

Configuration Manual

MSc Research Project
MSc Cyber Security

Pavan Kumar Chinta
Student ID: 23379685

School of Computing
National College of Ireland

Supervisor: Jawad Salahuddin

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student

Name: Pavan Kumar Chinta

Student ID: 23379685

Programme: MSc Cyber Security

Year: Year 1

Module: MSc (Research) Practicum

Lecturer: Jawad Salahuddin

Submission

Due Date: 28/01/2025

Project Title: Semantic Text Embedding Framework for Phishing Email Detection: Leveraging Pre-trained Language Models for Content-Based Threat Classification

Word Count: 217 **Page Count:** 5

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: Pavan

Date: 28/01/2025

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

Pavan Kumar Chinta
23379685

1. Authorizing Colab to access files in Google Drive.

Step 1 : Launching the colab environment (navigating to google colab website)

Step 2: Sign in into the google account

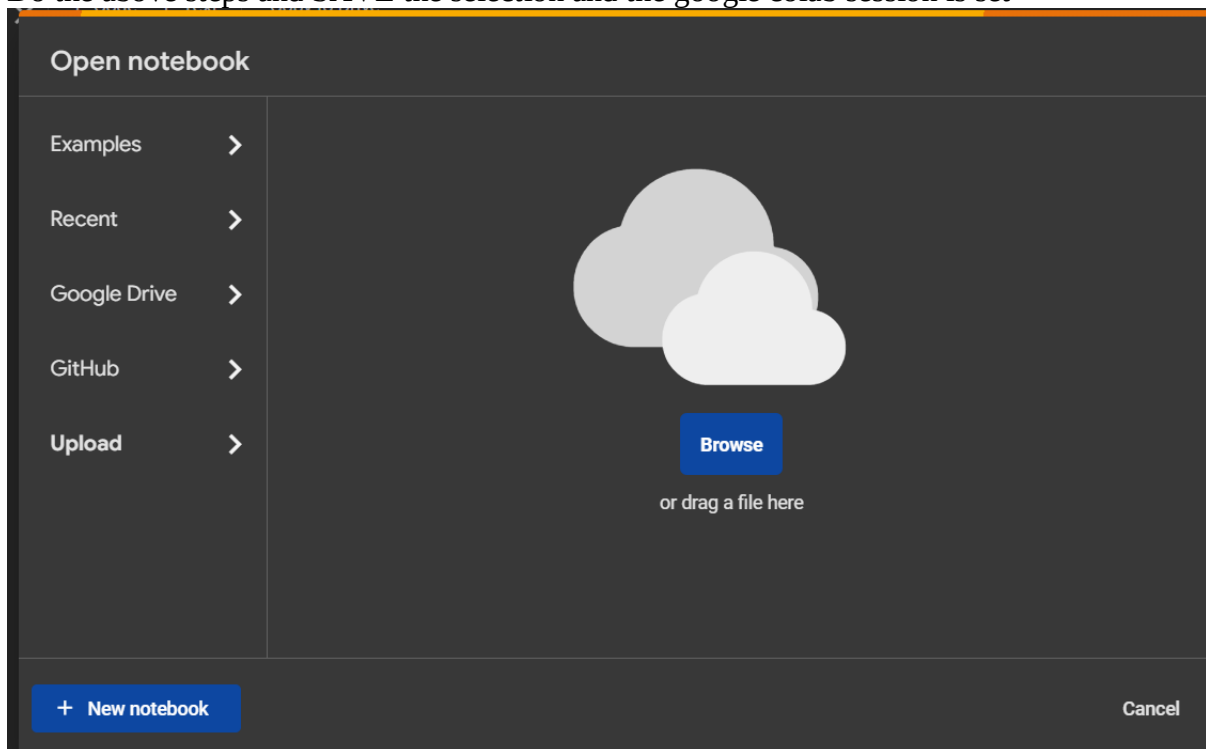
Step 3:uploading the ipynb file /notebook

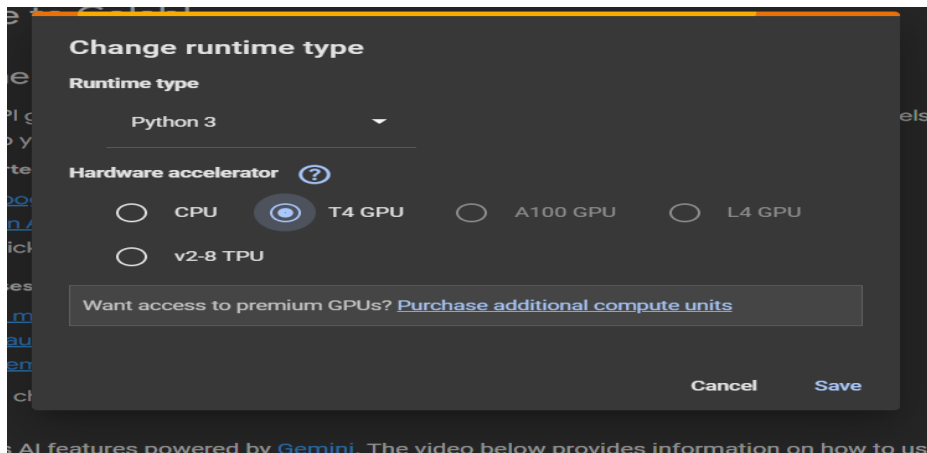
1. Upload the Embedding preparation file first
2. Now Same thing do and upload Modeling training file

Step 4:Enabling the GPU Acceleration (in this change the runtime)

Step 5 : change the hardware accelerator to T4 GPU

Do the above steps and SAVE the selection and the google colab session is set





2. Beginning the session

Step 1: open the google colab session

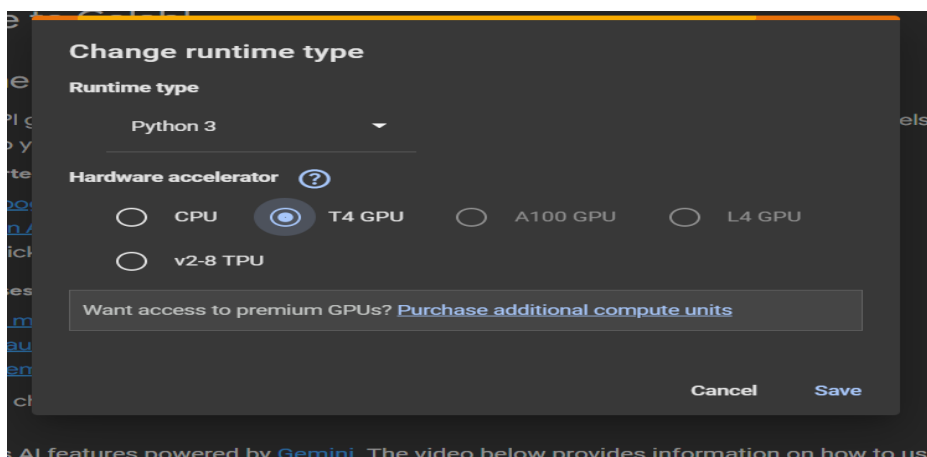
Step 2: Sign in to your Gmail

Step 3: start a new note book

Step 4 : Activate the GPU acceleration

Step 5: Create a folder on the website

1. Upload the Embedding preparation file first
2. Now Same thing do and upload Modeling training file
3. Step 6: You can see the results now



3. Reading the data into the notebook

Step 1: Load the data into the notebook (change the drive path)

```

1 import pandas as pd
2
3 # 1. Load the Parquet file
4 path = "/content/drive/MyDrive/Pavan_Cyber_Embeddings/all_email_datasets_clean_dedup.parquet"
5 df = pd.read_parquet(path)
6
7 # 2. Print just the column names
8 print("=== COLUMN NAMES ===")
9 for col in df.columns:
10     print(col)
11
12 # 3. Print a small sample of each column's data
13 print("\n=== SAMPLE VALUES PER COLUMN ===")
14 for col in df.columns:
15     print(f"\n--- {col} ---")
16     print(df[col].head(10).to_string(index=False))
17
18 # 4. (Optional) See full dataframe shape
19 print("\n=== DATAFRAME SHAPE ===")
20 print(df.shape)
21

```

Step 2: Mount to the google drive for saving

```

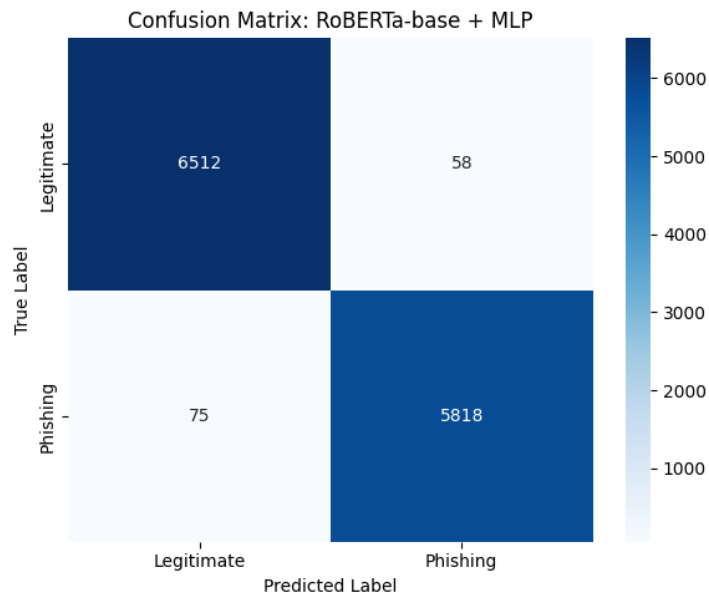
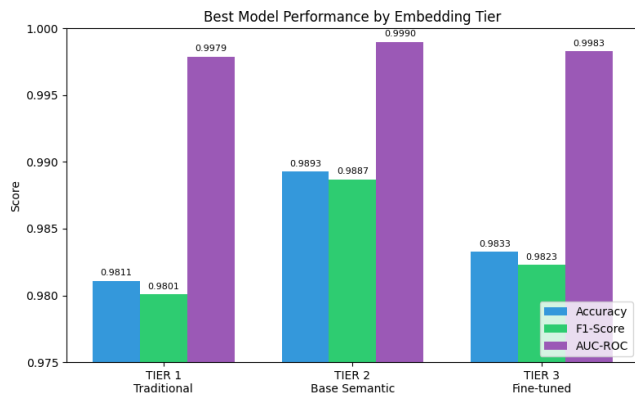
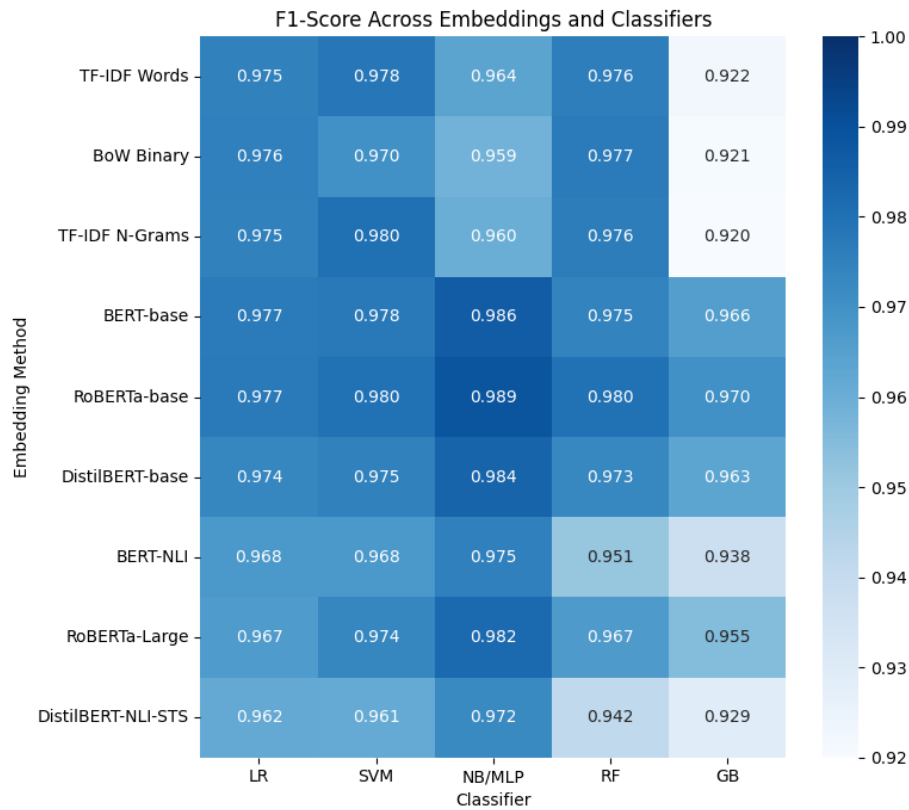
1 # Mount Google Drive
2 from google.colab import drive
3 drive.mount('/content/drive')
4
5 print("\n✓ Google Drive mounted successfully!")

```

... Mounted at /content/drive

✓ Google Drive mounted successfully!

4. Results



5 References

Google Colab. (n.d.). <https://colab.research.google.com/>

Schmitt, M. and Flechais, I. (2024) 'Digital deception: Generative artificial intelligence in social engineering and phishing', *Artificial Intelligence Review*, 57(12). Available at: <https://doi.org/10.1007/s10462-024-10973-2>.