

# Configuration Manual

MSc Research Project  
Cybersecurity

Michael O'Brien  
x20191359

School of Computing  
National College of Ireland

Supervisor: Ross Spelman

National College of Ireland

Project Submission Sheet – 2021/2022

Student Name: Michael O'Brien  
 Student ID: x20191359  
 Programme: Cybersecurity Year: 2022  
 Module: MSC Research Project  
 Lecturer: Ross Spelman  
 Submission Due Date: December 15<sup>th</sup> 2022  
 Project Title: Evaluate the use of Supervised Machine Learning Algorithms in the detection of phishing attacks (Configuration Manual).  
 Word Count: 658 (Excluding Cover Sheet and Project Submission Sheet)

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 references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: *Michael O'Brien*  
 Date: 01<sup>st</sup> December 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

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

# 1 Research Lab

A personal PC, MacBook Pro, was used to set up the testing lab for this research, with the following specification

- Processor – 2.6 GHz 6-core Intel Core i7
- Memory – 32 GB 2400 MHz ddr4
- Startup Disk – Macintosh HD
- Graphics – Intel UHD Graphics 630 1536 MB
- OS – macOS Monterey (Version 12.6)

## 2 Applications

### 2.1 Anaconda

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. It is an open-source package and environment management system that runs on Windows, macOS, and Linux. It quickly installs, runs, and updates packages and their dependencies. It also easily creates, saves, loads, and switches between environments on your local computer. It was created for Python programs, but it can package and distribute software for any language (Anaconda, 2022).

Version Used: 2022.10 (This comes with Python 3.9 and conda v22. 9.0)

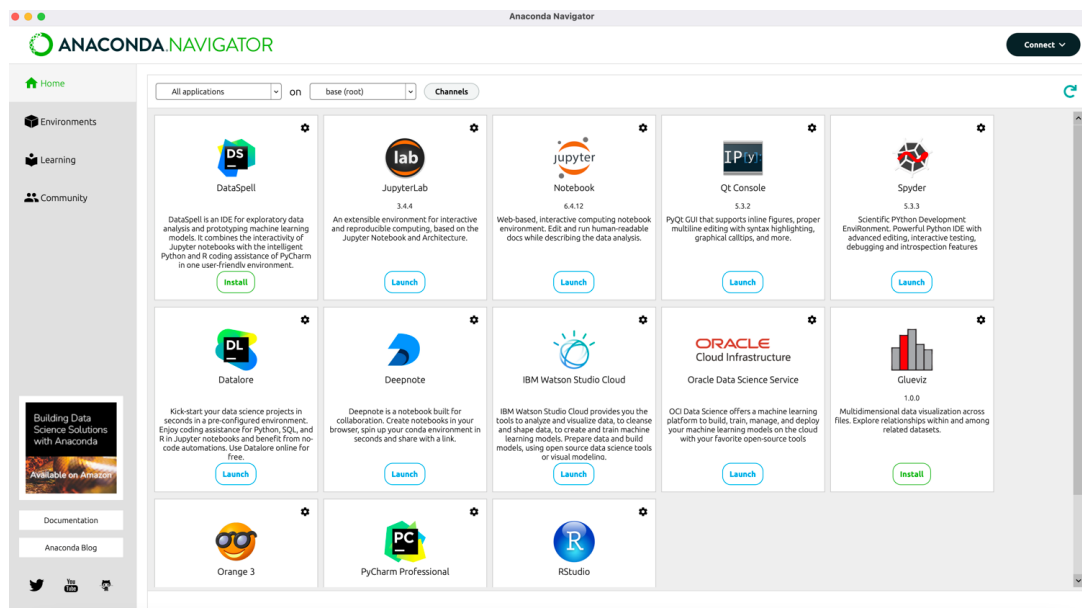


Figure 1: Anaconda Dashboard

### 2.2 Jupyter Notebook

Jupyter is an open-sourced project created to support interactive data science and scientific computing across programming languages. Jupyter offers a web-based environment for working with notebooks containing code, data, and text. Jupyter notebooks are the standard workspace for most data scientists (Anaconda, 2022).

Version used: 6.4.12



Figure 2: Jupyter Notebook Dashboard

## 2.3 Python (Pre-installed in Anaconda)

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed (Python, 2022).

Python 3 is a newer version of the Python programming language which was released in December 2008. This version was mainly released to fix problems that exist in Python 2. The nature of these changes is such that Python 3 was incompatible with Python 2. It is backward incompatible (guru99, 2022)

## 2.4 Conda (Pre-installed in Anaconda)

Conda is an open-source package and environment management system that runs on Windows, macOS, and Linux. Conda quickly installs, runs, and updates packages and their dependencies. It also easily creates, saves, loads, and switches between environments on your local computer. It was created for Python programs, but it can package and distribute software for any language (Anaconda, 2022).

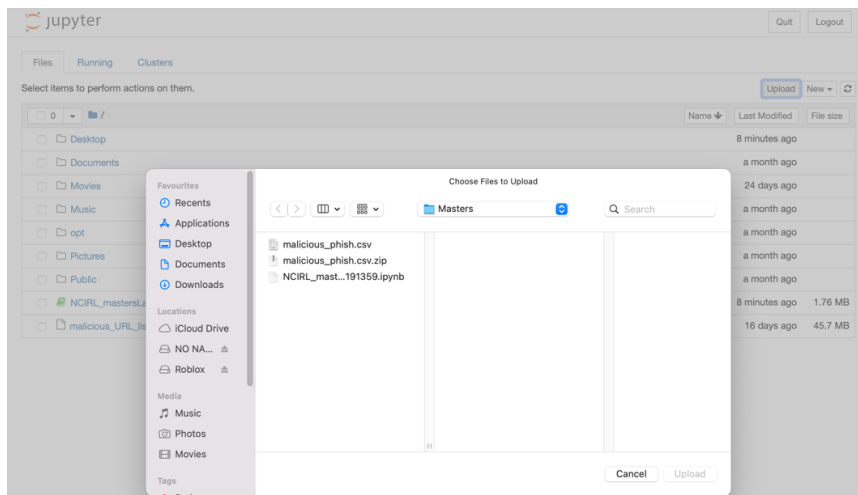
Version used: 22.9.0



## 2) Open Jupyter Notebook

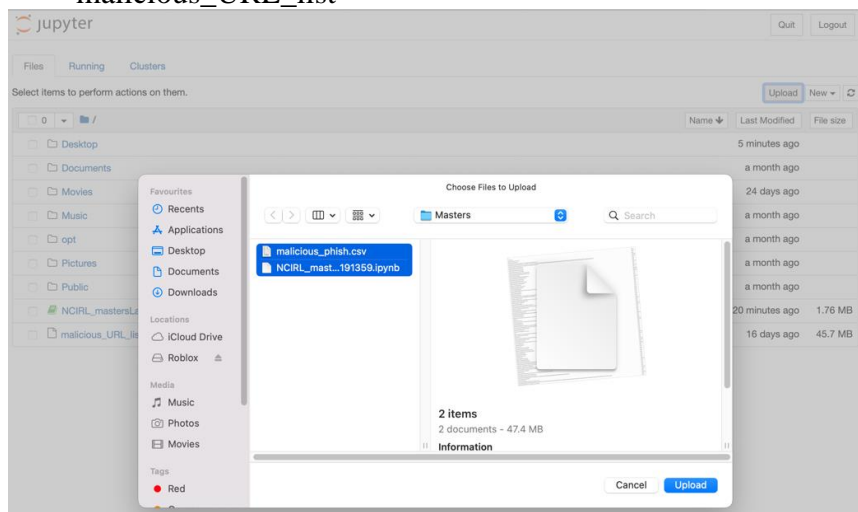


## 3) Click Upload



## 4) Select Files and click upload

- NCIRL\_mastersLab\_x20191359
- malicious\_URL\_list



## 5) Click on NCIRL\_mastersLab\_x20191359.ipynb



## 6) Start Running code

The screenshot shows the JupyterLab code editor with the following code and output:

```
In [9]: #importing required libraries
import itertools
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
import seaborn as sns
from wordcloud import WordCloud

In [15]: #importing dataset and viewing shape of dataset
data=pd.read_csv('malicious_URL_list.csv')
print(data.shape)
data.head()

(651200, 2)

Out[15]:
```

	url	type
0	br-icloud.com.br	malicious
1	mpdrad.com/music/fezz_kalkko.html	benign
2	bopsecrets.org/tesroth/or?1.htm	benign
3	http://www.garage-pirene.be/index.php?option=...	malicious
4	http://adventure-nicaragua.net/index.php?option=...	malicious

```
In [16]: #counting data types
data.type.value_counts()

Out[16]:
benign    428112
malicious 223088
Name: type, dtype: int64

In [17]: #defining the data types to create WordCloud
data_malicious = data[data.type == 'malicious']
data_benign = data[data.type == 'benign']
```

# 5 Bibliography

- Anaconda, 2022. *anaconda.com*. [Online]  
Available at: <https://www.anaconda.com/>
- Anaconda, 2022. *Installing on macOS*. [Online]  
Available at: <https://docs.anaconda.com/anaconda/install/mac-os/>  
[Accessed November 2022].
- guru99, 2022. *Python 2 vs Python 3*. [Online]  
Available at: <https://www.guru99.com/python-2-vs-python-3.html>
- Python, 2022. <https://www.python.org/doc/essays/blurb/>. [Online]  
Available at: <https://www.python.org/doc/essays/blurb/>