



National
College of
Ireland

Configuration Manual

MSc Research Project
Cyber Security

Nagasunder Rao Pawar Babu Rao Pawar
Student ID: X20107668

School of Computing
National College of Ireland

Supervisor: Michael Pantridge

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Nagasunder Rao Pawar Babu Rao Pawar

 X20107668

Student ID:
 Cyber Security 2021

Programme: **Year:**
 MSc Research Project

Module:
 Michael Pantridge

Lecturer:
Submission Due Date: 16th Aug 2021

Project Title:
 Detection of Phishing URL using Machine Learning

Word Count: 603 **Page Count:** 7

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.

Signature: Nagasunder Rao Pawar Babu Rao Pawar

 16th Aug 2021

Date:

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):	

Configuration Manual

Nagasunder Rao Pawar Babu Rao Pawar
X20107668

1 Introduction

This Configuration Manual paper contains information about the project's technical resources, technologies, and tools. This guidebook also includes a step-by-step implementation guide. The method includes downloading and installing the required programs and services, as well as the basic configuration required to keep the project running properly.

2 Environment [Hardware and Software]

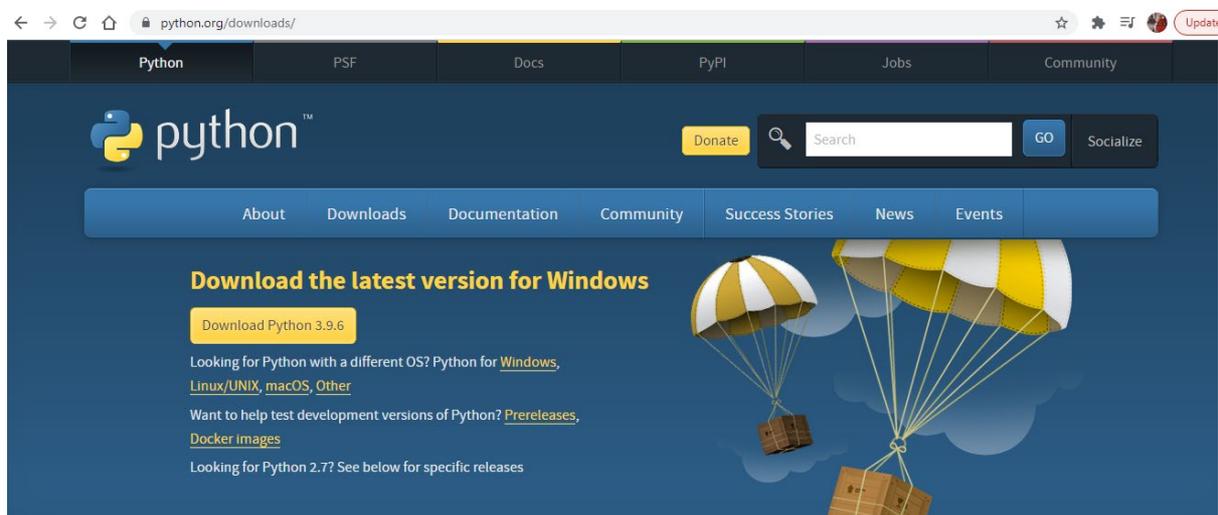
The proposed solution is implemented with below specification and configuration.

- Processor: Intel i5
- Speed: 2GHz
- Memory: 8GB RAM
- Programming language: Python
- Environment: Jupyter Notebook, Anaconda

3 Download and Installation

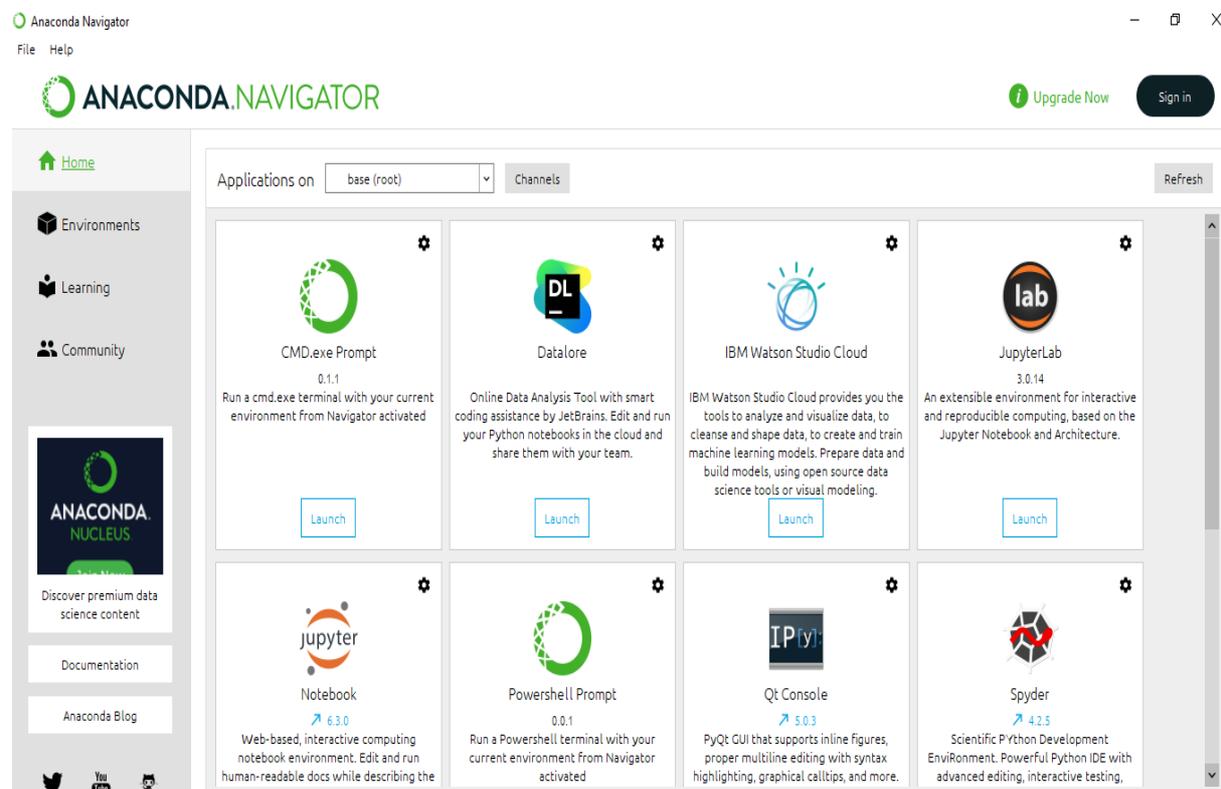
3.1 Python

For development purposes, the most recent version of Python was downloaded from the official website. It's free to download and use because it's Open Source.[1]



3.2 Anaconda

Anaconda is a free and open-source distribution of the Python and R programming languages that simplifies package management for data science and machine learning applications. For package management, the most recent version of Anaconda was downloaded from the official website.[2]



3.3 Jupyter Notebook

The Jupyter Notebook is a web-based open-source program that allows you to create and share documents with live data, visualizations, calculations, and other features. It's used for cleaning and transforming data, numerical simulation, data visualization, statistical modelling, and machine learning, among other things.

Select items to perform actions on them.

Upload New ↻

0	Name	Last Modified	File size
0	/		
	3D Objects	10 months ago	
	anaconda3	a month ago	
	Contacts	10 months ago	
	Documents	3 months ago	
	Downloads	seconds ago	
	Favorites	11 days ago	
	Links	10 months ago	
	Music	10 months ago	
	OneDrive	11 days ago	
	Postman	2 months ago	
	RyanairTask	2 months ago	
	Saved Games	10 months ago	
	Searches	10 months ago	
	source	3 months ago	

4 Development

Step 1: Importing basic packages

```
#importing basic packages
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Step 2: Loading the dataset

```
#Loading the data
data0 = pd.read_csv('final.csv')
data0.head()
```

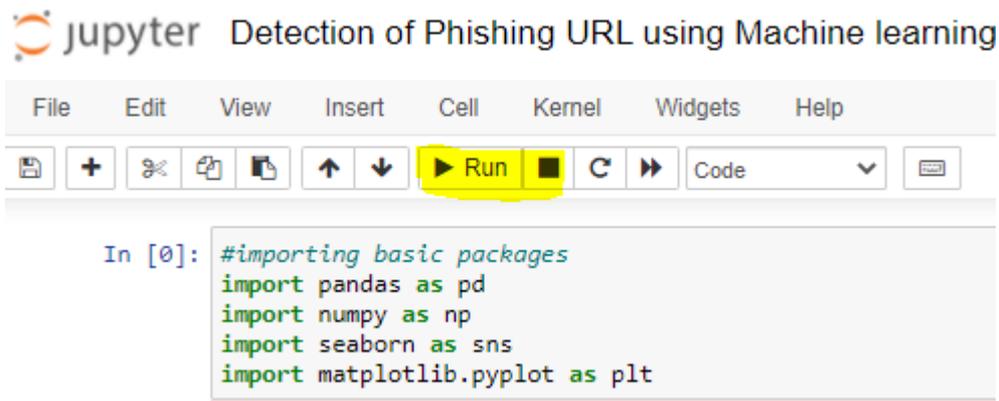
Step 3: Dataset Information

```
#Information about the dataset
data0.info()

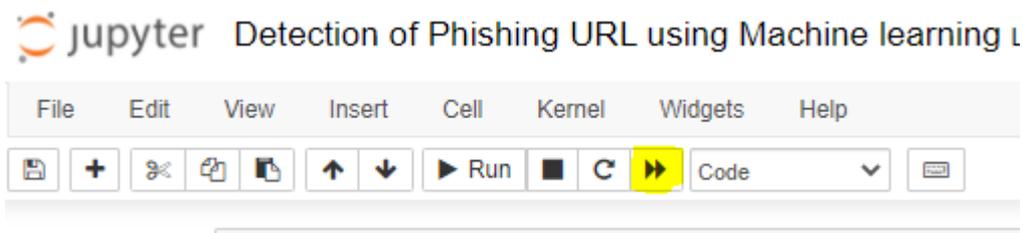
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype
---  ---             
0   Domain               10000 non-null  object
1   Have_IP              10000 non-null  int64
2   Have_At              10000 non-null  int64
3   URL_Length           10000 non-null  int64
4   URL_Depth            10000 non-null  int64
5   Redirection          10000 non-null  int64
6   https_Domain        10000 non-null  int64
7   TinyURL              10000 non-null  int64
8   Prefix/Suffix        10000 non-null  int64
9   DNS_Record           10000 non-null  int64
10  Web_Traffic           10000 non-null  int64
11  Domain_Age           10000 non-null  int64
12  Domain_End           10000 non-null  int64
13  iFrame               10000 non-null  int64
14  Mouse_Over           10000 non-null  int64
15  Right_Click          10000 non-null  int64
16  Web_Forwards         10000 non-null  int64
17  Label                10000 non-null  int64
```

Step 4: Execution of Code

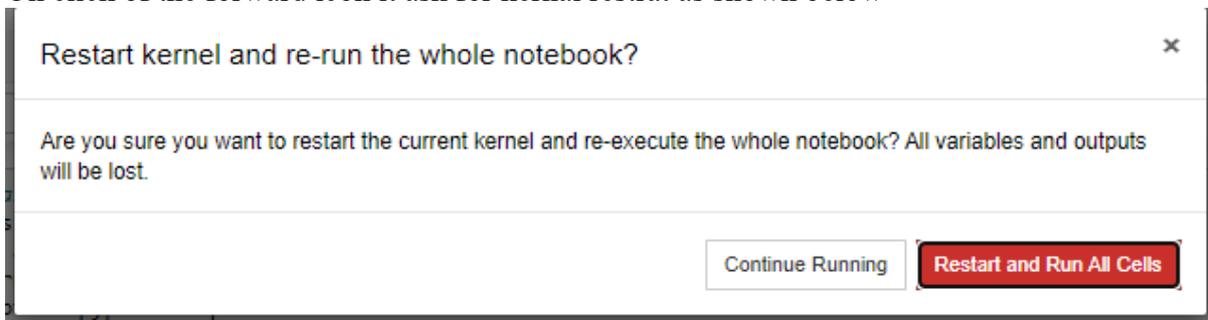
In notebook the bit of code can be executed instead of whole project



Click on forward icon to execute the entire code



On click of the forward icon it ask for kernal restrat as shown below



Once the code is executed the comparison model of the accuracy will be reflected as below.

```
#Sorting the datafram on accuracy  
results.sort_values(by=['Test Accuracy', 'Train Accuracy'], ascending=False)
```

	ML Model	Train Accuracy	Test Accuracy
3	XGBoost	0.866	0.864
2	Multilayer Perceptrons	0.858	0.863
1	Random Forest	0.814	0.834
0	Decision Tree	0.810	0.826
4	AutoEncoder	0.819	0.818
5	SVM	0.798	0.818

References

- [1] Download Python [online] (2021) Python.org, available: <https://www.python.org/downloads/> [accessed 14 Aug 2021].
- [2] Anaconda | The World's Most Popular Data Science Platform [online] (2021) Anaconda, available: <https://www.anaconda.com/> [accessed 14 Aug 2021].