

Report

Member

- 資工三 110590018 劉承翰 - 程式撰寫、文件撰寫
- 資工三 110590011 劉承軒 - 文件撰寫

Env

```
Python 3.10.12
```

Package

```
1. sklearn
2. re
3. pandas
```

data

From Sentiment Analysis Dataset.csv (第一次功課的資料集)

```
X_train, X_test, y_train, y_test = train_test_split(X_train, y_train,
test_size=0.2, random_state=42)
```

Result

```
(NLP_env) → HW1 git:(main) X python3 main.py
ItemID Sentiment SentimentSource SentimentText
0 1 0 Sentiment140 is so sad for my APL frie...
1 2 0 Sentiment140 I missed the New Moon trail...
2 3 1 Sentiment140 omg its already 7:30 :O
3 4 0 Sentiment140 .. Omgaga. Im sooo im gunna CRy. I'...
4 5 0 Sentiment140 i think mi bf is cheating on me!!! ...
are you ok?
Positive
Accuracy: 0.7953522549830072
Precision: 0.7886707816641643
Recall:0.8076025130810551
F1: 0.7980243826195687
```

Issue

再讀入csv時報了一個錯誤，是有關columns數不對的問題，原本只有4欄的值，但資料及突然多出一個5欄的資料-.-。加入on_bad_lines='skip'後略過有問題的data，成功解決。

```
dataset = pd.read_csv("Sentiment Analysis Dataset.csv", on_bad_lines='skip')
```