

Configuration Manual

MSc Research Project
Data Analytics

Abhinav Thapa
Student ID: 20259409

School of Computing
National College of Ireland

Supervisor: Dr. Abdul Razzaq

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Abhinav Thapa
Student ID:	20259409
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Dr. Abdul Razzaq
Submission Due Date:	01/02/2023
Project Title:	Configuration Manual
Word Count:	324
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:	1st February 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
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.	<input type="checkbox"/>

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

Abhinav Thapa
20259409

1 Access the Primary Dataset and necessary files

1. The given link to Google drive is shared here for access to relevant project artifacts. Click me!

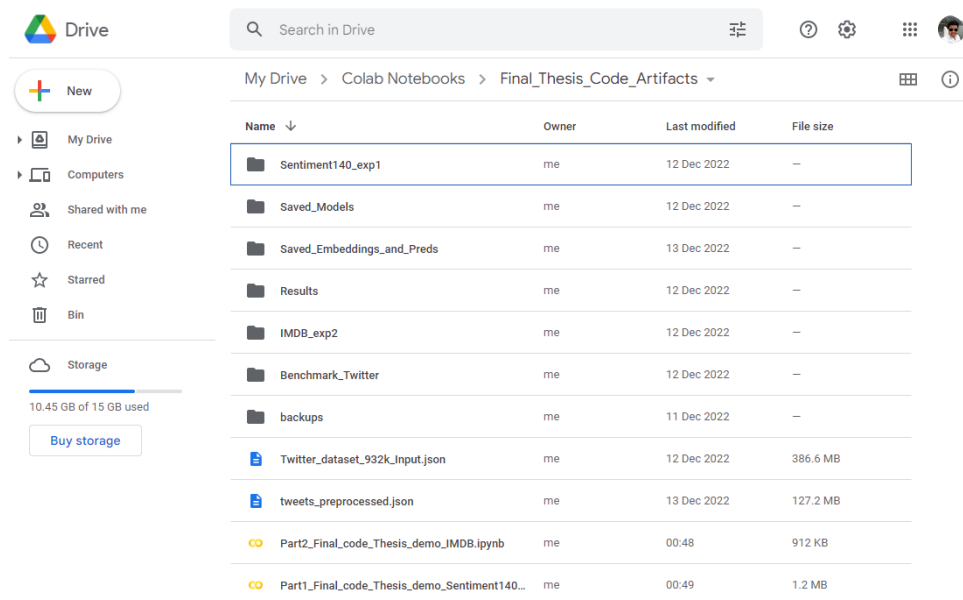


Figure 1: Google Drive location for relevant project files

2. Explore the shared drive location for Primary dataset ('Twitter_dataset_932k_Input.json') and relevant files.
3. Models and predictions were saved for reuse in Saved_Models folder.
4. Colab notebooks starting with names Part1 and Part2 are the main project development codes. You can start here.

2 Google Colab Pro+ IDE for development

2.1 Choose the right Colab Membership for you

For this development project Pro+ membership was subscribed however, regular plans may allow execution (although execution time may vary).

Choose the Colab plan that's right for you

Whether you're a student, a hobbyist or a ML researcher, Colab has you covered
Colab is always free of charge to use, but as your computing needs grow, there are paid options to meet them.

[Restrictions apply. Learn more here.](#)

The image shows three subscription plans for Google Colab:

- Pay As You Go:** No subscription required. Only pay for what you use. Offers faster GPUs and the ability to upgrade to more powerful premium GPUs. Pricing: €9.25 for 100 compute units, €42.25 for 500 compute units.
- Colab Pro (Current plan):** €9.25/month. Includes 100 compute units per month (expiring after 90 days), faster GPUs, and more memory.
- Colab Pro+:** €42.25/month. Includes 500 compute units per month (expiring after 90 days), faster GPUs, more memory, and background execution.

Figure 2: Google Colab Subscription for development and testing

2.2 Setup Google Colab IDE for Code Testing

1. Download the project files from the above mentioned drive link onto your google drive and update the drive paths in the code.
2. Open colab notebooks 'Part1_Final_code_Thesis_demo_Sentiment140.ipynb' and 'Part2_Final_code_Thesis_demo_IMDB.ipynb' for code setup. Both the code can be run independently for results. Sections have been designed properly to help with code understanding.

The screenshot shows a Google Colab notebook interface with the following content:

- Files:** A sidebar showing a file explorer with folders like 'drive', 'MyDrive', and 'Colab Notebooks'.
- Code:** A code cell titled '1. Installing Dependencies' containing:

```
[5] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt
time: 2.14 s (started: 2022-12-15 02:05:54 +00:00)

[1] !pip install tqdm
!pip install stopwords
!pip install nltk
!pip install tweet-preprocessor

!pip install ipython-autotime
%load_ext autotime
```
- Resources:** A panel on the right showing 'You are subscribed to Colab Pro+', 'Python 3 Google Compute Engine backend (GPU)', and RAM/Disk usage.

Figure 3: Google Colab Subscription for development and testing

3. Install Dependencies using 1st code cell to mount drive and necessary dependencies

for testing.

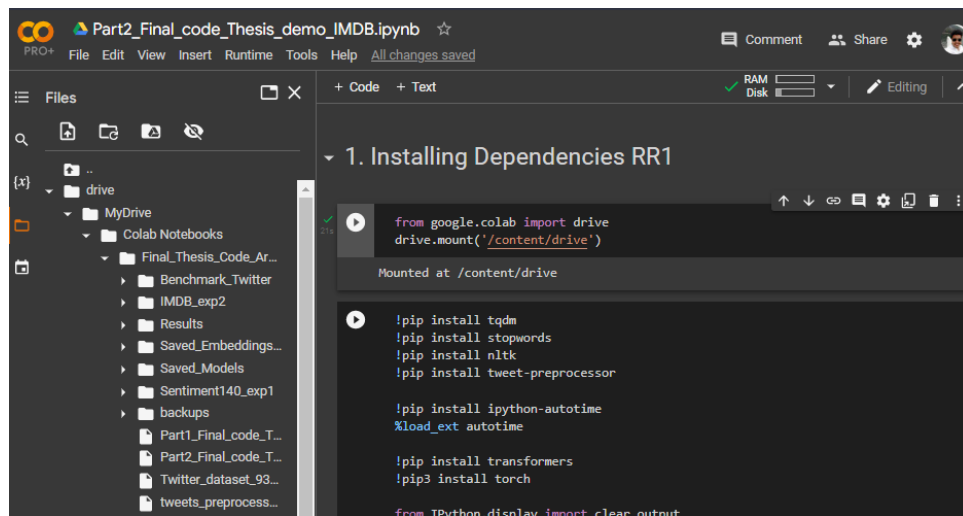


Figure 4: Google Colab Subscription for development and testing

3 Code Execution Steps

1. The whole code has been designed to run with 'Run all' feature in Google Colab, however it is advised to skip some steps to avoid long execution cycles. Comments and Hints have been marked throughout the code for testing.
2. The commented read commands can be uncommented to boost execution time instead of running section 3 Benchmarking step in each code.
3. All is set! You can now run the code section by section for Testing.