IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12746

Android Malware Prediction using Efficient Learning Approach for Cybersecurity

Preeti Simolya[1], Prof. Sushma Kushwaha[2]

M.Tech Scholar, Department of Computer Science and Engineering, Swami Vivekanand College of Science and Technology, Bhopal, India[1]

Assistant Professor, Department of Computer Science and Engineering, Swami Vivekanand College of Science and Technology, Bhopal, India[2]

Abstract— Android malware is a term used to refer to malicious software that is designed to infect a certain kind of device, in this instance smartphones that run the Android operating system. Malware is able to thrive in an environment that is made feasible by Android's less secure platform. This platform includes the Play Store, from which users are able to download programmes, as well as the ability for Android users to side load content from the internet. Both of these features allow for the distribution of malware. This study presents the use of machine learning methods for the purpose of predicting malicious android software. Additionally, performance enhancements are given. The simulation is run using Python Sypder 3.7, which is the programme that is used to carry it out. The outcomes of the simulation indicate that there has been an improvement in the standard of the performance indicators.

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

I. INTRODUCTION

Mobile malware is a term used to describe malicious software that is designed to infect wirelessly equipped mobile devices. This kind of software has the potential to put the system to a halt as well as result in the loss of confidential information or its public dissemination. Since the usage of wireless phones and PDA networks has gotten more widespread and since these technologies have become more advanced, it has become more difficult to ensure their safety and security against electronic attacks in the form of viruses and other types of malware. There is a wide variety of harmful software, including computer viruses, worms, Trojan horses, ransomware, spyware, adware, rogue software, wiper, and scareware. Some of the more common types of malicious software are included below.

Antivirus software, firewalls, applying frequent updates to limit the risk of zero-day attacks, safeguarding networks from intrusion, having regular backups, and isolating infected systems are some of the most effective defence measures against malware. Other techniques include securing networks from intrusion, having regular backups, and isolating infected systems. On the other hand, the efficacy of various defence measures varies widely depending on the kind of malware that is being protected against. At this time, malicious software is being designed specifically with the goal of circumventing the detection algorithms of antivirus software.

Figure 1: Android Malware

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12746

The growth of applications for the internet of things (IoT) in numerous sectors of life, including as detection, medical care, remote monitoring, and other areas, is driving the world to undergo a deep change. This is leading the world to undergo a profound transformation. The Internet of Things (IoT) is becoming a reality thanks in large part to the tight collaboration taking place between Android devices and applications. In recent times, there has been an increase in the prevalence of malware as well as other types of cyber assault on mobile devices that use the Android operating system. It is also tough to conduct testing since the Android operating system is so widely used in devices that are connected to the Internet of Things. This makes it more difficult to verify whether or not a malicious programme is there. This research offers a novel framework that combines the benefits that can be gained from AI processes as well as those that can be gained from blockchain technology in order to work on the problem of locating malware on Android IoT devices. As a component of this larger body of work, this structure was developed. Malware that may be installed on mobile devices is one of the most critical threats to computer security in the current day. In addition, dangerous software designed for mobile devices undergoes frequent updates, which leads in the emergence of new dangers. However, while modern security solutions defend mobile devices, in general, from recognised dangers, such devices are still vulnerable to dangers that have not yet been identified. In this paper, an investigation of the utilisation of evolutionary computation methodologies is carried out. Both the creation of new forms of mobile malware that are able to effectively dodge anti-malware systems that are based on static analysis and the improvement of security solutions that can be applied automatically against these forms of mobile malware are accomplished with the help of these approaches. Since a long time ago, the co-evolutionary arms race mechanism has been considered a potential option for both the creation of a system that is more resistant to new attacks and the testing of the system itself.

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.

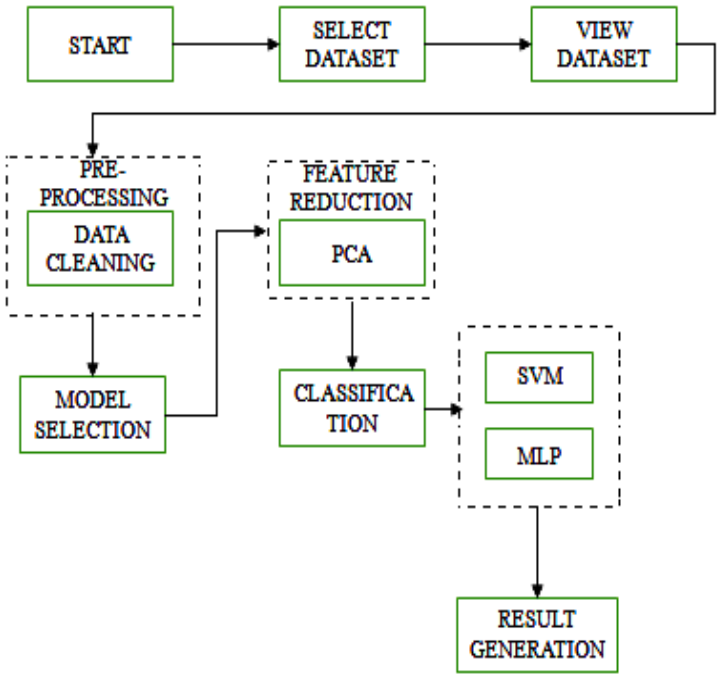


Figure 2: Flow Chart

The proposed system gives rise to a number of drawbacks, all of which are addressed by the new model that has been suggested. This method will improve the accuracy of the findings obtained from machine learning by using a machine learning algorithm to identify malicious software inside an Android dataset. The performance of the overall classification results is improved as a consequence of this. Using data from Android devices to make predictions about malicious software helps improve accuracy. The primary goal is to construct an AIML-based model for the prediction of android mobile malware, with an emphasis on improving the model's performance parameters. As a result, the primary aim of this study effort is to develop an effective method for detecting malware on android based on the dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12746

SVM, which stands for support vector machine, is a well-known machine learning method that may be used for the prediction of Android malware. The Support Vector Machine (SVM) is a technique for supervised learning that may be used to classification and regression problems. In the context of Android malware prediction, support vector machines (SVM) may be used to categorise mobile apps as either dangerous or benign on the basis of a collection of characteristics that are retrieved from the programme.

In order to utilise SVM for the prediction of malware on Android, you must first extract a collection of characteristics from each programme. These features may contain information about the application, such as the permissions that were sought, the API calls that were performed, and other metadata. After the features have been retrieved, an SVM model may be trained using those features. After that, the SVM model may be used to the task of determining whether new apps are harmful or not.

Utilising SVM as a predictive tool for Android malware comes with a number of important benefits. SVM is a robust and accurate machine learning technique that is capable of dealing with high-dimensional data and is resistant to overfitting to a certain extent. Additionally, it is able to function effectively with minimal datasets, which is often the case with Android malware prediction. MLP, or multilayer perceptron, is an additional well-known machine learning approach that has the potential to be used for Android malware prediction. The multilayer perceptron (MLP) is a specific sort of artificial neural network that is able to learn non-linear correlations between the data that are input and the labels that are output. To begin using MLP for Android malware prediction, we must first extract a collection of characteristics from each application, in a manner similar to how the SVM method works. These features may contain information about the application, such as the permissions that were sought, the API calls that were performed, and other metadata. After the features have been retrieved, they may be fed into an MLP model to be utilised as input. After that, the MLP model may be trained to determine if an application is malicious or benign based on its behaviour.

The capability of MLP to capture complicated correlations between input characteristics is one of the advantages of using this technique for Android malware prediction. MLP is also capable of dealing with high-dimensional data and has the ability to function well with tiny datasets, both of which are common scenarios in Android malware prediction.

III. RESULT AND ANALYSIS

Python Spyder 3.7 is used in order to carry out the algorithm's implementation that has been suggested. We are able to employ the functions that are made accessible in the spyder environment for different approaches such as SVM and MLP thanks to the sklearn, numpy, pandas, matplotlib, pyplot, seaborn, and os library.

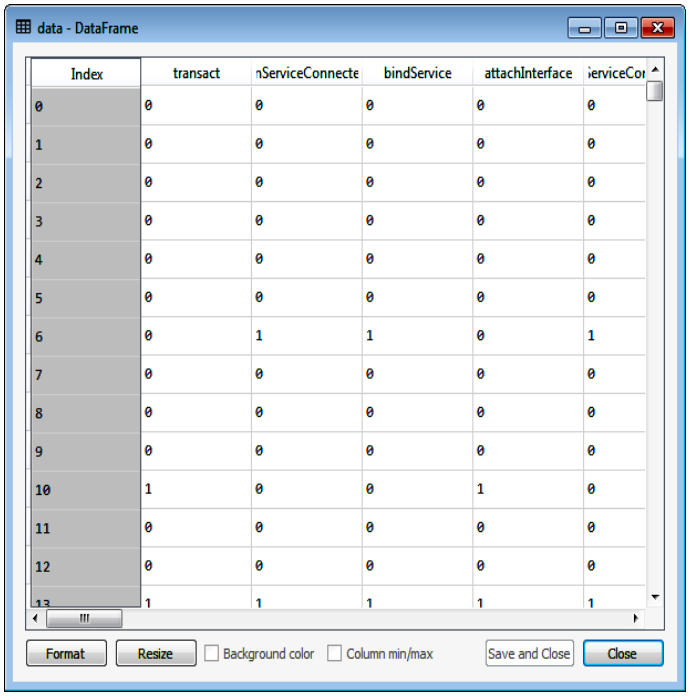


Figure 3: Dataset

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12746

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.

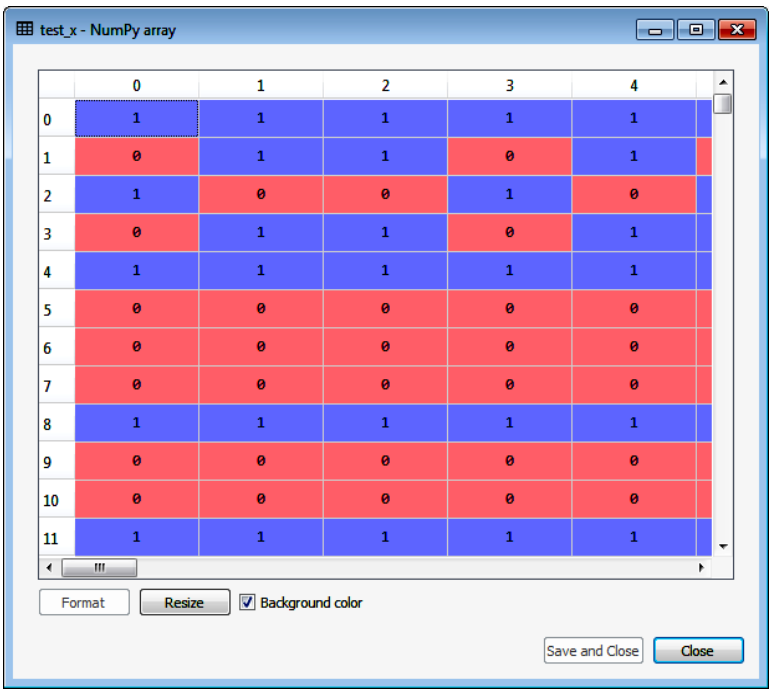


Figure 5: Test Data

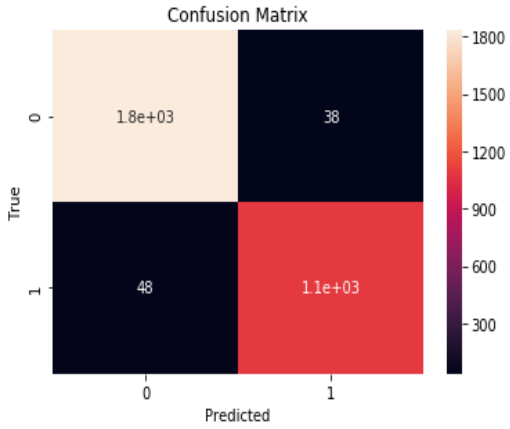
Figure 5 is showing the x test of the given dataset. The given dataset is divided into the 20-30% part into the train dataset.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

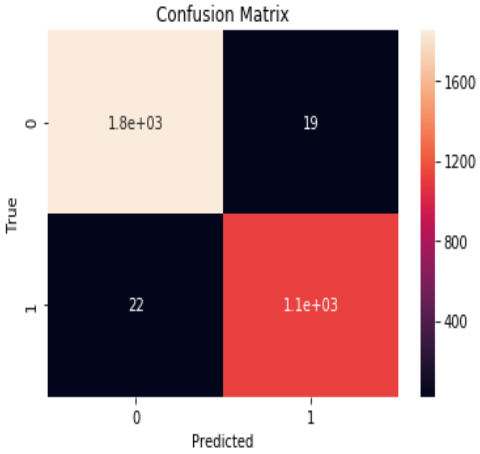
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 7, July 2023

DOI: 10.17148/IJARCCE.2023.12746



(a)



(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 %

DOI: 10.17148/IJARCCE.2023.12746

IV. CONCLUSION

This paper proposes a machine learning technique, along with an analysis of performance, for predicting malware on android devices. Machine learning classifiers are being used for the goal of this inquiry in order to produce predictions about android malware. The information about Android malware is utilised as input data and is placed through the process of pre-processing. During the phase of the technique known as preprocessing, the dataset should be scrubbed, and the label encoding should be implemented. The previous level of accuracy, which was reached at 94.92%, has been raised to 99.00% as a result of the work that is being advised. Previously, the accuracy level was 94.92%. In comparison, the mistake rate for the strategy that is currently being used is 5.08%, whereas the recommended alternative has an error rate of just 1%.

REFERENCES

[1] H. Alamro, W. Mtouaa, S. Aljameel, A. S. Salama, M. A. Hamza and A. Y. Othman, "Automated Android Malware Detection Using Optimal Ensemble Learning Approach for Cybersecurity," in IEEE Access, vol. 11, pp. 72509-72517, 2023, doi: 10.1109/ACCESS.2023.3294263.

[2] 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 2022, doi: 10.1109/TNSE.2020.2996379.

[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.