Avoidance of Malware Attacks in Android Platform using Deep Learning on Benchmark Datasets

Supervisor

Dr.Bashar M.Nema

Submitted by

Ahmed R. Al-Mhanawi

Abstract:

- this study, **will start with the Android APK datasets**. Both benign and malignant APKs will be sent. This study focuses on finding and extracting the signatures deep inside the APKs, which will make it possible to assemble a training dataset. In all, will analyses **500 APK files**, and we expect that around half of the files will be completely harmless, while the other half will be malicious. Then comes the verification of the permissions in each APK and what impact they have. Once it's been cleaned, a dataset will be prepared for training the model in order to do prediction.

- To complete predictive analytics, any random APK outside of the 500 APKs is selected and used. Then, one can tell what the odds are of having harmful code spots in the new APK being analyses. Using machine learning, the results of various prediction measures are tracked using time, cost, accuracy, and other measures. To conduct a comparison study, we use machine learning to integrate our predictions.

Introduction

- Researchers at Zimperium company have discovered a new virus that poses as a system update programme, which makes it difficult to find. Installed, it controls Android phones, stealing data, texts, and pictures, etc. According to the researchers, after hackers get control, they can record audio and phone calls, take photos, steal messages and files, and access instant messenger accounts. Hackers can also investigate the user's browser, stealing search history and bookmarks. They can see what the user is copying to the clipboard, and even gather device information.

Android Package (APK)

An Android Package (APK) is the file format of the Android operating system that the OS uses for both the installation and distribution of mobile applications. .apk, which is typically created using Android SDK, is an extension used by the Android Operating System to signify application files

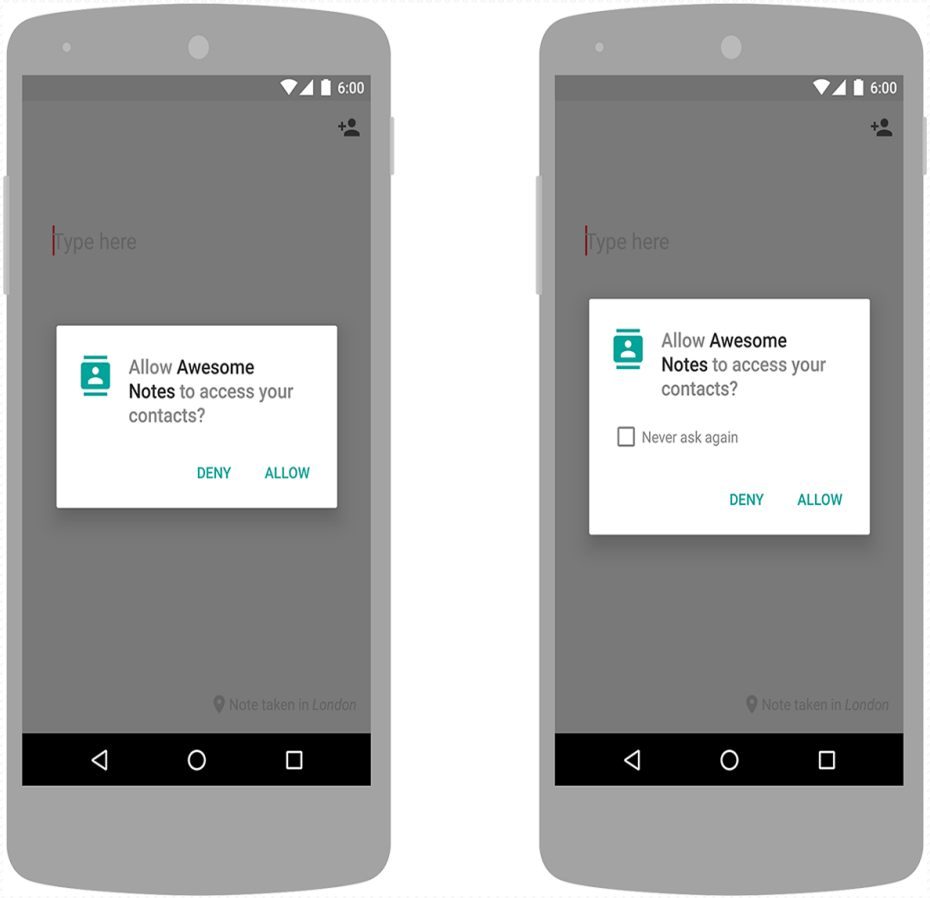


Figure 1: Permissions in APK

Table 1: Traditional structure of Android APK

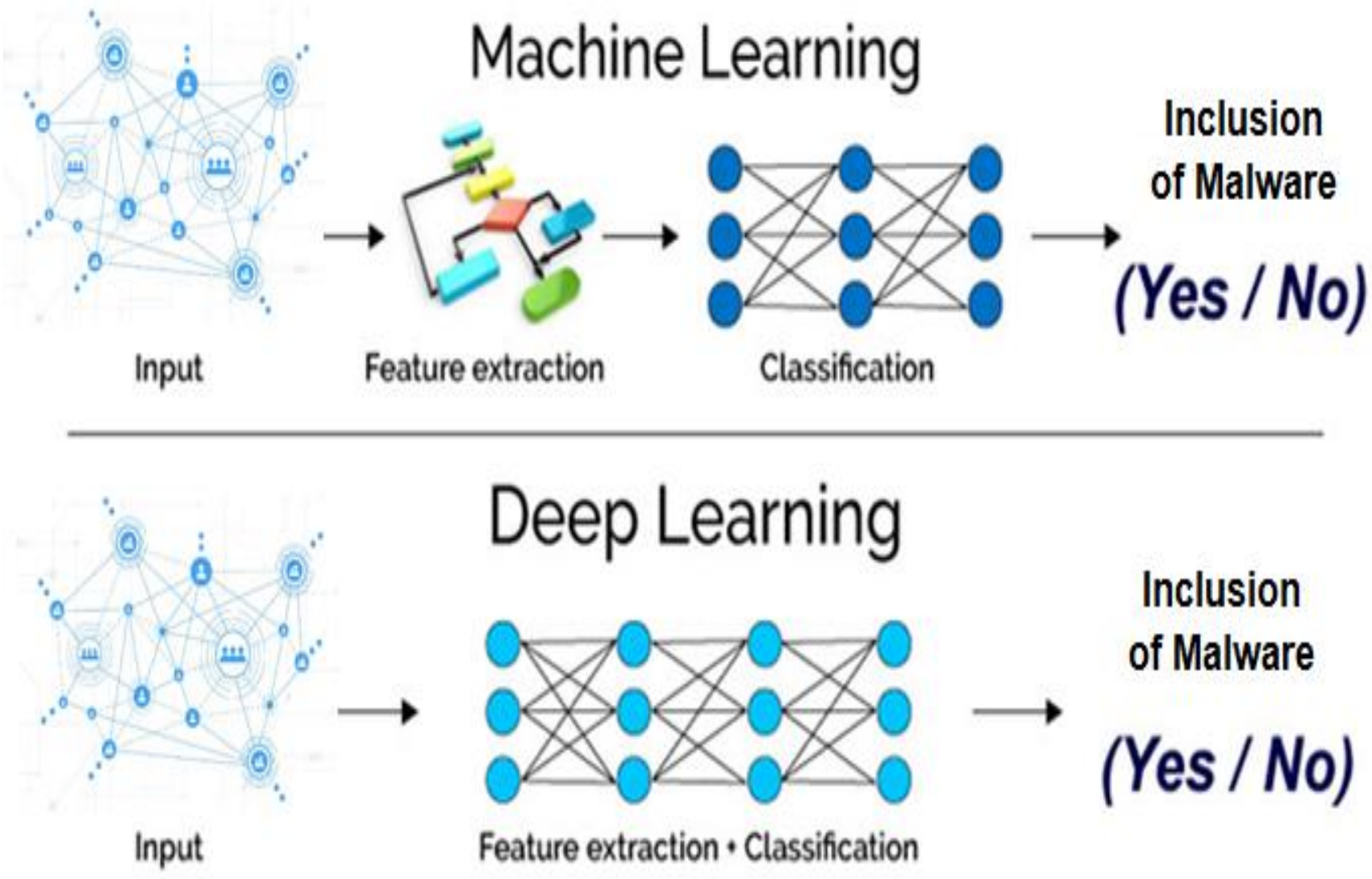
APK File / Folder	Description
assets/	Optional Folder for Asset Manager
classes.dex	Compiled Code of Application
res/	Resources without compilation
resources.arsc	Application resources
META-INF/	Information of Metadata
AndroidManifest.xml	XML format Manifest File
lib/	Optional Folder with Compiled Code

Malware Attack

- Malware, in many forms, may be developed to wreak havoc on a system and achieve a number of sinister goals, such as destroying the system, gaining financial benefit, or accessing the system in an illicit manner, leading to degraded security or perhaps system information leakage . There are many other ways harmful packets may be sent, such as Trojans, Rootkits, Beast, Suspicious packers, Scare ware, Evasion, Backdoors, Key loggers, Trojan Spy, Trojan Game Thief, Browser Hijacker, Ransom ware, Rogue Software, Botnets, and a plethora of other methods. Malware may spread in two ways. Polymorphic malware utilizes new code each time it replicates keeping its original coding intact and always seeming different. A large selection of IDS tools are available, including free solutions that can both categories assaults (using PCAP Files) as well as network traffic

Deep Learning

- Deep learning is a subfield of machine learning and it becomes the leader in this domain recently. Although deep learning has been given various descriptions and definitions in existing literature, two key elements are: (1) including multiple and non-linear layers to process input data; (2) learning feature representation at increasing high-level abstract layers. And therefore, unlike traditional neural networks, deep neural networks constructed by a more complex and deeper structure are enabled feature transformation and selection automatically.


Machine Learning
Inclusion of Malware
(Yes / No)
Input
Feature extraction
Classification
Deep Learning
Inclusion of Malware
(Yes / No)
Input
Feature extraction • Classification

- **Research Statement and Goal**
- While the vulnerabilities are growing in the Android ecosystem, with the large Android smartphone market globally, consumers get their applications from a multitude of sites. Identifying APK vulnerabilities is easier if you have a method to display the model and technique.

- **Experimental Results and Outcome:**
- First of all the fetching APK files from Online Repositories is done so that the permission based dataset can be generated. Further the dataset is trained on the machine learning and deep learning model.

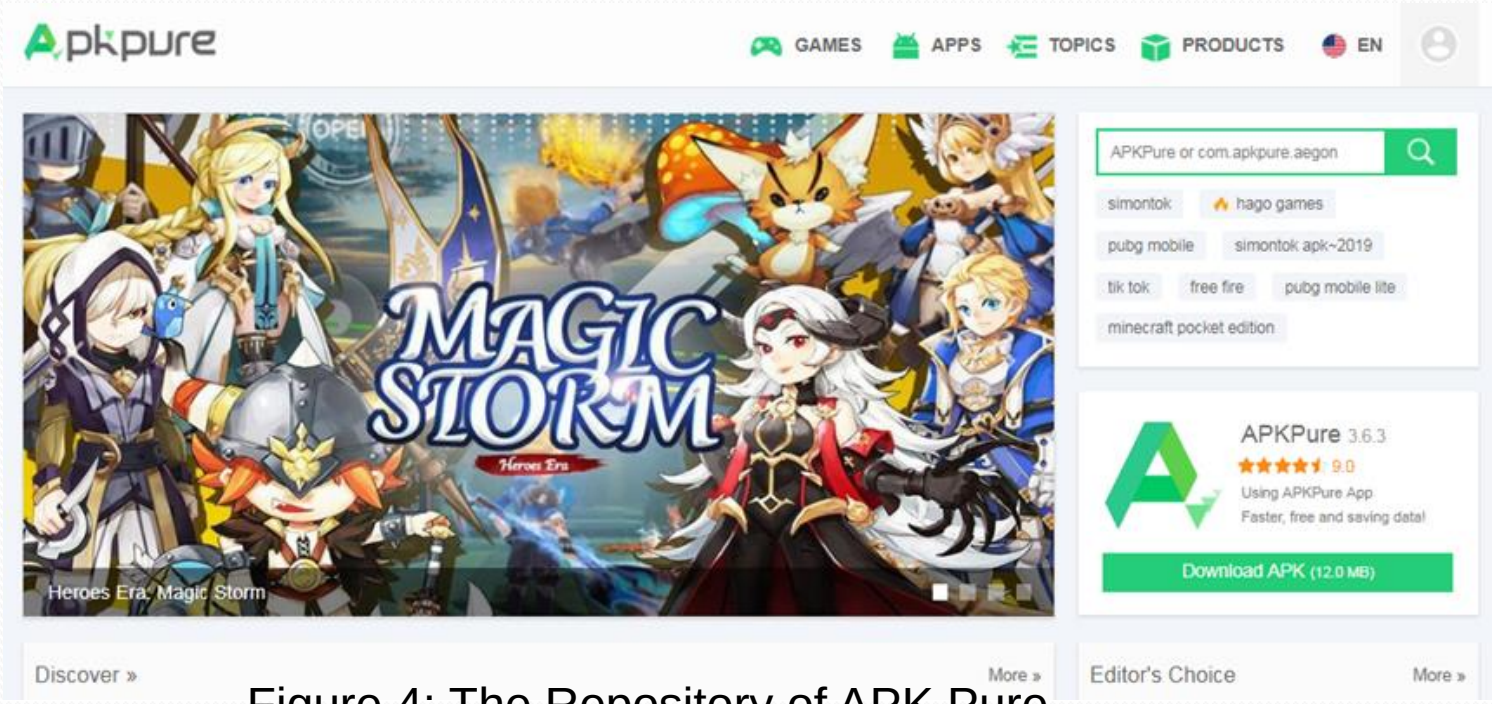


Figure 4: The Repository of APK Pure

- The APKPure repository is having enormous Android Apps which are analyzed using APK Tool so that the analytics on their permissions and suspicious parameters can be evaluated.

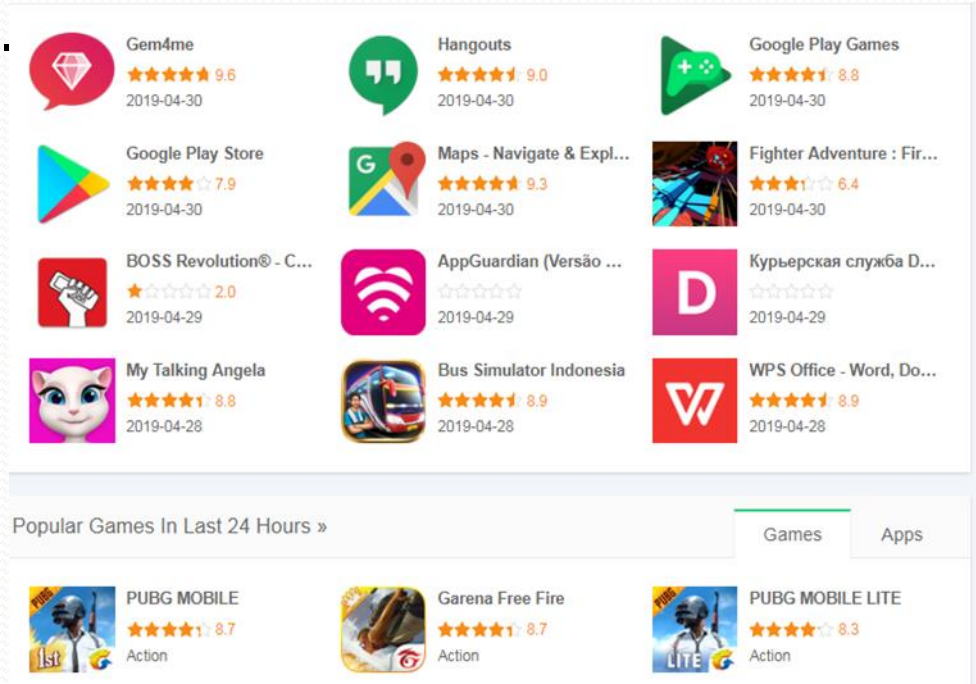


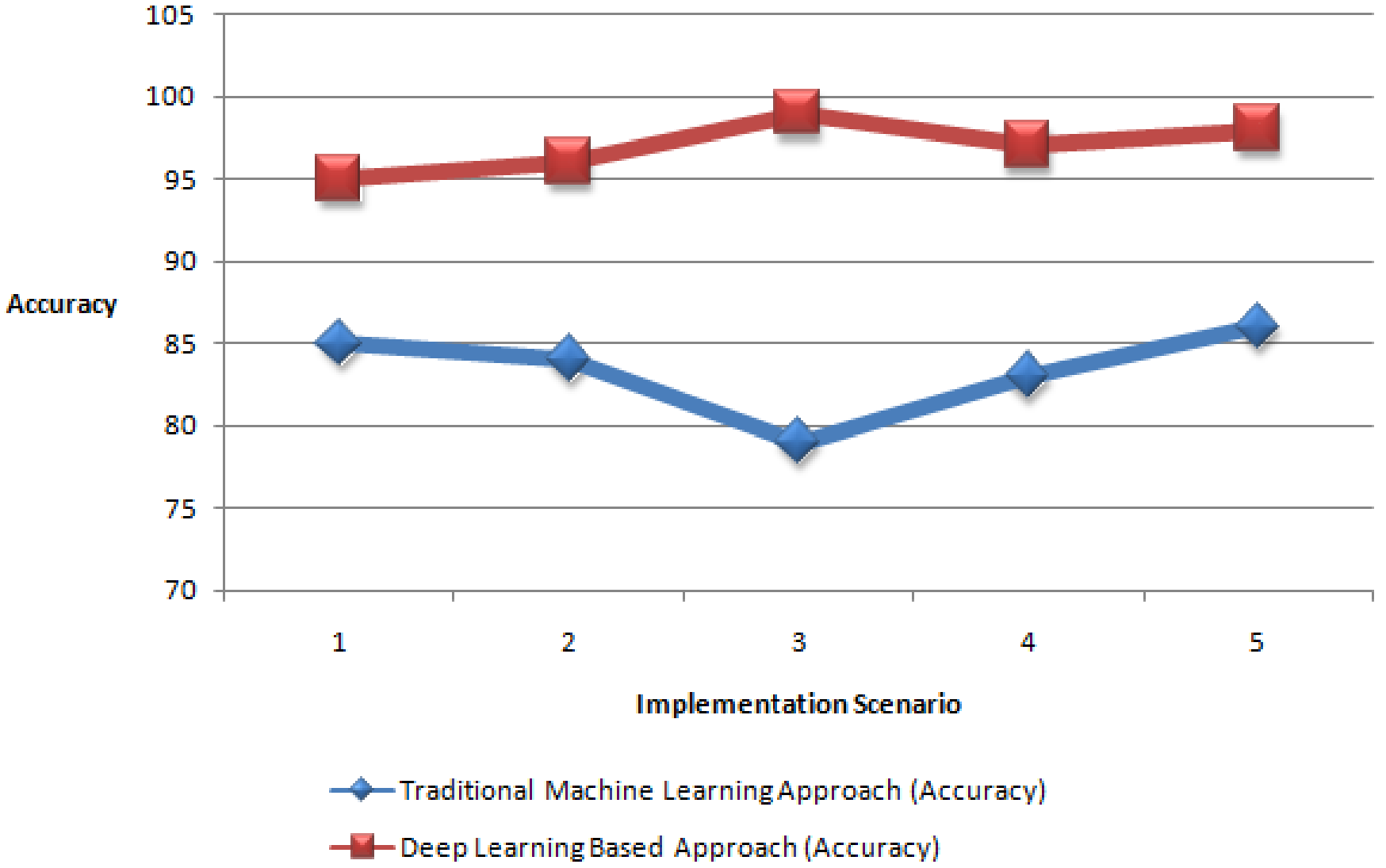
Figure 5: Apps in APKPure Repository

Table 2: Formation of Dataset for Implementations

APK File	Suspicious (1-Malignant, 2-Benign)	Call	Bluetooth	SMS	Contact	E-mail	Phone
1	**2**	0	0	1	0	0	1
2	**1**	1	1	1	1	0	1
3	**1**	0	0	0	1	1	1
4	**1**	1	1	1	1	1	1
5	**2**	1	1	1	1	1	1

Table 3: Evaluation of Parameters

Execution Scenario	Traditional Machine Learning Approach (Accuracy)	Deep Learning Based Approach (Accuracy)
1	85	95
2	84	96
3	79	99
4	83	97
5	86	98


105
100
95
90
85
80
75
70
Accuracy
1
2
3
4
5
Implementation Scenario
Traditional Machine Learning Approach (Accuracy)
Deep Learning Based Approach (Accuracy)

- From the tables above the performance of deep learning and traditional machine learning on the parameter of execution time. The execution time of RF is found less than the classical approach of traditional machine learning in all five attempts and therefore the RF proves to be more optimized as compared to traditional machine learning.

Conclusion:

- The presented work focuses on the deep learning based implementation for the advanced malware detection and predictions with higher degree of accuracy.
- By applying specialized methods, it was discovered that employing classification approaches on the Android APK makes it possible to forecast with a high level of certainty by analyzing footprints and signatures from APK. The methods provided help get results that cover many factors while also ensuring that each process is aware of how to operate efficiently. For more security and privacy in apps like Android APKs, developers may utilize Block chain Technology.

Thank you