National College of Ireland

Configuration Manual

MSc Research Project
Master of Science in Cyber Security

Kousic Bhadan Janarthanan
Student ID: x20177313

School of Computing
National College of Ireland

Supervisor: Dr Vanessa Ayala-Rivera

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. Kousic Bhadan Janarthanan ……

Student ID: …… x20177313……..……..

Programme: …… MSc Cyber Security …… **Year:** …2021-2022...

Module: …… MSc Research Project ……..……

Lecturer: …… Dr Vanessa Ayala-Rivera…..……..……

Submission Due Date: ……16/12/2021……..……..……

Project Title: … Preventing the Inbound Malware Using Recommendation Algorithm

Word Count: ……858…… **Page Count:** ……..…6……..……..……

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: ……

Date: ……

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

Kousic Bhadan Janarthanan
X20177313

1 Introduction

Machine learning and data analysis techniques are used in this study. It involves number key prerequisites in order for deployment to perform properly. This setup manual describes the step-by-step approach for deploying this research model. In this section, we discussed the software, hardware, integration framework, and storage architecture used in this experiment.

2 System Configuration

2.1 Hardware Requirements

Our study model is deployed in the Google Collaboratory environment, which allows any Python code to be executed in a browser window. For run our model training data set, we needs lots of Computational resources and GPU capacity. As a result, the Google Collaboratory environment offers limited free RAM and GPU to execute the code, as well as twelve hours of uninterrupted GPU usage, that's more than enough to perform the entire training phase.

The following are the hardware specifications for the host system:

- Processor: Intel Core i3 – 2.30 GHz
- RAM: 8 GB – DDR3
- Hard Drive: 160 GB SSD

3 Software Requirements

Python 3.6 and data analysis libraries are used to run this project. This prototype code is composed in a Google Collaborative notebook that is hosted by the TensorFlow framework. Google Collaborate is a user-friendly environment that allows for step-by-step code execution and the installation of any data analysis modules. The majority of machine learning libraries are preconfigured in Google Collaboratory. In our model, we employed a classification algorithm to explore malware in the data samples that we chose. Before beginning the model deployment, the dataset must be uploaded to Google Drive, as detailed in detail in the below section.

The major software and operating system which require for deployment mentioned below:

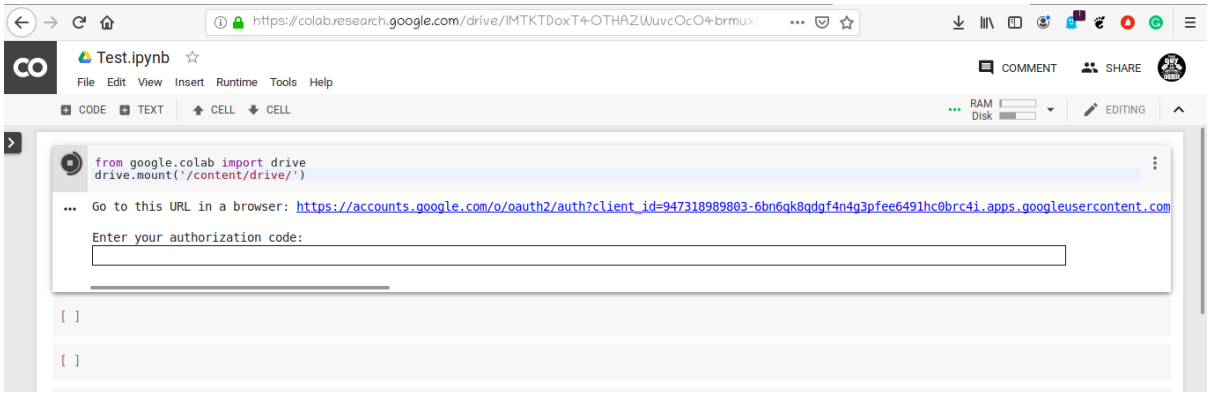
- Operation System: Windows 10 - 64 bit
- Web Browser: Microsoft Edge / Google Chrome

3.1 Loading out Data set into Google Collaboratory

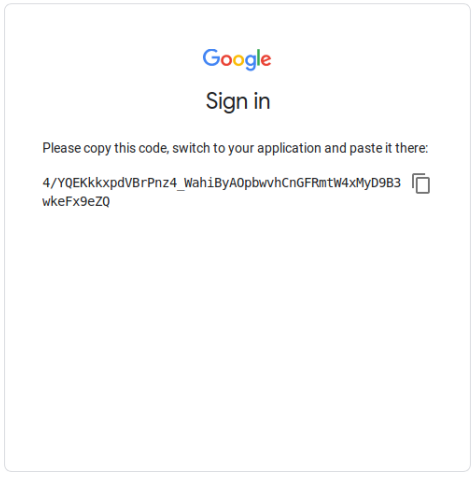
- First and foremost, upload our data set file to Google Drive.
- Then in the Collaboratory console, execute the following code.

```
from google.colab import drive

drive.mount('/content/googledrive')
```



- Navigate to the highlighted blue link.



- Copy the authentication token and paste it into the token section box.
- Our Google Drive is now accessible at /content/googledrive/My Drive/.

The majority of analysis libraries are pre-installed, making it simple for using; libraries that are not pre-loaded here may be loaded with a simple command.

```
pip install package_name
```

```
[ ] pip install imblearn

    Requirement already satisfied: imblearn in /usr/local/lib/python3.7/dist-packages (0.0)
    Requirement already satisfied: imbalanced-learn in /usr/local/lib/python3.7/dist-packages (from imblearn) (0.8.1)
    Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn->imblearn) (1.1.0)
    Requirement already satisfied: scikit-learn>=0.24 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn->imblearn) (1.0.1)
    Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn->imblearn) (1.19.5)
    Requirement already satisfied: scipy>=0.19.1 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn->imblearn) (1.4.1)
    Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn>=0.24->imbalanced-learn->imblearn) (3.0.0)
```

Once the datasets is set up, ensure the packages listed here are installed in the Collaboratory environment because they are utilized in our model.

- Matplotlib - 3.2.2
- Sklearn - 1.0.2
- Imblearn - 0.8.1
- Numpy - 1.19.5
- Seaborn - 0.11.2
- Pandas - 1.1.5

4 Implementation and Evaluation

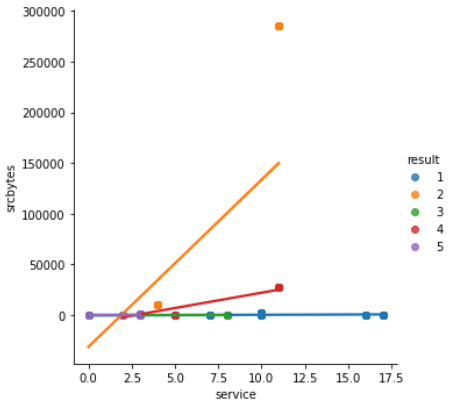
Initial Stage is Data pre-processing.

DataFrame can describe, convert objects to integers, and do a variety of other things using the pandas package.

```
[ ] df.describe()
```

	duration	srcbytes	dstbytes	land	wrongfragment	urgent	numfailedlogins	loggedin	rootshell	numroot
count	60938.000000	6.093800e+04	6.093800e+04	60938.000000	60938.000000	60938.000000	60938.000000	60938.000000	60938.000000	60938.000000
mean	10.746890	7.480419e+02	3.582470e+03	14.193935	16.794233	0.994402	0.007781	0.111512	163.519462	231.014556
std	536.641023	3.416999e+04	3.616603e+04	53.988190	54.906234	0.059121	0.080519	0.234405	102.753185	60.727626
min	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	1.050000e+02	1.460000e+02	1.000000	2.000000	1.000000	0.000000	0.000000	50.000000	253.000000
50%	0.000000	2.260000e+02	5.700000e+02	4.000000	5.000000	1.000000	0.000000	0.000000	245.000000	255.000000
75%	0.000000	3.020000e+02	2.507000e+03	11.000000	15.000000	1.000000	0.000000	0.120000	255.000000	255.000000
max	54451.000000	6.291668e+06	5.203179e+06	511.000000	511.000000	1.000000	1.000000	1.000000	255.000000	255.000000

Then using Seaborn library is possible to visualize the graphical representation of our dataset.



Once all of the cells have completed the execution, we can use the theoretical accuracy rate to determine the evaluation result of our study. Accuracy is defined as the number of right predictions from total number of projections.

```
[ ]  knn.score(X_test,y_test)

0.9998594639062178
```

The performance of our classification model may then be determined using the confusion matrix.

```
Confusion Matrix
[[48609     3     0     0     0     0]
 [    0 48717     0     0     0     0]
 [    0     0 48534     0     0     0]
 [    0     0     0 48663     0     0]
 [    0     0     0     0 48620     0]
 [    0     0     0     0     0 48594]]
Test Set: 291740
Accuracy = 99.99897168711867 %
```

References

Analytics Vidhya. (2020). *Use Google Colab for Deep Learning and Machine Learning Models*. [online] Available at: https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/.

colab.research.google.com. (n.d.). *Google Colaboratory*. [online] Available at: https://colab.research.google.com/?utm_source=scs-index# [Accessed 9 Dec. 2021].

Quantitative Trading Ideas and Guides - AlgoTrading101 Blog. (2021). *Google Colab - A Step-by-step Guide - AlgoTrading101 Blog*. [online] Available at: https://algotrading101.com/learn/google-colab-guide/.

TensorFlow. (n.d.). *Install TensorFlow 2*. [online] Available at: https://www.tensorflow.org/install.