

Android Mobile Malware Detection using Ensemble Learning

MSc Research Project
Cyber Security

Janish Ottakanchirathingal

Student ID: 19120788

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

**National College of Ireland
Project Submission Sheet
School of Computing**



Student Name:	Janish Ottakanchirathingal
Student ID:	19120788
Programme:	Cyber Security
Year:	2020
Module:	MSc Research Project
Supervisor:	Vikas Sahni
Submission Due Date:	15/08/2020
Project Title:	Android Mobile Malware Detection using Ensemble Learning
Word Count:	360
Page Count:	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

I agree to an electronic copy of my thesis being made publicly available on NORMA the National College of Ireland's Institutional Repository for consultation.

Signature:	
Date:	17th August 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Android Mobile Malware Detection using Ensemble Learning

Janish Ottakanchirathingal
19120788

1 Introduction

This document presents the details on how to execute the Android mobile malware detection code. The entire code has been written in the Python programming language.

2 System Requirements

- It is assumed that the user of this document have basic python programming knowledge and we need python3(version '3.8.*') installed on the system in order to run this code. Latest version of python can be downloaded from python official website [1])
- Anaconda Version 3.* should be downloaded and installed on the PC. Anaconda software Individual edition can be downloaded from Anaconda official website [2]. Make sure that this software have Jupyter Notebook IDE installed in order to run the python code.

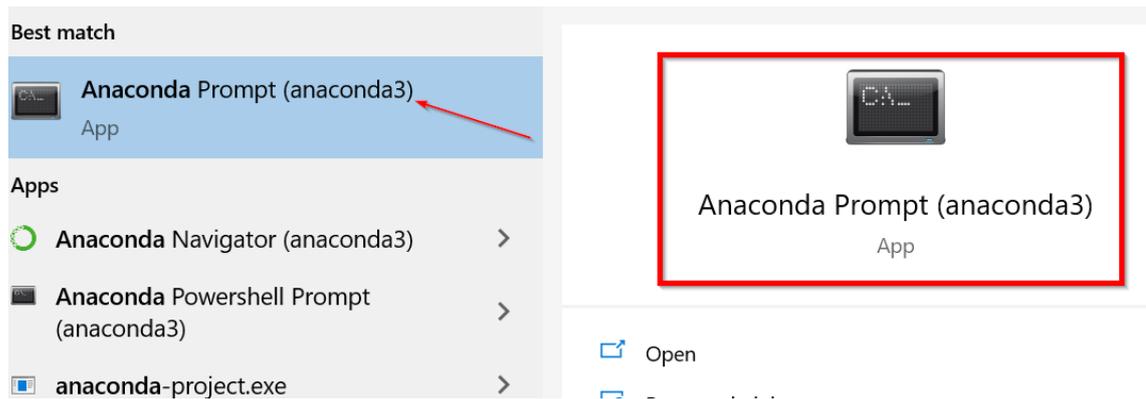
2.1 Packages needed to install and run the code

The following packages or library is needed to run the code:

1. **Pandas**: Used for data manipulation and analysis
2. **Numpy**: Used for array related operations and numerical operations
3. **Matplotlib**: Used for plotting the graphs
4. **Scikit Learn**: Used for Machine learning algorithms.

2.2 How to install a package :

- To install a package first open Anaconda Prompt.



- the command to install a package or library in Anaconda prompt
 - pip install {package name} E.g.: pip install sklearn

2.2.1 Numpy

Command: pip install numpy

```
(base) C:\WINDOWS\system32>pip install numpy
Requirement already satisfied: numpy in c:\users\janish\anaconda3\lib\site-packages (1.19.1)
```

2.2.2 Pandas

Command: pip install pandas

```
(base) C:\WINDOWS\system32>pip install pandas
Requirement already satisfied: pandas in c:\users\janish\anaconda3\lib\site-packages (1.1.0)
Requirement already satisfied: numpy>=1.15.4 in c:\users\janish\anaconda3\lib\site-packages (from pandas) (1.19.1)
Requirement already satisfied: pytz>=2017.2 in c:\users\janish\anaconda3\lib\site-packages (from pandas) (2019.3)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\janish\anaconda3\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: six>=1.5 in c:\users\janish\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.14.0)
```

2.2.3 Matplotlib

Command: pip install matplotlib

```
(base) C:\WINDOWS\system32>pip install matplotlib
Requirement already satisfied: matplotlib in c:\users\janish\anaconda3\lib\site-packages (3.1.3)
Requirement already satisfied: pyparsing!=2.0.4,!<2.1.2,!<2.1.6,>=2.0.1 in c:\users\janish\anaconda3\lib\site-packages (from matplotlib) (2.4.6)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\janish\anaconda3\lib\site-packages (from matplotlib) (1.1.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\janish\anaconda3\lib\site-packages (from matplotlib) (2.8.1)
```

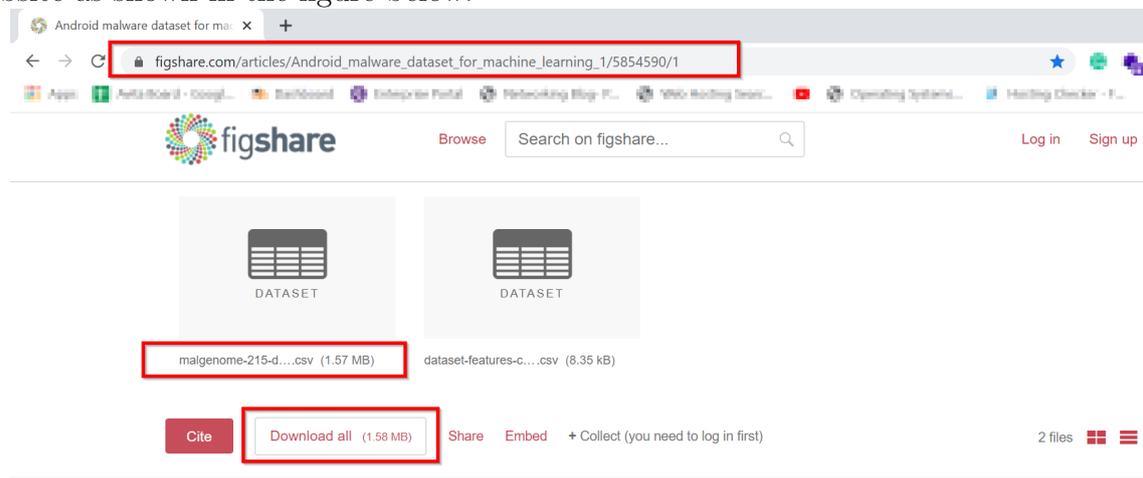
2.2.4 Scikit Learn

Command: pip install sklearn

```
(base) C:\Users\Janish> pip install sklearn
Collecting sklearn
  Downloading sklearn-0.0.tar.gz (1.1 kB)
Requirement already satisfied: scikit-learn in c:\users\janish\anaconda3\lib\site-packages (from sklearn) (0.22.1)
Requirement already satisfied: numpy>=1.11.0 in c:\users\janish\anaconda3\lib\site-packages (from scikit-learn->sklearn) (1.18.1)
Requirement already satisfied: joblib>=0.11 in c:\users\janish\anaconda3\lib\site-packages (from scikit-learn->sklearn) (0.14.1)
Requirement already satisfied: scipy>=0.17.0 in c:\users\janish\anaconda3\lib\site-packages (from scikit-learn->sklearn) (1.4.1)
Building wheels for collected packages: sklearn
  Building wheel for sklearn (setup.py) ... done
  Created wheel for sklearn: filename=sklearn-0.0-py2.py3-none-any.whl size=1320 sha256=f2688751798f2dc7011d945e7aa258e27f2beb4797674f9b6f86efc15056bfc1
  Stored in directory: c:\users\janish\appdata\local\pip\cache\wheels\46\ef\c3\157e41f5ee1372d1be90b09f74f82b10e391eaacc8f22d33e
Successfully built sklearn
Installing collected packages: sklearn
Successfully installed sklearn-0.0
```

3 Data Sources

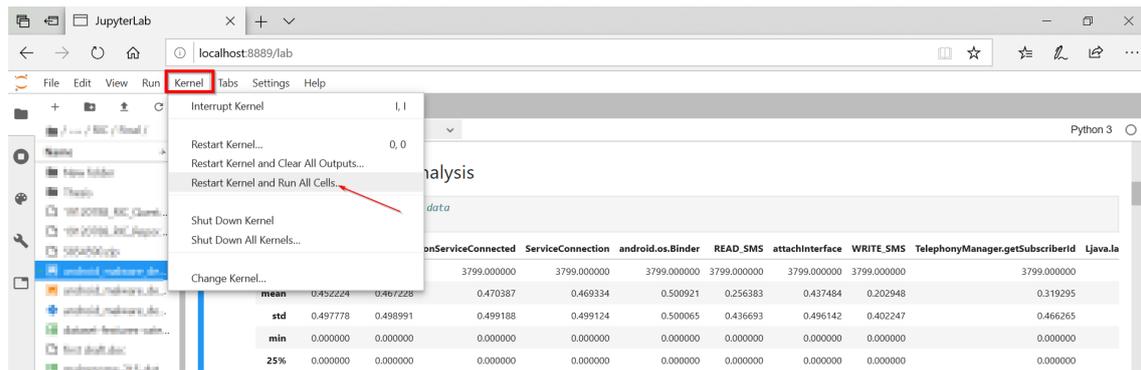
The dataset used for implementation is Malgenome mobile malware dataset provided by North Carolina State University. The dataset can be downloaded from figshare [3] website as shown in the figure below:



After downloading the Malgenome dataset in CSV format, put the data file in the same folder as the code ipynb file.

4 Code Execution

To execute the code, run the Jupyter notebook program, or open Anaconda Navigator and Launch Jupyter notebook. Jupyter notebook come pre-installed with Anaconda. From the directory structure, move to the folder where the code file is stored. On the Kernel tab select Restart and Run all option, all the code blocks will execute one by one presenting the output in the cell below.



References

- [1] "Download python latest version," <https://www.python.org/downloads/>.
- [2] "Anaconda individual edition," <https://www.anaconda.com/products/individual>.
- [3] S. Yerima, "Android malware dataset for machine learning 1," https://figshare.com/articles/Android_malware_dataset_for_machine_learning_1/5854590/1.