Configuration Manual

MSc Research Project
MSc Cyber Security

Bhargav Chowdary Rayankula
X21138508

School of Computing
National College of Ireland

Supervisor: Dr. Arghir-Nicolae Moldovan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Bhargav Chowdary Rayankula
Student ID:	x21138508
Programme:	MSc Cyber Security **Year:** 2023
Module:	MSc Research Project
Lecturer:	Dr. Arghir-Nicolae Moldovan
Submission Due Date:	29th May 2023
Project Title:	An Evaluation and Performance study on BODMAS dataset for Malware Analysis

Word Count: 393 **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: Bhargav Chowdary

Date: 29th May 2023

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

Bhargav Chowdary Rayankula
X21138508

1 Introduction

The handbook contains documentation on all of the relevant tools and technologies that are needed to put the research model into action. The handbook is broken up into a few different parts for your convenience. In Section 2, we discuss the necessary configuration of the environment. In Section 3, we discuss the instruments and programmes that were used, and in Section 4, we discuss the actual execution of the project

2 Environment Setup

Below mentioned configuration was used to implement the model.

- Processor: MacOS m1
- Memory: 8GB RAM
- Programming language: Python3
- Python Environment: Jupyter Notebook, Google collab

3 Tools and Software Used

Software from the list below was utilized to put the model into practice.
• For programming purposes, Python 3.9.13 was utilized.

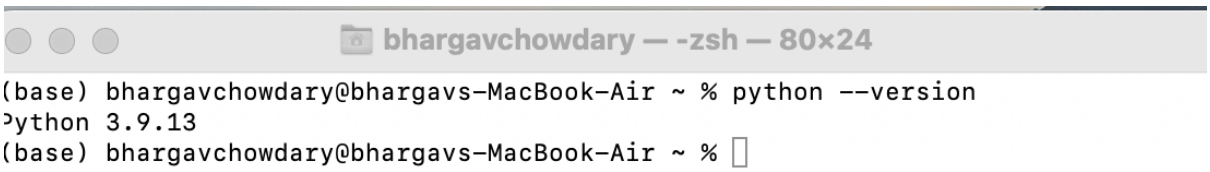

Figure 1. Python Version

- For performing operations on datasets, we have been using jupyter, a free and open-source Python programme. The Python kernel is used to perform processing and other operations on the datasets

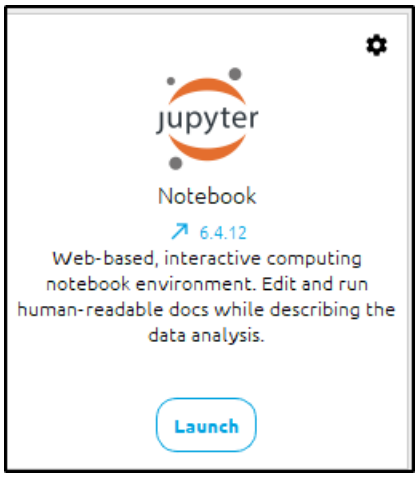

Figure 2. Python Jupyter

- For building models and evaluating accuracy scores we have used Google collab since we had less computation resources

4 Implementation of the Model

Step 1: - We have used google collab, to use python notebook files. Go to https://colab.research.google.com/ URL and sign in with your account

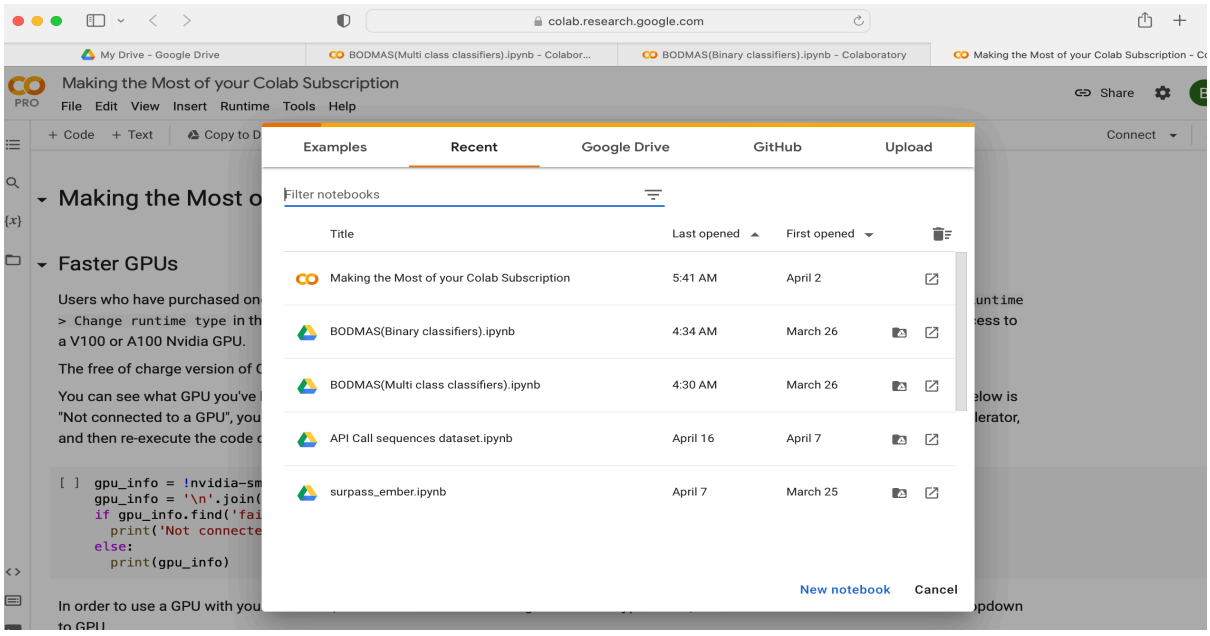

Figure 3. Google collab

Step 2: - Open the file “BODMAS (Binary classifiers).ipynb” and BODMAS (Multi class classifiers).ipynb in the google collab.

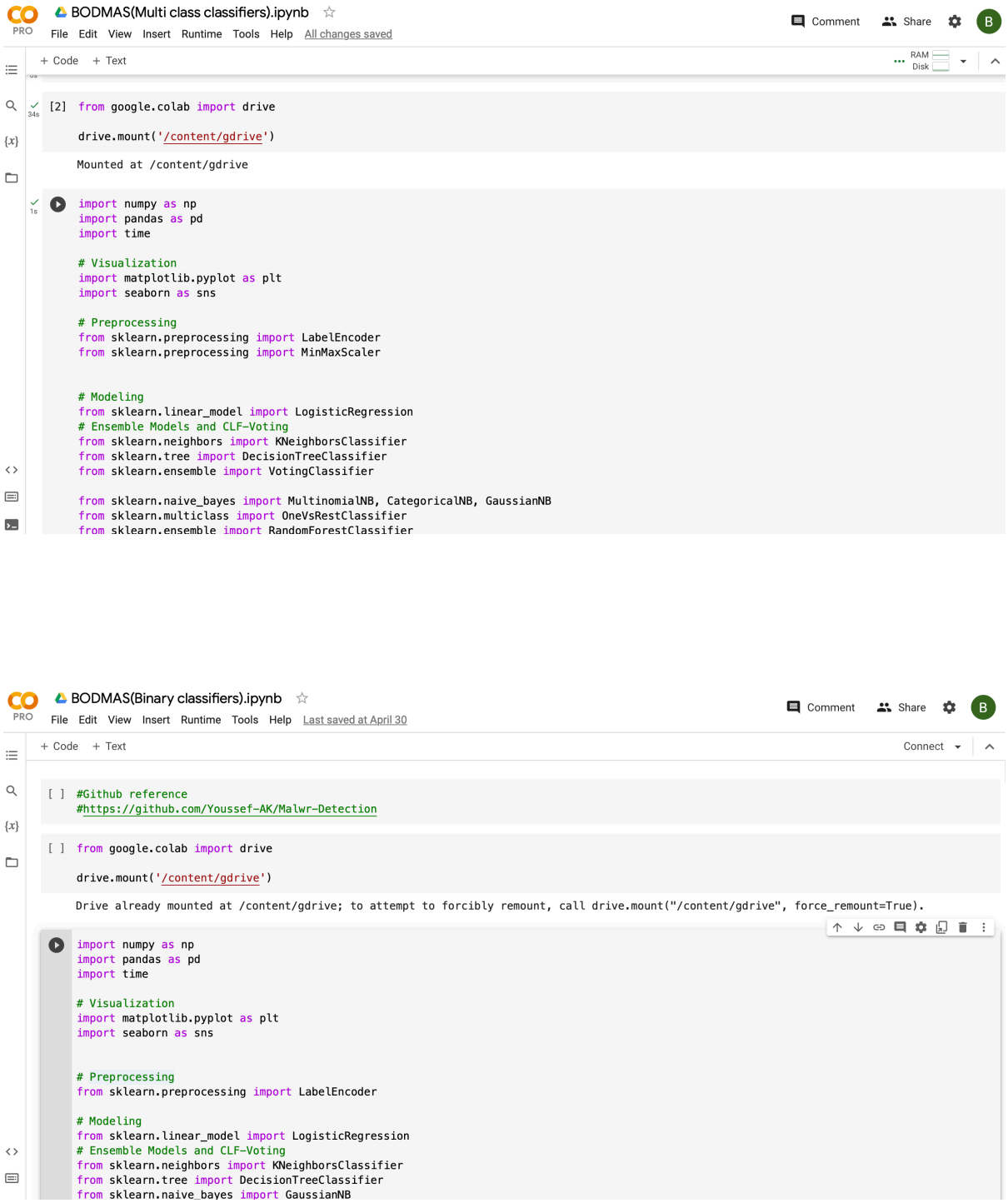

Figure 4. Ipynb file launched in Google collab

Step 3: - From here, we begin the model's implementation. We begin by importing all the model's required libraries.

```
import numpy as np
import pandas as pd
import time

# Visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Preprocessing
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler


# Modeling
from sklearn.linear_model import LogisticRegression
# Ensemble Models and CLF-Voting
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier

from sklearn.naive_bayes import MultinomialNB, CategoricalNB, GaussianNB
from sklearn.multiclass import OneVsRestClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import SGDClassifier



import xgboost as xgb
from xgboost import XGBClassifier
```

Figure 5. Libraries Imported for the Model

Step 4: -. After importing the libraries connect to the Google Compute Engine backend (GPU) and make sure connection is established or not.

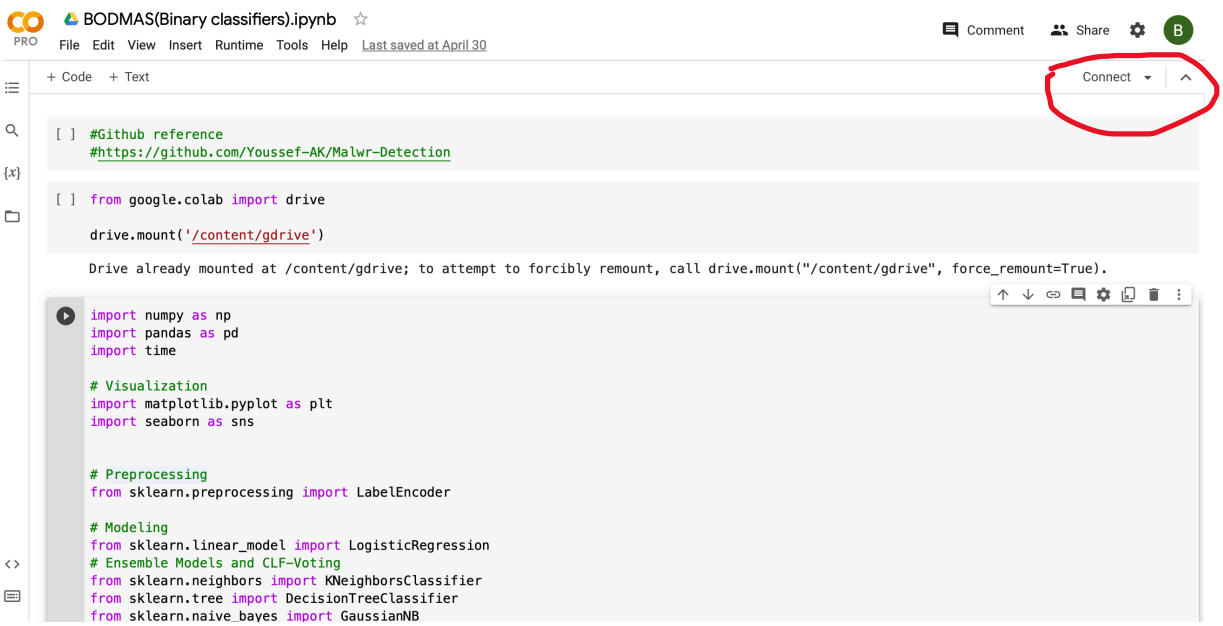

Figure 6. Connection to Network GPU

Step 5: - We use the pandas libraries to load the dataset, which is stored on Google Drive and accessible through Google Collab.

```
[1] #Github reference
    #https://github.com/Youssef-AK/Malwr-Detection

[2] from google.colab import drive

    drive.mount('/content/gdrive')

    Mounted at /content/gdrive

    import numpy as np
    import pandas as pd
    import time
```

Figure 7. Loading the Dataset from google drive

Step 6: - Features extracted from the dataset.

Figure 8. Dataset Features Extracted

Step 6: - Training the pretrained model with the test data and getting accuracies

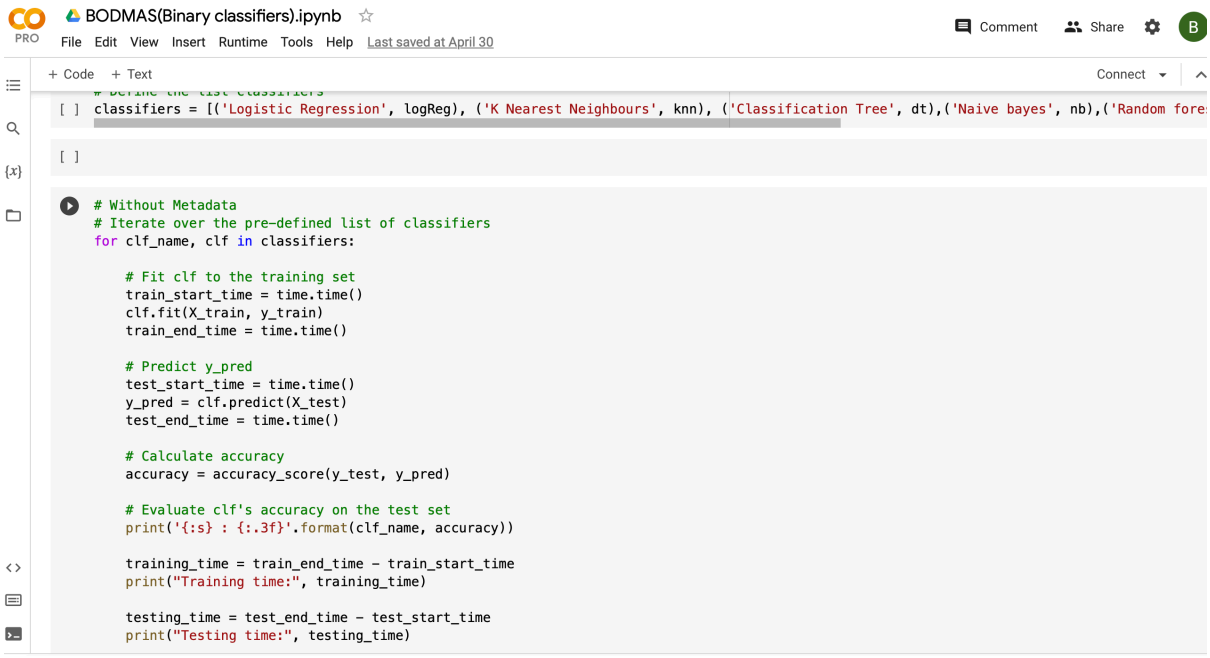

Figure 9. Evaluation

Step 7: - If we click Restart and run all in the Google Collab interface from the file menu in runtime, users can execute the code.

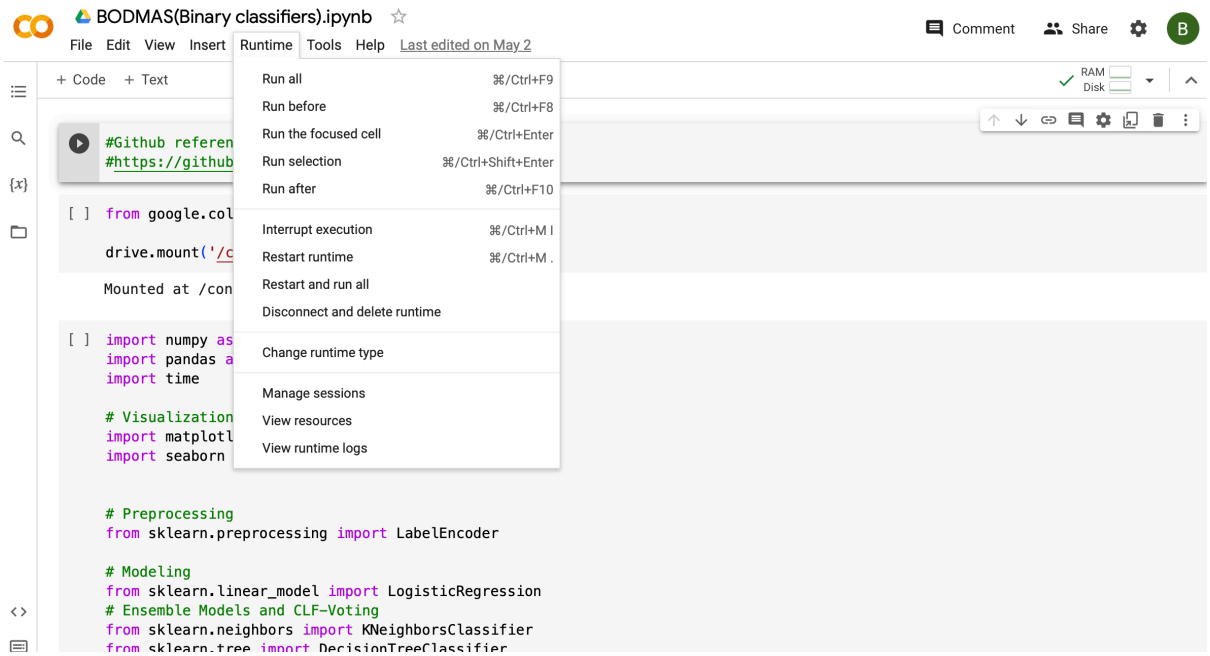

Figure 10. Code Execution in Google collab