International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 04, April 2023)

An Efficient Machine Learning Technique for Android Malware Prediction

Shivani Tiwari[1], Prof. Roopali Soni[2]

[1]*Research Scholar,* [2]*Associate Professor, Department of Computer Science Engineering, Oriental College of Technology, Bhopal, India*

***Abstract*— Malicious software that targets a particular kind of device, in this case Android devices, is referred to as Android malware. Malware is able to flourish in an environment that is made possible by Android's less secure platform, which includes the Play Store, from which users may download programs, and the ability of Android users to side load material from the internet. The prediction of android malware using machine learning techniques is presented in this research, along with performance improvements. Python Sypder 3.7 is the program that is used to carry out the simulation. The results of the simulation demonstrate an increase in the quality of the performance metrics.**

***Keywords*— Android, SVM, MLP, Malware, Artificial Intelligence, Secuiry, Attack, Cyber.**

I. INTRODUCTION

Malicious software that targets wirelessly-enabled devices is referred to as mobile malware. This kind of software is capable of bringing down the system and causing the loss or disclosure of sensitive information. As wireless phones and PDA networks have been more widely used and as their sophistication has increased, it has become harder to guarantee their safety and security against electronic assaults in the form of viruses and other forms of malware. There are many different kinds of malicious software, such as computer viruses, worms, Trojan horses, ransomware, spyware, adware, rogue software, wiper, and scareware. Antivirus software, firewalls, applying regular patches to reduce the risk of zero-day attacks, securing networks from intrusion, having regular backups, and isolating infected systems are some of the most effective defense strategies against malware. However, the effectiveness of these defense strategies varies depending on the type of malware being defended against. Malware is now being developed with the intention of evading the detection algorithms of antivirus software.

Figure 1: Android Malware

The proliferation of applications for the internet of things (IoT) in several spheres of life, including as detection, medical care, remote monitoring, and other areas, is causing the world to undergo profound transformation. Android devices and apps are collaborating closely in order to bring the Internet of Things (IoT) into reality. Malware and other forms of online attack on mobile devices running the Android operating system have been on the rise recently. In addition, the widespread usage of the Android operating system in Internet of Things devices makes it difficult to perform tests to determine whether or not a piece of malware is there. This work provides an innovative structure that consolidates the advantages of both AI processes and blockchain technology to work on the malware location for Android IoT devices. This structure was created as part of this body of work. Mobile malware is one of the most significant challenges to computer security in the modern day. In addition, malicious software for mobile devices is constantly being updated, which results in the introduction of new threats. Yet, whereas current security solutions protect mobile devices, in general, against known threats, these devices are still susceptible to hazards that have not yet been discovered. The use of evolutionary computation techniques is investigated in this study.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 04, April 2023)

These techniques are used for developing new variants of mobile malware that are able to successfully evade anti-malware systems that are based on static analysis, as well as for developing better security solutions against them automatically. The co-evolutionary arms race mechanism has long been seen as a possible choice for the development of a more robust system against new assaults and for the testing of the system.

II. Proposed Methodology

Focusing on scientific approach to assess how assistance is acknowledged in the public arena, we created malware prediction model displaying framework. Machine learning classifiers incorporate SVM and MLP are utilized in the planning of the framework.

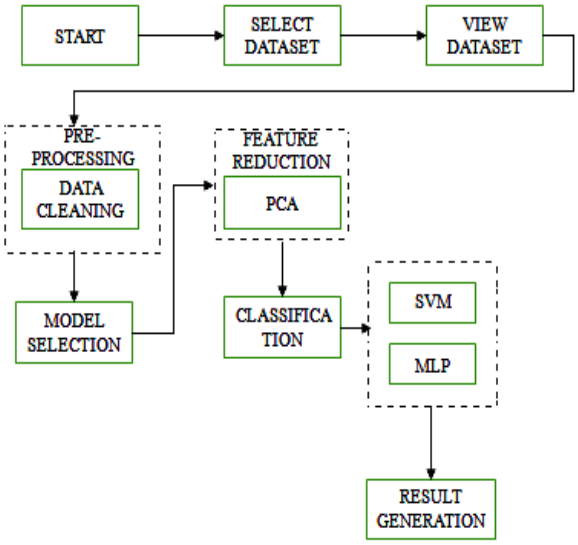


Figure 2: Flow Chart

The proposed model is introduced to overcome all the disadvantages that arise in the existing system. This system will increase the accuracy of the machine learning results by detecting malware from android dataset using machine learning algorithm. It enhances the performance of the overall classification results. Predict the malware from android data is to find the accuracy more reliable. The main objective is to develop the AIML based model to prediction of the android mobile malware prediction with the improvement in the performance parameters. Therefore an efficiently detect the malware in android from the dataset is the prime objective of this research work.

Support Vector Machine- SVM is a popular machine learning algorithm that can be used for Android malware prediction. SVM is a supervised learning algorithm that can be used for classification and regression tasks. In the context of Android malware prediction, SVM can be used to classify mobile applications as either malicious or benign based on a set of features extracted from the application.

To use SVM for Android malware prediction, first extract a set of features from each application. These features can include information such as permissions requested, API calls made, and other metadata about the application. Once the features are extracted, can use them to train an SVM model. The SVM model can then be used to classify new applications as either malicious or benign.

There are several advantages of using SVM for Android malware prediction. SVM is a powerful and accurate machine learning algorithm that can handle high-dimensional data and is relatively immune to overfitting. It can also work well with small datasets, which is often the case with Android malware prediction.

Multilayer Perceptron- MLP is another popular machine learning algorithm that can be used for Android malware prediction. MLP is a type of artificial neural network that is capable of learning non-linear relationships between input features and output labels.

To use MLP for Android malware prediction, we need to first extract a set of features from each application, similar to the SVM approach. These features can include information such as permissions requested, API calls made, and other metadata about the application. Once the features are extracted, they can be used as input to an MLP model. The MLP model can then be trained to predict whether a given application is malicious or benign.

One advantage of using MLP for Android malware prediction is its ability to capture complex relationships between input features. MLP can also handle high-dimensional data and can work well with small datasets, which is often the case with Android malware prediction.

III. Result And Analysis

The implementation of the proposed algorithm is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like SVM and MLP etc.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 04, April 2023)

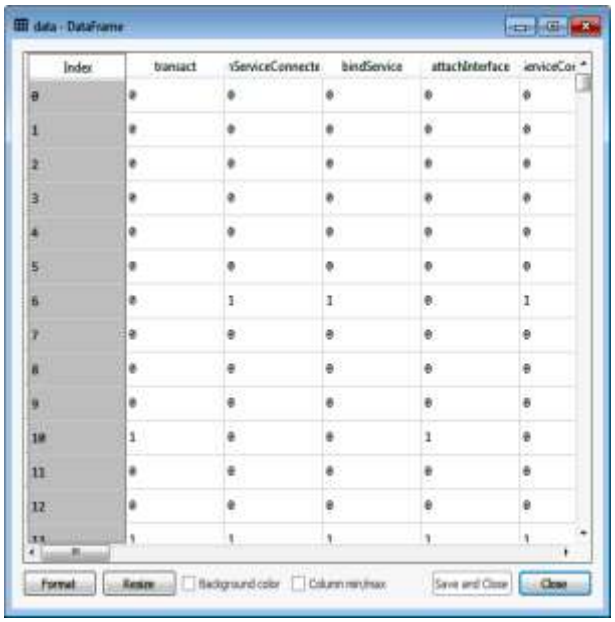


Figure 3: Dataset

Figure 3 is is showing the dataset in the python environment. The dataset have various numbers of rows and column. The features name is mention in each column.

Figure 4: Train data

Figure 4 is showing the x train of the given dataset. The given dataset is divided into the 70-80%% part into the train dataset.

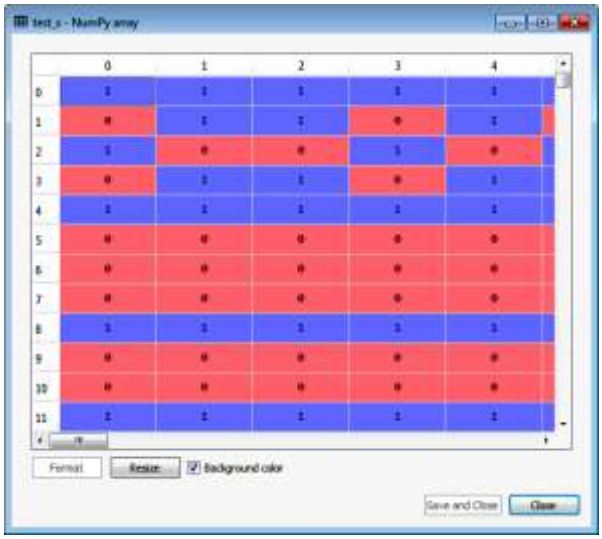
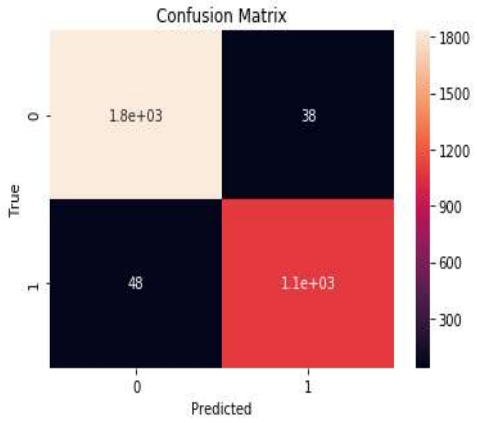


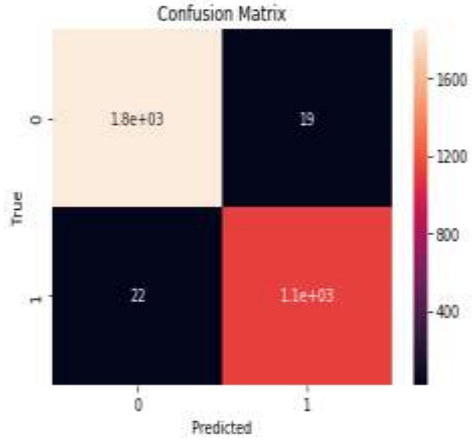
Figure 5: Test Data

Figure 5 is showing the x test of the given dataset. The given dataset is divided into the 20-30% part into the train dataset.



(a)

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 04, April 2023)



(b)

Figure 6: Confusion Matrix (a) SVM (b) MLP

Figure 6 is showing the CM matrix of the SVM and MLP classification technique. It is an N x N matrix used for evaluating the performance of a classification model.

Table 1:
Result Comparison

Sr. No.	Parameters	Previous work [1]	Proposed Work
1	Precision	97.04 %	99 %
2	Recall	96.94 %	98 %
3	F_Measure	96.99 %	99 %
4	Accuracy	94.92 %	99.00 %
5	Error Rate	5.08 %	1 %
6	Specificity	84.08 %	99.4 %
7	Area under the ROC Curve (AUC)	90.45 %	99.83 %

IV. Conclusion

This research presents a machine learning approach for predicting malware on android devices, along with a study of performance. For the purpose of this investigation, machine learning classifiers are used to make predictions about android malware. The information about Android malware is used as input data and is placed via the pre-processing procedure.

In the pre-processing procedure, the dataset should be cleaned up, and the label encoding should be applied. The previous level of accuracy, which was obtained at 94.92%, has been improved to 99.00% by the work that is being suggested. The suggested method has an error rate of 1%, while the present approach has an error rate of 5.08%.

REFERENCES

[1] H. Zhu, Y. Li, R. Li, J. Li, Z. You and H. Song, "SEDMDroid: An Enhanced Stacking Ensemble Framework for Android Malware Detection," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 984-994, 1 April-June 2021, doi: 10.1109/TNSE.2020.2996379.

[2] A. Alzubaidi, "Recent Advances in Android Mobile Malware Detection: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 146318-146349, 2021, doi: 10.1109/ACCESS.2021.3123187.

[3] H. Kato, T. Sasaki and I. Sasase, "Android Malware Detection Based on Composition Ratio of Permission Pairs," in IEEE Access, vol. 9, pp. 130006-130019, 2021, doi: 10.1109/ACCESS.2021.3113711.

[4] C. Li et al., "Backdoor Attack on Machine Learning Based Android Malware Detectors," in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2021.3094824.

[5] L. Gong, Z. Li, H. Wang, H. Lin, X. Ma and Y. Liu, "Overlay-based Android Malware Detection at Market Scales: Systematically Adapting to the New Technological Landscape," in IEEE Transactions on Mobile Computing, doi: 10.1109/TMC.2021.3079433.

[6] I. Almomani et al., "Android Ransomware Detection Based on a Hybrid Evolutionary Approach in the Context of Highly Imbalanced Data," in IEEE Access, vol. 9, pp. 57674-57691, 2021, doi: 10.1109/ACCESS.2021.3071450.

[7] F. Mercaldo and A. Santone, "Formal Equivalence Checking for Mobile Malware Detection and Family Classification," in IEEE Transactions on Software Engineering, doi: 10.1109/TSE.2021.3067061.

[8] L. N. Vu and S. Jung, "AdMat: A CNN-on-Matrix Approach to Android Malware Detection and Classification," in IEEE Access, vol. 9, pp. 39680-39694, 2021, doi: 10.1109/ACCESS.2021.3063748.

[9] L. Gong et al., "Systematically Landing Machine Learning onto Market-Scale Mobile Malware Detection," in IEEE Transactions on Parallel and Distributed Systems, vol. 32, no. 7, pp. 1615-1628, 1 July 2021, doi: 10.1109/TPDS.2020.3046092.

[10] W. Yuan, Y. Jiang, H. Li and M. Cai, "A Lightweight On-Device Detection Method for Android Malware," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 9, pp. 5600-5611, Sept. 2021, doi: 10.1109/TSMC.2019.2958382.

[11] K. Liu, S. Xu, G. Xu, M. Zhang, D. Sun and H. Liu, "A Review of Android Malware Detection Approaches Based on Machine Learning," in IEEE Access, vol. 8, pp. 124579-124607, 2020, doi: 10.1109/ACCESS.2020.3006143.

[12] D. Li and Q. Li, "Adversarial Deep Ensemble: Evasion Attacks and Defenses for Malware Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3886-3900, 2020, doi: 10.1109/TIFS.2020.3003571.