GODS 4.0 Hackathon

Presented by:

- AYARI Fahmi

Contents:

Fun Fact

3rd consecutive 2rd place in GoDS Hackathons

Objective

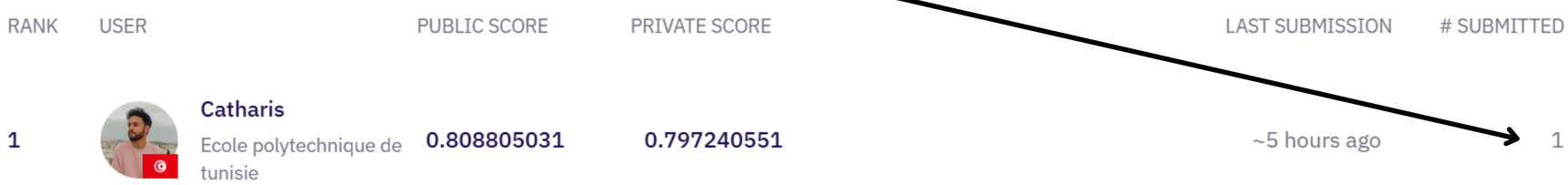
Classifying text entries from online discussion sinto mental-health issues categories

Dataset

- Train:
  - 20k labeled texts
- Test:
  - 2.5k texts
- Labels :
  - 5 Labels that are related to mental health

Approaching the competition

- 20K texts in training data vs 720 texts in public leaderboard => heavily rely on CV



RANK	USER	PUBLIC SCORE	PRIVATE SCORE	LAST SUBMISSION	# SUBMITTED
1	Catharis Ecole polytechnique de tunisie	0.808805031	0.797240551	~5 hours ago	1

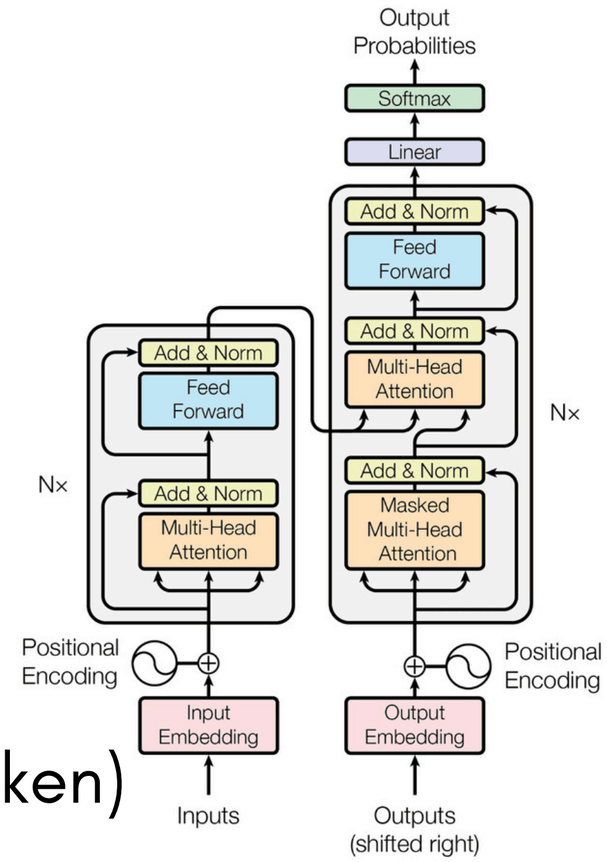
- Use StratifiedKFold to validate models
- Build a robust but simple ensemble strategy
- Optimize final weights based on out of folds (20k rows)

Modeling

1s Approach:

- Classical NLP
- Concatenate title and content
- Apply some preprocessing to the input text
- Fit TFIDF Vectorizer on all data
- Fit Logistic Regression & Predict

Modeling

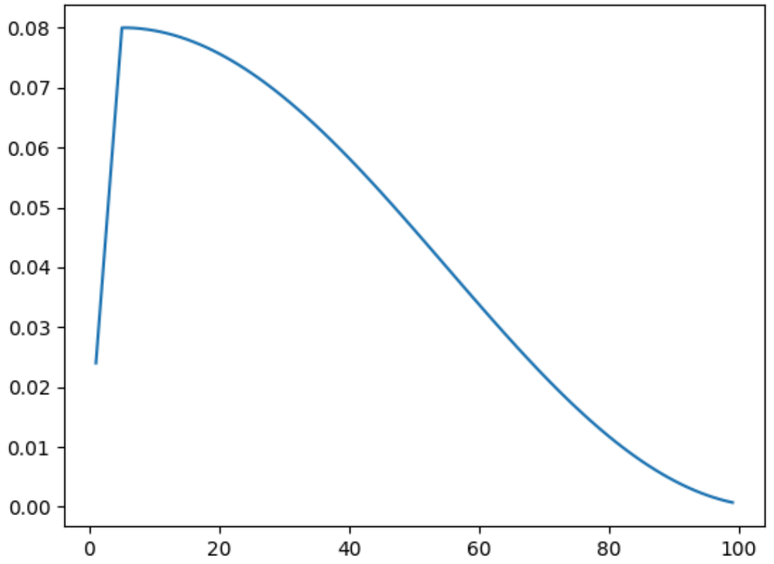


2nd Approach:

- Use Transformers (DebertaV3 Base & Large)
- Concatenate title and content (seperated by SEP Token)
- Added special tokens like \n \n
- Mean Pooling
- Train using various optimizations strategies like mixed precision, gradient accumulation ...

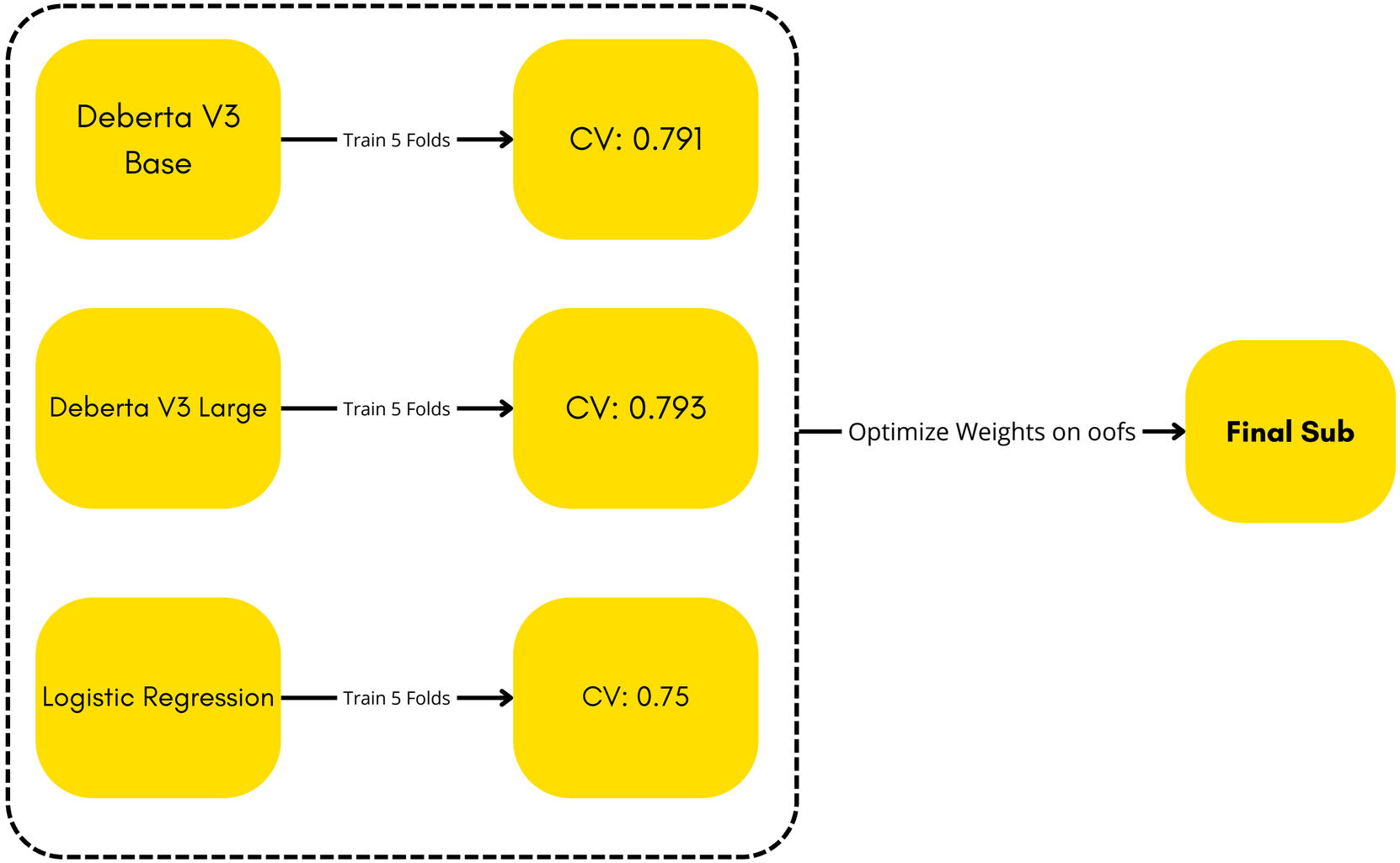
Shared Hyperparameters

- lr: 1e-5
- loss: BCE
- Folds: 5
- Epochs: 2
- Optimizer:
  - Adam
  - Weight decay 1e-6
- Scheduler: Cosine with warmup
- Mixed Precision
- Gradient Accumulation



Cosine with warmup scheduling

Final Submission



Thank You