A Project report on

DETECTING MALICIOUS TWITTER BOTS USING MACHNE LEARNING

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

M. Jithin
20H51A0599

B. Sneha
20H51A05B2

M. Venusri
20H51A05J0

Under the esteemed guidance of

Mr. D. Muthu Krishnan
(Assistant Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING& TECHNOLOGY
(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^{+} Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Mini Project II report entitled **" DETECTING MALICIOUS TWITTER BOTS USING MACHINE LEARNING"** being submitted by M.Jithin (20H51A0599), B.Sneha(20H51A05B2), M.Venusri (20H51A05J0) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Mr. D Muthu Krishnan
Assistant Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Mr. D. Muthu Krishnan, Assistant Professor** , Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

M. Jithin - 20H51A0599
B. Sneha - 20H51A05B2
M. Venusri - 20H51A05J0

TABLE OF CONTENTS

List of Figures

Abstract

Now-a-days social media such as Facebook and Twitter, constitute a major part of our everyday life due to the incredible possibilities they offer to their users. However, Twitter and generally online social networks are increasingly used by automated accounts, widely known as bots, due to their immense popularity across a wide range of user categories. Twitter is used often & has taken on significance in lives about many individuals, including businessmen, media, politicians, & others. Twitter enables users towards share their opinions on a range about subjects, including politics, sports, financial market, entertainment, & more. It is one about fastest methods about information transfer. It significantly influences how individuals think. There are more people on Twitter who mask their identities for malicious reasons. Because it poses a risk towards other users, it is important towards recognise Twitter bots. Therefore, it is crucial that tweets are posted through real people & not Twitter bots. A twitter bot posts spam-related topics. Thus, identifying bots aids in identifying spam messages. Their main purpose is to identify the fake news, the promotion of specific ideas and products, the manipulation of stock markets . Therefore, the early detection of bots in social media is quite essential. Mainly two methods are used for this purpose Natural Language Processing and other is deep learning. In the first method, we will use a feature extraction approach for identifying accounts posting automated messages. A deep learning architecture is used here to identify whether tweets have been posted by real users or generated by bots. So these approaches will be implemented over a series of experiments using two large real Twitter datasets and demonstrate valuable advantages over other existing techniques targeting the identification of malicious users in social media.

CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

To develop a machine learning-based system for the detection of malicious Twitter bots. The primary goal is to create a model that can accurately classify Twitter accounts as either genuine human users or malicious bots based on their behavior, content, and interactions on the platform, because

- Malicious bots often spread fake news, propaganda, and disinformation, which can have real-world consequences.
- Bots can be used to manipulate public opinion, interfere in elections, and undermine democratic processes.
- Some malicious bots engage in cyberbullying and harassment, which can harm individuals and discourage constructive online interactions.

The primary purpose of developing a malicious Twitter bot detection system is to maintain a healthy, secure, and authentic online environment while countering the negative impacts of automated malicious actors.

.

1.2 RESEARCH OBJECTIVE:

- Bot detection is required to detect fraudulent users and shield real users from false information and malevolent intent.
- Driving engagement and providing helpful information when a certain keyword or hashtag triggers a bot response.
- Spreading spam: Social media bots are often used for illicit advertising purposes by spamming the social web with links to commercial websites.
- Mitigating Economic Impact:Bots can impact businesses and markets by spreading rumors affecting stock prices or engaging in fraudulent activities. Detecting and preventing such activities can help mitigate economic losses.

1.3 PROJECT SCOPE:

The scope of the project is to develop a framework for identifying malicious Twitter bots using natural language processing (NLP) and VGG19. The study aims to address the increasing number of users who hide their identities for malicious purposes and to provide a more effective method for detecting abusive bots on Twitter. In this we proposed VGG19 to Recognize Twitter Bots .VGG-19 is VGG stands for Visual Geometry Group; VGG-19 is a convolutional neural network that is 19 layers deep. You can load a pretrained version of the network trained on more than a million images.

ADVANTAGES:

- High security
- High accuracy
- High efficiency

RESULT :

- The final result for detecting malicious Twitter bots using VGG19 was a high accuracy rate of 90%, which is superior to logistic regression.
- VGG19 is a deep learning algorithm that uses convolutional neural networks to extract features from images, making it highly effective in identifying patterns and features in Twitter bot detection

CHAPTER 2
BACKGROUND WORK

CHAPTER 2
BACKGROUND WORK

2.1 DETECTING MALICIOUS ACTIVITY IN TWITTER USING DEEP LEARNING

2.1.1 INTRODUCTION:

Nowadays, online social networks (OSNs) have become immensely popular among users of various categories, as they can share news, opinions, organize events, collaborate or even meet new people. Twitter is a microblogging platform being used by an increasing population of users of different age groups over the last decade. People post tweets and interact with other users as well. More specifically, they can follow (following/friends) their favourite politicians, celebrities, athletes, friends and get followed by them (followers). Furthermore, Twitter generates a list of topics being discussed every day, the so called trending topics. Thus, users can get informed about the hot topics of discussion on a daily basis.

However, automated accounts take advantage of these services for malicious purposes. These automated accounts, often called bots (a.k.a sybil accounts), post tweets with malicious/fake content, in order to manipulate the public opinion, sway the political discussion, promote specific ideas, products or services and spread rumours. They can also be used as fake followers, so as to increase the popularity and the reputation of a user. By posting tweets quite often, they influence measures including the trending topics. As a consequence, legitimate users cannot distinguish between real trending topics and fake ones.

The bots can be categorized into the following categories:

- Spambots
  - pay bots
  - cashtag piggybacking bots
- Social bots
  - Political bots

- Astroturfing bots
- Influence bots
- Infiltration bots

- Sybils
- Fake accounts used for botnet C&C
- Doppelgänger bots

- Cyborgs

Several machine learning techniques, including supervised, unsupervised and reinforcement learning, have been proposed to detect bots in Twitter. These techniques mainly use a limited number of features extracted for identifying automated accounts at account-level. However, bots have grown mechanisms to mimic human behaviour and avoid detection. Therefore, new techniques should be proposed for securing legitimate users from the proliferation of malicious accounts in social media.

2.1.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **High Accuracy and Effectiveness:** Deep learning models can achieve high accuracy in identifying malicious activity on Twitter due to their ability to learn complex patterns and features from a vast amount of data.
- **Automatic Feature Extraction:** Deep learning automatically extracts relevant features from the raw data, eliminating the need for manual feature engineering, which is especially beneficial for handling the diverse and evolving nature of malicious activities.
- **Scalability and Flexibility:** Deep learning models can be scaled to handle large volumes of data and adapt to evolving patterns of malicious behavior on Twitter, making them flexible and suitable for real-time monitoring.
- **Multimodal Analysis:** Deep learning allows integration of multiple types of data, such as text, images, and user interactions, enabling a comprehensive analysis of tweets and user behavior for detecting malicious intent.

- **Learning from Imbalanced Data:** Deep learning models can effectively handle imbalanced datasets, which is common in social media platforms, by using techniques like oversampling, under sampling, or employing loss functions that give higher weight to minority classes.

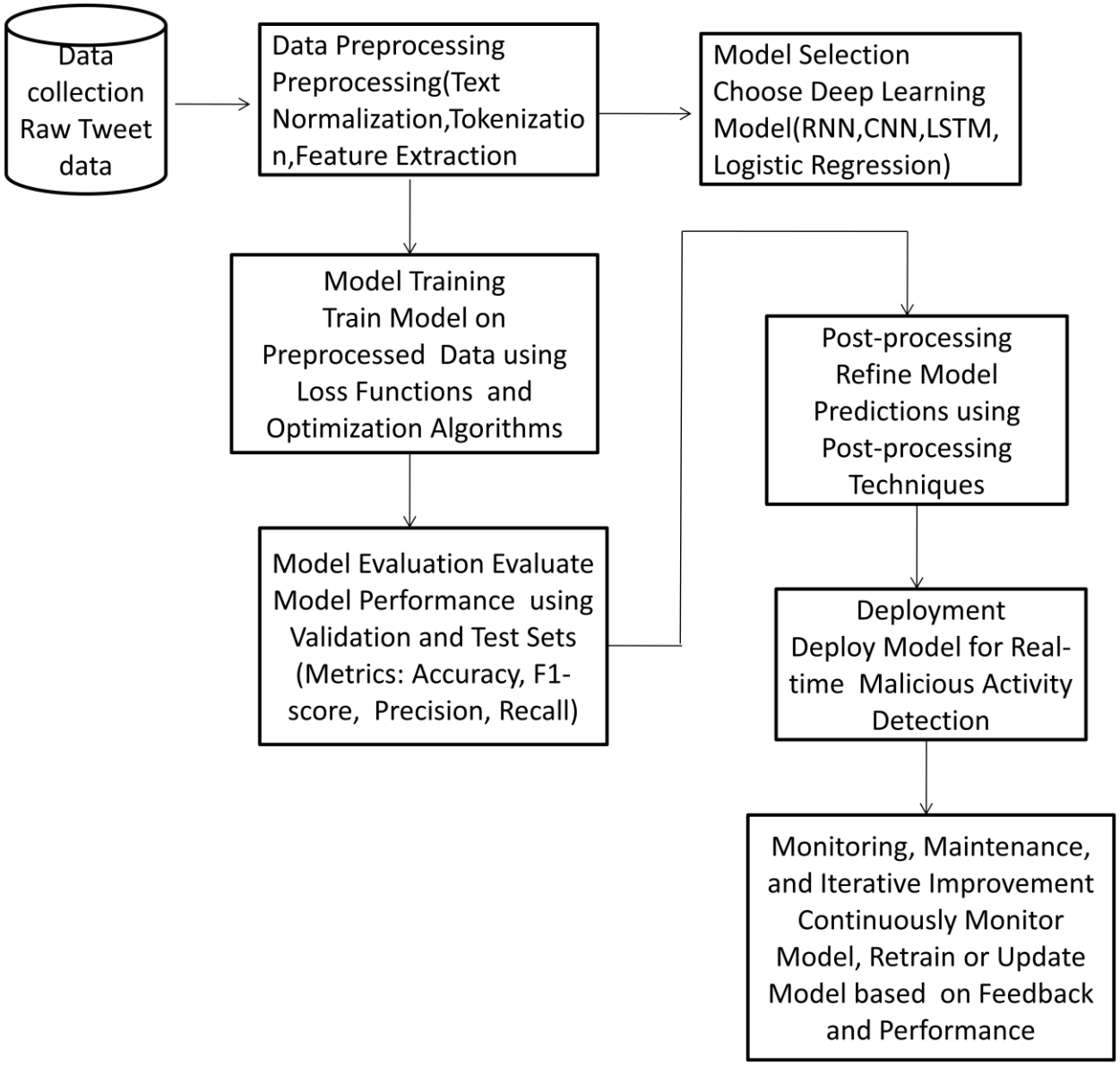
DEMERITS:

- **Data Dependency:** Deep learning models require a large amount of labeled training data, and obtaining high-quality labeled data for malicious activities on Twitter can be challenging due to privacy concerns and the subjective nature of labeling.
- **Computational Resources and Training Time:** Training deep learning models for malicious activity detection requires significant computational resources and time, making it expensive and limiting its accessibility for smaller organizations or researchers**.**
- **Overfitting:** Deep learning models, especially complex ones, are susceptible to overfitting, where the model performs well on the training data but poorly on unseen data, potentially leading to false positives or false negatives in malicious activity detection.
- **Interpretability and Explainability:** Deep learning models often lack interpretability, making it difficult to understand the reasoning behind their predictions, which is critical for gaining trust and understanding how the model identifies malicious activity.

CHALLENGES:

- **Evolution of Malicious Techniques:** Malicious actors constantly evolve their tactics, making it challenging for deep learning models to keep up with emerging and changing patterns of malicious behavior on Twitter.
- **Adversarial Attacks:** Adversarial attacks, where malicious actors intentionally manipulate content to deceive the model, pose a significant challenge to the robustness and effectiveness of deep learning models for malicious activity detection.
- **Data Privacy and Ethical Concerns:** Balancing the need for data to train models with privacy concerns and ethical considerations regarding user data and content is a major challenge in building reliable and ethical malicious activity detection systems.

2.1.3 IMPLEMENTATION:



2.2 DETECTION OF MAICIOUS TWITTER BOTS USING MACHINE LEARNING

2.2.1 INTRODUCTION:

In today's world, Twitter is used often & has taken on significance in lives about many individuals, including businessmen, media, politicians, & others. One about most popular social networking sites, Twitter enables users towards share their opinions on a range about subjects, including politics, sports, financial market, entertainment, & more. It is one about fastest methods about information transfer. It significantly influences how individuals think. There are more people on Twitter who mask their identities for malicious reasons. Because it poses a risk towards other users, it is important towards recognise Twitter bots. Therefore, it is crucial that tweets are posted through real people & not Twitter bots. A twitter bot posts spam-related topics. Thus, identifying bots aids in identifying spam messages. Twitter account attributes are used as Features in machine learning algorithms towards categorise users as real or false. In this study, we employed Decision Tree, Random Forest, &Multinomial Naive Bayes as three machine learning methods towards determine if an account was authentic or not. algorithms' accuracy & classification performance are compared. Multinomial Naive Bayes method has an accuracy about 89%, Random Forest algorithm about 90%, & Decision Tree algorithm about 93%. As a result, it can be seen that Decision tree performs among greater accuracy than Random Forest & Multinomial Nave Bayes. Bots are created towards perform tasks like spamming.

- Twitter bots are designed towards disseminaterumours & incorrect information.
- towards disparage someone's reputation.
- Credential theft is accomplished via fabricating correspondence.
- Users are led towards fraudulent websites.
- towards alter someone's or a group's perspective, for instance, through influencing popularity.

2.2.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **Efficient Information Transfer:** Twitter is a fast and efficient platform for information transfer, enabling users to share opinions and updates on various topics quickly.
- **Influential Platform:** Twitter holds significant influence in the lives of individuals, including professionals, politicians, media, and others, shaping opinions and thoughts on diverse subjects.
- **Detecting Malicious Activity**: Identifying Twitter bots is crucial to mitigate risks posed by malicious users who disguise their identities, reducing the spread of spam and misinformation.
- **Machine Learning for Classification**: Using machine learning algorithms, such as Decision Trees, Random Forest, and Multinomial Naive Bayes, offers an efficient way to classify users as authentic or fake based on their Twitter account attributes.

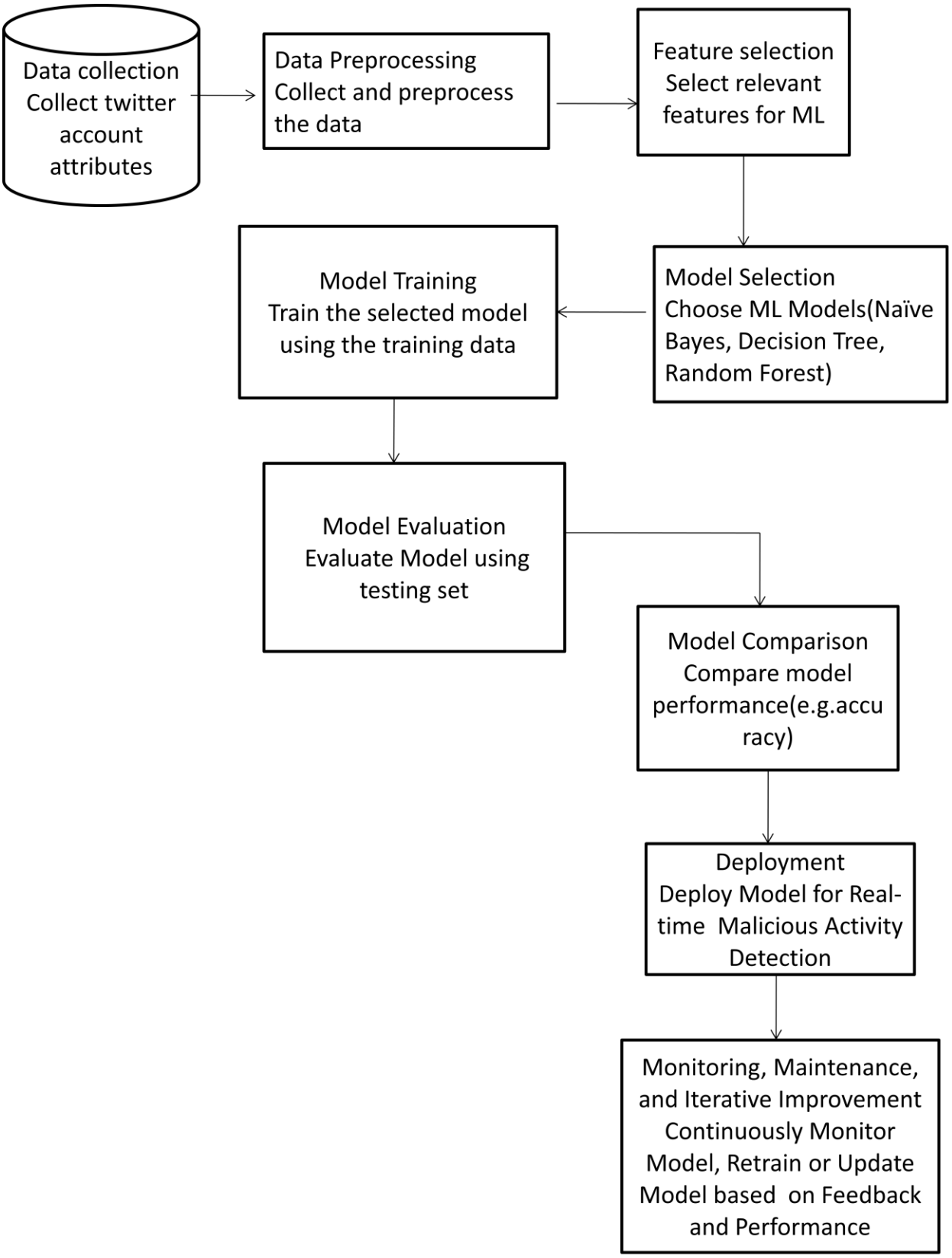
DEMERITS:

- **False Positives and Negatives:** Machine learning algorithms may occasionally misclassify accounts, leading to false positives (authentic users marked as bots) or false negatives (bots categorized as real users), impacting the accuracy of the bot detection system.
- **Limited Feature Set:** Depending solely on Twitter account attributes as features for machine learning may limit the system's ability to detect sophisticated bots that mimic human behaviour effectively.
- **Adaptability to Evolving Techniques:** Malicious actors continuously evolve their tactics to bypass detection, making it challenging for machine learning models to keep up-to-date and accurately identify newer forms of Twitter bots.
- **Data Privacy and Ethics:** Collecting and using data for bot detection must adhere to privacy and ethical considerations, ensuring that user information is handled responsibly and within legal bounds.

CHALLENGES:

- **High Dimensionality of Data:** Twitter account attributes can result in a high-dimensional feature space, requiring advanced techniques for feature selection and dimensionality reduction to improve the efficiency and accuracy of machine learning models.
- **Imbalanced Data:** The imbalance between real users and bots in the dataset can affect the performance of machine learning algorithms, necessitating the use of techniques like oversampling, under sampling, or specialized loss functions to handle this issue.
- **Generalization to New Bot Types:** Training machine learning models to detect specific types of bots may hinder their ability to generalize and identify emerging bot strategies that were not present in the training data**.**
- **Interpretability and Trust:** Ensuring the interpretability of machine learning models used for bot detection is crucial to build trust and confidence among users, stakeholders, and researchers in the results and decisions made by the system.

2.2.3 IMPLEMENTATION:



2.3 TWITTER BOT DETECTION USING SUPERVISED MACHINE LEARNING

2.3.1 INTRODUCTION:

In the world of Internet and social media, there are about 3.8 billion active social media users and 4.5 billion people accessing the internet daily. Every year there is a 9% growth in the number of users and half of the internet traffic consists of mostly bots. Bots are mainly categorized into two categories: good and bad bots; good bots consist of web crawlers and chat bots whereas bad bots consist of malicious bots which make up 20% of the traffic, the reason they are not good is that they are used for nefarious purposes, they can mimic human behavior, they can impersonate legal traffic, attack IoT devices and exploit their performance. Among all these concerns, the primary concern is for social media users as they represent a large group of active users on the internet, they are more vulnerable to breach of data, change in opinion based on data. Detection of such bots is crucial to prevent further mishaps. We use supervised Machine learning techniques in this paper such as Decision tree, K nearest neighbors, Logistic regression, and Naïve Bayes to calculate their accuracies and compare it with our classifier which uses Bag of bots' word model to detect Twitter bots from a given training data set.

2.3.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **Information Access and Sharing:** Internet provides a vast repository of information, empowering users to access knowledge and share their ideas, opinions, and experiences with a wide audience.
- **Business Opportunities and Marketing:** Social media platforms offer businesses a powerful medium for marketing and engaging with potential customers, driving sales and brand recognition.

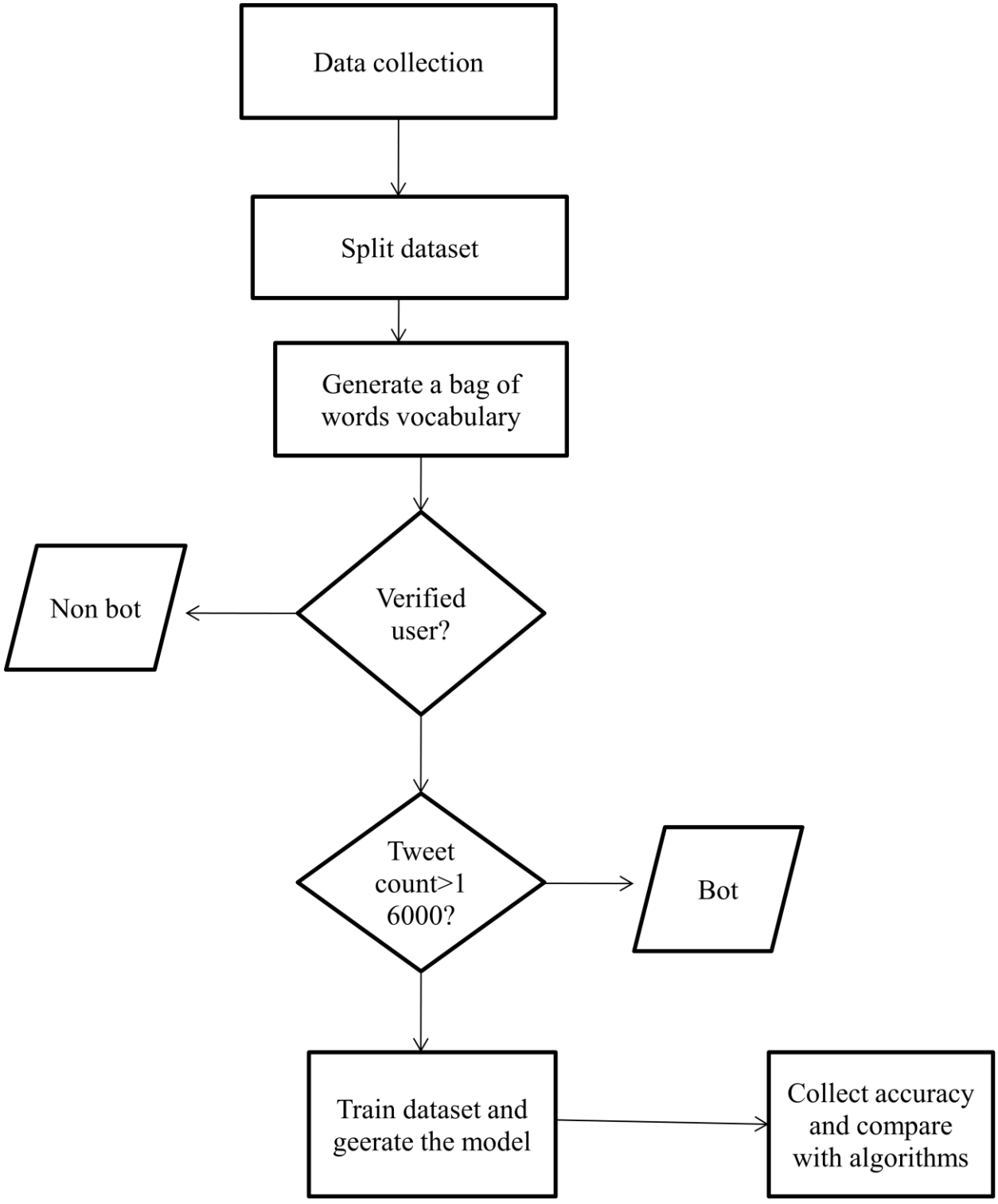
DEMERITS:

- **Privacy and Data Breach:** The vast amount of personal data shared on social media raises concerns about privacy and the potential for data breaches, leading to identity theft and misuse of personal information.
- **Misinformation and Fake News:** The spread of misinformation and fake news is a significant problem on social media, influencing public opinion, causing confusion, and potentially inciting harmful actions.

CHALLENGES:

- **Bot Detection and Mitigation:** As mentioned, distinguishing between good and bad bots is a significant challenge, requiring advanced machine learning techniques to detect and mitigate malicious bot activities.
- **Regulation and Governance:** Creating effective regulations to manage internet and social media usage while preserving freedom of expression is a complex challenge, requiring a delicate balance.
- **Technological Advancements and Security:** Rapid technological advancements make it difficult to keep up with security measures and stay ahead of potential threats, necessitating continuous innovation and adaptation

2.3.3 IMPLEMENTATION:



CHAPTER 3
RESULTS AND DISCUSSION

CHAPTER 3
RESULTS AND DISCUSSION

3,1 PERFORMANCE METRICS:

DETECTION OF MAICIOUS TWITTER BOTS USING MACHINE LEARNING

When evaluating machine learning models such as Decision Tree, Random Forest, and Multinomial Naive Bayes for identifying Twitter bots, several performance metrics can be used to assess their effectiveness. Here are common performance metrics:

Accuracy: Accuracy is the proportion of correctly classified instances (both true positives and true negatives) out of the total instances. It's a general measure of model correctness.

Precision: Precision is the ratio of correctly predicted positive observations (true positives) to the total predicted positives (true positives + false positives). It assesses the accuracy of the positive predictions.

Recall (Sensitivity): Recall is the ratio of correctly predicted positive observations (true positives) to the all observations in actual class (true positives + false negatives). It measures the ability of the model to identify all relevant cases within the data.

F1-Score: F1-Score is the weighted average of precision and recall. It's the harmonic mean of precision and recall and provides a balance between them.

Area Under the ROC Curve (AUC-ROC): AUC-ROC measures the model's ability to distinguish between the positive and negative classes. It represents the area under the receiver operating characteristic curve.

DETECTING MALICIOUS ACTIVITY IN TWITTER USING DEEP LEARNING

When using deep learning techniques to detect malicious activity on Twitter, the following performance metrics can be employed to evaluate the effectiveness of the model:

1.ACCURACY AND PRECISION:

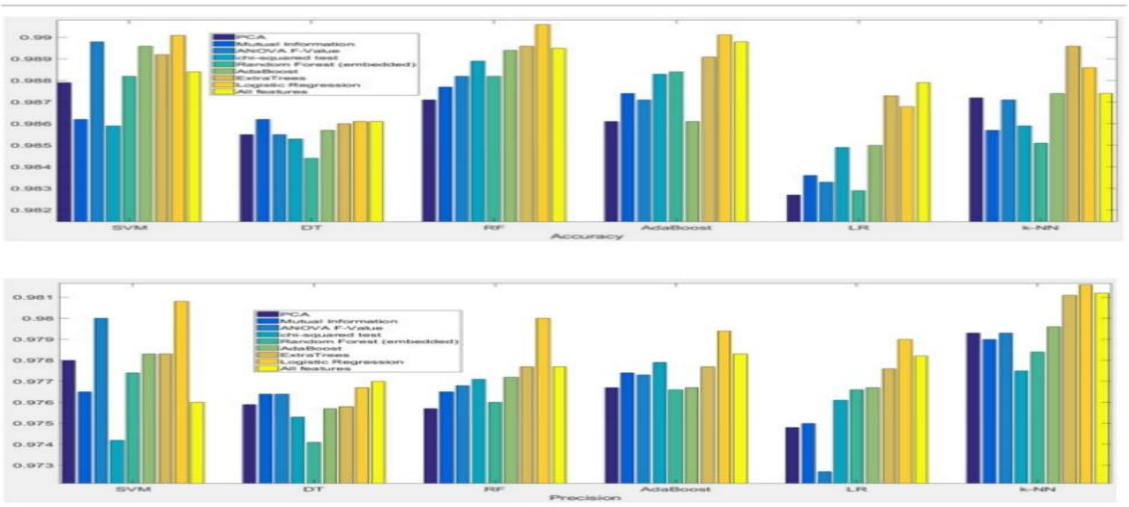


FIG 3.1.1

2.RECALL,F1-SCORE AND AUROC:

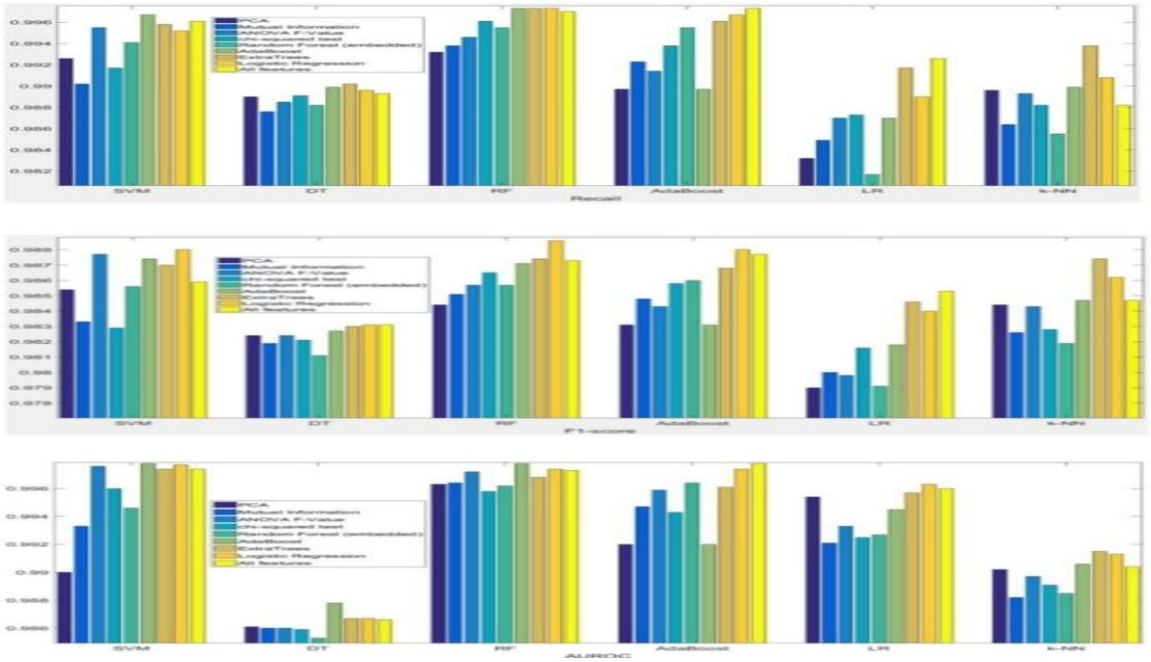


FIG 3.1.2

TWITTER BOT DETECTION USING SUPERVISED MACHINE LEARNING :

Accuracies for various classifier

Accuracy		DT	KNN	LR	NB
90%-10%	Tr	88.17	86.46	66.4	40.30
	Te	87.1	86.42	65	43.57
80%-20%	Tr	88.5	86.1	68.3	40.68
	Te	87.2	82.6	67.4	39.7
70%-30%	Tr	88.2	86	69	40
	Te	88.4	82	67	38
60%-40%	Tr	86.7	86.20	69.65	40.56
	Te	86.6	82.36	67.71	38.04

AUC score for various classifier

Accuracy		DT	KNN	LR	NB
90%-10%	Tr	95.7	94.9	72.5	62.3
	Te	95.5	91.9	67.8	61.4
80%-20%	Tr	95.8	94.7	73.4	62.4
	Te	94.3	88.8	73.3	60.5
70%-30%	Tr	95.70	94.5	71.98	63.6
	Te	94.9	88.2	70.5	61.1
60%-40%	Tr	94.9	94.4	73.5	61.4
	Te	93.0	88.3	69.4	61.6

CHAPTER 6
CONCLUSION

CHAPTER 4

CONCLUSION

CONCLUSION:

- Developing a malicious Twitter bot detection system using machine learning is vital for preserving the integrity of online platforms. It addresses critical issues such as the spread of fake news, cyberbullying, and threats to digital democracy, making online spaces safer and more reliable.
- We created a method for automatically spotting Twitter bots as compared to logistic regression, the best model for train data was VGG19's bag of words approach due to its superior accuracy.
- Consequently the twitter bots were successfully recognized by applying word algorithms to real-time data

REFERENCES

REFERENCE

- Van Der Walt, Estée, & Jan Eloff. "Using machine learning towards detect fake identities: bots vs humans." IEEE Access 6 (2018): 6540-6549.
- Sever Nasim, Mehwish, Andrew Nguyen, Nick Lothian, Robert Cope, & Lewis Mitchell. "Real- time detection about content polluters in partially observable Twitter networks." arXiv preprint arXiv:1804.01235 (2018).
- Khalil, Ashraf, Hassan Hajjdiab, & Nabeel Al- Qirim. "Detecting Fake Followers in Twitter: A Machine Learning Approach." International Journal about Machine Learning & Computing 7,no.6(2017).
- Wetstone, Jessica & Sahil R. Nayyar. “I Spot a Bot Building a binary classifier towards detect bots on Twitter.” (2017).