DETECTION OF MALICIOUS BOTS IN TWITTER NETWORK

Akash J[1], Ravikumar N[2], Shishir Somapur[3], Sumit Rathod[4] , Mr. Sandesh R[5]

8th Sem Student, Dept of CSE, ATME College of Engineering, Mysuru, Karnataka, India[1]

8th Sem Student, Dept of CSE, ATME College of Engineering, Mysuru, Karnataka, India[2]

8th Sem Student, Dept of CSE, ATME College of Engineering, Mysuru, Karnataka, India[3]

8th Sem Student, Dept of CSE, ATME College of Engineering, Mysuru, Karnataka, India[4]

Assistant professor, Dept of CSE, ATME College of Engineering, Mysuru, Karnataka, India[5]

ABSTRACT

Malicious social bots generate fake tweets and automate their social relationships either by pretending like a follower or by creating multiple fake accounts with malicious activities. Moreover, malicious social bots post shortened malicious URLs in the tweet in order to redirect the requests of online social networking participants to some malicious servers. Hence, distinguishing malicious social bots from legitimate users is one of the most important tasks in the Twitter network. In this project we are detecting Twitter bots within the network using machine learning algorithms, namely Naive Bayes, Random Forest, and Decision Trees. We collected and pre-processed a comprehensive dataset of bot and genuine user profiles, extracting features like posting frequency, content patterns, and follower relationships for classification. Through extensive experimentation, we compared the accuracy, precision, recall, and F1- score of these algorithms.

Keyword: - *Malicious Social Bots, Spearman's Correlation, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Logistic Regression, Cross Validation, Ensemble.*

1. INTRODUCTION

In the digital era, social media platforms serve as vital hubs for communication, information dissemination, and community engagement. However, alongside genuine users, these platforms also harbour a growing population of malicious social bots—automated accounts programmed to spread misinformation, manipulate public opinion, and engage in disruptive behaviours. To combat this pervasive threat and safeguard the integrity of online interactions, our project proposes an innovative approach: ensemble stacking. By combining the predictive capabilities of diverse machine learning models such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Naive Bayes, with Logistic Regression as the meta-model, we aim to develop a robust system for detecting malicious social bots. Through the strategic integration of cross-validation techniques, we seek to optimize the performance of our model, ensuring its effectiveness across varied datasets and real-world scenarios. By enhancing our understanding of bot detection and fortifying online security measures, our project endeavours to foster trust, transparency, and authenticity within digital communities.

2. LITERATURE SURVEY

[1] "Detection of Malicious Social Bots using ML technique in Twitter Network" [2022]
Authors: V N P Sai Siri Dantu, Jhansi Devi Telu, Padma Sree Kuncham, Gurudatta Pilla
This paper aimed to design a framework by considering the features set to be evaluated the trust value of each online social network account and to detect the malicious social bots in the Twitter network effectively and efficiently. The authors discusses the use of deep-learning algorithms to extract hidden patterns in text for spam detection on Twitter. The paper focus on detecting malicious social bots based on the LA model with the Naive Bayes algorithm with URL based features. It presents a framework that combines text-based and metadata-based features to detect malicious accounts on Twitter. The proposed system uses the LA-MSBD (Learning Automata

Malicious Social Bot Detection) algorithm by integrating a Naïve-Bayes algorithm model with a set of URL-based feature extraction to distinguish between malicious bots and legitimate users. In this research work, it is an extension for the existing system that is based on trust computational model of direct and indirect trust and the proposed system is an enhancement of the existing system. It also mentions the prevalence of automated bot accounts on Twitter and the potential benefits of using Twitter bots for customer support.

[2] "Random Forest Twitter Bot Classifier" [2019]
Authors: James Schnebly and Shamik Sengupta
The paper addresses the issue of detecting Twitter bot accounts, which are automated accounts that spread malicious content or generate fraudulent popularity for political and social figures. The authors propose a set of attributes for a Random Forest classifier that achieves high accuracy (90.25%) and generalizability in detecting bot accounts. The paper provides a practical solution for detecting Twitter bot accounts, which can help in combating the spread of malicious content and fraudulent popularity on social media platforms. The proposed Random Forest classifier, with its high accuracy and generalizability, can be deployed directly after training to accurately identify and label bot accounts. By using a combination of accessible features from the Twitter API and derivative ratio features, the classifier offers valuable insights into the behaviour and characteristics of bot accounts. The use of the Random Forest algorithm helps prevent overfitting and creates a model that can be effectively used for detecting bot accounts on Twitter. The findings of this paper can be applied by Twitter and other social media platforms to develop more effective strategies for identifying and removing bot accounts, thereby improving the overall user experience and reducing the spread of false information.

[3] "Detection of Malicious Social Bots using Learning Automata with URL features in Twitter Network" [2022]
Authors: M. Divya, D. Abhishek, K. Naresh, P. Sparsha
A Learning Automata-based Malicious Social Bot Detection (LA-MSBD) model is proposed by integrating a trust computation model with URL-based features for identifying trustworthy participants (users) in the Twitter network. In this the authors will be giving 13 parameters as input to the SVM algorithm which process these inputs and return a single digit either 0 or 1. Experimentation has been performed on two Twitter data sets, and the results illustrate that the proposed algorithm achieves improvement in precision, recall, F-measure, and accuracy compared with existing approaches for MSBD. The proposed LA-MSBD algorithm achieves the advantages of incremental learning. Two Twitter data sets are used to evaluate the performance of our proposed LA-MSBD algorithm. The experimental results show that the proposed LA-MSBD algorithm achieves up to 7% improvement of accuracy compared with other existing algorithms. The results illustrate that the proposed algorithm achieves improvement in precision, recall, F-measure, and accuracy compared with existing approaches for MSBD.

[4] "Twitter bot detection using supervised machine learning" [2021]
Authors: A Ramalingaiah1, S Hussaini1, S Chaudhari
The paper educates us about Bots, Bots can be categorized as good or bad, with bad bots being responsible for malicious activities such as impersonating human behaviour, attacking IoT devices, and exploiting performance. Detecting and preventing the activities of these bots is crucial, especially for social media users who are more vulnerable to data breaches and manipulation. Supervised machine learning techniques are used to detect Twitter bots, comparing the accuracy of different algorithms with a classifier using a Bag of Bots word model. In this paper the dataset used for identifying bots contains various attributes like URL, description, friends, followers, etc. The dataset is pre-processed by removing data imbalance and applying the Spearman coefficient for feature independence. Decision tree, logistic regression, K nearest neighbour, and Naïve Bayes algorithms are implemented to identify bots, and the algorithm with the highest accuracy is tested for real-time data. Among different train-test data sets, the 90%-10% split yields the highest accuracy and AUC scores, while the 80%-20% split is considered optimal due to its close accuracy and AUC scores to the 90%-10% split, low variance, and computational efficiency for training moderately large datasets.

[5] "Detection of Malicious Social Bots Using Learning Automata with URL Features in Twitter Network" [2020]
Authors: Rashmi Ranjan Rout, Greeshma Lingam, and D. V. L. N. Somayajulu
Malicious social bots, which impersonate real users, engage in a range of harmful activities, including the dissemination of spam content and phishing attacks through shortened URLs. To combat this threat, the paper presents the Learning Automata-based Malicious Social Bot Detection (LA-MSBD) algorithm. This algorithm integrates a trust computation model with URL-based features to identify trustworthy users on Twitter. The trust model incorporates two key parameters: direct trust, calculated using Bayes' theorem, and indirect trust,

determined through the Dempster-Shafer theory (DST). The paper's experimental results, based on two Twitter datasets, demonstrate the superior performance of the LA-MSBD algorithm compared to existing methods. This approach not only enhances precision, recall, F-measure, and accuracy in malicious social bot detection but also offers a promising solution to the challenge of identifying and mitigating these threats within the Twitter network.

3. METHODOLOGY

The Methodology and the steps involved in the proposed system is shown in the form of block diagram in the below figure-1.

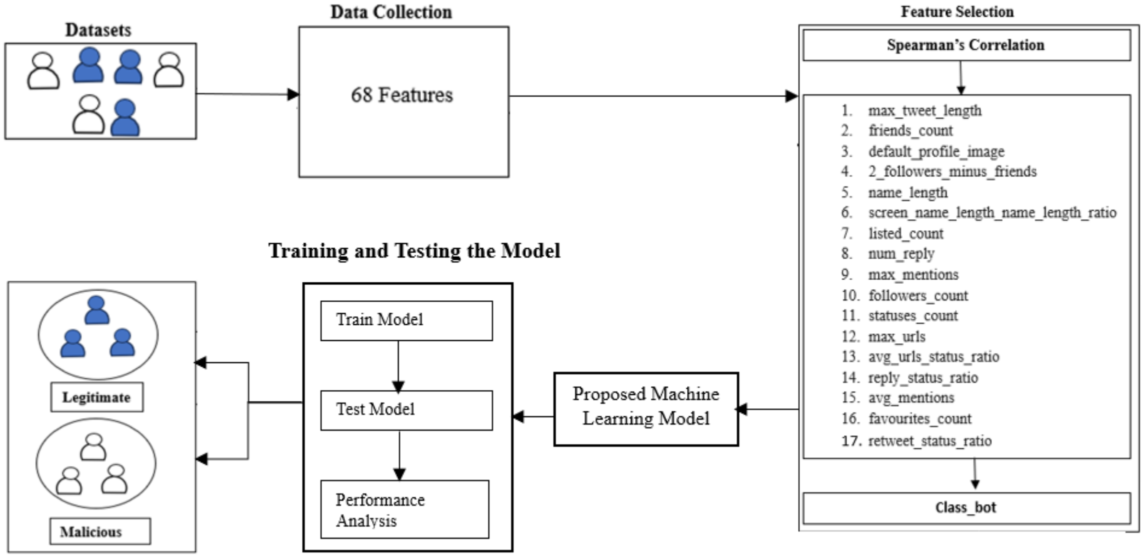


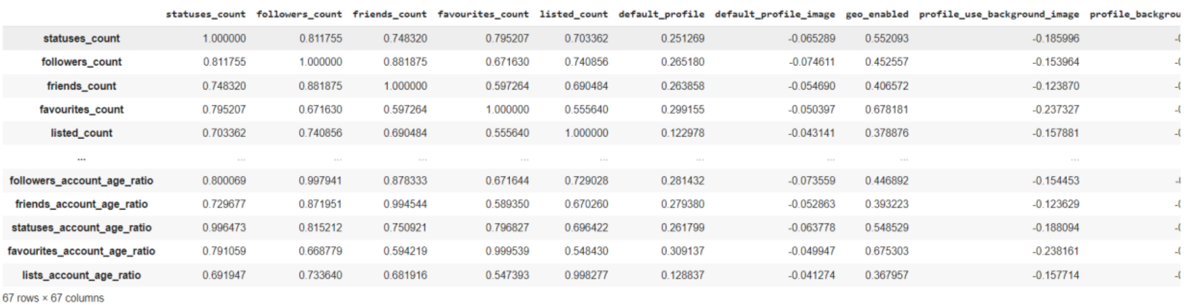
Figure-1: System Architecture of Detection of Malicious Bots in Twitter Network

3.1 Data Collection

Data collection involves gathering data from various sources, such as social media platforms' APIs (e.g., Twitter, Facebook), web scraping tools, or third-party datasets. It includes information about user profiles, posts, comments, likes, shares, and network connections. Techniques such as data crawling, API requests, or data streaming may be employed to collect real-time or historical data.

3.2 Feature Selection

Relevant features are selected from the data to capture patterns indicative of bot behavior. Spearman's correlation method is used as feature selection technique to select the most relevant features. The function takes two real-valued samples as arguments and returns both the correlation coefficient in the range between -1 and 1. where 1 indicates a perfect positive relationship, -1 indicates a perfect negative relationship, and 0 indicates no relationship**.** For the threshold value of 0.7, important 17 features are selected out of 67 features.



	statuses_count	followers_count	friends_count	favourites_count	listed_count	default_profile	default_profile_image	geo_enabled	profile_use_background_image	profile_backgrou
statuses_count	1.000000	0.811755	0.748320	0.795207	0.703362	0.251269	-0.065289	0.552093	-0.185996	-
followers_count	0.811755	1.000000	0.881875	0.671630	0.740856	0.265180	-0.074611	0.452557	-0.153964	-
friends_count	0.748320	0.881875	1.000000	0.597264	0.690484	0.263858	-0.054690	0.406572	-0.123870	-
favourites_count	0.795207	0.671630	0.597264	1.000000	0.555640	0.299155	-0.050397	0.678181	-0.237327	-
listed_count	0.703362	0.740856	0.690484	0.555640	1.000000	0.122978	-0.043141	0.378876	-0.157881	-
...	...	...	...	...	...	...	...	...	...	
followers_account_age_ratio	0.800069	0.997941	0.878333	0.671644	0.729028	0.281432	-0.073559	0.446892	-0.154453	-
friends_account_age_ratio	0.729677	0.871951	0.994544	0.589350	0.670260	0.279380	-0.052863	0.393223	-0.123629	-
statuses_account_age_ratio	0.996473	0.815212	0.750921	0.796827	0.696422	0.261799	-0.063778	0.548529	-0.188094	-
favourites_account_age_ratio	0.791059	0.668779	0.594219	0.999539	0.548430	0.309137	-0.049947	0.675303	-0.238161	-
lists_account_age_ratio	0.691947	0.733640	0.681916	0.547393	0.998277	0.128837	-0.041274	0.367957	-0.157714	-

67 rows × 67 columns

Figure-2: Correlation Matrix

4. MACHINE LEARNING MODELS

4.1 Decision Tree

Decision tree classifier is a popular machine learning language algorithm used for both classification and regression tasks. It works by recursively partitioning the data into subsets based on the features, creating tree like structure where each internal node represents a decision based on a feature, and each leaf node represents the outcome.

4.2 Random Forest

Random forest classifier is an machine learning method used for classification tasks in machine learning. It operates by constructing multiple decision trees during the training phase, each tree is trained individually and outputs the class that is the mode of the classes predicted by the individual trees.

4.3 Support Vector machine

Support vector machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates different classes in the input data. It maximizes the distance to point an either class. This distance is called as margin, the points that falls exactly on the margin are called as supporting vectors.

4.4 Naïve Bayes

A Naive Bayes classifier is a type of probabilistic classifier based on Bayes' theorem with an assumption of independence between features. It is commonly used in machine learning for classification tasks such as spam filtering and sentiment analysis. The algorithm calculates the probability of each class given a set of input features and predicts the class with the highest probability.

4.5 Logistic Regression

Logistic regression is used for binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and 1. The concept of the threshold value is used, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

4.6 Ensemble Technique: - Stacking Approach

Ensemble Stacking considers heterogeneous weak learners, learns them in parallel, and combines them by training a meta-learner to output a prediction based on the different weak learner’s predictions. Here decision tree, random Forest, SVM and naive bayes are used as base model and logistic regression is used as meta model. This model is trained and tested using cross validation method.

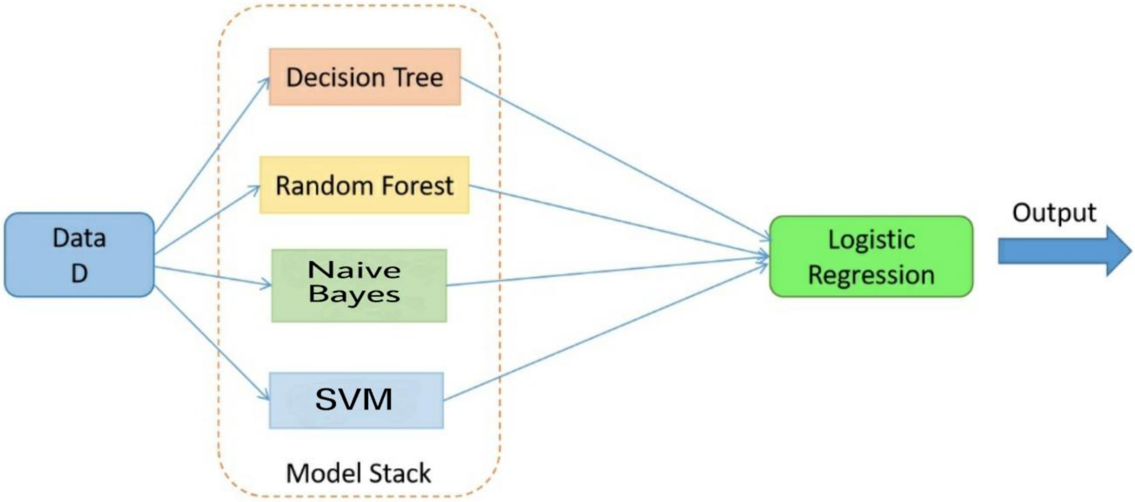


Figure-3: Stacking Model

5. RESULTS

The performance comparison of individual machine learning models with the stacking model is shown in the table-1**.**

Table-1: Performance Comparison Table

Machine Learning Models	Accuracy (in %)	Precision (in %)	Recall (in %)	F1 Score (in %)
Decision Tree	97.98	98.10	98.20	98.15
Random Forest	98.40	98.80	98.20	98.50
SVM	91.53	88.36	98.90	93.33
Naïve Bayes	96.96	96.40	98.60	97.49
Stacking Model	99.28	99.50	99.30	99.40

6. CONCLUSION AND FUTURE SCOPE

In conclusion, our research has demonstrated the effectiveness of ensemble stacking in detecting malicious social bots. By combining the predictive abilities of diverse machine learning models, including Decision Trees, Random Forest, SVM, and Naive Bayes, with Logistic Regression as the meta-model, we have developed a robust and accurate system for identifying bot behavior in social media platforms. Through rigorous experimentation and validation, we have shown that our ensemble stacking approach outperforms individual models with accuracy of 99.28 %. The integration of cross-validation techniques further enhances the model's generalization performance by overcoming over fitting and ensuring its effectiveness across different datasets and scenarios. Looking ahead, the future scope of our project extends to several applications of deep learning techniques, such as neural networks and recurrent neural networks (RNNs), holds promise for capturing more complex patterns in bot behavior and improving detection accuracy. Additionally, exploring the integration of natural language processing (NLP) techniques for analyzing textual content can further enhance the model's capabilities.

7. REFERENCES

[1] V N P Sai Siri Dantu, Jhansi Devi Telu, Padma Sree Kuncham, Gurudatta Pilla: "Detection ofMalicious Social Bots using ML technique in Twitter Network", June 2022.

[2] James Schnebly and Shamik Sengupta: "Random Forest Twitter Bot Classifier" 2019.

[3] M. Divya, D. Abhishek, K. Naresh, P. Sparsha: “Detection of Malicious Social Bots usingLearning Automata with URL features in Twitter Network”

[4] A Ramalingaiah1, S Hussaini1, S Chaudhari: “Twitter bot detection using supervised machinelearning” Jan 2021.

[5] Rashmi Ranjan Rout, Greeshma Lingam, and D. V. L. N. Somayajulu: “Detection of MaliciousSocial Bots Using Learning Automata with URL Features in Twitter Network” 2020.

[6] G. Lingam, R. R. Rout, and D. V. L. N. Somayajulu, “Adaptive deep Q-learning model for detecting social bots and influential users in online social networks,” Nov 2019.

[7] D. choi, j. Han, S. Chum, E, Rappos, S. Robert, and T. T. Kwon, “Bit.ly/Practice: Uncovering content publishing and sharing through URL shortening services,” Dec 2018.

[8] D. R. Patil and J. B. Patil, “Malicious URL detection using decision tree classifier and majority voting technique,” Mar 2018.

[9] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, “A framework for real-time spam detection in Twitter,” Jan 2018.

[10] A. K. Jain and B. B. Gupta, “A machine learning based approach for phishing detection using hyperlinks information,” May 2019.