Behavioural Similarities Analysis and Detection of Bot Accounts in Twitter

Divyanshi Bhojak, Tarushi Jat

***Abstract*— As technology is getting more advanced every single day for making human lives more convenient, we have defrauders who exploit these tools for their personal or professional benefits. We have identified one such problem where an individual or group of defrauders tries to manipulate people's opinion and decision on social media by posting such content in bulk that satisfies their goal but can harm the general public to meet their personal benefit with the help of spam bots. Therefore, in this paper we have worked on analyzing the similarities and differences in the behaviour of spam bots and legitimate accounts taking the social media platform, Twitter as the base. Furthermore, we have also worked on detection of spam bot accounts and legitimate accounts in twitter using machine learning algorithms.**

Keywords: **spam-bot accounts, legitimate accounts, bot detection, twitter**

I. INTRODUCTION

In past few years, we have seen that social media has become so powerful such that it can manipulate anything happening in the world, either positively or negatively. We have seen several cases in the past where social media triggered even the biggest decisions of politics, share market, an individual's life and many more. All these are possible just by people posting their opinions on social media platforms. Twitter is one such very powerful social media platform where people post tweets to put their perspective in front of the world. But, the platform has been exploited many times in the past to manipulate people's decisions with the help of spam bots. Spam bots are automatic programs that are designed with an aim to post spam contents in large amounts to spread misinformation and propaganda style messages to favour some group or an individual. It was estimated in the research by Carnegie Mellon[1] University that bots are involved in upto 20% of the conversations on social media, especially pertaining to elections and other political issues. During US elections in 2016, it was found out that 13,512 russian linked accounts posted automatically generated election related contents favoring Donald Trump in the elections. There are many more cases in everyday life about spam bots posting misleading information to promote certain products, ecommerce advertisements and many more that can harm our society.

Thus, in this paper we have worked on finding out to what extent the actions or behaviour of spambot accounts are similar with the actions or behaviour of legitimate accounts. To capture the actions of the accounts in twitter, we have utilized the concept of Digital DNA Sequence as described in [1]. We have considered several parameters like follower ratio, the number of accounts posted tweets containing matching strings after capturing the tweets with the help of Digital DNA Sequence, etc to analyze the similarities between the two. Furthermore, We have applied bot detection model using various machine learning algorithms to detect an account to be a spam-bot account or a legitimate account. In short, our paper has two main objectives as described below:

i) to analyze the behavioural similarities between bot users and legitimate users in twitter

ii) to implement a bot detection model that will classify an account to be a bot account or a legitimate account.

II. LITERATURE SURVEY

As in recent times we are facing a huge pandemic due to Covid, and in many of the places there is partially or fully lockdown applied. Due to lockdown many people are suffering so some of them want relaxation in the terms. To achieve this Social media serves as a better platform to push the government. In Recent article [1] they have lifted the situation on twitter # trends and many tweets in which nearly half of the accounts were spam bots. So this is a big issue we are facing that needs to be under control. There are various researches that have been done and some ongoing research to detect spambot in social media platforms. In this paper [2] they have designed some techniques as an automated supervised machine learning solution which utilises web navigation behaviour to detect web spambots. They have given weightage to user behaviour to differentiate between web bot or legitimate content. G. Qingqing and S. Torsten have proposed in there paper [3] a different method to detect spam contents, they have divided their work in two different parts one is a classifier to catch a large portion of spam in our data, and the other is design several heuristics to decide if a node should be relabeled based on the pre classified result and knowledge about the neighborhood. In [4] they have used five different supervised learning methods to accomplish their task of statistical filtering, vector machine, AdaBoost, and maximum entropy model, Naive Bayes classifier, and aggressive feature pruning. In [5] authors have proposed a different way to model the behaviours of the users by digital DNA sequencing which performs better as a state of the art model. In the paper [6] they have proposed a supporting technique they have first analyzed how spammers overpass the detection system, and secondly they have utilized 24 features that are helping in detecting spam accounts. Out of these thesis and works We have incorporated digital DNA sequencing in our paper.

III. Methodology

A. Dataset Description

The dataset taken for the implementation task contains the over 9.5 millions tweets and user details, broadly divided into four major categories of over 3.5 thousands legitimate users and over 4.9 thousands of three groups of twitter bot users that are Bot User Group 1, Bot User Group 2 and Bot User Group 3. Data contains every action for all the group of twitter users such as tweet, retweet, reply. Legitimate users contain over 3.4 thousand twitter accounts of genuine users. Bot User Group 1 comprises 991 twitter accounts of political background community to influence or tarnish the views of people or influence online public opinion by those in power. Bot User Group 2 comprises 3457 twitter accounts for product promotions by enabling automated bots for brand promotions, users and product endorsements. Bot User Group 3 comprises 464 twitter accounts for e-commerce advertisements by driving traffic, advertising and making the brand visible to the masses.

B. Behavioural Similarities Analysis

With the help of behavioural similarities analysis, we aim to find out how different groups of bot accounts are similar or different with that of legitimate accounts group. We have done several analysis which are described below:

1. Frequency Statistics on the Distribution of Followers and Friends

The distribution of the frequency of friends and followers for each of the group of users from our dataset is calculated on per account basis. The Fig. 1 shows the frequency distribution for followers of bot users from group 1 and Fig. 2 shows its frequency distribution for friends. The Fig. 3 shows the frequency distribution for followers of bot users from group 2 and Fig. 4 shows its frequency distribution for friends. The Fig. 5 shows the frequency distribution for followers of bot users from group 3 and Fig. 6 shows its frequency distribution for friends. The Fig. 7 shows the frequency distribution for followers of legitimate users and Fig. 8 shows its frequency distribution for friends.

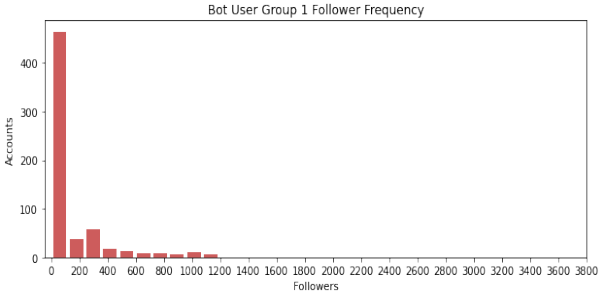

Fig. 1: Follower Frequency - Bot Users Group 1.

One thing that can be observed from the frequency distribution of friends and followers is that almost all of the bot accounts from all the three groups have a limited

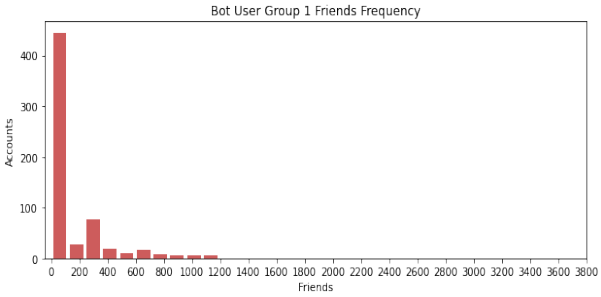

Fig. 2: Friend Frequency - Bot Users Group 1.

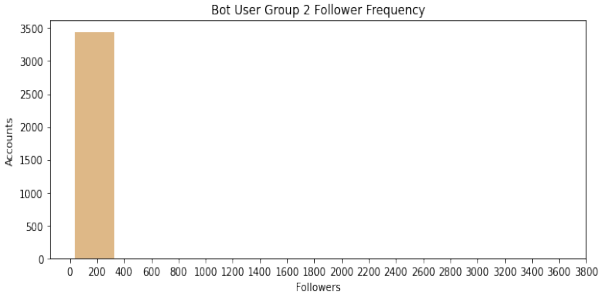

Fig. 3: Follower Frequency - Bot Users Group 2.

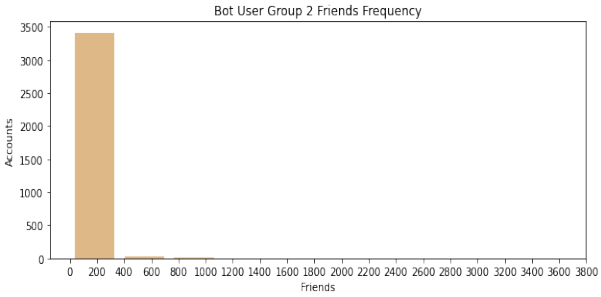

Fig. 4: Friend Frequency - Bot Users Group 2.

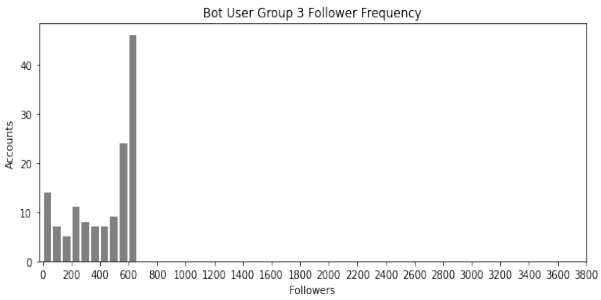

Fig. 5: Follower Frequency - Bot Users Group 3.

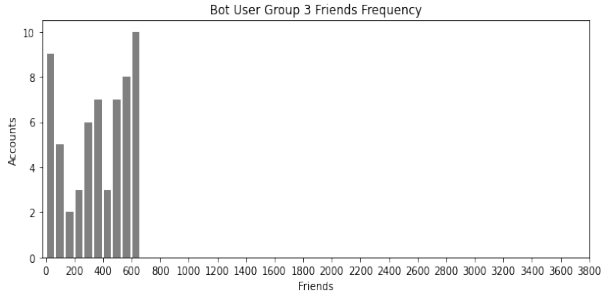

Fig. 6: Friend Frequency - Bot Users Group 3.

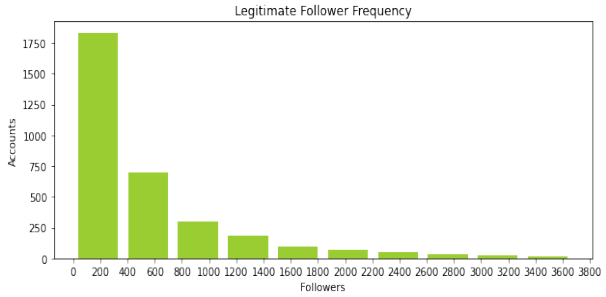

Fig. 7: Follower Frequency - Legitimate Users.

number of friends and followers relative to that of legitimate users' followers and friends. Whereas, in the frequency distribution of followers and friends of legitimate users, it can be seen that almost half of the legitimate users have small amounts of followers as well as friends and the remaining half of the legitimate users have relatively high numbers of followers and friends. To conclude more from the number of followers and friends we have calculated the follower ratio.

2. Follower Ratio

Follower ratio is defined as the number of followers that a user have relative to the number of accounts the user is following. If the number of accounts the user is following is very high than the number of accounts following the user, then the follower ratio will be relatively small and the chances of account to be a bot account will be high. We have computed the follower ratio of all four groups of twitter accounts and represented in Table I.

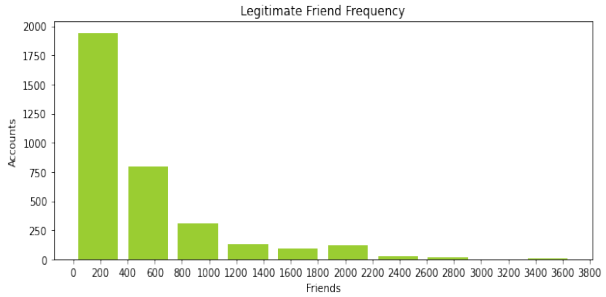

Fig. 8: Friend Frequency - Legitimate Users.

TABLE I: Follower Ratio

Account Type	Follower Ratio
Legitimate Account	0.5027722347999382
Bot Account Group - 1	0.5237114689719357
Bot Account Group - 2	0.11543068901989084
Bot Account Group - 3	0.3778966696976445

Follower ratio of bot accounts that belong to group 2 and group 3 is relatively small when compared to follower ratio of accounts that belong to legitimate user groups. On the other side, bot accounts that belong to group 1 shows similar behaviour to that of legitimate account users group in terms of follower ratio.

3. Digital DNA Sequence

To analyze the behavioural similarities between the bot accounts and the legitimate users accounts, we first need to capture the actions made by them on twitter. To capture the actions of these accounts, we are utilizing the concept of Digital DNA Sequence which is a row vector of characters that represents the user's behaviour by encoding each action of the user with a certain base technique. We have used three base techniques to capture the actions of the accounts and all the three techniques are aimed to capture different kinds of actions or behaviour of accounts. The three techniques are described as follows:

i) B3 type - Encoding with the help of this base technique is aimed to capture the type of the tweet posted on twitter by the account. This encoding technique captures three types of tweets and these three types are tweet, reply and retweet. For this encoding scheme, we utilize the three alphabets which are A, C and T to capture the type of the tweet. The alphabet A represents the tweet, the alphabet C represents the reply and the alphabet T represents the retweet.

ii) B3 content - Encoding with the help of this base technique is aimed to capture the tweet based on the content posted on twitter by the account, rather than capturing the tweet type as in B3 type encoding technique. This encoding technique captures the tweets based on three types of content which are present in tweets. For this encoding scheme, we utilize three alphabets which are N, E and X to capture the tweet content. Tweets that contains zero entities (i.e.,only plain text) are encoded with the alphabet N, tweets that contains one or more entities but of the same type are encoded with the alphabet E and tweets that contain one or more entities but of mixed types are encoded with the help of alphabet X.

iii) B6 content - Encoding with the help of this base technique is also aimed to capture the tweet based on the content posted on twitter by the account, rather than capturing the tweet type as in B3 type encoding technique.

This encoding technique captures the tweets based on six types of content which are present in tweets. For this encoding scheme, we utilize six alphabets which are N, U, H, M, D and X to capture the tweet content. Tweets that contains zero entities (i.e.,only plain text) are encoded with the alphabet N, to encode the tweets that contains one or more URLs we use the alphabet U, tweets that contains one or more hashtags are encoded with the alphabet H, tweets that contains one or more mentions are encoded with the help of alphabet M, tweets that contains one or medias (images, videos, etc.) are encoded with the alphabet D, and tweets that contain one or more entities but of mixed types are encoded with the help of alphabet X.

Following are the analysis for bot accounts and legitimate accounts based on base encoding technique B3 type. We have plotted the string length on x-axis and number of strings on y-axis for all four groups of users. The graph for legitimate user group is shown in Fig. 2, and we can see that nearly half of the legitimate users have very high number of interaction including tweets, retweets and replies whereas the other half accounts of legitimate users show a moderate level of interaction. This can be because the legitimate accounts of well known personalities usually have high number of interactions than the legitimate accounts of the normal user. Graph for bot accounts that belong to group 1 is shown in Fig. 3, graph for bot accounts that belong to group 2 is shown in Fig. 4 and graph for bot accounts that belong to group 3 is shown in Fig. 5. It can be seen in graph of bot accounts group that most of the bot accounts have high interaction, while few of them are just inactive. In case of bot accounts that belong to group 2 have much similarity with legitimate users accounts.

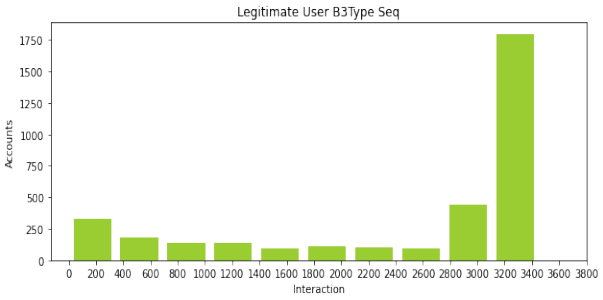

Fig. 9: Legitimate User B3 Type Sequence.

4. Longest Common Subsequence Curve

In order to observe the behavioural patterns between groups of users, we need to process our encoded sequence that we captured using the three Digital DNA encoding schemes. One of the possible paths to find out the similarities between our DNA sequences is to use the longest common substring. Suppose, we have two strings given, Si with length m and Sj with length n. Now the longest common substring i.e., LCS between them will be the largest string that is the subset of both the string Si and Sj. For example, consider the two strings Si = STOCK and Sj = SHOCK, so their

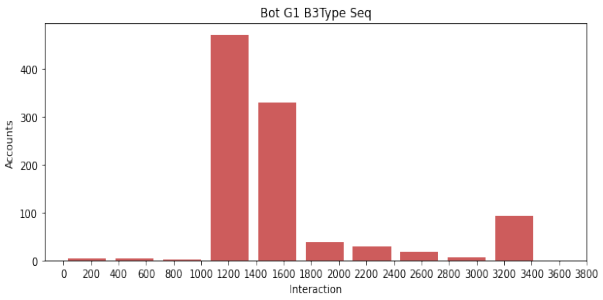

Fig. 10: Bot User Group-1 B3 Type Sequence.

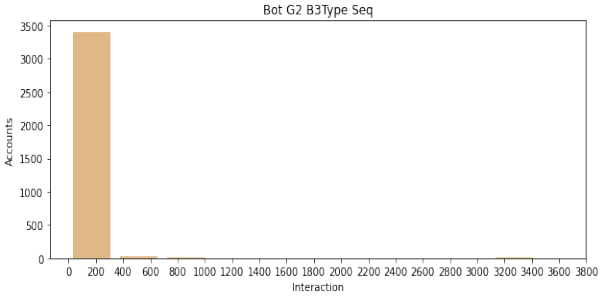

Fig. 11: Bot User Group-2 B3 Type Sequence.

LCS = OCK and its length is 3. This is the idea to find out the similarities between the DNA sequences using longest common subsequence. In our case, we will be considering k strings at a time and thus the time complexity of LCS would be O(k*(m+n)). It is possible for us to plot LCS curve by considering the relationship between the length of the LCS and the number of strings. This LCS curve will be the representation of the behavioural similarities among the group of twitter users, since LCS is an ordered sequence of substring lengths. The LCS curves that we have plotted for each group of users using B3 type encoding technique is represented in Fig. 13 for legitimate users, LCS curve for bot users group 1 is shown in Fig. 14, LCS curve for bot users group 2 is shown in Fig. 15 and LCS curve for bot users from group 3 is shown in Fig. 16.

Now, if we can find out one very single measure for the similarity analysis for the entire group of users then it will be helpful in interpreting the LCS curves. Therefore, the computation of area under the LCS curve (AUC) is used

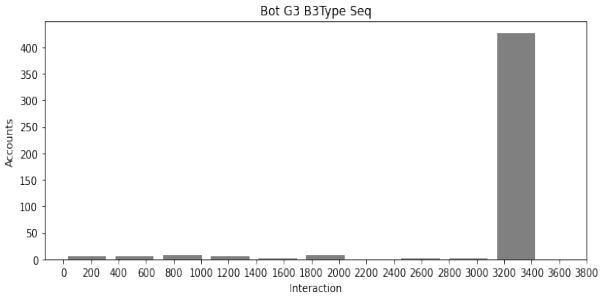

Fig. 12: Bot User Group-3 B3 Type Sequence.

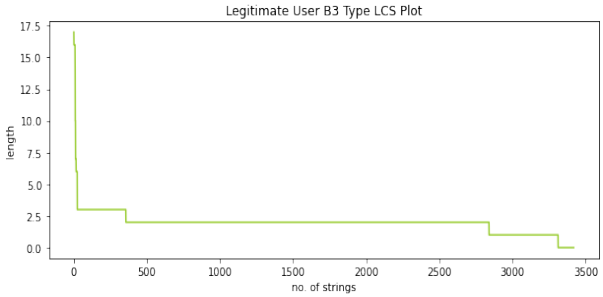

Fig. 13: LCS Curve - Legitimate Users.

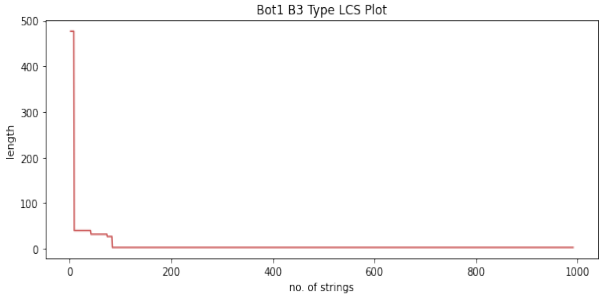

Fig. 14: LCS Curve - Bot Users Group 1.

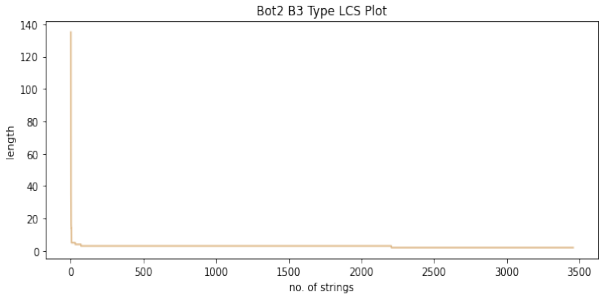

Fig. 15: LCS Curve - Bot Users Group 2.

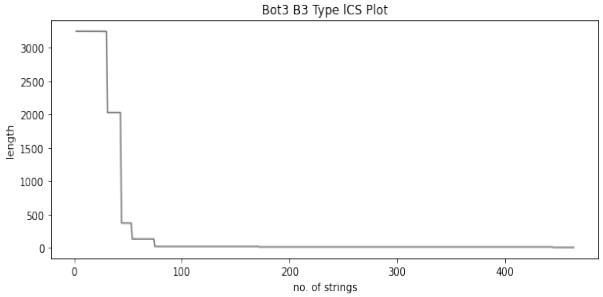

Fig. 16: LCS Curve - Bot Users Group 3.

here to perform the analysis.

We have calculated the area under the LCS curves, as shown in Table II, for all the four groups of users data that we have after processing the data with the help of all the three digital DNA sequence encoding techniques that was described above.

TABLE II: Area under LCS Curves

Account Type	Encoding Scheme	AUC
Legitimate Account	B3 Type	6674.83
Legitimate Account	B3 Content	41.02
Legitimate Account	B6 Content	41.02
Bot Account Group - 1	B3 Type	8316.76
Bot Account Group - 1	B3 Content	1138.01
Bot Account Group - 1	B6 Content	211.78
Bot Account Group - 2	B3 Type	9405.27
Bot Account Group - 2	B3 Content	3938.20
Bot Account Group - 2	B6 Content	17.82
Bot Account Group - 3	B3 Type	129978.02
Bot Account Group - 3	B3 Content	620.97
Bot Account Group - 3	B6 Content	620.97

A bigger value of AUC describes that there exists high similarity between the group of users, on the other hand a smaller value of AUC describes that there is low similarity between the group of users. Now after calculating the area under the LCS curve for each group of accounts and using each type of encoding technique, we have calculated the average area under the LCS curve which is shown in Table. III. In these results as shown in table III, we can observe that the accounts that belong to the legitimate user group have the least average AUC value. On the other hand, all the bot groups accounts have higher AUC. Therefore, it can be concluded here that accounts operated by humans have more differences in their behaviour than similarity whereas accounts that are operated by bots show high similarity in their behaviour.

TABLE III: Average AUC

Account Type	Average AUC
Legitimate Account	2252.29
Bot Account Group - 1	4453.76
Bot Account Group - 2	43739.99
Bot Account Group - 3	3222.18

IV. Experimental Analysis and Results

A. Data Visualization

For visualizing the data counts of user accounts with respect to each group. We plotted the histogram of account type of each group against the number of accounts. The plotting provides us insights of the strength of accounts present in each subsequent twitter group. Users denote the legitimate users and Bot-1, Bot-2 and Bot-3 the three automated bot groups. Fig. 5 is the plotted histogram where X-axis denote account type and y-axis denote number of accounts.

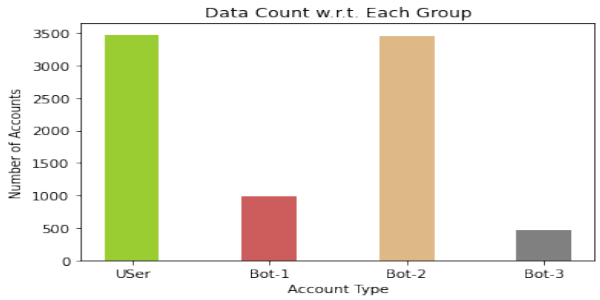

Fig. 17: Count of Data of Each Group.

B. *Data Pre-processing*

After the visualization of the dataset, we have processed the dataset for later implementation analysis. Then Data is pre-processed removing NULL, wrong and empty data values, then cleaning the data and normalizing it to feed into the various machine learning classifiers. The dataset is splitted into train and test data in a ratio of 80:20. Training data of 80% and test data of 20%. And lastly after this will be used for the implementation task.

C. *Implementation*

After getting the pre-processed data and deep insights on the behaviours similarities on account type, we will classify whether an account is spam or legitimate on evaluating various machine learning classifiers. Below are the description of each classifiers used for the task:

1. Decision Tree:
Decision trees are supervised algorithms where based on specific parameter data is continuously split to make predictions. After experimenting with decision trees we got a train accuracy of 88.51% and test accuracy of 86.25%. Below are the confusion matrix and ROC curve of the model.

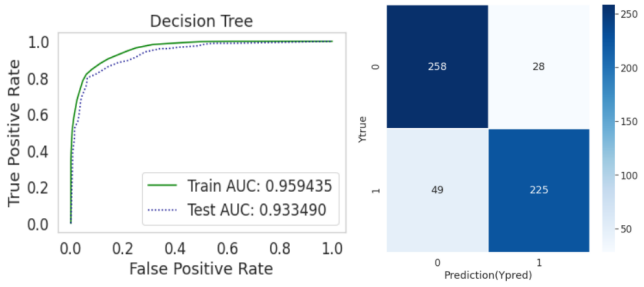

Fig. 18: Decision Tree Model

2.: XGBoost Model
eXtreme Gradient Boost is a decision-tree based algorithm that uses a boosting framework for prediction. After experimenting with XGBoost classifier we got a train accuracy of 87.78% and test accuracy of 86.31%. Below are the confusion matrix and ROC curve of the model.

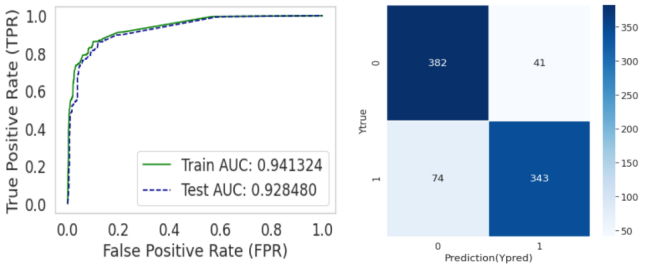

Fig. 19: XGBoost Model

3. Multinomial Naive Bayes Model
Multinomial Naive Bayes is a probabilistic learning paradigm that works on discrete multinomial distributions rather than on other distributions. After experimenting with multinomial classifier we got a train accuracy of 68.44% and test accuracy of 68.75%. Below are the confusion matrix and ROC curve of the model.

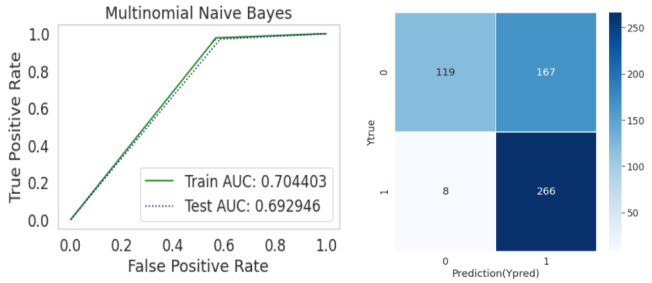

Fig. 20: Multinomial Naive Bayes Model

4. Bernoulli Naive Bayes Model
Bernoulli Naive Bayes is based on Bayes theorem works on Bernoulli distribution and used for discrete data. After experimenting with bernoulli NB classifier we got a train accuracy of 68.01% and test accuracy of 69.64%. Below are the confusion matrix and ROC curve of the model.

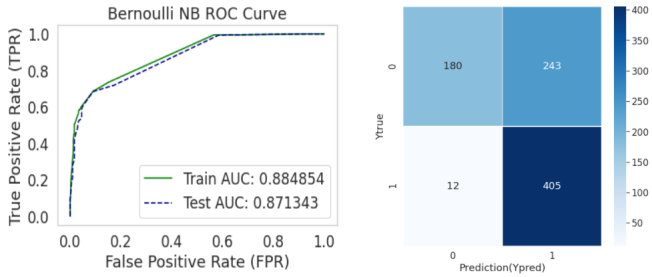

Fig. 21: Bernoulli Naive Bayes Model

5. Bagging Model
Bagging ensemble technique used bootstrap sampling of the training data to improve the estimation of one by combining others. After experimenting with bagging classifier we got a

train accuracy of 88.51% and test accuracy of 88.25%. Below are the confusion matrix and ROC curve of the model.

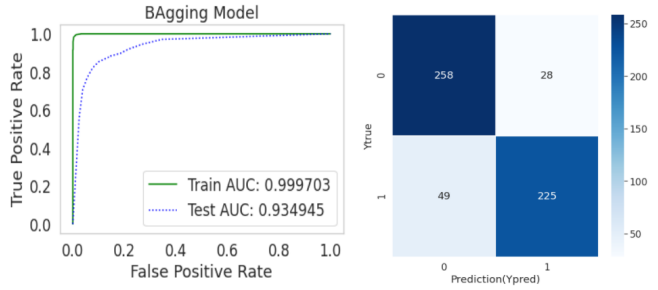

Fig. 22: Bagging Model

TABLE IV: Classification Report of Classifiers

Model	Accuracy	ROC Score	Precision	Recall
Decision tree	0.8625	0.86	0.86	0.86
XGBoost	0.8630	0.862	0.87	0.86
Multinomial NB	0.687	0.6934	0.78	0.69
Bernoulli NB	0.6964	0.6983	0.78	0.70
Bagging	0.8625	0.8616	0.86	0.86
Voting Classifier	0.88571	0.8757	0.89	0.89

6. Ensemble Technique

Ensemble Technique combines several base models in order to produce one optimal predictive model and for this we used Hard Majority voting classifier to employ all the above mentioned classifiers to make predictions. Hard Voting Classifier in which each and every classifier votes for a class label and class label with the majority of votes wins. In our case, if decision tree, multinomial naive bayes and bernoulli naive bayes voted for legitimate account type and XGBoost and bagging model voted for spambot account type, so as majority classifier voted for spambot account type so it will predict type as spambot. After experimenting with decision trees we got a train accuracy of 93.47% and test accuracy of 88.57%.

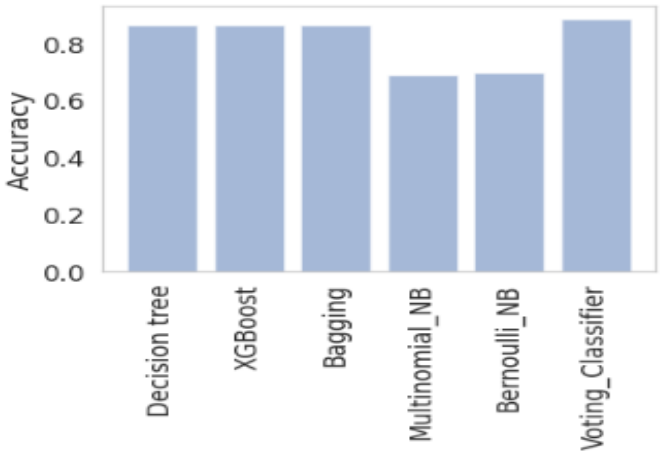

Fig. 23: Voting Classifier.

CONCLUSION

One of the foremost problems in social media platforms like Twitter is the large number of social or automated bots which are controlled by automated agents, generally used for malicious activities, political benefits, and commercial advertisements. In this project, we seek deep insights in the behavioral properties and similarities on the dataset we used for the experiments. We generated a digital DNA sequence from the various user twitter actions and then computed the similarities of the same using the longest common substring of the generated DNA sequence. Consequently, twitter bot detection is of great importance and we addressed this problem by classifying this using various machine learning algorithms and lastly applied ensemble techniques using majority voting classifier where voting classifier gave the highest accuracy of 88.57%. We observed fundamental properties of behaviours of online users and despite human behaviours being heterogeneous it is not at all uniformly random.

Further future work requires optimization and improvement in bots classification techniques. In digital DNA sequence generation, we can use B3 Interaction and B6 Interaction base encoding techniques that generates the DNA sequences with respect to the popularity level of the peers with whom a given user interacts We can also use B3 account-type encoding technique to generate digital DNA sequences that represents the age of the accounts.

REFERENCES

[1] Stefano Cresci Et al, Emergent properties, models, and laws of behavioral similarities within groups of twitter users, published in 2019.

[2] R. Agrawal, R. Srikant, Mining sequential patterns, in: ICDE'95, IEEE, 3–14, 1995.

[3] M. Arnold, E. Ohlebusch, Linear time algorithms for generalizations of the longest common substring problem, Algorith-mica 60 (4) (2011) 806–818.

[4] A. Bessi, Towards the Modeling of Behavioral Trajectories of Users in Online Social Media, CoRR abs/1611.05778, URLhttp://arxiv.org/abs/1611.05778.

[5] A.-L. Barab asi, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[6] A. Cali'o, R. Interdonato, C. Pulice, A. Tagarelli, Topology-driven Diversity for Targeted Influence Maximization withApplication to User Engagement in Social Networks, IEEE Transactions on Knowledge and Data Engineering.

[7] L. Chi, K. Hui, Color set size problem with applications to string matching, in: Combinatorial Pattern Matching, Springer,230–243, 1992.

[8] L. Cao, In-depth behavior understanding and use: the behavior informatics approach, Information Sciences 180 (17) (2010)3067–3085.

PLAGIARISM REPORT

Team 8 - Deepu, Divyanshi, Tarushi.pdf

ORIGINALITY REPORT

5%	4%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, Maurizio Tesconi. "Emergent properties, models, and laws of behavioral similarities within groups of twitter users", Computer Communications, 2020 Publication	1%
2	academic.odysci.com Internet Source	1%
3	www.iit.cnr.it Internet Source	1%
4	www.forbes.com Internet Source	1%
5	Submitted to University of Sussex Student Paper	<1%
6	espace.curtin.edu.au Internet Source	<1%
7	Submitted to City University Student Paper	<1%