

Configuration Manual

MSc Research Project
MSc in Data Analytics

Ajay Gurralla
x23427931

School of Computing
National College of Ireland

Supervisor: Sallar Khan

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Ajay Gurralla
.....
x23427931
Student ID:

Programme: MSc in Data Analytics
.....
MSc Research Project
Year: 2025-2026
.....

Module:
Sallar Khan
Lecturer:
Submission Due Date: 28-01-2026
.....

Project Title: Fine tuning GPT-3.5 for hate speech detection with continuous severity
scoring
.....
170
Word Count: **Page Count:**8.....

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: Ajay Gurralla
.....
28-01-2026
Date:

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

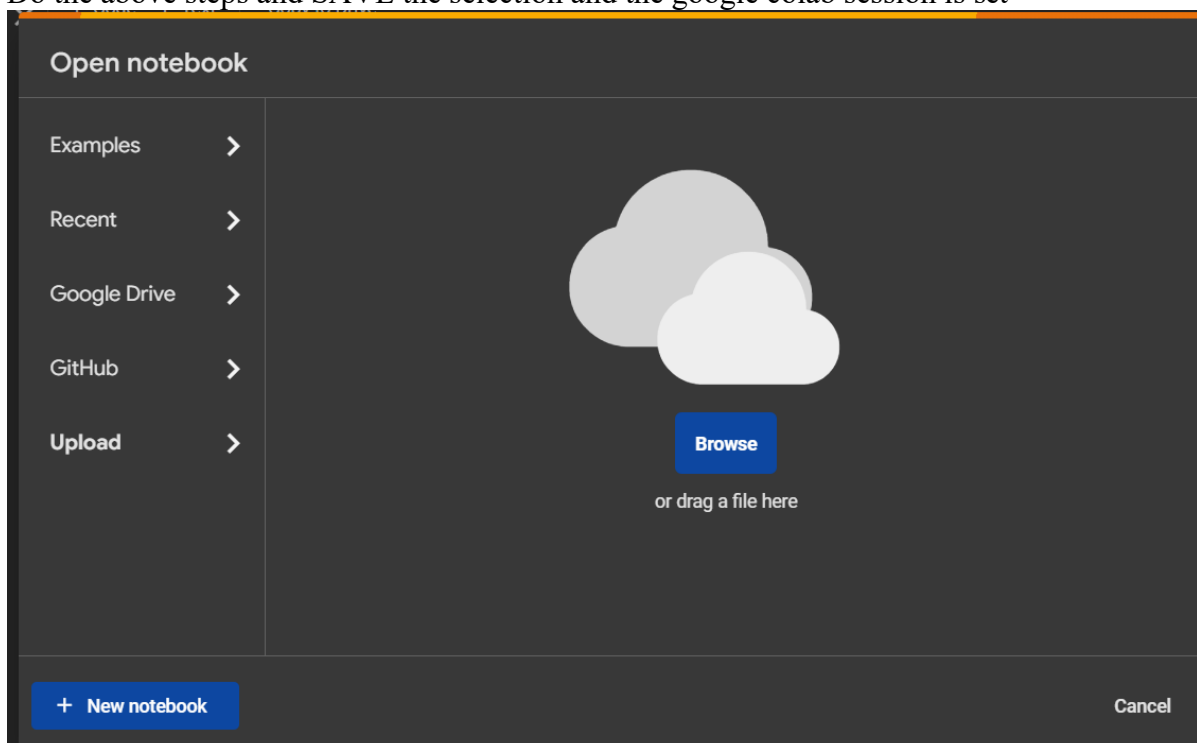
Ajay Gurralla

X23427931

1. Connecting a Colab session to Google Drive for file access

- 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
- 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. Starting the session

Step 1: open the google colab session

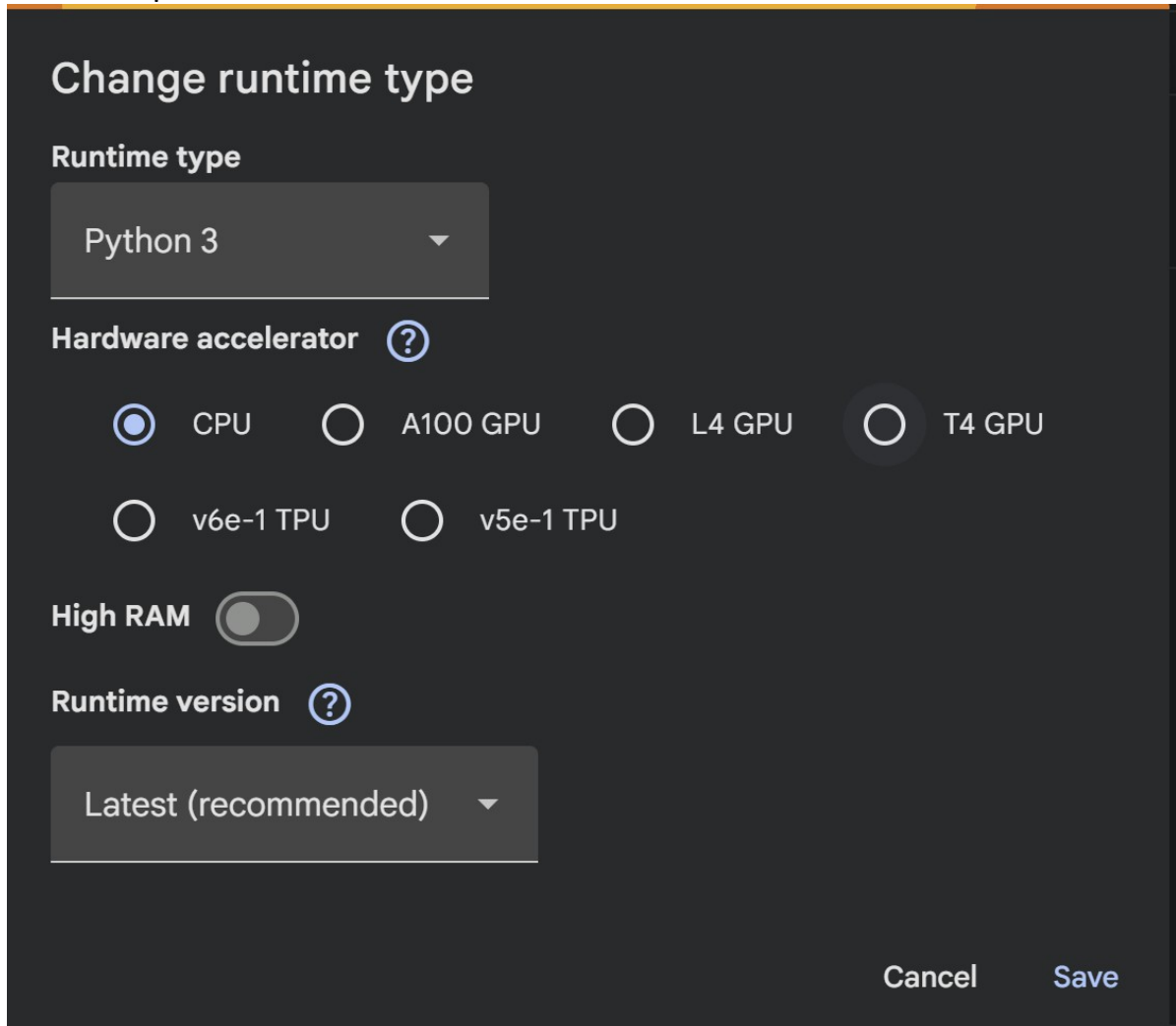
Step 2: Sign in to your Gmail

Step 3: start a new note book

Step 4: Activate the GPU acceleration

Step 5: Create a folder on the website

1. Upload the Fine tune model batch



The screenshot shows the 'Change runtime type' dialog box in Google Colab. It has a dark gray background with white text. At the top, the title 'Change runtime type' is in a larger font. Below it, the 'Runtime type' is set to 'Python 3' in a dropdown menu. The 'Hardware accelerator' section has a help icon (?) and five radio button options: 'CPU' (selected), 'A100 GPU', 'L4 GPU', 'T4 GPU', and 'v6e-1 TPU'. The 'High RAM' toggle switch is currently turned off. The 'Runtime version' section has a help icon (?) and a dropdown menu set to 'Latest (recommended)'. At the bottom right, there are 'Cancel' and 'Save' buttons.

Change runtime type

Runtime type

Python 3 ▼

Hardware accelerator (?)

☒ CPU ☐ A100 GPU ☐ L4 GPU ☐ T4 GPU

☐ v6e-1 TPU ☐ v5e-1 TPU

High RAM ☐

Runtime version (?)

Latest (recommended) ▼

Cancel Save

2. Now Upload the Fine tune 3.5 model

Change runtime type

Runtime type

Python 3 ▼

Hardware accelerator (?)

☒ CPU ☐ A100 GPU ☐ L4 GPU ☐ T4 GPU

☐ v6e-1 TPU ☐ v5e-1 TPU

High RAM ☐

Runtime version (?)

Latest (recommended) ▼

Cancel Save

3. Now upload last file Baseline Model comprehensive

Change runtime type

Runtime type

Python 3 ▼

Hardware accelerator ⓘ

☐ CPU ☐ A100 GPU ☐ L4 GPU ☒ T4 GPU

☐ v6e-1 TPU ☐ v5e-1 TPU

High RAM ☒

Runtime version ⓘ

Latest (recommended) ▼

Cancel Save

Step 6 : Now Run the code

3. Dataset loading

Step 1: load the the data (HARDCODED CONFIGURATION - MODIFY THESE VALUES AS NEEDED)

Step 2: change the API KEY

```

# =====
# API Configuration
API_KEY = "sk-proj-YHEvG0TL31raUJJCQgAYwQ8JCqt-VeshFwVCRZK-x4dL4Q_TivobTe8A0hu8N_yTgb5xoPnOZIT3B1bkFJWu0KlclR9fsJKNcz5czVTRelH_QJ1R6LaSkEg51S1zaC22176h"
FINETUNED_MODEL = "ft:gpt-3.5-turbo-1106:personal:hategpt:Cf850amh"

# File Configuration
TEST_FILE = "ajay_test.jsonl"
OUTPUT_CSV = "evaluation_results.csv"
CHECKPOINT_FILE = "evaluation_checkpoint.json"

# Batch Processing Configuration
BATCH_SIZE = 200          # Number of samples per batch
MAX_WORKERS = 50          # Number of concurrent API calls
MAX_RETRIES = 3           # Maximum retries for failed API calls
RETRY_DELAY = 1           # Initial retry delay in seconds
RATE_LIMIT_DELAY = 0.05  # Delay between API calls (seconds)

# Initialize OpenAI client
client = OpenAI(api_key=API_KEY)

print("Configuration loaded successfully!")
print(f"Model: {FINETUNED_MODEL}")
print(f"Test file: {TEST_FILE}")
print(f"Batch size: {BATCH_SIZE}")
print(f"Max workers: {MAX_WORKERS}")

```

Step 3 Uploading the test file

```

1 # Upload the test file if not already present
2 if not os.path.exists(TEST_FILE):
3     print(f"Please upload your test file: {TEST_FILE}")
4     uploaded = files.upload()
5     print(f"Uploaded files: {list(uploaded.keys())}")
6 else:
7     print(f"Test file found: {TEST_FILE}")

... Test file found: ajay_test.jsonl

```

4. Results

```

-----
CLASSIFICATION METRICS
-----
...
Overall Accuracy: 0.7068 (70.68%)

Per-Category Metrics:
-----
Category    Precision    Recall    F1-Score    Support
-----
HATE        0.6992      0.7119    0.7055      2926
ABSTAIN     0.5327      0.5464    0.5395      3340
NOT_HATE    0.8312      0.8086    0.8198      5116
-----
Macro Avg   0.6877      0.6890    0.6882      11382

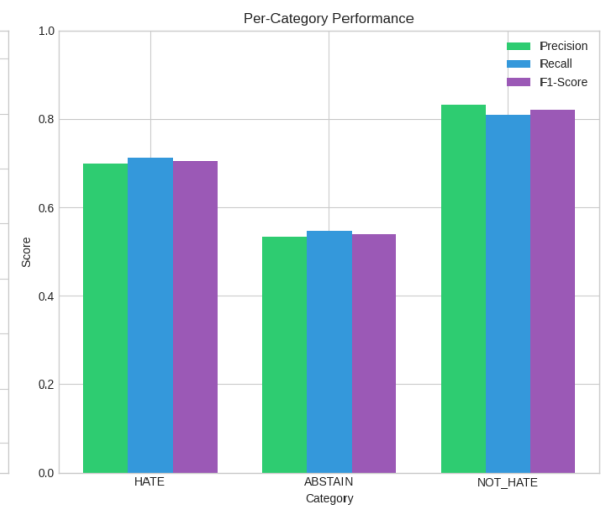
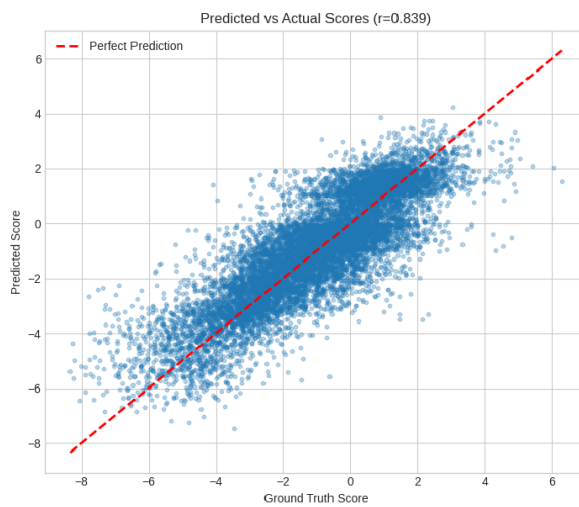
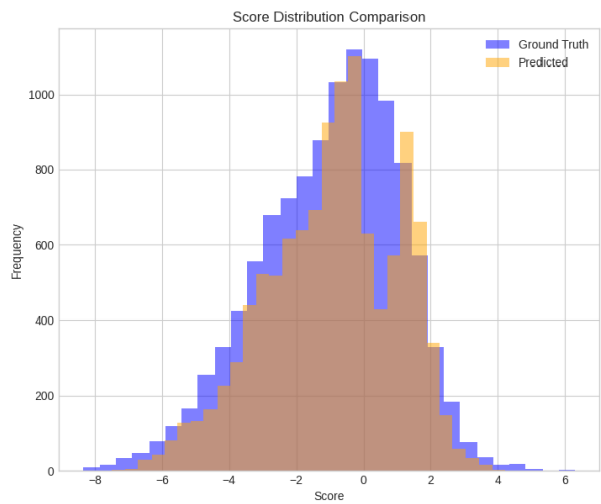
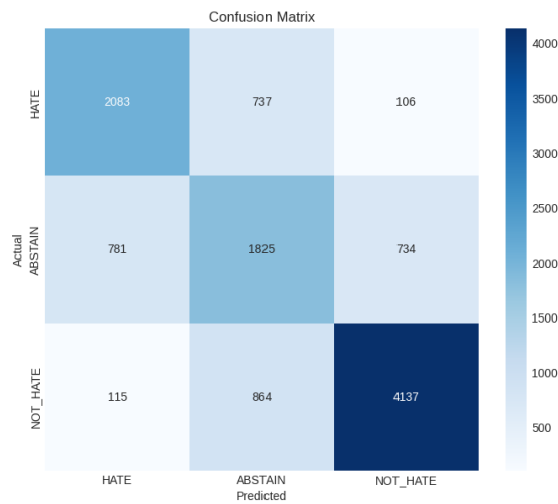
Confusion Matrix:
-----
Actual \ Pred  HATE    ABSTAIN    NOT_HATE
-----
HATE           2083        737        106
ABSTAIN         781       1825        734
NOT_HATE        115        864       4137

Detailed Classification Report:
-----
              precision    recall  f1-score   support

   HATE               0.70       0.71       0.71       2926
  ABSTAIN             0.53       0.55       0.54       3340
  NOT_HATE            0.83       0.81       0.82       5116

   accuracy               0.69
  macro avg              0.69
 weighted avg              0.71

```



...

Comparison with Fine-Tuned GPT-3.5:

Model	Accuracy	Macro F1	MAE	Correlation
BiLSTM	0.6052	0.5791	1.1540	0.6998
BERT-base	0.6813	0.6621	0.9448	0.8163
RoBERTa-base	0.6825	0.6605	0.9295	0.8233
GPT-3.5 (tuned)	0.7068	0.6882	0.8862	0.8390


```

=====
..  COMPREHENSIVE MODEL COMPARISON
=====

Classification Performance:
-----
Model            Accuracy    Macro F1    HATE F1    ABSTAIN F1    NOT_HATE F1
-----
BiLSTM           0.6052      0.5791      0.5851      0.4165      0.7356
BERT-base        0.6813      0.6621      0.6845      0.5069      0.7950
RoBERTa-base     0.6825      0.6605      0.6802      0.5008      0.8006

Regression Performance:
-----
Model            MAE          RMSE         R2 Score    Correlation
-----
BiLSTM           1.1540       1.4929       0.4885      0.6998
BERT-base        0.9448       1.2344       0.6503      0.8163
RoBERTa-base     0.9295       1.2189       0.6590      0.8233

Best Performers:
-----
Best Accuracy:   RoBERTa-base (0.6825)
Best F1-Score:   BERT-base (0.6621)
Best MAE:        RoBERTa-base (0.9295)
Best Correlation: RoBERTa-base (0.8233)

```

5. References

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

Linguistics: EMNLP 2023 (pp. 6116–6128). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-emnlp.407>

Sachdeva, P., Barreto, R., von Vacano, C., & Kennedy, C. J. (2022). The measuring hate speech corpus: Leveraging rasch measurement theory for data perspectivism. In *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022* (pp. 83–94). European Language Resources Association. <https://aclanthology.org/2022.nlperspectives-1.11>