2024 Fall INLP
HW2 Announcement

TA : Yu-Chen Kang
connie.cs12@nycu.edu.tw

Task introduction

- **Topic** : Multi-label Classification in NLP

- **Purpose** : Apply BERT-based models to predict the concern in tweets

- You can use any external package you want, e.g. Pytorch, Keras, scikit-learn

- Requirement :
1. Submit a report and your source code to E3
2. Upload your submission to Kaggle

Dataset

COVID-19 anti-vaccine tweets labelled with various specific anti-vaccine concerns in a multi-label setting.

- [HW2_dataset.zip](#)
  - train.json
  - val.json
  - test.json
  - sample_submission.csv

Data Introduction

Each object in the json file contains information about a tweet and its associated labels, including:

- 'ID': unique identifier of the tweet
- 'tweet': the content of the tweet
- 'labels': concerns that have been identified within the tweet with the specific
  terms causing the concerns and their positions. Each tweet can have multiple concerns!

```
{'ID': '1349576125454852097',
  'tweet': '@4mishee @ScottMorrisonMP can keep his vaccine 💉 I'll take covid with its " non existent " symptoms over some b
ig pharma no liability or indemnity lab 🧪 cooked vaccine thank you.',
  'labels': {'pharma': [{'index': 0,
     'start': 19,
     'end': 25,
     'terms': 'big pharma no liability or indemnity'}],
   'unnecessary': [{'index': 0,
     'start': 13,
     'end': 17,
     'terms': 'non existent " symptoms'}]}},
```

Kaggle Submission Format

Your model is expected to classify the concern of the tweets in 'test.json' file, and then upload your model's predictions to Kaggle. the submission format should be :

- A 1977*13 .csv file, first row for column name and the last 1976 rows for your result (**'1' indicates that the tweet has the concern, '0' signifies that it does not**)
- First row must match the one shown in the sample_submission.csv, make sure the order is correct !

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	index	ineffective	unnecessary	pharma	rushed	side-effect	mandatory	country	ingredients	political	none	conspiracy	religious	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	0	0	0	
4	2	0	0	0	0	0	0	0	0	0	0	0	0	

Kaggle Submission (80%)

- [Kaggle link](#)
- Display team name : <student ID>
- Public leaderboard is calculated with approximately 50% of the test data, private leaderboard is calculated with the other 50%, so the final standings may be different. You can only view your private leaderboard score after the competition has ended.
- There is a simple baseline and a strong baseline, beat them to get higher score.

Public   Private

This leaderboard is calculated with approximately 50% of the test data. The final results will be based on the other 50%, so the final standings may be different.

#	Team	Members	Score	Entries	Last
🚶	Strong Baseline		0.54106		
🚶	Simple Baseline		0.35804		

Kaggle Submission (80%)

- Kaggle link
- Display team name : <student ID>
- Public leaderboard is calculated with approximately 50% of the test data, private leaderboard is calculated with the other 50%, so the final standings may be different. You can only view your private leaderboard score after the competition has ended.
- There is a simple baseline and a strong baseline, beat them to get higher score.
- The scoring metric is macro-F1, higher score means better performance.
- You can submit at most 5 times each day and choose 2 of the submissions to be considered for the private leaderboard, or will otherwise default to the best public scoring submissions.

Change your team name

Please remember to change the team name to <student ID>, or there will be a deduction of 5 points for HW 2.

Overview   Data   Discussion   Leaderboard   Rules   **Team**   Submissions   Settings

Your Team

Everyone that competes in a Competiton does so as a team - even if you're competing by yourself. Learn more.

General

TEAM NAME

yourID

This name will appear on your team's leaderboard position.

Report Submission (20%)

Answer the following 4 questions :

1. How did you do to preprocess your data from dataset ?

2. How were the model's hyperparameters chosen ? Did you perform hyperparameter tuning ? If so, what were the specific steps and results?

3. In your experimental results, which categories of concerns were the most difficult to predict? And which categories were these concerns most often misclassified as?

4. Building on the previous question, what methods have you tried in your experiment to improve the models' ability to more accurately identify the concerns expressed by users? Please describe both the successful and unsuccessful cases.

Please answer the questions in detail to get the full point of each question.

Grading policy

- Kaggle (80%)
    - 30% based on the public leaderboard score and 70% based on the private leaderboard score
    - Basic score :
            Over strong baseline : 60
            Over simple bassline : 45
            Under simple baseline : 30
    - Ranking score:
            20-(20/N)*(ranking-1), N=numbers of people in the interval

- Report (20%)
    - 5% for each quesiton

You will receive 0 points if you do not submit the source code.

E3 Submission

Submit your source code and report to E3 before 11/20(Wed) 23:59, no late submissions will be accepted !

Format :

- HW2_<student ID>.zip
  - source code : HW2_<student ID>.py or HW2_<student ID>.ipynb
  - report : HW2_<student ID>.pdf

If you have any question about HW2, please feel free to contact with TA: Yu-Chen Kang

through email connie.cs12@nycu.edu.tw

Have Fun !