

Configuration Manual

MSc Academic Internship CyberSecurity

Saishankar Murali Student Id: - x18174990

School of Computing National College of Ireland

Supervisor: Prof. Imran Khan



National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Saishankar Murali		
Student ID:	X18174990		
Programme:	MSc in Cybersecurity	Year: 2019	- 2020
Module:	MSc Academic Internship		
Lecturer: Submission Due	Mr. Imran Khan		
Date:	17.08.2020		
Project Title:	Detection of malware in a file using machine	earning with KNN model	
Word Count:	998 words	Page Count: 8 pages	

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:	Saishankar	Murali

Date: 17.08.2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	
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.	

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):	

1.Introduction:

This article will give us a glance on how the implement is done and executed on the system and also how it can be done on other systems. This prototype or the model that has been implemented is created to detect the malware in a system, this prototype will detect the malware in a large scale. In this we will be discussing how to configure the system to run the proposed model in the respective system. This proposed method performs well and the accuracy achieved is good on the basis of the dataset which is in a large scale. In the below mentioned sections we will be talking about how to install the softwares and system version so that the model is compatible and is executed successfully.

2. Configuration of System:

The below mentioned detail consists of the software and hardware requirements to implement the model.

2.1. Hardware Configuration:

- Operating system: Windows 7 or Later
- Processor: CPU cores should be 2 or more
- System: It is compatible to 32- bit and 64-bit
- Hard disk: 256GB to 4TB
- SSD: 256 GB or more (Not Compulsory)
- RAM: More than 2GB or 2GB

2.2. Software Configuration:

This section involves all the software that needs to be installed before we start the implementing the model and further.

Tool	Version	Illustration
Python (32bit or 64-bit)	3.8	Python programming is used
		for creating the model and
		how should the model work
Microsoft Excel	2019	Microsoft excel is used to
		open the datasets and
		create a new dataset
		accordingly.
Anaconda	5.3.0	Anaconda is used because
		many IDE related to data
		science are available on one
		platform
Jupyter Notebook	6.0.3	Jupyter Notebook is used
		because it gives us live
		output and gives us a better
		understanding of work and
		also has visualization.

1.https://www.python.org/downloads/release/python-380/

2. <u>https://www.anaconda.com/products/individual</u>

3. Installations:

This section will tell you how to install all the software, step by step installations of each is explained below in detail

3.1. Python Install:

3.8.0 is the latest version of python and it is available in their official website; you can download any system compatible version such as 32-bit or 64-bit

🦆 pyt	hon™			Donate	Q Search	GO Socialize
5 7 7						
	About	Downloads	Documentation	Community Suco	ess Stories News	Events
Do Look Linux Want Dock	wnload Python 3 ing for Python w «/UNIX, <u>Mac OS X</u> t to help test dev <u>ser images</u>	3.8.5 /ith a different OS? (, <u>Other</u>	Version for Win Python for <u>Windows</u> , s of Python? <u>Prereleases</u> , specific releases	dows		
Active Pythor	n Releases					
		hon Developer's G	uide.			
	tion visit the Pytl	hon Developer's G	uide. First released	End of s	upport	Release schedule
For more informat	tion visit the Pytl			End of s 2024-1		Release schedule PEP 569
For more informat	tion visit the Pytl Maintena		First released		0	
For more informat Python version 3.8	tion visit the Pyth Maintena bugfix		First released	2024-1	0 6-27	PEP 569
For more informat Python version 3.8 3.7	tion visit the Pytl Maintena bugfix security		First released 2019-10-14 2018-06-27	2024-1 2023-0	0 6-27 2-23	PEP 569 PEP 537

3.2. Anaconda Install:

The anaconda software is totally free and you can download it from their official website and do the registration part and run it on the respective system. The below figure shows from where you can install and how.

	Anaconda Installer	S
Windows 🕊	MacOS 🗯	Linux 🛆
Python 3.8 64-Bit Graphical Installer (466 MB)	Python 3.8 64-Bit Graphical Installer (462 MB)	Python 3.8 64-Bit (x86) Installer (550 MB)
32-Bit Graphical Installer (397 MB)	64-Bit Command Line Installer (454 MB)	64-Bit (Power8 and Power9) Installer (290 MB)

3.3. Microsoft Excel:

Most of the Laptops and desktop comes with inbuilt Microsoft Installed in which you get Excel installed and use it. There are some laptops and desktops who don't have excel pre-installed. Excel is not a free software; you will need to pay for it if it is not installed in the respective laptop or desktop. The below figure shows how and where to install Excel.

Microsoft Hom	e Devices 🗸 Software 🗸 Games & entertainment 🗸 Deals Shop Business Students & parents Gift Cards All Microsoft 🧸	Search $ ho$ Cart 몇 Sign in (8
	Excel Microsoft Corporation	€135.00 Buy now
X	For 1 PC or Mac Turn data into useful insights Compatible with Windows 10 or macOS More	Downloadable products will be delivered to you by a download link becoming available after completing purchase.
Get Excel plus all the Office a	sps and 1 TB cloud storage with Microsoft 365 for one low monthly price.	For up to six users For one user

4. Working:

In this section the explanation on how to start with the implementation as a beginner perspective.

Step 1: Run the anaconda Software. The below figure shows the user interface of anaconda.



Step 2: Click on the Jupyter Notebook which is the 3rd from the left. The below figure shows you the interface of the Jupyter notebook.

💭 jupyter	Quit	Logou
Files Running Clusters		
ielect items to perform actions on them.	Upload	New -
	Name Name Python 3	e
C 3D Objects	Other:	
AndroidStudioProjects	Text File	
the ansel	Folder	
Cisco Packet Tracer 7.2	Terminal	_
Contacts	a month ago	
Creative Cloud Files	a year ago	
Desktop	6 minutes ago	
Documents	a month ago	
Downloads	13 minutes ago	
Co Favorites	a month ago	
	a month ago	
Cim Microsoft	a year ago	
Comparison	a month ago	
CheDrive	12 days ago	
C Pictures	a month ago	
Saved Games	a month ago	
Searches	a month ago	
Ci Videos	16 hours ago	
C VirtualBox VMs	5 months ago	
C Zotero	2 months ago	
B Untitled.ipynb	2 months ago	72 E
Dutitled1.ipynb	a month ago	1.18 kE
🗌 🥔 Untitled2.ipynb	Running 9 days ago	55.3 kE

In the above figure there is an option called "New" in which you need to select the language as python 3. Then a new interface will be opened and you can proceed with the further coding process.

Step 3: After the installation is complete. coding part is started, but before building the model or train the dataset we need to import some packages that we will need. The below figures show the packages that are been installed in python language.

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [13]:	import pandas as pd
	import numpy as np
	<pre>from sklearn.model_selection import train_test_split</pre>
	<pre>from sklearn.ensemble import RandomForestClassifier</pre>
	import seaborn as sns

Step 4: The next step is to load the dataset and also to give the command to train them by splitting them into two parts which is Training dataset and Test dataset. The figure below shows how we have loaded the dataset and the split.

Step 5: After splitting the data, we then further move to the coding part where we tell what and how the model should perform the detection based on the individual model that we have compared. The below figure shoes how the detection took place and how much malware and benign files have been detected.

```
Confusion matrix, without normalization [[2486 3849] [3887 6821]]
```

