Programming assignment 3 report

資管二 李訓佑 b11705022

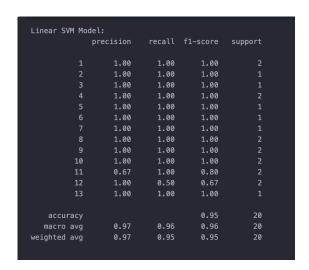
Env:

Run on my own computer, the original data (the data from hw1) is under ./PA1-data/python = ">=3.9.0,<3.11" tensorflow-macos = " 2 .15.0" keras-bert = " 0 .89.0" scikit-learn = " 1 .3.2"

程式邏輯:

- 1. 先將所有的資料放入x_train, y_train(都沒切)
- 2. 我使用的BERT model是'./uncased_L-8_H-512_A-8' = BERT-Base
- 3. 資料放入BERT model,取出他的embeddings,根據講義只取第一個也就是cls
- 4. 將embeddings轉換切成x_train,x_test,y_train,y_test
- 5. 放入sklearn的SVM模型並且訓練
- 6. 做predict, 輸出prediction到prediction2.csv

SVM model output:



我們可以看見他f1 score是很高的,跟其他bert model 比起來。

Kaggle score:

最上面的是使用BERT-base, 第二坨使用medium,最後是BERT-mini可以看到他們的分數差距很大,差了整整0.04。

0	prediction2.csv Complete - 9h ago	0.89861	V
0	prediction2.csv Complete -21h ago	0.83888	
0	prediction2.csv Complete - 21h ago	0.83888	
0	prediction2.csv Complete - 21h ago	0.83888	
\otimes	prediction2.csv Complete - Id ago	0.85416	