



Boom AI Assessment

Verlon Salaysay

#1- Fine-tune a HuggingFace model for text classification by following this documentation. Below are the specifications:

- Make use of Google Colab's free GPU to train a HuggingFace model
- Follow the documentation from start to finish
- Be able to answer questions about each piece of code during the interview
- Demonstrate fine-tuning using the sample dataset provided.
- Bonus points: Use another text classification dataset to perform fine-tuning

Solution Reference:

https://github.com/verlonsalaysay/boomai-textclassification/blob/main/VPS_Text_Classification_on_GLUE.ipynb

Summary of Results:

```
✓ [27] trainer.train()
37m
/usr/local/lib/python3.9/dist-packages/transformers/optimization.py:391:
  warnings.warn(
You're using a DistilBertTokenizerFast tokenizer. Please note that with a
[2219/2675 3:05:12 < 38:05, 0.20 it/s, Epoch 4/5]

Epoch  Training Loss  Validation Loss  Matthews Correlation
1         0.525000         0.553606         0.404296
2         0.349800         0.503014         0.489065
3         0.233600         0.550874         0.553008
4         0.171800         0.782780         0.538470
[2675/2675 3:43:00, Epoch 5/5]

Epoch  Training Loss  Validation Loss  Matthews Correlation
1         0.525000         0.553606         0.404296
2         0.349800         0.503014         0.489065
3         0.233600         0.550874         0.553008
4         0.171800         0.782780         0.538470
5         0.133700         0.824053         0.530857

TrainOutput(global_step=2675, training_loss=0.27141547158499746, metrics=
We can check with the evaluate method that our Trainer did reload the best model
```

```
▶ for n, v in best_run.hyperparameters.items():
    setattr(trainer.args, n, v)

trainer.train()

Some weights of the model checkpoint at distilbert-base-uncased were not
- This IS expected if you are initializing DistilBertForSequenceClassification
- This IS NOT expected if you are initializing DistilBertForSequenceClassification
Some weights of DistilBertForSequenceClassification were not initialized
You should probably TRAIN this model on a down-stream task to be able to
/usr/local/lib/python3.9/dist-packages/transformers/optimization.py:391:
  warnings.warn(
[428/428 15:53, Epoch 4/4]

Epoch  Training Loss  Validation Loss  Matthews Correlation
1         No log         0.639090         0.141975
2         No log         0.659621         0.320045
3         No log         1.319887         0.336993
4         No log         1.386572         0.323012

TrainOutput(global_step=428, training_loss=0.28272013797938267, metrics=
```