

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:

Linear SVM Model:					
	precision	recall	f1-score	support	
1	1.00	1.00	1.00	2	
2	1.00	1.00	1.00	1	
3	1.00	1.00	1.00	1	
4	1.00	1.00	1.00	2	
5	1.00	1.00	1.00	1	
6	1.00	1.00	1.00	1	
7	1.00	1.00	1.00	1	
8	1.00	1.00	1.00	2	
9	1.00	1.00	1.00	2	
10	1.00	1.00	1.00	2	
11	0.67	1.00	0.80	2	
12	1.00	0.50	0.67	2	
13	1.00	1.00	1.00	1	
accuracy			0.95	20	
macro avg	0.97	0.96	0.96	20	
weighted avg	0.97	0.95	0.95	20	

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。

 prediction2.csv Complete · 9h ago	0.89861	<input checked="" type="checkbox"/>
 prediction2.csv Complete · 21h ago	0.83888	<input type="checkbox"/>
 prediction2.csv Complete · 21h ago	0.83888	<input type="checkbox"/>
 prediction2.csv Complete · 21h ago	0.83888	<input type="checkbox"/>
 prediction2.csv Complete · 1d ago	0.85416	<input type="checkbox"/>