Configuration Manual

MSc Research Project
MSc. Cyber Security

Safvaan Shadab Nakid
Student ID: X20180527

School of Computing
National College of Ireland

Supervisor: Dr Vanessa Ayala-Rivera

Student Name:	Safvaan Shadab Nakid		
Student ID:	X20180527		
Programme:	MSc Cyber Security	**Year:**	2021
Module:	MSc Research Project/Internship		
Lecturer:	Dr Vanessa Ayala-Rivera		
Submission Due Date:	16/12/2021		
Project Title:	EVALUATION AND DETECTION OF CYBERCRIMINAL ATTACK TYPE USING MACHINE LEARNING		
Word Count:	579	**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.

<u>ALL</u> 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.

Signature: SAFVAAN SHADAB NAKID

Date: 16/12/2021

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.

Configuration Manual

Safvaan Shadab Nakid
Student ID: X20180527

1 Introduction

This configuration manual consists of information required to load and test the prediction of the dataset in a new local environment. A step-by-step implementation guide is also detailed in the following sections.

2 Hardware and Software Environment

Table 1: Specification Table

Model attributes	Information Values
Processor	AMD Ryzen 5- 3550H
RAM Size	25 GB
Internal Memory	256 GB
Environment	Local – Windows 10.
Programming Language	Python
Programming environment	Jupyter Notebook
Libraries required	• Pandas • NumPy • Seaborn • Sklearn (Scikit) • Matplotlib

The version attributes for the libraries used in the research are as shown in Table [2] below. The said information can then be used in future events to recreate the project.

Table 2: Library Version Information

Libraries Used	Version
Pandas	1.2.4
NumPy	1.20.1
Seaborn	0.11.1
Sklearn (Scikit)	0.24.1
Matplotlib	3.3.4

3   Implementation

    A. Installing Python and Jupyter Notebook

    I. Download the latest version of python from the URL mentioned in the footnote of this page.[1]

        a. Run the setup once it's downloaded and select an appropriate directory for running the software. Make sure to have enough space on the local disk for future file saves.

    II. Jupyter Notebook

The Jupyter notebook is an open-source programming environment mainly used for data mining and classification procedures. Since the project is developed on this platform, it needs to be installed to run the programs systematically. It can be installed from the URL mentioned in the footnote[2].

    III. Running the code

- Import the necessary required libraries using the code given below.

```python
import pandas as pd  # to read the dataset and store it in dataframe
import numpy as np # for numeric operation
import seaborn as sns # for visualization
import matplotlib.pyplot as plt # for visualization
```

Figure 1: Importing Libraries

- Import the dataset from the stored directory into the Jupyter environment. Make sure to specify the proper directory path while importing and specify the proper read method in the pandas statement.

Table 3: Importing Dataset

Code Snippet
df = pd.read_excel (r'D:\"Your path"\cyberattack.xlsx', sheet_name='dataset') print (df)

- Executing the code

There are three ways in which the code can be run in the local environment. If you click on the cells tab at the top, a dialogue box like figure [2]. From there you can either select to run all cells or selectively run cells using any of the options specified.

---

[1] https://www.python.org/downloads/
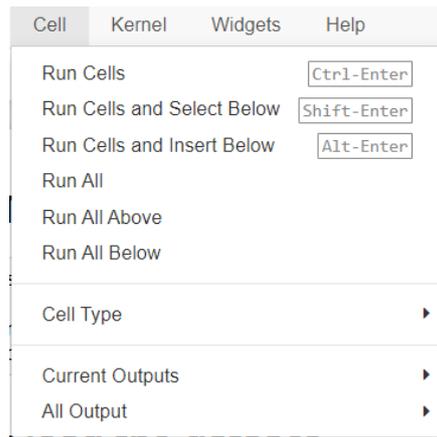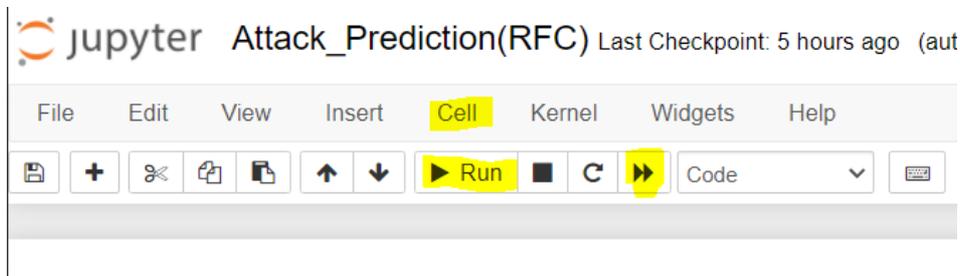[2] https://jupyter.org/install.html

Figure 2: Cell Tab

- Another way the code can be executed is by clicking on Run or by clicking on two arrows which will run the entire environment in one go, the only difference being that this will also restart the kernel, so be careful of unsaved changes.
- A possible error that can be encountered while code execution is during Logistic Regression classification. An error of *"value Error: Infinity, NaaN, or object value detected instead of float64"* would pop up. This has nothing to do with the code or the dataset, but it is mainly because of restarting the kernel frequently. To solve this error please use the following code.

```
profiling.drop("Unnamed: 43",axis=1,inplace=True)
df.drop("Unnamed: 43", axis=1,inplace=True)
```

Figure 3: Error Correction

- The tester is at liberty to change variable and method declarations at their own will, just make sure to call the newly declared variables correctly.
- Prediction can be made using the following code:
  - It is necessary to first load the attack prediction index into the model. Thereafter, a new test variable must be declared and the suitable classifier must be called.

Code:	Snippet
Loading the index	predattack.drop("Attack type",axis=1,inplace=True) attacktype= predattack.iloc[6] [The tester can here change the parameter of iloc from 6 to any number of their choosing. Make sure to not exceed the dataset limit]
Creating a test set	test = pd.Series(attacktype).values.reshape(1,-1)
Calling the classifier	randomforestclass.predict(test)