National College of Ireland

Configuration Manual

MSc Research Project
MSc in Artificial Intelligence

Ayush
Student ID: x22186590

School of Computing
National College of Ireland

Supervisor: Dr. Rejwanul Haque

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ayush
Student ID:	22186590
Programme:	Msc in Aritificial Intelligence **Year:** 2023-2024
Module:	MSc Research Project
Lecturer:	Dr. Rejwanul Haque
Submission Due Date:	31/01/2024
Project Title:	Configuration Manual

Word Count: 595 Page Count: 3

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: 31st January 2024

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):	

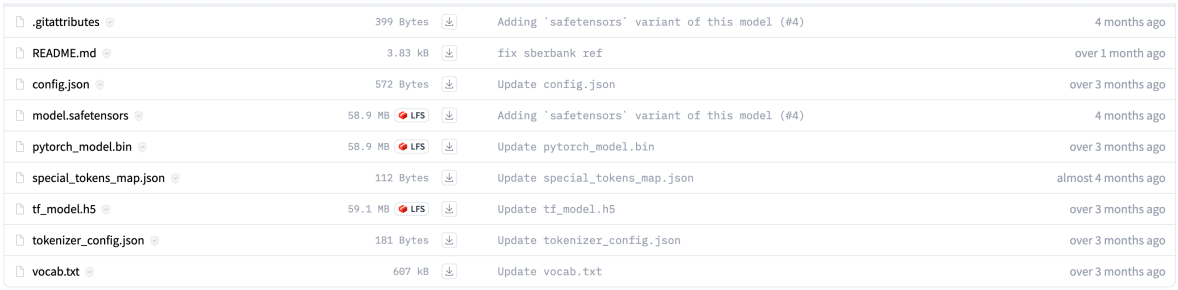
Configuration Manual

Ayush

22186590

1. To access the Primary Dataset and essential files

- For the project, please follow the provided link to the Google Drive: Click Here to access.



.gitattributes	399 Bytes	Adding `safetensors` variant of this model (#4)	4 months ago
README.md	3.83 kB	fix sberbank ref	over 1 month ago
config.json	572 Bytes	Update config.json	over 3 months ago
model.safetensors	58.9 MB LFS	Adding `safetensors` variant of this model (#4)	4 months ago
pytorch_model.bin	58.9 MB LFS	Update pytorch_model.bin	over 3 months ago
special_tokens_map.json	112 Bytes	Update special_tokens_map.json	almost 4 months ago
tf_model.h5	59.1 MB LFS	Update tf_model.h5	over 3 months ago
tokenizer_config.json	181 Bytes	Update tokenizer_config.json	over 3 months ago
vocab.txt	607 kB	Update vocab.txt	over 3 months ago

Figure 1: Drive location for relevant project files

- Browse the shared drive for the primary dataset, now named 'hindi_financial_text' and other pertinent files.
- Models and their predictions are stored in the folder.
- The Colab notebooks titled 'Part1.ipynb' and 'Part2.ipynb' contain the core project development code. Begin your exploration here.

2. Google Colab Pro+ IDE for development

2.1 Select the Appropriate Colab Membership Plan

For this development project, a Pro+ membership was used. However, standard membership plans might also support the execution of the project, albeit with potential variations in execution time.

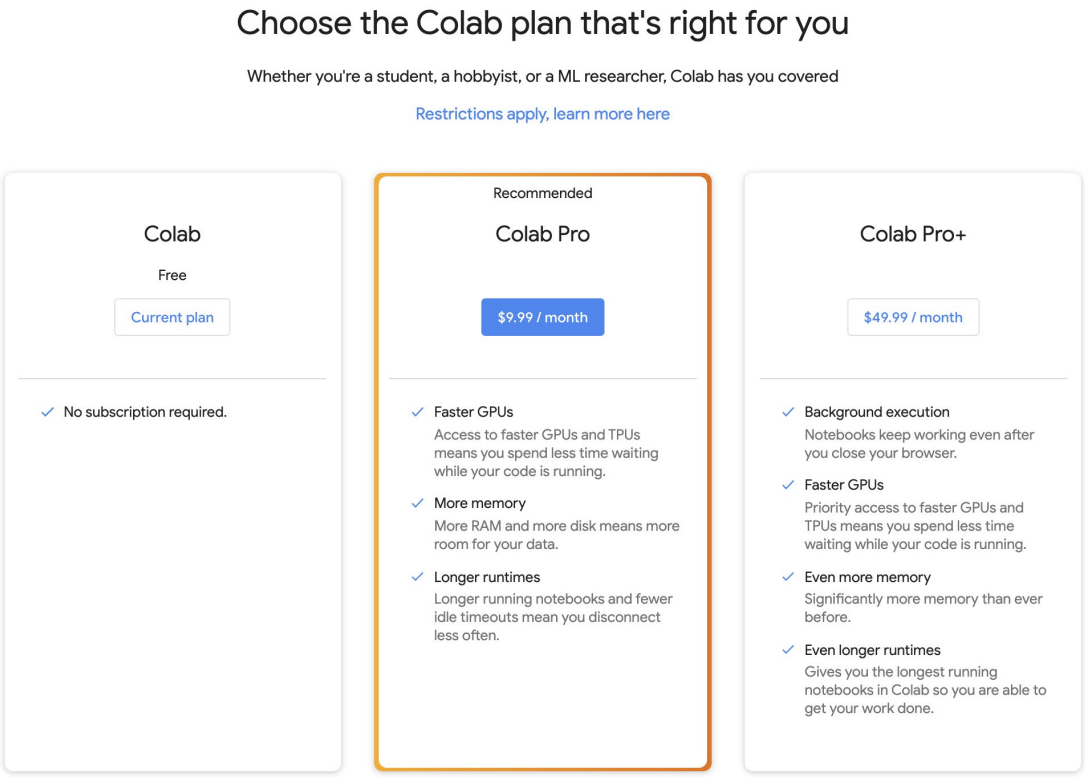


Figure 2: Google Colab Subscription description used for the project.

2.2 Setup Google Colab IDE for Code Testing

- To begin, download the project files from the provided drive link to your Google Drive. Once downloaded, make sure to update the file paths in the code to match their new locations on your drive.
- Open the Colab notebooks titled 'Part1 Final code Thesis demo Part1.ipynb' and 'Part2 Final code demo initial.ipynb' to set up the code. These notebooks can be run independently to obtain results. Each section is structured to facilitate an easy understanding of the code.
- Use the first code cell to install dependencies, mount your drive, and set up necessary components for testing.

```
model = t.get_classifier()
learner = ktrain.get_learner(model, train_data=trn, val_data=evalr, batch_size=6)
learner.fit_onecycle(1.2e-4, 5)
```

```
begin training using onecycle policy with max lr of 0.00012...
Train for 1135 steps, validate for 71 steps
Epoch 1/5
1135/1135 [==============================] - 183s 161ms/step - loss: 0.8110 - accuracy: 0.5574
Epoch 2/5
1135/1135 [==============================] - 158s 140ms/step - loss: 0.6197 - accuracy: 0.6855
Epoch 3/5
1135/1135 [==============================] - 158s 140ms/step - loss: 0.3896 - accuracy: 0.8490
Epoch 4/5
1135/1135 [==============================] - 159s 140ms/step - loss: 0.1814 - accuracy: 0.9423
Epoch 5/5
1135/1135 [==============================] - 159s 140ms/step - loss: 0.0817 - accuracy: 0.9766
<tensorflow.python.keras.callbacks.History at 0x7f682acde240>
```

3 Code Execution Steps

- The code is structured for use with the 'Run all' feature in Google Colab, but it's recommended to bypass certain sections to minimize long execution times. Throughout the code, there are comments and hints that guide you on which parts to skip during testing.
- To expedite the execution process, consider uncommenting the read commands. This approach is more efficient than executing the 'Benchmarking' step found in each section of the code.
- Everything is ready! You are now set to execute the code, section by section, for testing purposes.