National College of Ireland

Configuration Manual

MSc Research Project
MSc Cyber Security

Suraj Wandhare
Student ID: 18157432

School of Computing
National College of Ireland

Supervisor: Ben Fletcher

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Suraj Wandhare
Student ID:	18157432
Programme:	MSc Cyber Security
Year:	2019
Module:	MSc Research Project
Supervisor:	Ben Fletcher
Submission Due Date:	03/02/2020
Project Title:	Configuration Manual
Word Count:	538
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.

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:	3rd February 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):	

Configuration Manual

Suraj Wandhare
18157432

1 Introduction

1.1 Purpose of this document

This document contains the required software tools and settings in order to successfully run the Phishing detection application.

1.2 Document Structure

Section	Purpose
General Information	This module explains the general prerequisites and software requirements for the application
Deployment Procedure	This module explains how to set up and run the application

2 General Information

2.1 Objective

The main objective of this application is to provide an interface in which the user enters a URL and the application based on the modules, returns the result if it is a phishing website.

The application consists of three modules. The first is the URL unshortener, which takes in a short URL and returns it in unshortened form. The file 'list.txt' contains the list of top 400+ url shortening services to compare from.

Second is the database check, the file 'verified_online.json' contains the 30000+ urls classified as phishing taken from PhishTank.com.

Third is the machine learning algorithm. The trained data is saved in the file 'finalised_model.sav' so there is no need to re-train the data. The file 'rf_model.py' contains the necessary code for the prediction using the trained model. The files dataset2, dataset3 are the training and testing datasets respectively. The application uses URL lexical analysis and the file 'feature_extractor.py' contains the required code with comments wherever necessary.

2.2 Architecture Requirements

This section describes required software installations and dependencies to run the application.
Python: Python3 is required for running the application. It can be downloaded online[1] from the python official website.
Python Dependencies: Additional python packages and modules required for running the application are

- numpy
- pandas
- sklearn

All of the above can be installed using pip. pip is a standard package management system to install and python packages.[2]

3 Deployment Procedure

3.1 Running the application

Make sure all the files are in the same folder. Open command prompt inside the folder and run the 'gui.py' file.

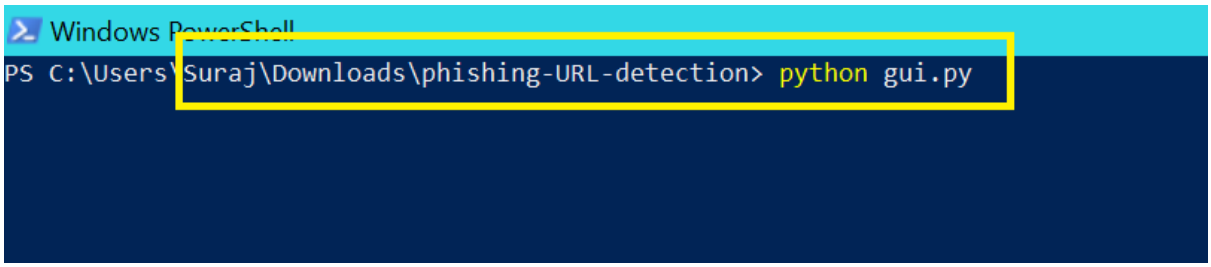

Figure 1: python gui.py

After a few seconds a window will show up as shown in figure 2
Enter the malicious URL into the text field provided and press check. The application will show two kinds of results.

1. Confidence Score: Meaning the application used the machine learning model to predict the webiste is legitimate or not. Example screenshots are 3 4
2. Phish ID: The phishtank database will return the Phish ID of the URL if it is present in the database. Example 5

The application is built in such a way to display all messages for the processes that are running in the background in the console from where the file gui.py is runnning. The figure 6 shows how to debug the application if it stops abruptly.

[1] `https://www.python.org/downloads/`
[2] `https://pip.pypa.io/en/stable/installing/`

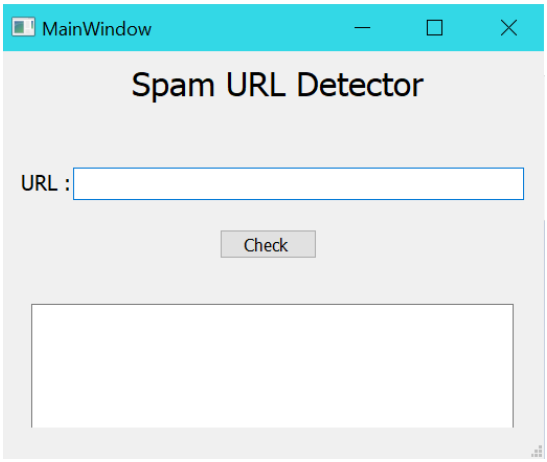

Figure 2: Main Window

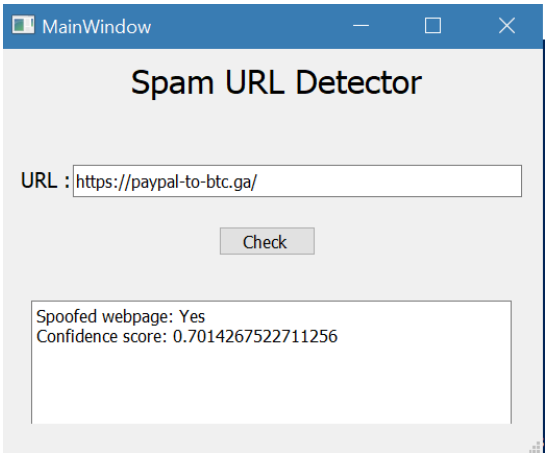

Figure 3: Result 1

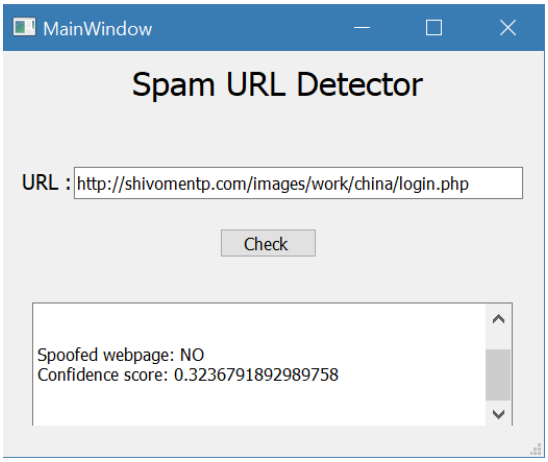

Figure 4: Result 2

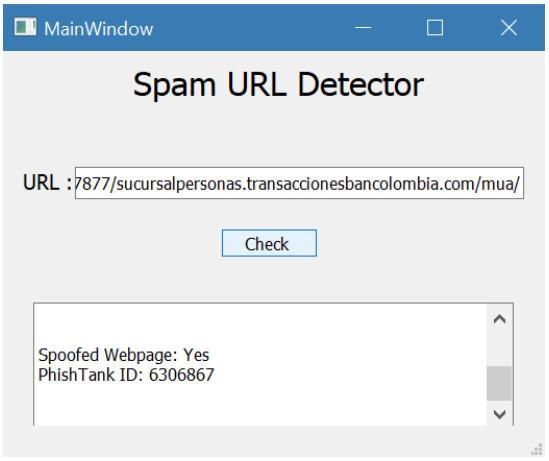

Figure 5: Result 3

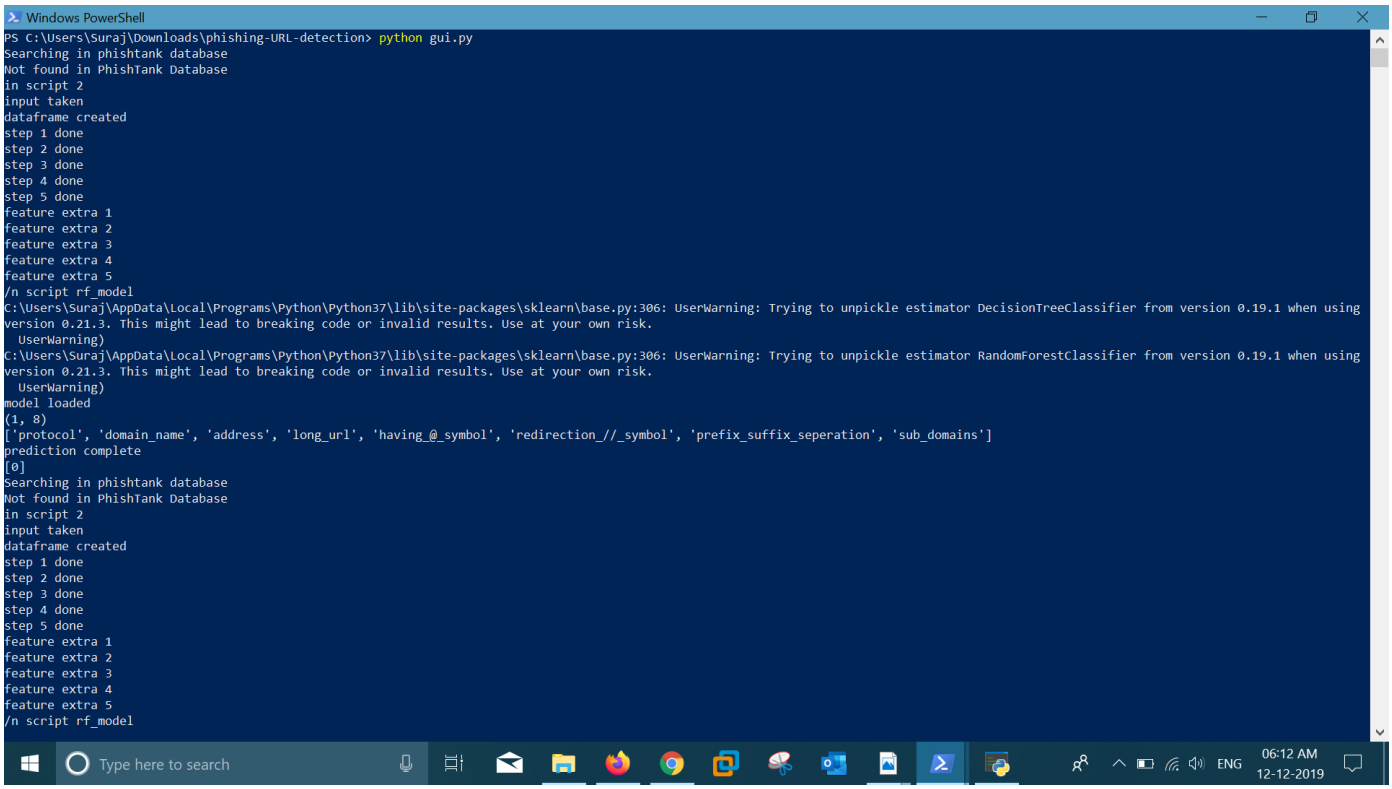

Figure 6: Background tasks