



# Android Malware Prediction using Machine Learning technique with Performance Analysis

<sup>1</sup>Anu Malik, <sup>2</sup>Harshita Kamboj, <sup>3</sup>Sarvendra Kumar\*

<sup>1,2,3</sup>Neelkanth Institute of Technology, Meerut, India

e-mail: \*[sarvendraricky@gmail.com](mailto:sarvendraricky@gmail.com); [anumalik1707@gmail.com](mailto:anumalik1707@gmail.com)

**Abstract** : Android malware is malicious software that targets a specific type of device: the Android device. Android's less secure platform, such as its Play Store where applications are downloaded, and users' ability to side load content from the internet creates an environment where malware can thrive. This paper presents android malware prediction using machine learning technique with performance analysis. The simulation is performed using python synder 3.7 software. The simulation results show the improvement in the performance parameters.

**IndexTerms** – Android, Malware, Artificial intelligence, PCA, MLP, AUC, SVM.

## I. INTRODUCTION

Mobile malware is malicious software that targets wireless-enabled device, by causing the collapse of the system and loss or leakage of confidential information. As wireless phones and PDA networks have become more and more common and have grown in complexity, it has become increasingly difficult to ensure their safety and security against electronic attacks in the form of viruses or other malware. Many types of malware exist, including computer viruses, worms, Trojan horses, ransomware, spyware, adware, rogue software, wiper, and scareware. The defense strategies against malware differs according to the type of malware but most can be thwarted by installing antivirus software, firewalls, applying regular patches to reduce zero-day attacks, securing networks from intrusion, having regular backups and isolating infected systems. Malware is now being designed to evade antivirus software detection algorithms.



Figure 1: Android Malware [Google image]

Internet of things (IoT) is changing this world with its developing applications in different parts of life like detecting, medical care, remote checking, etc. Android gadgets and applications are working hand to hand to acknowledge dreams of the IoT. As of late, there is a fast expansion in dangers and malware assaults on Android-based gadgets. Besides, because of broad double-dealing of the Android stage in the IoT gadgets makes an errand testing of getting such sort of malware exercises. This work presents an original structure that consolidates the benefits of both AI procedures and blockchain innovation to work on the malware location for Android IoT gadgets. Mobile malware is one of today's greatest threats in computer security. Furthermore, new mobile malware is emerging daily that introduce new security risks. However, whilst existing security solutions generally protect mobile devices against known risks, they are vulnerable to as yet unknown risks. In this study, the use of evolutionary computation techniques are investigated, both for developing new variants of mobile malware which successfully evades anti-malware systems

based on static analysis and for developing better security solutions against them automatically. Co-evolutionary arms race mechanism has always been considered a potential candidate for developing a more robust system against new attacks and for system testing.

## 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.

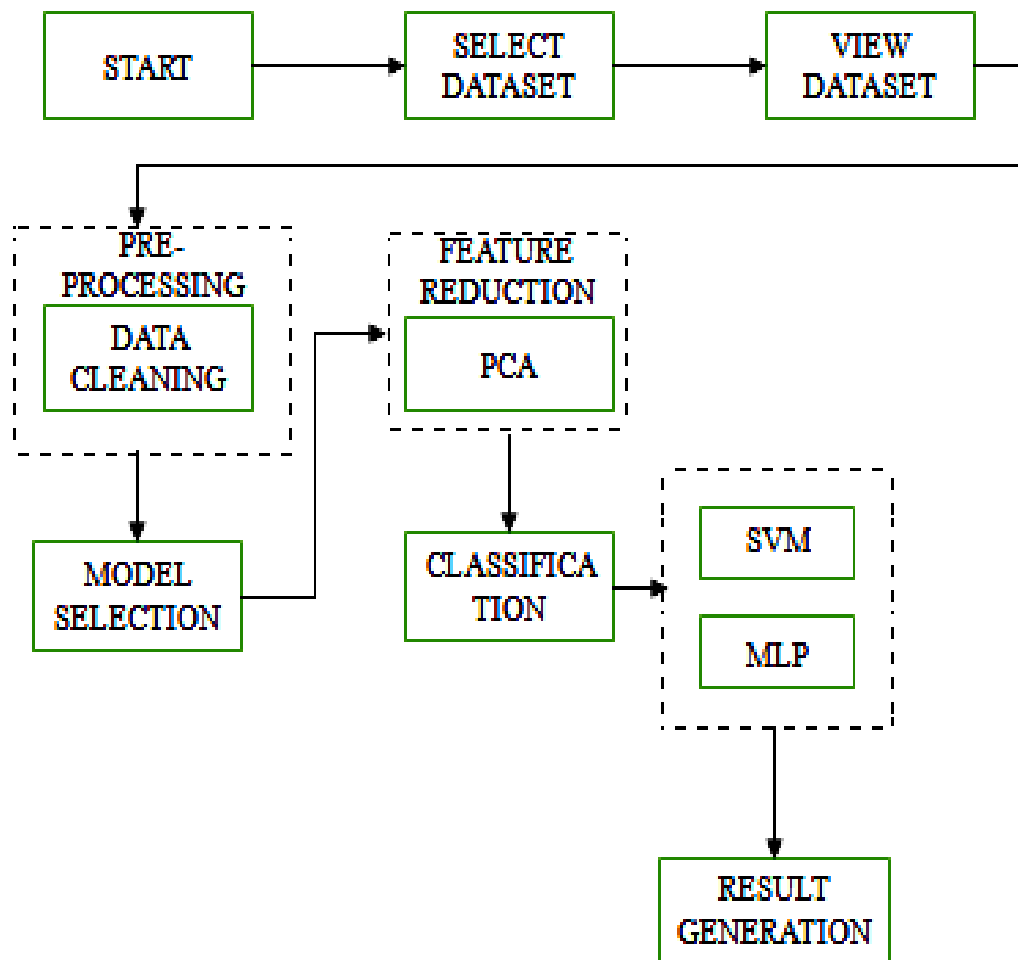


Figure 2: Flow Chart

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.

## 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.

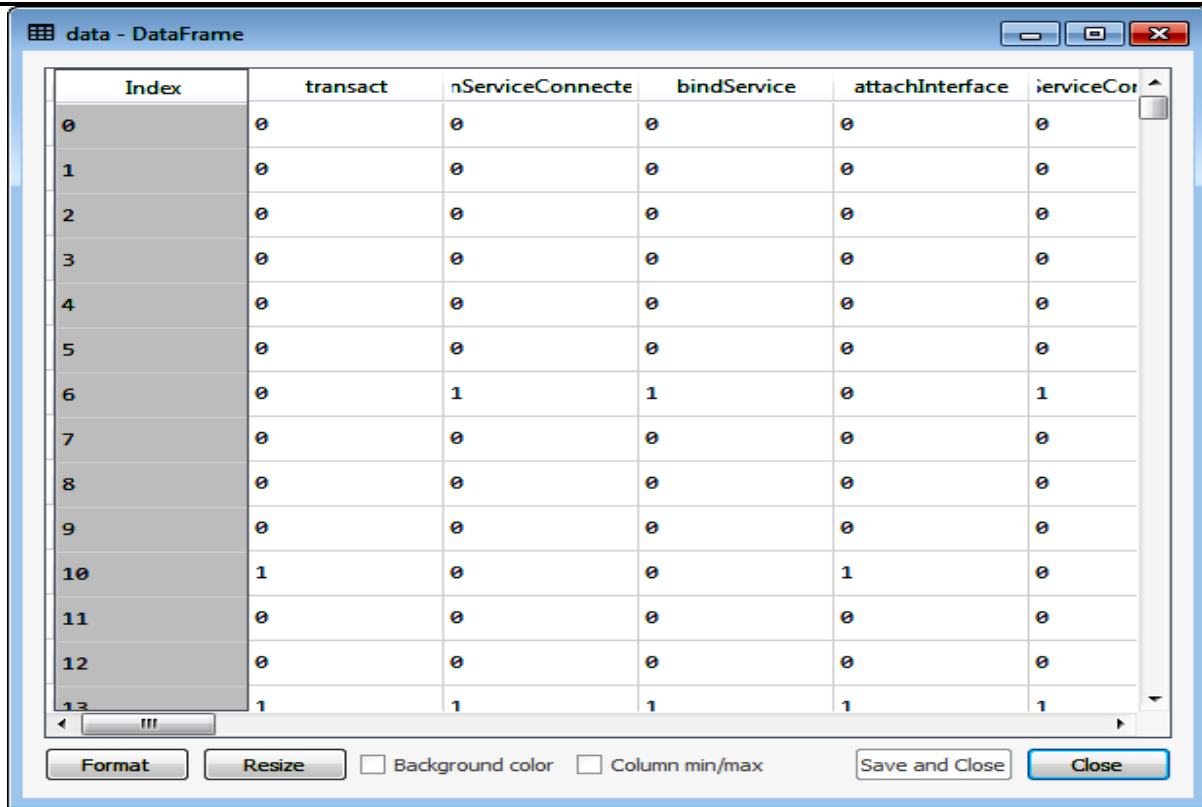


Figure 3: Dataset

Figure 3 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.

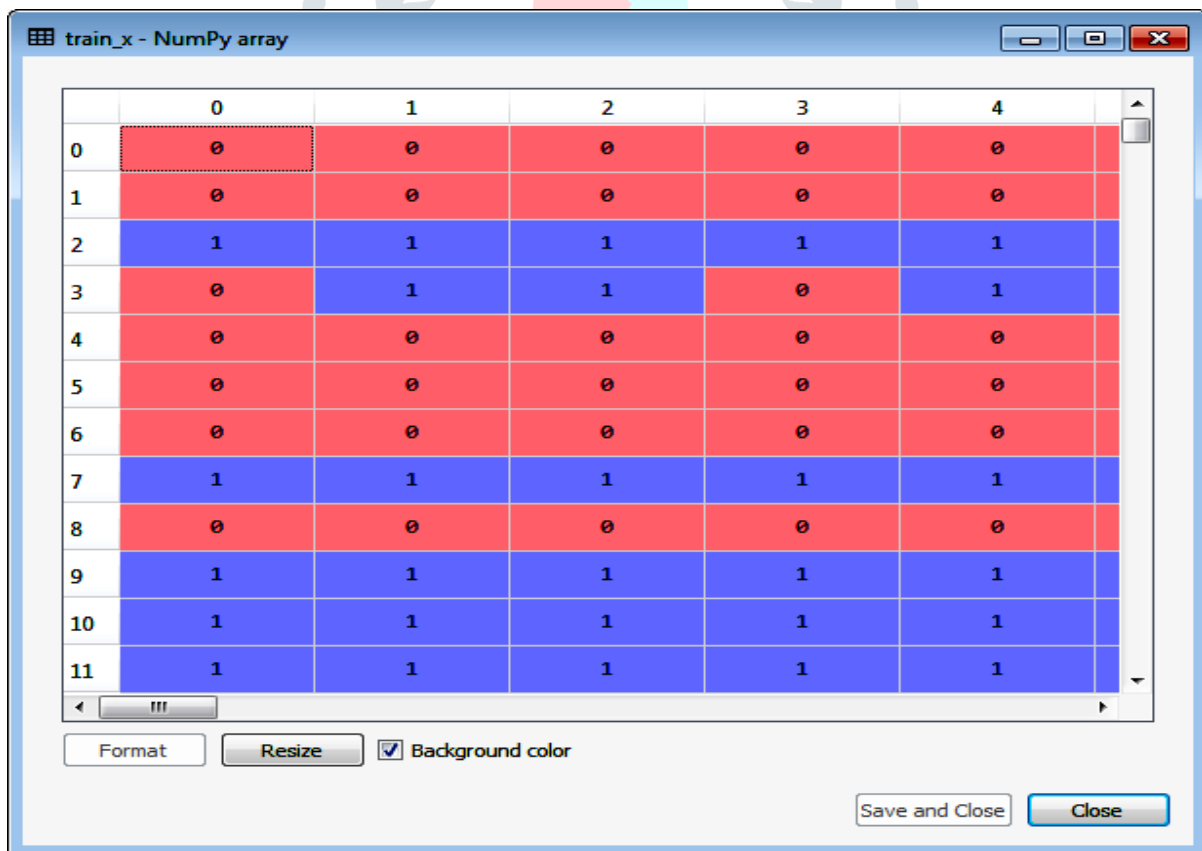


Figure 4: Train data

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.

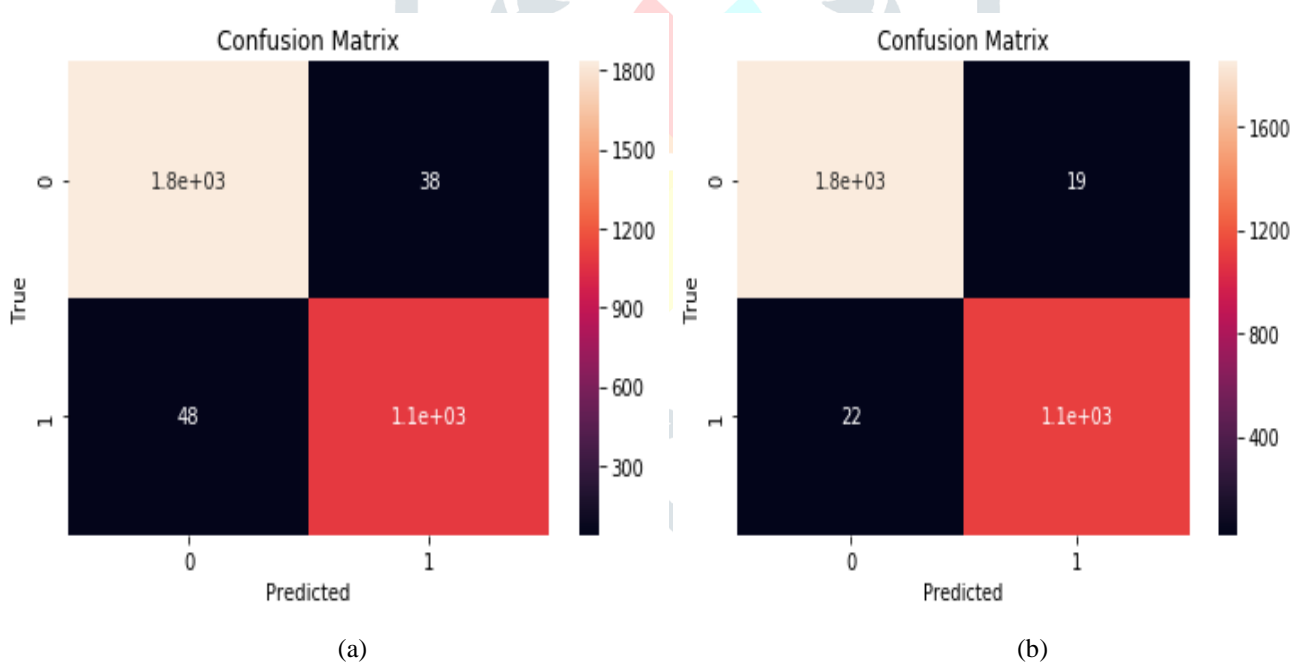


Figure 6: Confusion Matrix (a) SVM (b) MLP

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 paper presents android malware prediction using machine learning technique with performance analysis. In this study, the machine learning classifiers are predicting the android malware. The Android Malware data is taken as input data and applied into pre-processing method. In pre-processing method clean the dataset and apply the label encoding. The overall accuracy achieved by the proposed work is 99.00 % while previous it is achieved 94.92 %. The error rate of proposed technique is 1 % while 5.08 % in existing work.

#### 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.