Configuration Manual

MSc Internship

Cyber Security

Mohammed Afnan Ikram

Student ID: 18189725

School of Computing

National College of Ireland

Supervisor: Ross Spelman

National College of Ireland

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Nam	Mohammed Afnan Ikram		
Student ID:	18189725		
Programme:	Cyber Security	**Year**	2019-20
Module:	MSc Internship		
Supervisor:	Ross Spelman		
Submission Due Date:	17 September, 2020		
Project Title:	Phishing Detection System using Dueling Network		
Word Count:	651 **Page Count** -8		

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: 17, September 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):	

Phishing Detection System using Dueling Network

Mohammed Afnan Ikram

Student ID: 18189725

1. Summary

The proposed research work is done on phishing detection system with dueling network. This document gives details of all the software/tools and setting required for successfully execution of this program.

2. Structure

Information	purpose
Basic Information	This section describes the basic requirements of tools and software needed for the application
Process of Deployment	This section describes how the application should be setup and run.

3. Basic Information

The objective of this program is to fulfil the purpose of phishing detection with high accuracy using dueling network. The program consists of different modules which are briefly explained in the below deployment section.

4. Architecture Requirement

This program is a python code, so to run this program, python is required which can be installed from python official website. Python 3.8 is preferred as it is comparatively more stable. There are few additional python packages (numpy, pandas, seaborn) are also required if older version is used. OR Anaconda3 (preferred) is installed which is freely available on its official website. Anaconda3 is recommended as it is user friendly and give access to Jupiter notebook and spyder, which can also be used for executing the code. Going forward, further explanation and examples are given in Jupiter notebook; which can be used from Anaconda or could be installed separately from Jupiter official website. I preferred using Jupiter as it is web based, gives interactive environment and human readable doc can easily be run or edited.

5. Deployment procedure

5.1. Execute Application

Make sure all the files (including dataset) are in the same folder. It will automatically detect the dataset file in that folder, else we'll have to mention dataset file path separately in code (which is also shown in below explanation). Run juniper notebook and open python file "duelingNetworkforPhishing.ipynb" as shown in below figure1.

Figure1

Once code is open, import required libraries by typing the following commands. Also shown in figure2,[1][2]

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

Dueling Network To Detect Phishing

Importing required libraries

```
In [1]: import numpy as np
        import pandas as pd

        import matplotlib.pyplot as plt
        %matplotlib inline
        import seaborn as sns
        from sklearn.model_selection import train_test_split
        from sklearn.ensemble import RandomForestClassifier
        from sklearn.metrics import accuracy_score
```

Figure2

5.2. Importing Dataset

Import the dataset by mentioning the dataset set file name as shown in below figure3, though it automatically picks the dataset file, if it is saved in a same python folder.

Importing phising data

```
In [2]: data = pd.read_csv("dataset.csv")
        data.head()
```

Figure 3

5.3. Tuning

In this section of code, we can change random selection count, which is set to 10 for Q1 and 100 for Q2. We can change the reward points, which is currently set to 10; and we can also change the accuracy rate condition of rewards, which is currently set to 95% as shown in figure 4 [4].

Note: As everything is already set to its optimal, this part is optional, and just shows how we can make certain changes in data settings, going forward.

Dueling Q Network

Training the model

```
In [13]: # Setting data
num_data = len(X_train)
model1=RandomForestClassifier(n_estimators=10)
model1.fit(X_train,y_train)
model2=RandomForestClassifier(n_estimators=100)
model2.fit(X_train,y_train)
#Set Reward
def getReward(accuracy):
    if accuracy>=0.95:
        return 10
    else:
        return 0
```

Figure4

5.4. Code Execution

In Jupiter notebook, we can test and run code in chunks, instead of running the entire code at once. Just select the code which is intended to run and then click on single forward arrow icon as shown in below figure5.

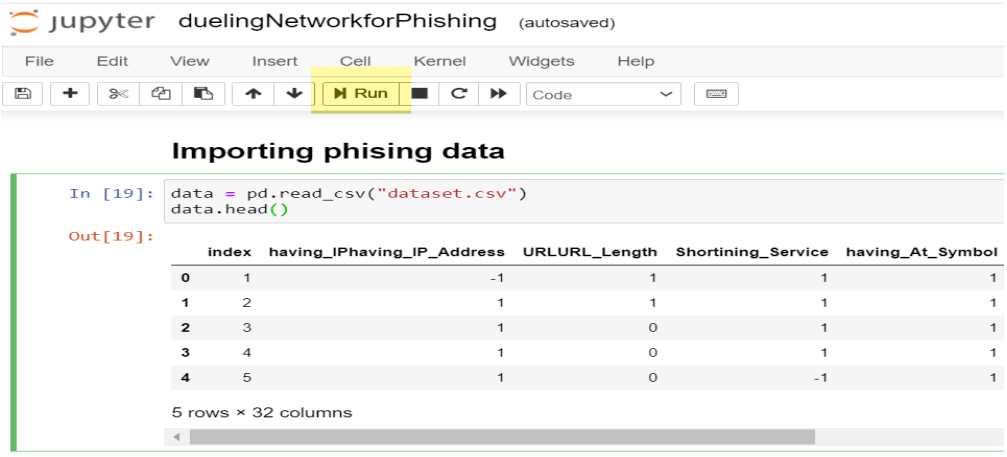

Figure5

Run the entire code at once to get accuracy result, clicking on the icon with two forward arrows as shown in the below figure6.

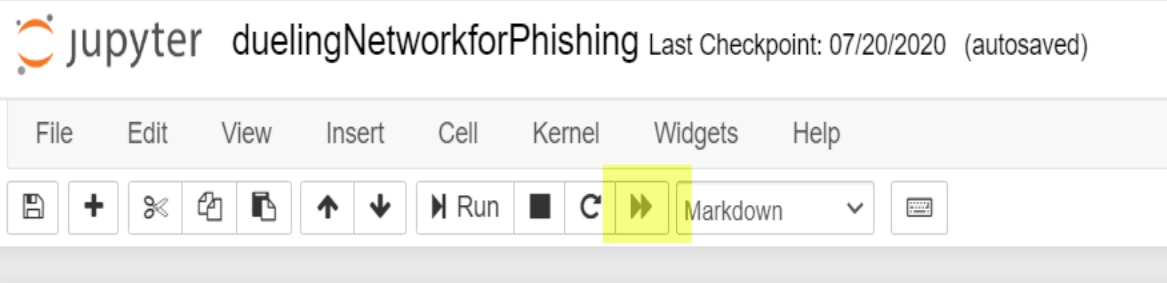

Figure6

Once clicked, a dialogue box will pop asking “Restart Kernel and re-run the whole notebook”
Click on “restart and run all cells” as shown in figure 7.

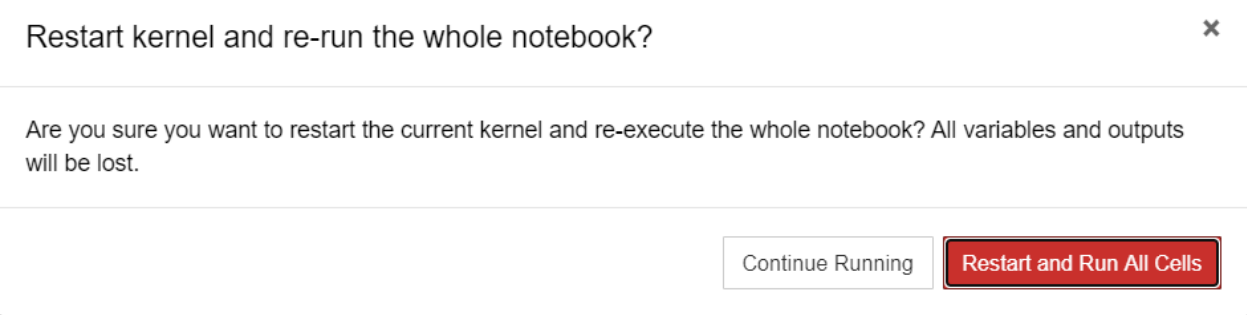

Figure7

Once the complete program is executed, the final accuracy result will be displayed in the end as shown in below figure8.

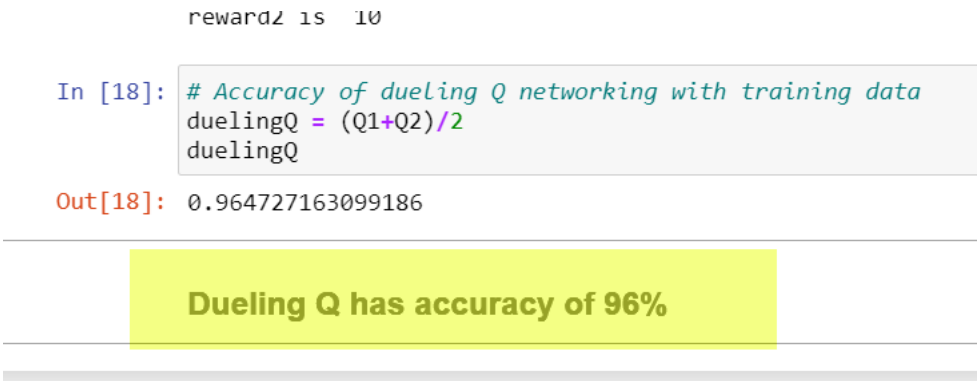

Figure 8

6. References

[1] W. McKinney, *Python for data analysis*, 1st ed. United States of America: O'Reilly books Media, Inc, 2013, pp. 1-377.

[2] T. Oliphant, *A guide to NumPy*. [Trelgol Publishing], 2006.

[3] H Thomas. “Data Input. In: An Introduction to Statistics with Python” *Springer,Cham*,2016, doi: https://doi.org/10.1007/978-3-319-28316-6_3

[4] [23]J. Boisberranger et al., "Scikit-learn: Machine Learning in Python", *Scikit-learn.org*, 2007. doi: https://scikit-learn.org/stable/search.html?q=randomforestclassifier+estimators.