A Project Report

On

Malware Detection using Machine Learning

Submitted in partial fulfillment of the

requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

DEGREE

Session 2023-24
in

By
Dhairya Bhardwaj(21SCSE1010505)
Vimal Maurya(21SCSE1011714)
Abhijeet Kumar(21SCSE1010930)

Under the guidance of
Dr. Alok Katiyar
SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA

INDIA

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **"Malware detection using machine learning"** in partial fulfillment of the requirements for the award of the B. Tech. (Computer Science and Engineering) submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of November-2023 to January-2024, under the supervision of Dr.Alok Katiyar, Department of Computer Science and Engineering, of School of Computing Science and Engineering , Galgotias University, Greater Noida.

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Dhairya Bhardwaj(21SCSE1010505)

Vimal Maurya(21SCSE1011714)

Abhijeet Kumar(21SCSE1010930)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. Alok Katiyar

CERTIFICATE

This is to certify that Project Report entitled "Malware detection using machine learning " which is submitted by **Dhairya ,Vimal ,Abhijeet** in partial fulfillment of the requirement for the award of degree B. Tech. in Department of School of Computing Science and Engineering Department of Computer Science and Engineering Galgotias University, Greater Noida, India is a record of the candidate own work carried out by him/them under my supervision. The matter embodied in this thesis is original and has not been submitted for the award of any other degree

Signature of Examiner(s) **SignatureSupervisor(s)**

Signature of Program Chair **Signature of Dean**

Date: 28/01/2024

Place: Greater Noida

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. 3rd Year. We owe special debt of gratitude to Professor Dr. Alok katiyar Department of Computer Science & Engineering, Galgotias University, Greater Noida, India *for his constant support and guidance throughout the course of our work. His/Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.*

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Signature:

Name :Dhairya bhardwaj

Roll No.:21SCSE1010505

Date :28/01/2024

Signature:

Name :Vimal Maurya

Roll No.:21SCSE1010714

Date :28/01/2024

Signature:

Name: Abhijeet Kumar

Roll No:21SCSE1010930

Date:28/01/2024

ABSTRACT

Nowadays there are lots of issues faced by internet users, and one of them is viruses and malware issue. In this upcoming generation there are lots of new malware in market that cause lots of harm to internet devices and users' data. One of them is Polymorphic malware, it is a type of new malicious software that is more adaptable than any other previous versions of that malwares. It modifies its signature traits constantly to escape itself to being identified. To identify this type of malware, we use various types of machine learning algorithms. We select that algorithm that gives best and highest accuracy to be used in the system. When we are identified particular malware on the computer system or networks, and hence providing the security to computer networks by different approaches. There are lots of algorithms and techniques we can use in malware detection like DT, CNN and SVM algorithms. Furthermore, different feature representation approaches, such as n-grams, histograms, and embeddings, are explored to enhance the representation and discriminatory capabilities of the machine learning models. The results showed with our dataset and used following techniques to detect malware are:- Random Forest, Decision Tree, Ada boost and Linear Regression.

Key Points With accuracy of following techniques: -

- Decision Tree: 99.08728721477725
- Random Forest: 99.41325606664252
- Ada boost: 98.54762767113364
- Linear Regression :60.5760783194729

These outputs are correctly significant, as we know nowadays malicious soft-ware in internet wo rld and day by day it's becoming common & complex. The objective of this study is to explore the application of machine learning algorithms and models in identifying and classifying malware with high accuracy and efficiency. Various types of machine learning techniques, such as supervised learning, unsupervised learning, and deep learning, are investigated and evaluated in terms of their performance in malware detection. In conclusion, the application of machine learning in malware detection offers promising results in terms of accurately identifying and classifying malware samples. The research presented in this abstract contributes to the ongoing efforts in developing effective and efficient malware detection mechanisms, thereby enhancing the security of computer systems and networks in the face of evolving cyber threats.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1)Due to increasing internet users, there is lots of sensitive information and data that we need to se cure. Nowadays, cyberattacks are most significantly increasing concern in current cyber world. One of them is Malware attack. It is a type of software that develop by hackers by some set of in strictions or program to harm a particular computer, system, organization and any other business.

2)Malwares are of different types like Trojans horses, ransomware, spyware, spyware, adware, rogue software, wipers, scareware, and so on. It's a type of software that runs on others system with thought their permission or consent. In this study, we demonstrated the harmful malware that corrupt pt the files and detect them on systems, and by that we improve the security of computer system and networks. Here, we use lots of machine learning techniques to compute the differences with all methods accuracy and proposed approach. Modules of Malware detection are responsible for collecting data and analyses it for train such that it needs to determine whether it has specific mal ware or not, it can harm our system or not for security concern.

3)Algorithms that have ability or tr ained by means of machine learning, they can improve or enhance their ability to predict the well being of that how they performed previously and make proper changes further. Worldwide, cybe rattacks has now become the serious concern for business, sensitive data of any organization that may steal the confidential information of particular system & harm them. Even, we notice every day we receive the fraud calls, messages and mails on our devices (i.e., smartphones, laptop etc.).

4)Fraudsters uses lots of harmful software in attempt to get access of private networks or systems t o transfer money, steal useful information and harm system. To keep safe networks and system fr om such fraudsters has become urgency for all business owners, private organization and govern ment entities. For that, we require some methods and techniques of machine learning. In this rese arch paper, our aim to discover and detailed info about detecting the malware and techniques to p rotect the private information from hackers by means of data mining and machine learning. 5)We a nalyze signature-based features to enhance the malware detection and classification. Our experim ents have shown proven results on some machine learning algorithms. New Generation malware has become increasingly common and major threat complex, as it uses constantly modification fo r not to be identified. It has become threat to security of modern websites. With the rapid growth of technology and the increasing interconnectedness of computer systems and networks, the threat of malware has

become a significant concern for individuals, organizations, and governments worldwide. Malware, short for malicious software, encompasses a broad range of malicious programs designed to exploit vulnerabilities, steal sensitive information, or disrupt normal system operations. Traditional signature-based detection methods and rule-based systems have struggled to keep pace with the rapidly evolving landscape of malware, necessitating the exploration of more advanced and adaptive approaches.

6) In recent years, machine learning has emerged as a powerful tool for detecting and combating malware. Machine learning algorithms possess the ability to analyze large volumes of data, learn patterns and behaviours, and make intelligent predictions. Leveraging this capability, researchers and security experts have turned to machine learning techniques to develop robust and efficient malware detection systems.
The use of machine learning in malware detection offers several advantages over traditional approaches. Firstly, machine learning models can automatically learn and adapt to evolving malware patterns, enabling them to detect previously unseen and zero-day threats. This adaptability helps address the challenge posed by polymorphic and metamorphic malware, which continuously modify their code to evade detection.

7)Secondly, machine learning enables the analysis of a multitude of features and attributes, capturing both static and dynamic characteristics of malware. These features can include file attributes, network behaviour, system calls, opcode sequences, and more. By considering a comprehensive range of features, machine learning models can effectively capture the unique characteristics of different malware families and accurately classify them.

Moreover, machine learning techniques can detect subtle patterns and correlations that might not be easily discernible to human analysts. This capability allows for the identification of complex and sophisticated malware campaigns that may employ obfuscation techniques to evade detection.

8) The objective of this paper is to provide an overview of the research and advancements in the field of malware detection using machine learning. We will explore the different machine learning algorithms and models employed for malware detection, including Random forest, Ada boost, regression tree and Decision tree.

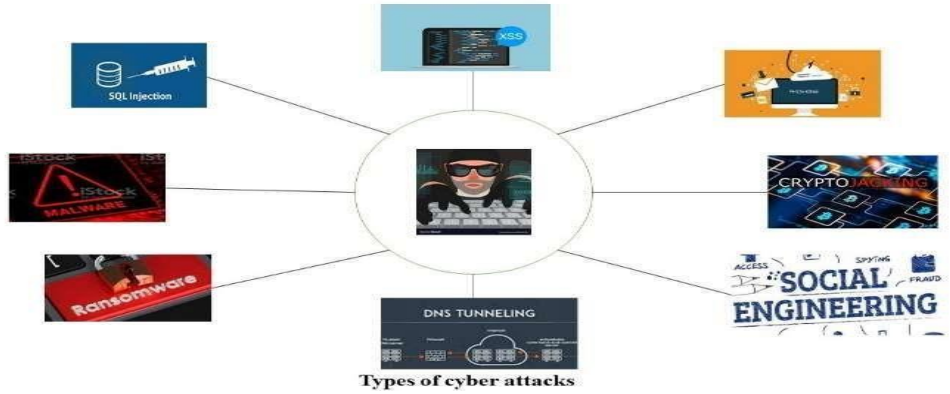

Fig 1.1 (Types of cyberattacks)

below depicts the types of cyberattacks in this digital world or cyberworld. Malware is a software developed by hackers or fraudsters with aim to harm other's networks and system without their consent. It has been increasing in different sectors like medical gear, IOT devices and environmental and industrial systems. Currently, Malware in the cyber is very hard to detect as it constantly updates its behavior. To overcome this situation and tackle this, we require wider range of defensive actions. In static analysis, it examines harmful files without actually running them on system. But dynamic analysis first takes account of tracking data flows, checking functions work, and take action to monitor their actions. Machine learning algorithms helps in static and behavior of particular malware, to identify the structure of malware, to increasing identify complex malware that can use signature-based method. But we can't rely singly on signature-based techniques of machine learning. Nowadays we have lots of successful malware detection techniques to detect malware.

1.1) Future of Machine learning in Malware detection

The future of machine learning in malware detection is highly promising, with ongoing advancements and research that will shape the field in several key ways:

1). Improved Accuracy: Machine learning algorithms will continue to evolve, leading to improved accuracy in malware detection. Researchers will explore new techniques, including ensemble learning, transfer learning, and deep learning architectures, to enhance the detection capabilities of models. By leveraging larger and more diverse datasets, models can learn from a wider range of malware samples and detect emerging threats more effectively.

2). Real-time Detection: The ability to detect malware in real-time is crucial in mitigating its impact. Future advancements will focus on developing lightweight and efficient machine learning models that can operate in real-time, providing instantaneous detection and response. This will enable the proactive identification and containment of malware before it can cause substantial damage.

3). Adversarial Detection: Adversarial attacks pose a significant challenge to machine learningbased malware detection systems. In response, researchers will explore techniques such as adversarial training and generative adversarial networks (GANs) to enhance the robustness of models against adversarial manipulation. By understanding and countering these attacks, machine learning models can become more resilient and reliable in the face of evolving threats.

4). Explainability and Interpretability: The lack of transparency in machine learning models has been a concern in the context of cybersecurity. Future research will focus on developing techniques to enhance the explainability and interpretability of machine learning models in malware detection. This will enable security analysts to understand the reasoning behind model decisions, aiding in the identification of false positives/negatives and improving overall trust and adoption of machine learning solutions.

5). Zero-day Threat Detection: Zero-day threats, which exploit previously unknown vulnerabilities, pose a significant challenge to traditional signature-based detection methods. Machine learning models have the potential to detect such threats by leveraging anomaly detection techniques and behavior analysis.

6). Integration with Ecosystems: Machine learning-based malware detection systems will become increasingly integrated into larger cybersecurity ecosystems. This integration will enable enhancing the overall threat intelligence and response capabilities. Integration with cloud-based security platforms and IoT security systems will also play a crucial role in providing comprehensive protection across diverse environments.

7). Continuous Learning and Adaptation: Malware evolves rapidly, requiring detection systems to adapt continuously. Future machine learning approaches will focus on developing selflearning systems that can dynamically update and retrain models based on new malware samples and evolving attack patterns. This continual learning will ensure that machine learning models remain effective and up-to-date in combating emerging threats.

CHAPTER 2

LITERATURE SURVEY

The increasing number of computers, smartphones and other Internet-enabled devices gets more cyber-attack and more common threat. A large number of malware detection methods are made to tackle the issue of malware attacks activity. Common machine learning -based malware detection methods take much processing time, but help in merging new malware. Feature engineering may be less perspective to increasing demand of machine learning algorithms, such as deep learning. In this research, we have experiment various malware detection techniques. Researchers has developed the ways to check data samples by machine learning nad deep learning.

Armaan (2021) described and tested the accuracy of many techniques' models. As we all know that for an application to function there must have data to operate. At this time, the proliferation of malicious software program poses a sizeable chance to worldwide stability. In the 1990s, because the quantity of interconnected computer systems exploded, so did the superiority of malicious software program [23], which in the end brought about the widespread distribution of malware. Multiple protecting measures had been created in reaction to this phenomenon. Unfortunately, present day safeguards can't maintain up with current threats that malware authors have created to thwart safety programs. In current years, researchers' cognizance on malware detection studies has shifted in the direction of ML set of rules strategies. In this studies paper, we gift a protecting mechanism that evaluates 3 ML set of rules processes to malware detection and chooses the maximum suitable one.

S.No	Specs	Score
1	ref-cycles	1.296660e+07
2	stalled-cycles-backend-percent	3.017643e+06
3	bus-cycle	1.494372e+06
4	stalled-cycles-frontend-percent	9.583614e+05
5	cache-references	1.233573e+05

6	Instructions-per-cycle	6.668084e+04
7	cache-misses-percent	9.394599e+03
8	bracnhes	1.923075e+07
9	page-faults	3.900794e+03
10	Branch-misses-percent	2.306111e+03

Table:2.1

Chowdhury (2018) proposed a new malware detection method that uses machine learning classification technique.N-gram and API call capabilities were prior into this approach. Experimental evaluation confirmed the efficiency and dependability of this proposed technique. they are also focusing on merging all possible features to increase detection while decreasing negatives. Experimental results by Chowdhury approach are shown in table below. his approach was clearly clever.

classifier	Accuracy
Decision Tree	99.08%
Random Forest	99.41%
Ada boost	98.54%
Linear Regression	60.57%

Table:2.2

The technique wishes a workaround this is adaptable sufficient to address non-widespread data. To successfully control and save you destiny assaults, we need to examine malware and create new policies and styles withinside the shape of advent of malware kind. To locate styles, IT safety professionals may also use malware evaluation equipment. This equipment assists screen safety signals and save you malware attacks. If malware is dangerous, we need to put off it earlier than it transmits its contamination any further. Malware evaluation is turning into an

increasing number of famous because it facilitates corporations reduce the consequences of the developing quantity of malware threats and the growing complexity of the approach's malware may be used to attack.

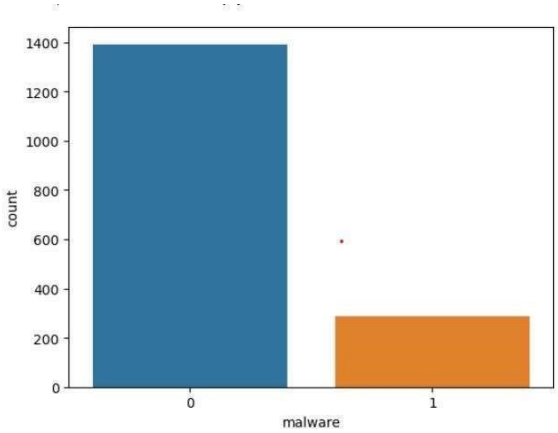

Fig 2.1 (Malware detection graph)

Malware continues to develop and flows at an increasing rate. Nur (2019) compared 3 classifiers techniques to analyze and check the accuracy of the ML classifier as it uses static analysis to extract features based on PE information. Malicious programs and their threats, or singly "malware" became increasingly common and internet continues to emerged. Their rapid growth over the internet has provided hackers or fraudsters with permission to a wide number of malware generation tools. Every day, malware reach increasing. Our study focused on machine learning techniques how they classify and works. It's recommended that ML systems be trained and tested to check a file has malware or not. Experimental results verifies that the random forest technique is preferred for data classify, has an accuracy of 99.41%. The main advantage is that user install a file or software as it will be checked before opening it and maintain validity.

RESEARCH PROBLEM

1)We can detect the malware potential using static and dynamic analysis.

2) In Static analysis, first a virus is disassemble using reverse engineering method, that focused on breaking malware binaries to find harmful strings.

3)However, dynamic analysis the monitoring of the dangerous software even if it operates in a controlled environment, such as virtual system.

4) Both methods have their own advantages and disadvantages; however, both are used of analyzing malware, both are best to use. and after reducing the number of dangerous features increase the accuracy of malware detection.

5) Now the researcher has more time to collect data and analyze it. But we have a concern most of characteristics are used to detect malware whereas fewer can also work same on it.

6)The manner of selecting which malicious functions to enforce starts off evolved with coming across feasible methods or algorithms.

7) We want answers which could each locate malware that has in no way been seen earlier than and substantially lessen the wide variety of traits which might be presently had to do so.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

The system design section provides a comprehensive overview of the architecture and components that constitute the malware detection system using machine learning. The system architecture encompasses three main components: the malware analysis engine, machine learning model(s), and a user interface for monitoring and reporting.

The malware analysis engine is responsible for scrutinizing files, network traffic, and system behavior to identify potential threats. This includes extracting relevant features from files, monitoring network packets for suspicious activities, and analyzing system processes for anomalies. The machine learning component involves both model training and real-time prediction. The system must support the training of machine learning models using labeled datasets, periodically updating them to enhance detection accuracy. Real-time prediction utilizes these trained models to assign probability scores to detected threats, ensuring swift and accurate identification of malicious activities.

The user interface is designed to facilitate user interaction. It includes a dashboard for monitoring system health and detected threats in real-time, promoting quick decision-making. Reporting functionalities generate detailed reports on identified malware, providing insights into threat severity, recommended actions, and overall system performance

CHAPTER 4

SYSTEM DESIGN AND METHODOLOGY

This project introduces many steps and components of machine learning flowchart for malware detection and classification, checks the challenges and have some limitations in its workflow and uses the latest innovations and trends in the field on deep learning techniques. Proposed research method of this study is stated below. To understand the complete concept of this proposed machine learning techniques for malware detection is depicted in figure below. has full flowchart start to finish.

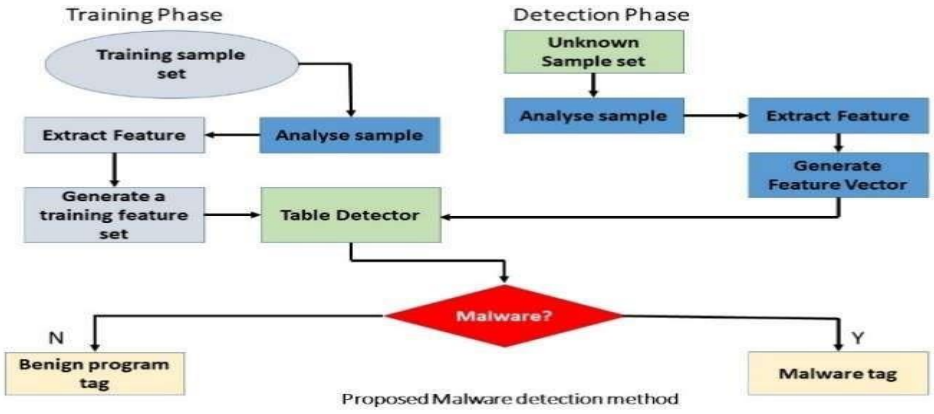

Figure. Proposed ML malware detection method

4.1) Dataset

This study relied completely on data provided by the Github for Cybersecurity. The dataset has many data files that comprise log data for different types of malware . These recovered log features may be used to train a broad diversity of models or algorithms. There are many distinct malware families were found in the samples. More than 138047 data points from different locations were included; the dataset had 57 columns and 138047 rows.

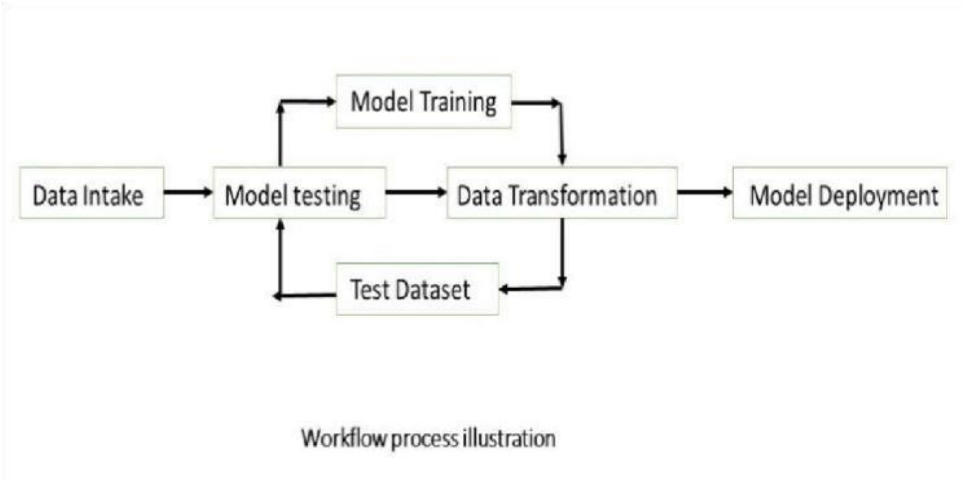

Fig 4.1 (Workflow process illustration)

4.2) Pre-processing

Data were stored in the file system as a binary code and files were unprocessed. As they are prepared in advance of our research. When we up pack the data flees, we require a protected environment like virtual machine (VM). Paid software which automatically unpack the compressed executables.

4.3) Features Extraction

Current generation datasets contain thousands of features. In recent years, features of datasets have grown, now it is required to innovate new machine learning algorithms as previous has become overfit. For this we develop or select fewer smaller set of features from large sets of features, it also works well to maintain accuracy. This research paper is to redefine the current dataset of both static and dynamic features, keep them which are helpful and eliminate that were not valuable for data analysis.

4.4) Feature Selection

Feature selection is performed after the feature extraction process which helps in discovery of more features. Feature selection has many advantages as it enhance the accuracy, simplify the model, and reduce overfitting, it's because it recognizes new features from a pool. As we know that Researchers bused many techniques in order to identify harmful code in software. But now feature selection technique is extremely used, also in this study.

SOURCE CODE FOR MALWARE DETECTION USING PYTHON

```
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import seaborn as sns
#from sklearn.svm import SVC
#from sklearn.model_selection import KFold
from sklearn import preprocessing
import matplotlib.pyplot as plt
import sklearn.ensemble as ek
from sklearn import tree, linear_model
from sklearn.feature_selection import SelectFromModel
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix
from sklearn.pipeline import make_pipeline
from sklearn import preprocessing
from sklearn import svm
from sklearn.linear_model import LinearRegression
import pandas_datareader as data


df=pd.read_csv('BT2051 Code.ipynb')
dataset = pd.read_csv('Malware_Detection_data.csv',sep='|', low_memory=False)
data=data.sample(frac=1).reset_index(drop = True)
dataset.head()
target_count = data.legitimate.value_counts()
print('Class 0:', target_count[0])
print('Class 1:', target_count[1])
count_class_0, count_class_1 = dataset.legitimate.value_counts()
df_class_0 = data[data['legitimate'] == 0]

df_class_1 = data[data['legitimate'] == 1]

df_class_1_over = df_class_1.sample(count_class_0, replace=True)

df_test_over = pd.concat([df_class_0, df_class_1_over], axis=0)

df_test_over.shape

sns.countplot(x='legitimate',data=df_test_over)

X=df_test_over.iloc[:,df_test_over.columns !='legitimate']

Y=df_test_over.iloc[:,df_test_over.columns =="legitimate"]
```

```
X.head()

from sklearn.feature_selection import SelectKBest

from sklearn.feature_selection import chi2

bestfeatures = SelectKBest(score_func=chi2, k=10)

fit = bestfeatures.fit(x,y)

dfscores = pd.DataFrame(fit.scores_)

dfcolumns = pd.DataFrame(X.columns)

featureScores = pd.concat([dfcolumns,dfscores],axis=1)

featureScores.columns = ['Specs','Score']

featureScores.nlargest(10,'Score')

from sklearn.ensemble import ExtraTreesClassifier

import matplotlib.pyplot as plt

model = ExtraTreesClassifier()

model.fit(X,Y)

print(model.feature_importances_) #use inbuilt class feature_importances of tree based classifiers

#plot graph of feature importances for better visualization

feat_importances = pd.Series(model.feature_importances_, index=X.columns)

feat_importances.nlargest(10).plot(kind='barh')

plt.show()

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,Y, test_size = 0.2, random_state=0)

from sklearn.tree import DecisionTreeClassifier

model = {"DecisionTree":DecisionTreeClassifier(max_depth=10),
```

```
        "RandomForest":ek.RandomForestClassifier(n_estimators=50),
        "Adaboost":ek.AdaBoostClassifier(n_estimators=50),
        "LinearRegression":LinearRegression()
}
results = {}
for algo in model:
    clf = model[algo]
    clf.fit(X_train,y_train)
    score = clf.score(X_test,y_test)
    print ("%s : %s " %(algo, score))
    results[algo] = score
```

CHAPTER 5

IMPLEMENTATION AND RESULTS

Training and testing are the two main phases of the classification process. Both harmful and safe files was sent, to train a system. Using Learning Algorithm automated classifiers were taught. All the classifiers become smarter i.e., Random Forest, Decision tree, Ad boost, Linear Regression with every set of data is annotated. During the testing phase, Collection of new files were sent by classifier some of them are harmful and some are not; It is determined by classifier whether the files were clean or infected. Figure Illustrate that Random Forest has highest accuracy (99.41%) and Decision tree (99.08%), that Linear Regression has lowest accuracy (60.57%). From the confusion matrix it is cleared that RF has highest accuracy than all the others machine learning algorithm or classifiers.

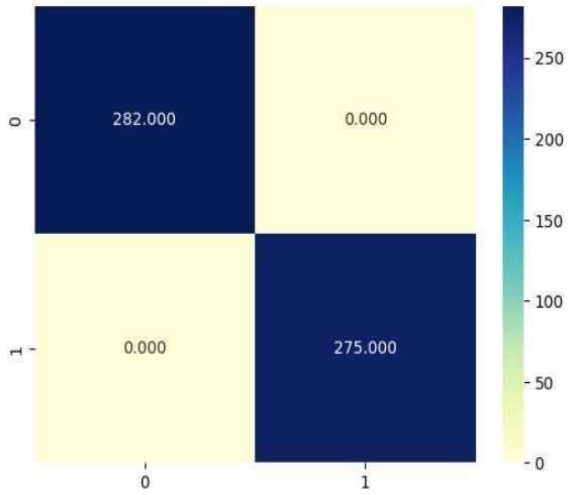

Fig 5.1(Confusion Matrix)

Using gathered malware and clean ware our suggested method for malware detection and categorization was evaluated experimentally. To examine malware and its characteristics we use classifiers (RF, DT, Ad boost, LS) or supervised machine learning. From the analysis of table 2 results, we get the result of classifiers accuracy as RF=99.41%, Ada boost= 98.54%, Decision tree=99.08%, and Linear regression= 60.57% from this it showed that RF is the best and optimal model for malware detection strategy and the second most optimal model for malware detection strategy is Decision tree with accuracy of 99.08% and Ada boost with accuracy of 98.54%.

CHAPTER 6

CONCLUSION

This paper shows that academician have contemporary shown increasing interest in machine learning algorithms solution for malware identification. We purposed three protective methods to detect malware and the most accurate one. The results show that compared with other classifiers, DT (99.08%), RF (99.41%), Ad boost (98.54%) and Linear regression (60.57%) performed well in term of detection accuracy. In this experiment, we quantify the Machine learning detection accuracy of classifiers which have the highest accuracy by comparison of other Machine learning (ML) classifiers. As a result of our experimental effort Machine learning algorithms can identify the different malwares now and we have seen that Decision Tree (DT) have the (99.41%) of accuracy which is highest in comparison of all other classifiers we have evaluated. Potentially to provide the accurate and highest accuracy of malware detection based on our cautious selected dataset have shown assurance in our experimental findings. And we have four Machine Learning Models (Decision Tree, Random Forest, Ad boost, Linear Regression) which were trained and checked their efficiency using the given dataset.

References

1) Nikam, U.V.; Deshmuh, V.M. Performance evaluation of machine learning classifiers in malware detection. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022; pp. 1–5. [CrossRef]

2) Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing. EAI Endorsed Trans. Create. Tech. 2022, 9, 172814. [CrossRef]

Sethi, K.; Kumar, R.; Sethi, L.; Bera, P.; Patra, P.K. A novel machine learning based malware detection and classification framework. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–13.

3) Abdulbasit, A.; Darem, F.A.G.; Al-Hashmi, A.A.; Abawajy, J.H.; Alanazi, S.M.; AlRezami, A.Y. An adaptive behavioral-based increamental batch learning malware variants detection model using concept drift detection and sequential deep learning. IEEE Access 2021, 9, 97180–97196. [CrossRef]

4) Feng, T.; Akhtar, M.S.; Zhang, J. The future of artificial intelligence in cybersecurity: A comprehensive survey. EAI Endorsed Trans. Create. Tech. 2021, 8, 170285. [CrossRef]

5) Sharma, S.; Krishna, C.R.; Sahay, S.K. Detection of advanced malware by machine learning techniques. In Proceedings of the SoCTA 2017, Jhansi, India, 22–24 December 2017.

6) Chandrakala, D.; Sait, A.; Kiruthika, J.; Nivetha, R. Detection and classification of malware. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–3. [CrossRef]

7) Zhao, K.; Zhang, D.; Su, X.; Li, W. Fest: A feature extraction and selection tool for android malware detection. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 714–720.

8) Akhtar, M.S.; Feng, T. Detection of sleep paralysis by using IoT based device and its relationship between sleep paralysis and sleep quality. EAI Endorsed Trans. Internet Things 2022, 8, e4. [CrossRef]

9) Gibert, D.; Mateu, C.; Planes, J.; Vicens, R. Using convolutional neural networks for classification of malware represented as images. J. Comput. Virol. Hacking Tech. 2019, 15, 15–28. [CrossRef]

10) Firdaus, A.; Anuar, N.B.; Karim, A.; Faizal, M.; Razak, A. Discovering optimal features using static analysis and a genetic search based method for Android malware detection. Front. Inf. Technol. Electron. Eng. 2018, 19, 712–736. [CrossRef]

11) Dahl, G.E.; Stokes, J.W.; Deng, L.; Yu, D.; Research, M. Large-scale Malware Classification Using Random Projections And Neural Networks. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing-1988, Vancouver, BC, Canada, 26–31 May 2013; pp. 3422–3426.

12) Akhtar, M.S.; Feng, T. An overview of the applications of artificial intelligence in cybersecurity. EAI Endorsed Trans. Create. Tech. 2021, 8, e4. [CrossRef]

Akhtar, M.S.; Feng, T. A systemic security and privacy review: Attacks and prevention mechanisms over IOT layers. EAI Endorsed Trans. Secur. Saf. 2022, 8, e5. [CrossRef]

13) Anderson, B.; Storlie, C.; Lane, T. "Improving Malware Classification: Bridging the Static/Dynamic Gap. In Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence (AISec), Raleigh, NC, USA, 19 October 2012; pp. 3–14.

14) Varma, P.R.K.; Raj, K.P.; Raju, K.V.S. Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms. In Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 10–11 February 2017; pp. 294–299.

15) Akhtar, M.S.; Feng, T. Comparison of classification model for the detection of cyber-attack using ensemble learning models. EAI Endorsed Trans. Scalable Inf. Syst. 2022, 9, 17329. [CrossRef]

16) Rosmansyah, W.Y.; Dabarsyah, B. Malware detection on Android smartphones using API class and machine learning. In Proceedings of the 2015 International Conference on Electrical Engineering and Informatics (ICEEI), Denpasar, Indonesia, 10–11 August 2015; pp. 294–297.

17) Tahtaci, B.; Canbay, B. Android Malware Detection Using Machine Learning. In Proceedings of the 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, 15–17 October 2020; pp. 1–6.

18) Baset, M. Machine Learning for Malware Detection. Master's Dissertation, Heriot Watt University, Edinburg, Scotland, December 2016. [CrossRef]

19) Akhtar, M.S.; Feng, T. Deep learning-based framework for the detection o cyberattack using feature engineering. Secur. Commun. Netw. 2021, 2021, 6129210. [CrossRef]

20) Altaher, A. Classification of android malware applications using feature selection and classification algorithms. VAWKUM Trans. Comput. Sci. 2016, 10, 1. [CrossRef]

Chowdhury, M.; Rahman, A.; Islam, R. Malware Analysis and Detection Using Data Mining and Machine Learning Classification; AISC: Chicago, IL, USA, 2017; pp. 266–274.