National College of Ireland

Detecting Phishing Emails with Machine Learning Algorithms

Configuration Manual

MSc Research Project

Cyber Security

Martin Casey

Student ID: x20256604

School of Computing

National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Martin Casey
Student ID:	X20256604
Program:	MSc Cybersecurity
Year:	2022
Module:	Research Project
Lecturer:	Vikas Sahni
Submission Due Date:	4th September 2022
Project Title:	Detecting Phishing Emails with Machine Learning Algorithms
Word Count:	
Page count	

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:
Date:

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 Program 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):	

Configuration Manual

Martin Casey
X20256604

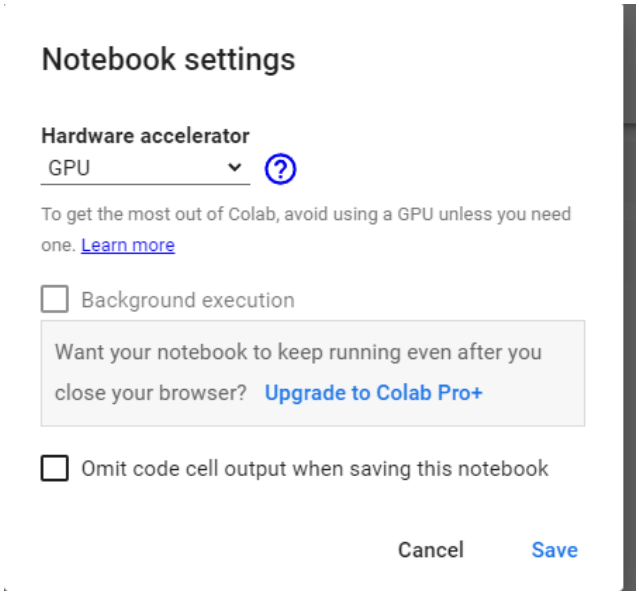
1 Introduction

The purposes of the configuration manual is to provide a step by step guide in the implantation of the project. The guide includes how to install the necessary software, download the required code from GitHub and download the dataset that is also in GitHub. For this implementation Google Colab was utilised, this was used to utilise the graphics card provided on the Colab instance.

2 Environment and Pre-Requisites

2.1 Pre-Requisites

1. Google account is needed to use google Colab this can be set up here https://www.google.com/account/about/
2. Navigate to Colab https://colab.research.google.com/ and create a new notebook.
3. An instance with a GPU needs to be utilised, this is configured under *Runtime* and *Change Runtime Type*



4. Check that the GPU is being utilised by running *!nvidia-smi* a similar out put to the below should be obtained

```
[ ] !nvidia-smi

Sat Aug  6 17:42:00 2022
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 460.32.03    Driver Version: 460.32.03    CUDA Version: 11.2     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
| N/A   56C    P8    11W /  70W |      0MiB / 15109MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

2.2 Colab Environment

Table 1 Shows specifications of the machine used to carry out the ML models

RAM	**14GB**
Hard Disk Space	100GB
CPU Count	2
CPU Info	Intel(R) Xeon(R) CPU @ 2.20GHz
GPU	Tesla T4
Python Libraries	• Pandas • Numpy • Matplotlib • Sklearn • Cuml • Rapids

3 Configuration and Implementation

1. Download the ipynb file from the GitHub repository (https://github.com/martincasey1/Masters_project) and upload it to Colab from the *File* tab. Report and other documents are also available at this repository.
2. Run the code cells from the first one to the ten in order waiting for each one to finish, this installs RAPIDS on the instance.
3. Upload the dataset from the GitHub repository ensuring the name is not changed.
4. Run the remaining code cells to output the results. Code cells should be run in order waiting for each one to run before the next one is run.