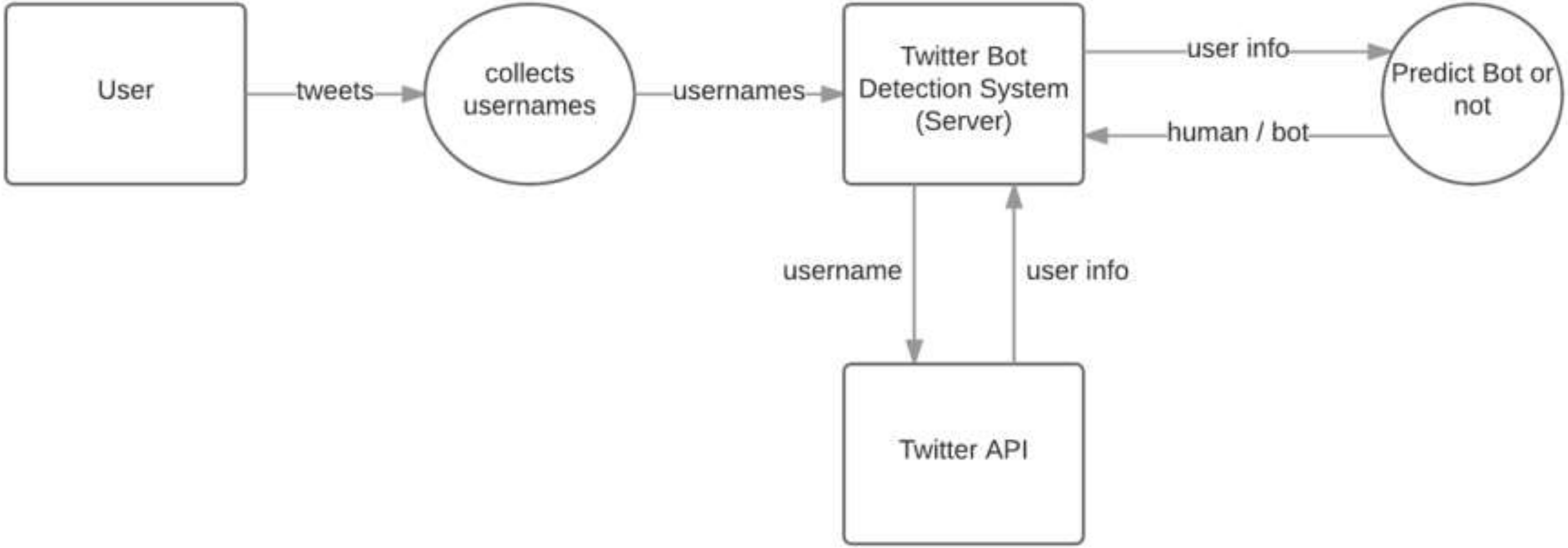


Fig: Architectural View of Deployment Phase

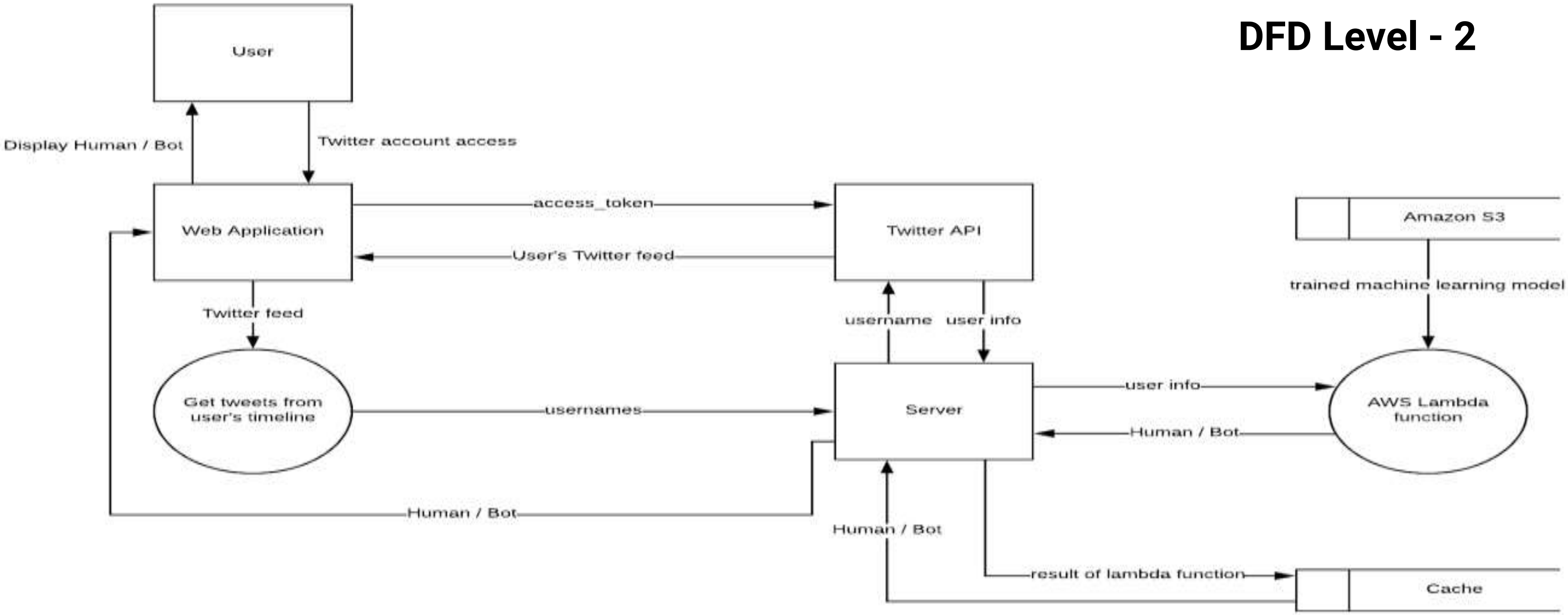
DFD Level - 0


User
Follower-Friend ratio
username
URL ratio
tweets
Bot/Human
Twitter bot detection system

DFD Level - 1


User
tweets
collects usernames
usernames
Twitter Bot Detection System (Server)
user info
Predict Bot or not
human / bot
username
user info
Twitter API

DFD Level - 2


User
Display Human / Bot
Twitter account access
Web Application
access_token
User's Twitter feed
Twitter API
Amazon S3
trained machine learning model
Twitter feed
username
user info
Get tweets from user's timeline
usernames
Server
user info
Human / Bot
AWS Lambda function
Human / Bot
Human / Bot
result of lambda function
Cache

Possible Problems and Suggestions

- **Problem**: Language of the tweet content can cause errors since our framework/model will tend to assign high bot scores to users who tweet in multiple languages.

  Solution : Ignore language dependent features.

- **Problem**: Determining which machine learning algorithm to use for our work that will deliver accurate results with higher precision.

  Solution : Decided to go for Logistic Regression and Random Forest and determine which algorithm solves our problem.

Possible Problems and Suggestions

- **Problem**: The real-time detection of bots may take longer due to delay in communication with server and running the trained algorithm

  Solution: Cache or store the result of a username search in the database to avoid running the algorithm for that username again

- **Problem** : Deploying trained machine learning model on server

  Solution : Using AWS Lambda and S3 for deployment and API gateway for sending the request and receiving the response.

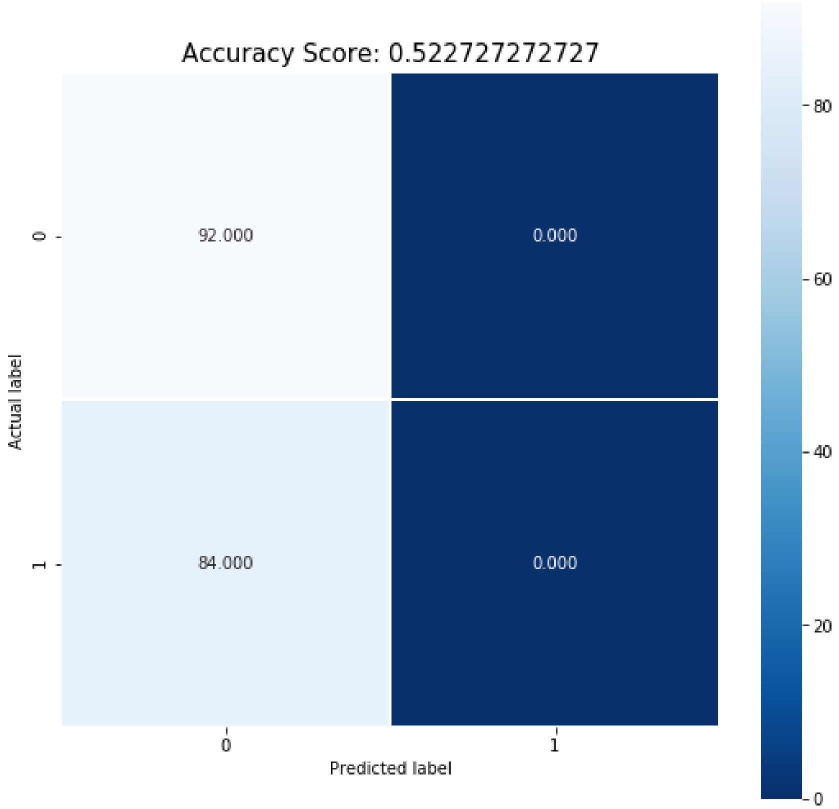
Technologies

Algorithms
Libraries
Tools

- Python
- Scikit-learn for machine learning algorithms
- AWS (Lambda, API Gateway, S3)
- Seaborn for graphing and visualization.
- MongoDB (Database)
- Python libraries like pickle
- Twitter API
- Redis (Cache Store)
- Nodejs (Back End)
- HTML, CSS, JS, JQuery ,Bootstrap, AJAX (FrontEnd)

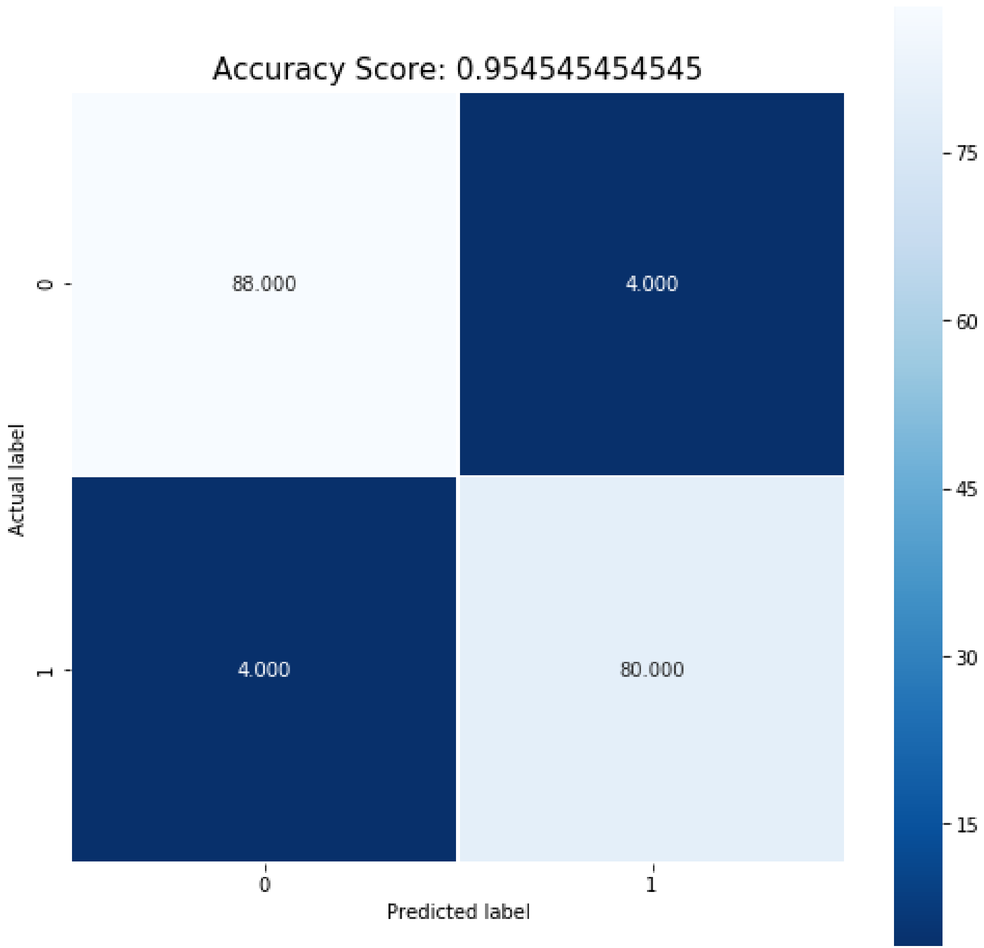
Implementation Results

	Precision	Recall	F1-score	Support
0 (human)	0.52	1.00	0.69	92
1 (bot)	0.00	0.00	0.00	84
avg/ total	0.27	0.52	0.36	176



Logistic Regression

	Precision	Recall	F1-score	Support
0 (human)	0.96	0.96	0.96	92
1 (bot)	0.95	0.95	0.95	84
avg/ total	0.95	0.95	0.95	176



Random Forest

Conclusion

Logistic regression underperforms even though it is known for its binary classification, and the reason for that its inflexibility to capture complex relationships and also tends to underperform when there are non-linear decision boundaries.Also, logistic regression are susceptible to outliers.

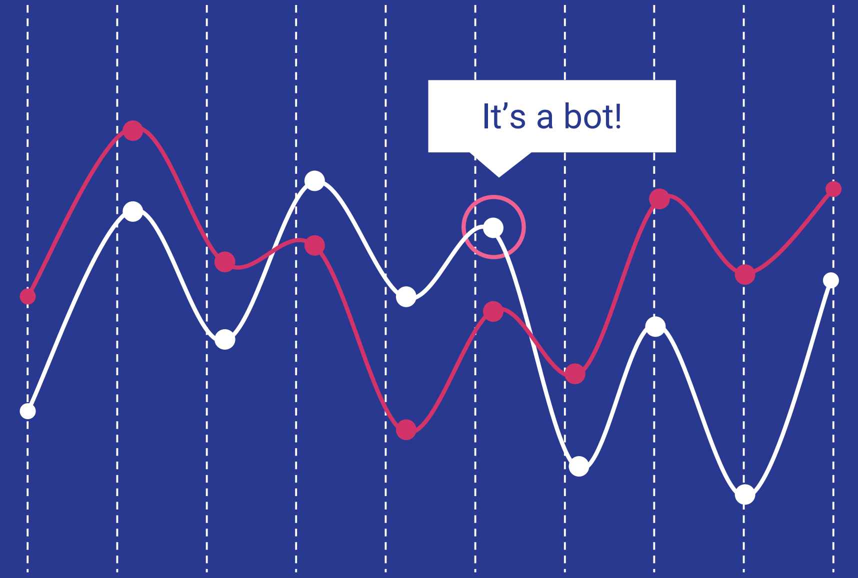
It must be noted that, in some cases, the boundary separating the bot and human is not sharp [3] and for logistic regression to perform its best the data points MUST be separable into two aforementioned regions by a linear boundary.

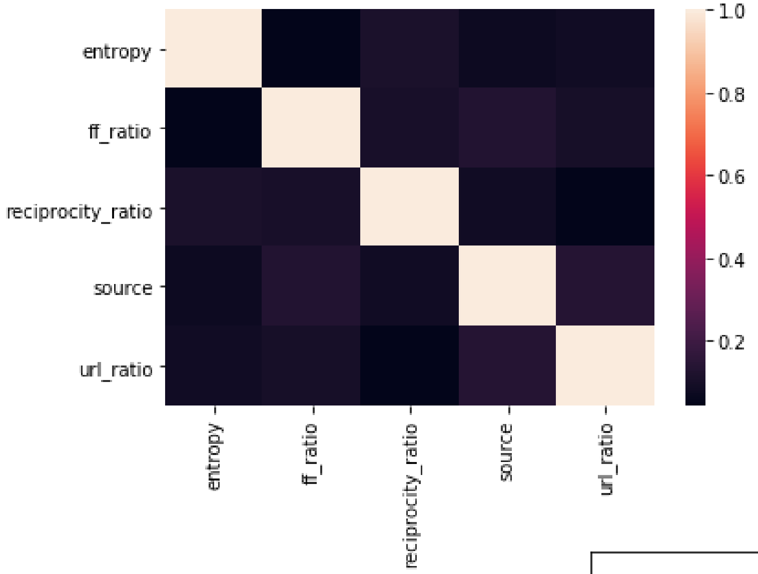
We can see that, Random Forest is one of the most effective and versatile machine learning algorithm and has higher classification accuracy (0.95)
The machine learning model will be trained using Random Forest algorithms to classify whether the given user is a bot or a human.

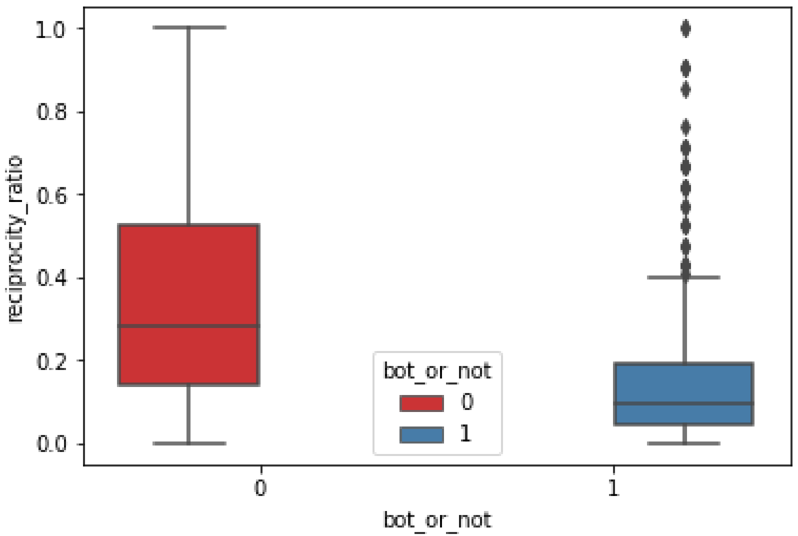
Visualization

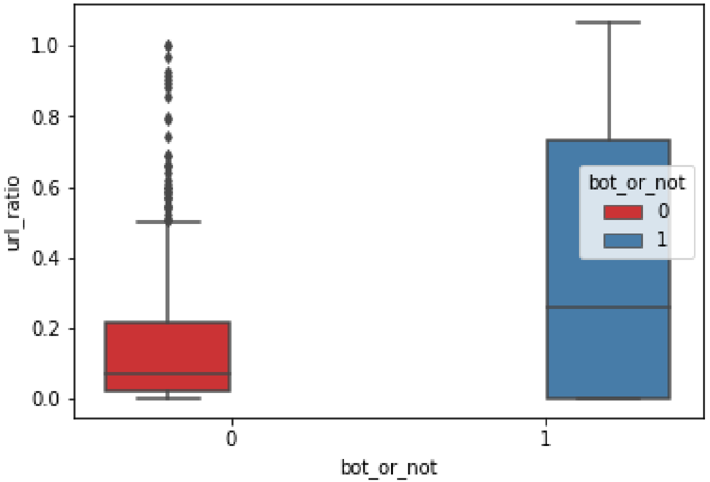
Analysis Results – Random Forest

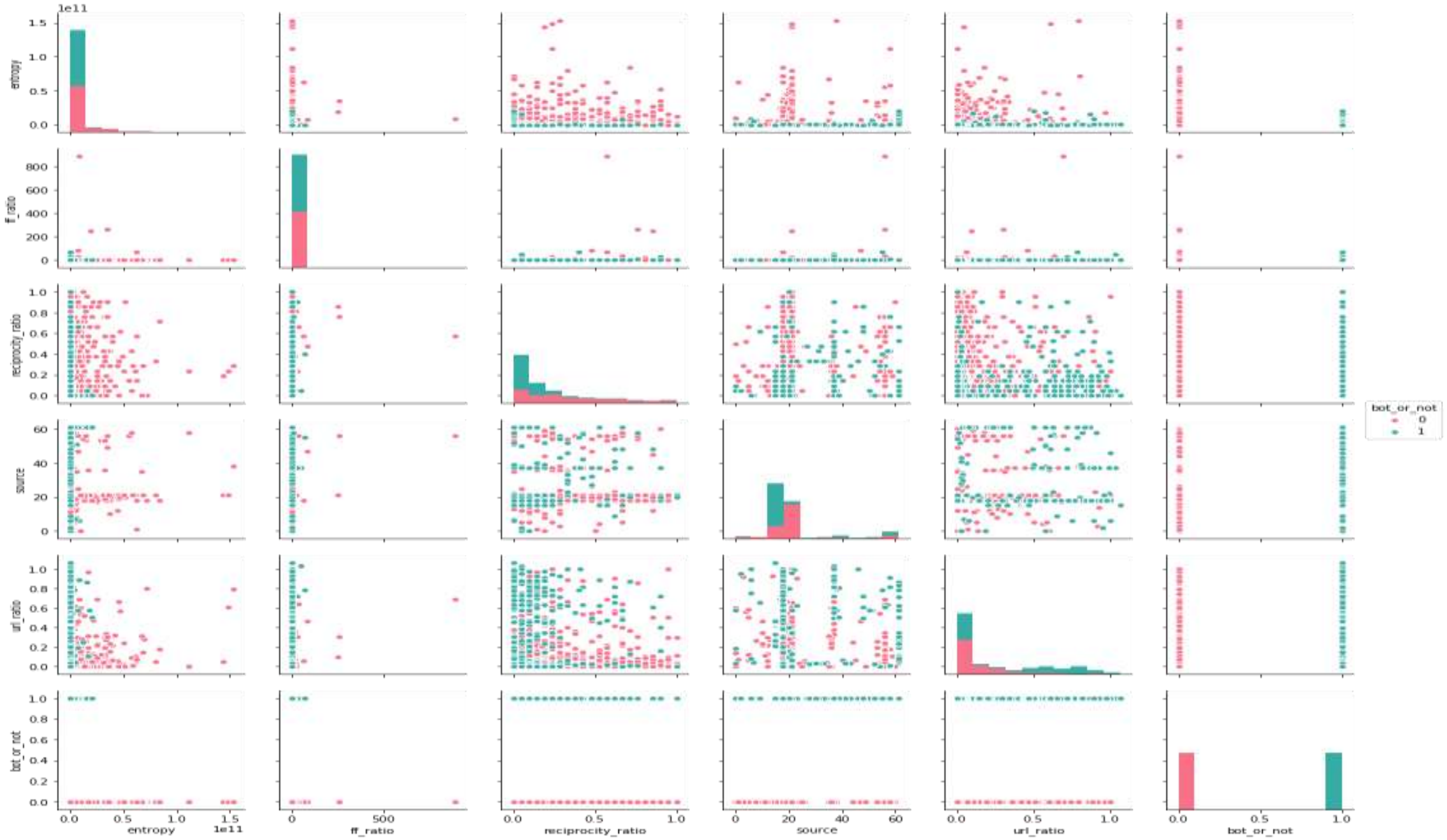
(Heat Map, Box Plot, Pair Plot)




entropy
ff_ratio
reciprocity_ratio
source
url_ratio
1.0
0.8
0.6
0.4
0.2


reciprocity_ratio
1.0
0.8
0.6
0.4
0.2
0.0
bot_or_not
0
1
0
1
bot_or_not


url_ratio
1.0
0.8
0.6
0.4
0.2
0.0
bot_or_not
0
1
0
1
bot_or_not


entropy
ff_ratio
reciprocity_ratio
source
url_ratio
bot_or_not
1e11
bot_or_not
0
1

Handcrafted APIs

1. GET /api/classifyUserName

1. POST /api/extractUserData

1. GET /api/fetchTweets

Applications / Future Work

1. Our model/framework will be able to identify whether a twitter user is a bot or a human.

1. We can extend our work to other social media platforms like facebook,etc.

1. Our work will safeguard oneself and an organization from false information, malicious contents and ensure their brand value.

1. Our project work can also be utilized to sort/identify human online traffic from bot activity.

References

[1] Title: Who is Tweeting on Twitter: Human, Bot, or Cyborg?

Authors: Zi Chu, Steven Gianvecchio, Haining Wang and Sushil Jajodia

URL: http://www.pensivepuffin.com/dwmcphd/syllabi/hcde530_wi17/twitter_readings/chu-who.tweets.ACSAC2010.pdf

[2] Title: A New Approach to Bot Detection: Striking the Balance Between Precision and Recall

Authors: Fred Morstatter, Liang Wu, Tahora H. Nazer, Kathleen M. Carley and Huan Liu

URL: http://www.public.asu.edu/~fmorstat/paperpdfs/asonam16.pdf

[3] Title: Online Human-Bot Interactions: Detection, Estimation, and Characterization

Authors: Onur Varol, Emilio Ferrara, Clayton A.Davis, Filippo Menczer, Alessandro Flammini

URL: https://arxiv.org/pdf/1703.03107.pdf

[4] Title: Using AWS for deploying trained model

Author: Jose Miguel Arrieta

URL: https://datascience.com.co/creating-an-api-using-scikit-learn-aws-lambda-s3-and-amazon-api-gateway-d9d10317e38d

Thank You