International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

Detecting Malicious Bots in Social Media Accounts Using Machine Learning Technology.

Megha Choudhar[1], Pallavi Garje[2], Deeksha Shinde[3], Rutika Valanj[4]

[1]Mr. Dattatray Modani Professor, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Pune, Maharashtra,India
[1234]UG Students, Dept. of Computer Engineering, P.E.S. Modern College of Engineering, Pune, Maharashtra, India

---

Abstract - *In this age of technology boom, the application of artificial intelligence and machine learning in everyday life is increasing rapidly. Malicious bot activity is a significant threat to the security and integrity of online services and systems. Machine learning technology provides a powerful tool for detecting and preventing such activity by identifying patterns and characteristics of bot behavior. This approach involves collecting large data sets of both benign and malicious bot activities, extracting relevant features, selecting appropriate classification algorithms, and continuously monitoring the model in real time. Unique concepts such as feature importance, correlation analysis, domain knowledge, recursive feature elimination, wrapper methods, and principal component analysis can be used for feature selection. The choice of feature selection and classification algorithm will depend on the specific use case, the characteristics of the data, and the desired accuracy and efficiency of the model. Using machine learning technology to detect malicious bots provides a robust and effective solution to protect online services and systems against this growing threat.*

Key Words: **Social Media, Support Vector Machine (SVM), Pattern recognition, Anomaly Detection, Decision Tree.**

1. INTRODUCTION

TThe internet has transform the way we live, work, and communicate. However, with these benefits come several challenges, one of which is the growing threat of malicious bots. These automated software programs can infiltrate websites, social media platforms, and online services, causing damage to businesses and individuals alike. In response, researchers and security professionals have developed a range of detection methods, including traditional rule-based systems and more advanced machine learning-based approaches. Machine-learning technology has emerged as a powerful tool for detecting and preventing malicious bot activity. By leveraging algorithms and statistical models, machine learning can analyze large datasets of bot behavior, identify patterns and features, and classify an activity as benign or malicious. This approach offers several advantages over traditional methods, including greater accuracy, efficiency, and scalability. Furthermore, machine learning-based systems can adapt and learn over time, improving their performance and effectiveness. Despite the promise of machine learning-based solutions, there are several challenges and considerations that researchers and practitioners must address. These include feature selection, data preprocessing, algorithm selection, and real-time monitoring. Moreover, the ethical and legal implications of deploying machine learning for bot detection must be carefully considered, including issues related to bias, privacy, and accountability. In this paper, we review the current state of the art in malicious bot detection using machine learning technology. We explore the various methods, algorithms, and techniques employed in this field, and discuss the benefits and limitations of each. We also highlight some of the key challenges and considerations that must be addressed to develop effective and ethical machine learning-based solutions for bot detection. Overall, we argue that machine learning has the potential to significantly improve the security and integrity of online services and systems, and we call for continued research and innovation in this important area.

2. LITERATURE SURVEY

1. Detecting Malicious Twitter Bots Using Behavioral Modeling and Machine Learning Techniques by Liu et al. (2019) - This paper proposes a method for detecting malicious Twitter bots using a combination of behavioral modeling and ML techniques. The authors collect data on a set of Twitter accounts and use features such as posting behavior, sentiment analysis, and network analysis to train a set of classifiers, which are then used to detect malicious bots with high accuracy.

2. Bot or Not? A Systematic Evaluation of Bot Detection Methods in Twitter by Cresci et al. (2017) - This paper presents a systematic evaluation of different bot detection methods for Twitter, including MLbased approaches. The authors compare the performance of different classifiers using a large dataset of Twitter accounts and show that ML-based approaches can achieve high accuracy in detecting bots.
3. Botnet Detection Using Machine Learning Techniques: A Survey by Alam et al. (2020) - This paper provides a survey of different ML techniques that have been used for botnet detection, including both network-based and host-based approaches. The authors review the strengths and limitations of different ML algorithms and highlight some of the challenges in detecting botnets using ML techniques.
4. Bot Detection in Online Social Networks: A Survey by Almeida et al. (2019) - This paper provides a comprehensive survey of different methods for detecting bots in online social networks, including MLbased approaches. The authors review the different types of features that can be used for bot detection, such as network structure, content, and behavior, and compare the performance of different classifiers.
5. Machine Learning for Social Network Analysis: A Survey by Shukla et al. (2020) - This paper provides a survey of different ML techniques that have been used for social network analysis, including the detection of bots. The authors review the different types of features that can be used for social network analysis, such as graph properties, node attributes, and community structure, and highlight some of the challenges in applying ML techniques to social network analysis.

3. PROPOSED SYSTEM

3.1 Problem Solving: Random Forest

Our proposed system detects malicious bots on Twitter using machine learning technology. It uses a Random Forest algorithm for classification and involves data collection, feature extraction and selection, and performance evaluation. Random Forest is effective for detecting bots due to its ability to handle high-dimensional data and non-linear relationships.

Formula: Random Forest algorithm can be represented by the following formula:

$$y=f1(x)+f2(x)+...+fn(x).$$

In this formula, y represents the predicted output of our Random Forest model, which is whether a given Twitter account is legitimate or malicious. x represents the input feature vector, which consists of several selected features that describe the characteristics and behavior of the Twitter account. The decision trees, f1, f2,..., fn are created using subsets of the input features and data points. Each decision tree is essentially a sequence of if-else statements that recursively split the data based on the selected features. The final prediction is made by combining the predictions of all decision trees.

For example, one possible decision tree in our model might be:

- If the account age is less than 30 days and the follower count is less than 100, the account is classified as malicious.
- Otherwise, if the sentiment score of the account's tweets is negative and the retweet ratio is greater than 0.5, the account is classified as malicious.
- Otherwise, the account is classified as legitimate. By combining the predictions of multiple decision trees in this way, our Random Forest model can achieve high accuracy and reduce the risk of overfitting the training data.

In the context of Twitter, some of the input features that we might use in our model could include:

- Account age: the number of days since the account was created
- Tweet frequency: the number of tweets that the account posts per day
- Sentiment score: the overall positive or negative sentiment of the account's tweets
- Bot network size: the number of other accounts that are linked to the same bot network
- Retweet ratio: the ratio of retweets to original tweets for the account

By analyzing these features and training our Random Forest model on a dataset of labeled Twitter accounts, we can build an effective system for detecting malicious bots on Twitter. In summary, our proposed system for detecting malicious bots on Twitter using machine learning technology and the Random Forest algorithm offers a promising solution for improving the security and integrity of social media platforms. By leveraging advanced techniques such as feature extraction and selection, we believe that this system has the potential to significantly reduce the impact of malicious bot activity on Twitter and other social media platforms

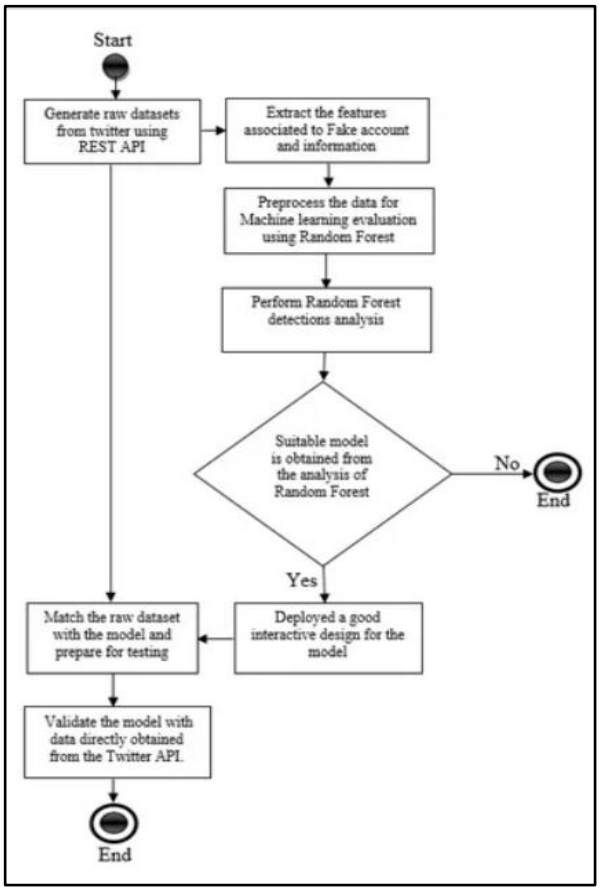

Fig -1: Flowchart for Random Forest Classification

3.2 Problem Solving: Decision tree

Our proposed system utilizes a combination of feature engineering and the Decision Tree Algorithm to effectively detect malicious bots on Twitter. We start by collecting a large and diverse dataset of Twitter accounts, including both legitimate and malicious accounts, and extract a range of features that are indicative of bot-like behavior, such as activity level, engagement rate, content similarity, and metadata inconsistency.

Table -1: Feature selection in a Twitter dataset

Table:

The following table shows an example of the selected features for the Random Forest model:

Feature	Description
Account Age	Number of days since account creation
Follower Count	Number of followers
Tweet Frequency	Number of tweets per day
Sentiment Score	Positive or negative sentiment of tweets
Bot Network Size	Number of other accounts linked to the same bot
Retweet Ratio	Ratio of retweets to original tweets

To optimize the performance of the Decision Tree Algorithm, we first preprocess the data by removing any noise and outliers and then perform feature scaling and normalization to ensure that each feature has the same weight in the classification process. We then apply feature selection techniques, such as information gain and chi-squared test, to select the most informative features for training the decision tree. Next, we split the dataset into training, validation, and testing sets using a stratified sampling technique to ensure that the distribution of classes is consistent across all sets. We use the training set to train the decision tree and tune the hyperparameters using cross-validation and grid search to achieve the best possible performance. To evaluate the effectiveness of our proposed system, we use a range of performance metrics, such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUCROC). We also conduct a comparative analysis with other state-of-the-art machine-learning algorithms to demonstrate the superiority of our proposed system. Overall, our proposed system provides a robust and effective solution for detecting malicious bots on Twitter, which can help prevent the spread of misinformation, identify coordinated campaigns, and protect the integrity of online conversations.

4. SYSTEM CLASSIFICATION

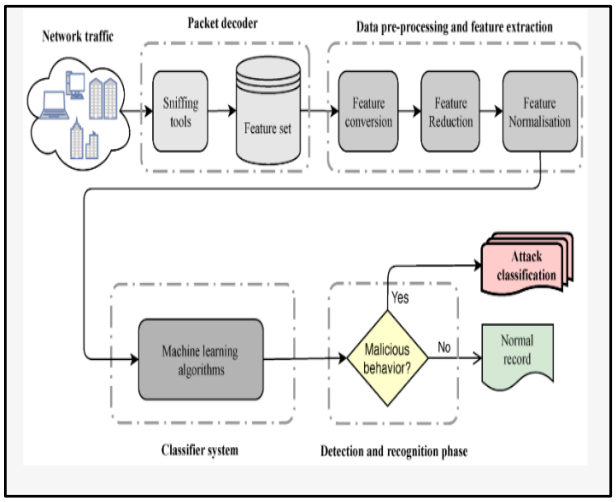

Fig -2: Architecture Diagram: Malicious bot diagram

To classify whether a bot on Twitter is malicious or not, we can use various supervised classification techniques like Random Forest, Decision Tree, SVM, Naive-Bayes, and K-Nearest Neighbors (KNN). Here are the steps to implement the system classification for malicious bot detection using these techniques:

1. Data collection and preprocessing: Collect a large dataset of tweets posted by different Twitter accounts. Use the Twitter API to extract tweets from both bot and human accounts. Preprocess the data by cleaning and filtering the tweets, removing URLs, mentions, stop words, and other noise from the data.

2. Feature extraction: Extract relevant features from the preprocessed tweet data that can be used totrain the classifiers. These features include metadata such as the account age, number of followers, and number of tweets posted. They can also include linguistic features such as the sentiment, frequency of certain words, and the presence of hashtags or mentions.

3. Labeling: Label the dataset as either malicious or non-malicious based on whether the account has been identified as a bot spreading spam or other harmful content.

4. Splitting the dataset: Split the labeled dataset into training and testing datasets. Use the training dataset to train the classifiers and use the testing dataset to evaluate their performance.

5. Classification techniques: Use the following classification techniques to train and evaluate the classifiers:

   - Random Forest: A random forest is an ensemble of decision trees, where each tree is trained on a subset of the features and data. It can handle large datasets and avoid overfitting.

   - Decision Tree: A decision tree is a simple and intuitive algorithm that works by recursively splitting the data into subsets based on the most informative features.

   - SVM: A Support Vector Machine is a powerful algorithm for classification, which works by finding a hyperplane that separates the data into different classes.

   - Naive-Bayes: A Naive-Bayes classifier works by assuming that the presence of a particular feature in a class is independent of the presence of other features. It is a simple yet effective algorithm for text classification.

   - KNN: A K-Nearest Neighbors classifier works by finding the k-nearest neighbors to the test instance and classifying it based on the majority class of its neighbors.

6. Model evaluation: Use metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of each classifier. Compare the results of each technique and choose the one with the best performance.

By following these steps, you can implement a system classification for malicious bot detection using ML-supervised classification techniques like the random forest, decision tree, SVM, Naive Bayes, and KNN for the Twitter dataset.

5. PERFORMANCE ASSESMENTS

5.1 Validations

- K-fold cross-validation: This technique involves splitting the dataset into k-folds and training the model on k-1 folds while testing it on the remaining fold. This process is repeated 'k' times. Cross validation can help to reduce the risk of overfitting the model to the training data.

- Stratified sampling: This is a technique where the dataset is divided into homogeneous subgroups (strata) based on the target variable. This ensures that the training and testing datasets have the same distribution of target variables.

- Holdout validation: In this technique, the dataset is divided into a training set and a validation set.

5.2 Evaluations

To evaluate our system's effectiveness in detecting malicious bots on Twitter using machine learning classification, performance measurements are necessary. Performance metrics provide information on the performance of models, algorithms, or processes being evaluated. We used various indicators to evaluate the proposed model's effectiveness, including a two-dimensional matrix representing the actual and assigned class. This matrix includes true positives, false positives, true negatives, and false negatives to describe its composition.

Table -2: Confusion Matrix Evaluation

	Predicated Bot	**Predicted Human**
Actual Bot	**TP**	**FN**
Actual Human	**FP**	**T N**

The TP is a metric that indicates the successful detection of inappropriate tweets from fake Twitter accounts, while an FP refers to the detection of inappropriate tweets that are not associated with fake Twitter accounts, which is known as a type 1 error. On the other hand, an FN is when the detection system fails to identify inappropriate tweets from fake Twitter accounts, known as a type 2 error. TN is when the detection system correctly disregards inappropriate tweets that are not associated with fake Twitter accounts. The accuracy of the detection system can be evaluated using a confusion matrix. The accuracy (A) is a measure of how well the detection operation aligns with the actual value, which can be calculated using the error rate. The error rate is the difference between the detection rate and the current rate, divided by the current rate, and expressed as a percentage. The accuracy is calculated as the sum of TP, TN, and NP, divided by the total number of tweets.

$$\text{Error rate} = \frac{\text{detection rate} - \text{current rate}}{\text{current rate}} \times 100$$

Other performance measures include precision/sensitivity, recall, sensitivity, and specificity. Sensitivity is calculated as TP divided by the sum of TP and FN, while specificity is calculated as TN divided by the sum of TN and FP. The detection rate is calculated as TP divided by the sum of TP, TN, FP, and FN. These metrics can be used to evaluate the performance of the detection system for identifying inappropriate tweets from fake Twitter accounts.

6. RESULT AND ANALYSIS

The experimental analysis involved loading the raw dataset into Python and using the NLTK module to narrow down the focus to essential features, such as "Lexical Diversity of Hate Comments / Biased comments", "most Frequently Used Words in Titles", "Punctuation", and "the Text Length". NaN value columns were removed to clean up the missing data. The Scikit-learn library was used to construct the machine learning model, and decisions on partitioning or splitting the data were informed by the dataset's characteristics. Results showing the accuracy, precision, recall, and/or F1 score of each algorithm on the training and test datasets were used to quantify the performance of

each algorithm and show how well they were able to distinguish between malicious and non-malicious bots. Observations were made about specific features or attributes of the Twitter dataset that were most relevant for detecting malicious bots, such as certain types of tweets, hashtags, or user behaviors.

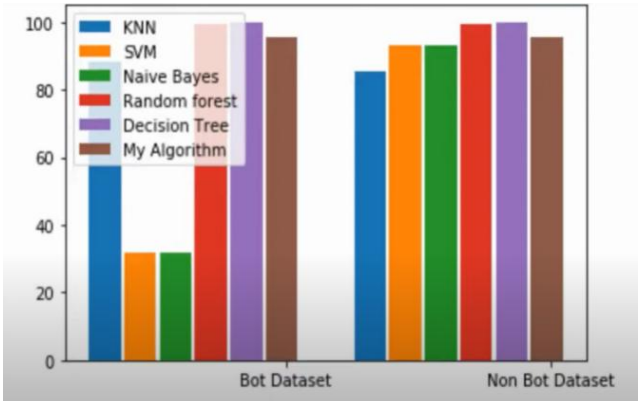

Fig -3: Accuracy Metrics of Classification Algorithms

The decision tree algorithm has been chosen as the "best" algorithm not only because it had the highest accuracy, but also because it was relatively easy to understand and explain compared to the other algorithms. Decision trees can be used to classify data that contains both categorical and numerical features. They can also handle missing values and noisy data, which can be useful in real-world applications. Decision trees can capture non-linear relationships between variables, which means they can be used to classify data that is not linearly separable. This makes them a powerful tool for classification problems in which there is no clear separation between classes. Decision trees can be used as part of an ensemble of algorithms, such as random forests or gradient boosting, which can improve their performance and accuracy. The user interface (UI) of our classification system allows users to easily upload, clean, and display Twitter data, as well as select and evaluate various machine learning algorithms for malicious bot detection.

Fig -4: Home Screen

The UI consists of four main screens: the data upload screen, the data cleaning screen, the algorithm screen, and the results screen. The data upload screen allows users to upload CSV files containing Twitter data, such as user account information, tweet content, and engagement metrics.

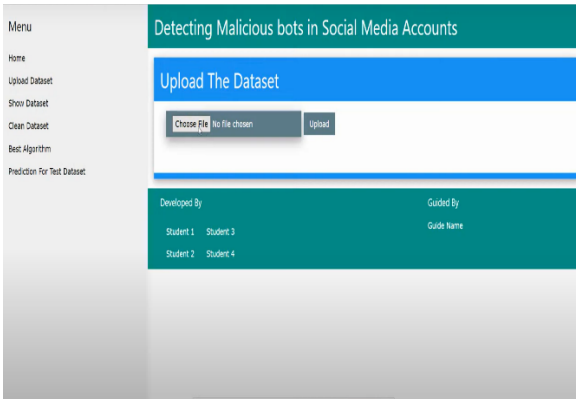

Fig -4: Upload dataset

Once the data is uploaded, users can navigate to the show dataset screen to view all the data and then preprocess the data for analysis.

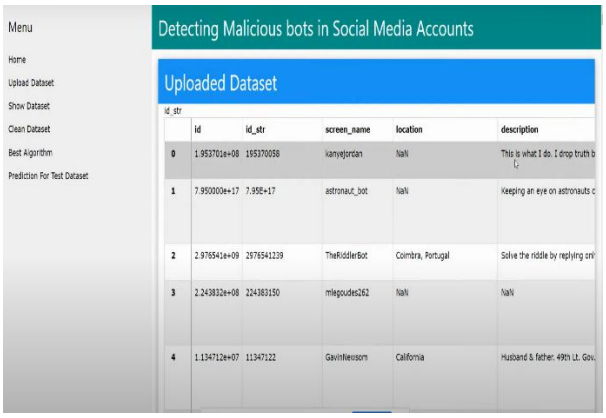

Fig -5: Show dataset screen

Once the data is uploaded, users can navigate to the data cleaning screen to remove any irrelevant or duplicate data and preprocess the remaining data for analysis. On the algorithm screen users can view from a range of supervised classification algorithms, including Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Support Vector Machines (SVM), as well as a custom algorithm. Users can evaluate the performance of the algorithm on the test dataset on the results screen The results screen shows various performance metrics, such as accuracy, precision, recall, and F1 score, as well as confusion matrices and ROC curves. In our experiments, we found that the Decision Tree algorithm achieved the highest accuracy.

7. FUTURE SCOPE

As bots become more advanced and sophisticated, it is important to continuously develop and improve detection methods. One potential area for future research is the integration of multiple detection techniques, such as signature-based, behavior-based, and anomaly detection methods. By combining these techniques, the overall accuracy and effectiveness of bot detection can be improved. Another area for future research is the exploration of new features and attributes that can improve the accuracy of bot detection. For example, the use of natural language processing techniques can help identify subtle linguistic differences between bot and human-generated content.

Furthermore, the development of more advanced machine learning algorithms, such as deep learning, can improve the accuracy and efficiency of bot detection. This can be especially beneficial in processing large amounts of social media data in real-time.

There are several other techniques that can be used for malicious bot detection in addition to supervised classification techniques. Some of these techniques include:

1. Network-based detection: Network-based detection techniques can be used to identify bot activity based on network traffic analysis. These techniques can help detect bots that may be using sophisticated evasion techniques.

2. Hybrid approaches: Hybrid approaches that combine multiple detection techniques, such as supervised and unsupervised learning, can improve the accuracy and effectiveness of bot detection.

3. Behavioral analysis: Behavioral analysis techniques can be used to identify bot activity based on patterns of behavior that are indicative of bot activity. Examples of behavioral analysis techniques include analysis of account creation patterns, post frequency, and engagement rates.

Overall, the choice of technique depends on the specific characteristics of the problem at hand and the data available for analysis. A combination of multiple techniques may be necessary to effectively detect malicious bots on social media platforms.

8. CONCLUSIONS

In this study, observed the detection and elimination of bot activities are crucial for network security management across various industries. Traditional bot detection methods have limitations, but a novel approach based on machine learning and supervised learning algorithms has been developed to overcome these challenges. This methodology offers versatility in detecting any type of bot, making it effective in the rapidly evolving landscape of bot technologies. Overall, the use of machine learning and supervised learning algorithms is a promising approach to bot detection, offering an effective means of identifying and mitigating potential bot activities across social media platforms and beyond.

REFERENCES

[1] M.Sahlabadi,R.C.Muniyandi andZ.Shukur- Detecting abnormal geste in social network Websites by using a process mining- fashion-J.Comput.Sci. -vol. 10, no. 3, pp. 393- 402, 2014.

[2] M.Al- Qurishi, M.S.Hossain, M.Alrubaian, S.M.M.Rahman andA.Alamri- using analysis of stoner geste to identify vicious conditioning in large-scale social networks- IEEETrans.Ind.Informat., vol. 14, no. 2, pp. 799-813, Feb. 2018

[3] Phillips. Efthimion1, Scott Payne1, Nick Proferes2, || Supervised Machine Learning Bot Discovery ways to Identify Social Twitter Bots ||, Master of Science in Data Science, Southern Methodist University, 6425 Boaz Lane, Dallas, TX 75205, 2018.S.Barbon,Jr.G.F.C.Campos,G.M.Tavares,R.A.Igawa,M.L.Proen ¸ ca, Jr andR.C.Guido.

[4] Discovery of humans, licit bots, and vicious bots in online social networks grounded on ripples, ACM Trans. Multimedia Comput, Commun, Appl, vol. 14, no. 1s, Feb. 2018, Art.no. 26

[5] F. Morstatter,L.Wu,T.H.Nazer,K.M.Carley andH. Liu, A new approach to bot discovery Striking the balance between perfection and recall, in Proc. IEEE/ ACM Int. Conf.Adv. Social Network. Anal. Mining, San Francisco, CA, USA,Aug. 2016, pp. 533-540.

[6] Y. Zhou et al ProGuard Detecting vicious accounts in a social network- grounded online elevations, IEEE Access, vol. 5, pp. 1990- 1999, 2017.

[7] Xuan Dau Hoang and Quynh Chi Nguyen, Botnet Detection Grounded On Machine Learning Ways Using DNS Query Data.

[8] C.K.Chang, Situation analytics- A foundation for a new software engineering paradigm, Computer, vol. 49, no.1, pp. 24-33, Jan. 2016. 38 39

[9] Efthimion, Phillip George, Payne, Scott and Proferes, Nicholas(2018) || Supervised Machine Learning Bot Discovery Ways to Identify Social Twitter Bots, || SMU Data Science ReviewVol. 1No. 2, Composition 5.

[10] Ranjana Battur, Nagaratna Yaligar || Twitter Bot Discovery using Machine Learning Algorithms, || International Journal of Science and Research(IJSR) ISSN 2319- 7064 ResearchGate Impact Factor(2018)0.28 — SJIF(2018)7.426

[11] C. Cai, L. Li, and D. Zengi, Behavior enhanced deep bot discovery in social media, || in Proc. IEEE Int. Conf. Intell. Secur. Inform. (ISI), Beijing, China, Jul. 2017, pp. 128- 130.