National College of Ireland

Configuration Manual

MSc Research Practicum Part-2
MSc in Data Analytics

Kunal Sonawane
Student ID: x23400617

School of Computing
National College of Ireland

Supervisor: Prof. Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Kunal Ramesh Sonawane
Student ID:	X23400617
Programme:	MSc in Data Analytics **Year:** 2025
Module:	Research Practicum-Part 2
Lecturer:	Prof. Vikas Sahni
Submission Due Date:	28/01/2026
Project Title:	Explaining Real-Time Phishing and Explainable AI: A Transformer Based Approach

Word Count: 839 **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: Kunal Sonawane

Date: 27/01/2026

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

Kunal Sonawane
Student ID: x23400617

1 Introduction

This Configuration Manual describes the entire configuration and implementation process of the Phishing Detection System created for implementation of this project. It also offers a clear and practical process of setting up the environment, dependencies, dataset preparation, and executing the machine-based learning and transformer-based models that are used to classify phishing. The manual describes all the necessary tools, libraries, and model setups, including RoBERTa embeddings, and classifiers, including Logistic Regression, SVM, Random Forest, and XGBoost etc. It should assist to replicate the findings of the system, learn its elements, and implement the phishing detection framework effectively and consistently.

2 System Requirement

In order to ensure the effective model processing and to reduce the time taken, it is important that one has the required hardware and software resources.

2.1 Hardware Requirement

The implementation is performed on the MSI laptop: The configuration of the system is as follows:

1. Processor: 11th Gen Intel(R) Core (TM) i5-11260H @ 2.60GHz (2.61 GHz)
2. Ram: 16.0 GB
3. Hard Disk: 256 GB SSD, 1 TB HDD
4. Operating System: Windows 11 Home 25H2 64-bit operating system, x64-based processor

2.2 Software Requirement

Before proceeding towards model building phase, the following software, libraries, and tools were installed in the system.

Software/ Library	Version
Google Chrome	142.0.7444.176
Python	3.12.12

Numpy	2.0.2
Pandas	2.2.2
Requests	2.32.4
MatplotLib	3.10.0
Seaborn	0.13.2
Tldextract	5.3.0
Sckit-Learn	1.6.1
Scipy	1.16.3
Transformer	4.57.2
Tqdm	4.67.1
Torch	2.9.0

3 Implementation

In this section we have provided the complete steps to run the project in any system.

3.1 Setting up Google Colab

1. Launch the google browser and go to the Google Colab.
2. Use your google account to login.
3. Create a new notebook.

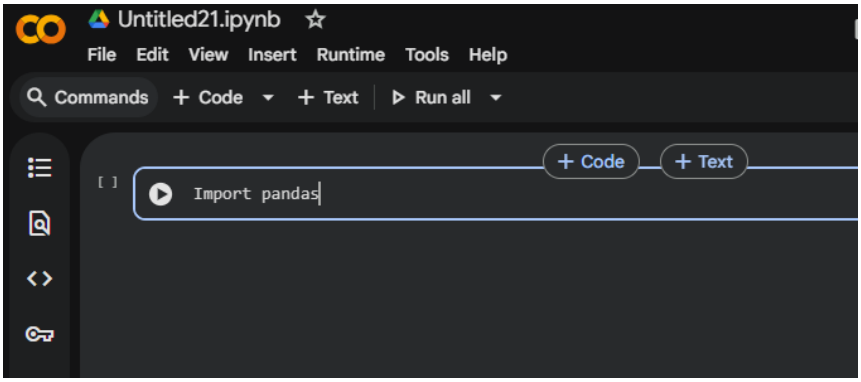

Fig 1. Google Colab Notebook Initial Outline

3.2 Enable the GPU

You can use the GPU which is provided by google colab itself for fast model training specially for deep leanrning models.

To select the GPU follow the path as follows:

Runtime → Hardware Accelator → Select the required GPU → Click on Save

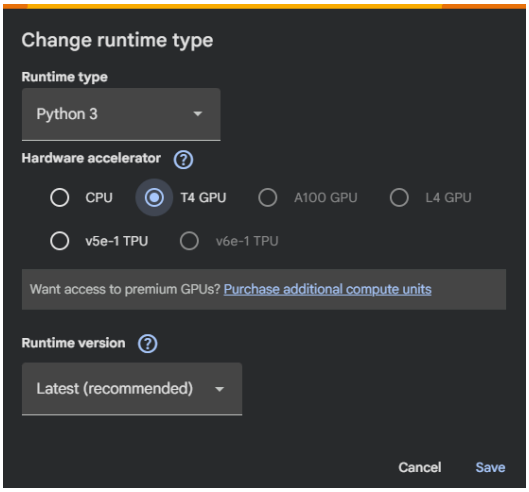

Fig 2: GPU Environment Selection

3.3 Install all the required Libraries

Install all the libraries along with the specified version which are mentioned in the above table.
At the start of the notebook import all the libraries using import statement

- Import pandas
- Import Numpy

3.4 Loading data

For this project we are using the URL Haus API which is freely available.

Fig 3: Data fetching form API

4 Execution

Run all the code cells from the Notebook

Execute each code cell sequentially from the notebook to replicate the data pre-processing, analysis, Model training, and Evaluation

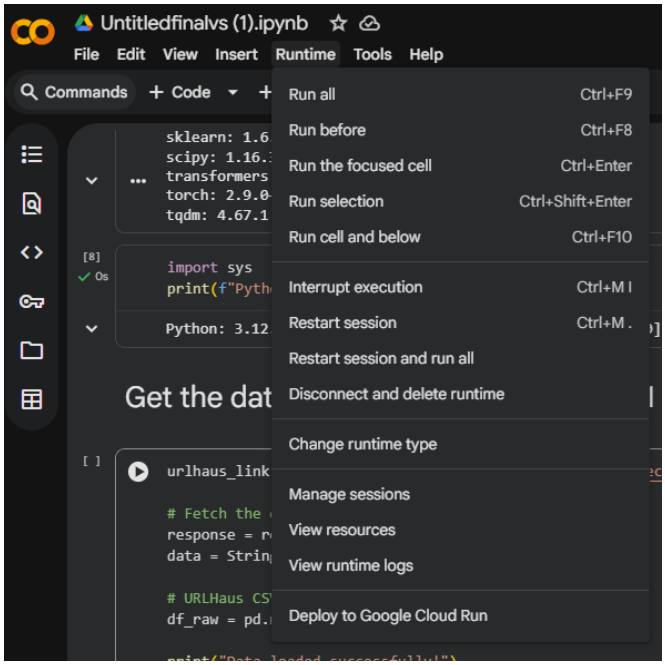

Fig 4: Option to run cell sequentially

5 Conclusion

The given Configuration manual assists in configuring the environment, checking the availability of GPUs, loading and processing data, extracting Bert and other machine learning and deep learning models, and successfully resolving such problems as class imbalance and the utilization of advanced models. This arrangement is very important because the project objectives are to identify phishing URL with a high level of performance.

References

Google Colaboratory. (2025). Welcome to Colaboratory. [Online]. Available: https://colab.research.google.com/

Pandas, "Python Data Analysis Library," *Pydata.org*, 2018. Available: https://pandas.pydata.org/

Matplotlib, "Matplotlib: Python Plotting — Matplotlib 3.1.1 Documentation," *Matplotlib.org*, May 30, 2024. Available: https://matplotlib.org/

seaborn, "seaborn: statistical data Visualization — seaborn 0.9.0 documentation," *Pydata.org*, 2012. Available: https://seaborn.pydata.org/

Scikit-learn, "scikit-learn: Machine Learning in Python," *Scikit-learn.org*, 2024. Available: https://scikit-learn.org/stable/

Numpy, "NumPy," *Numpy.org*, 2024. Available: https://numpy.org/