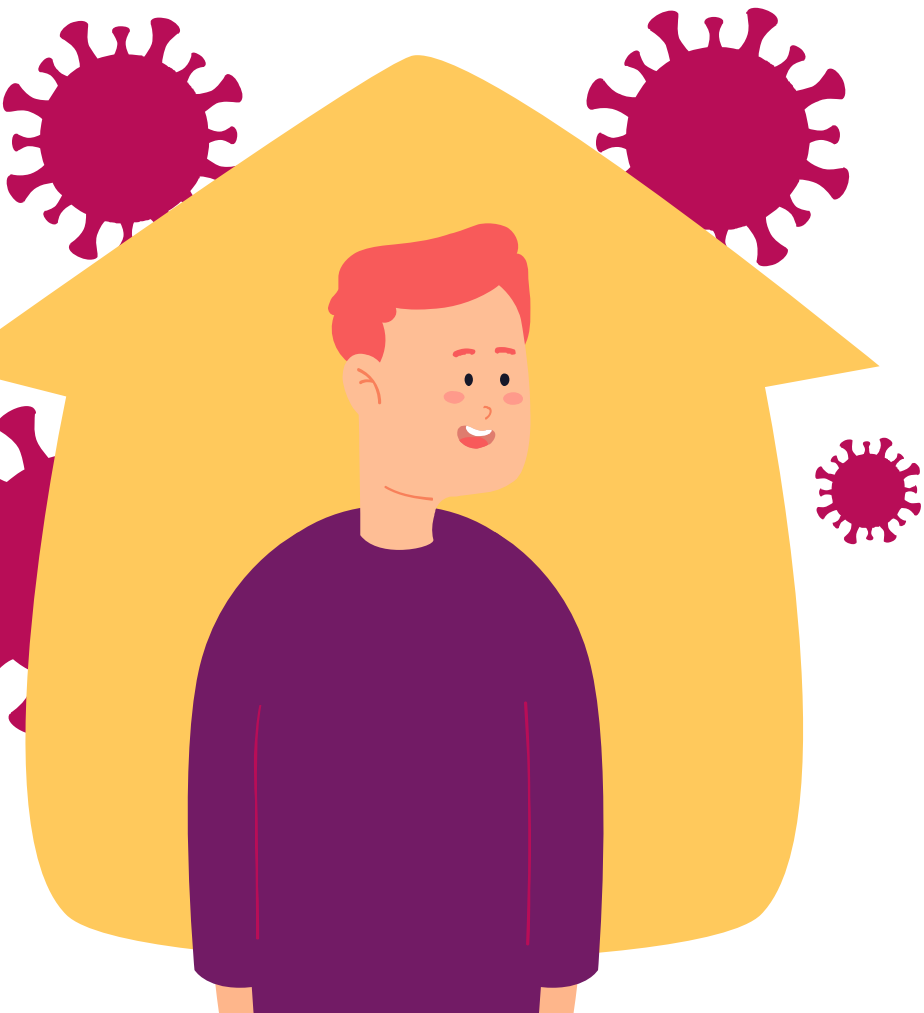
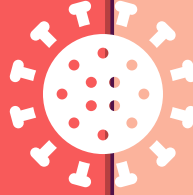


Covid-19

疫情下民眾情緒變化與政治人物
公眾形象、政策推動之關聯

人工智慧應用組 _ LCJ_team



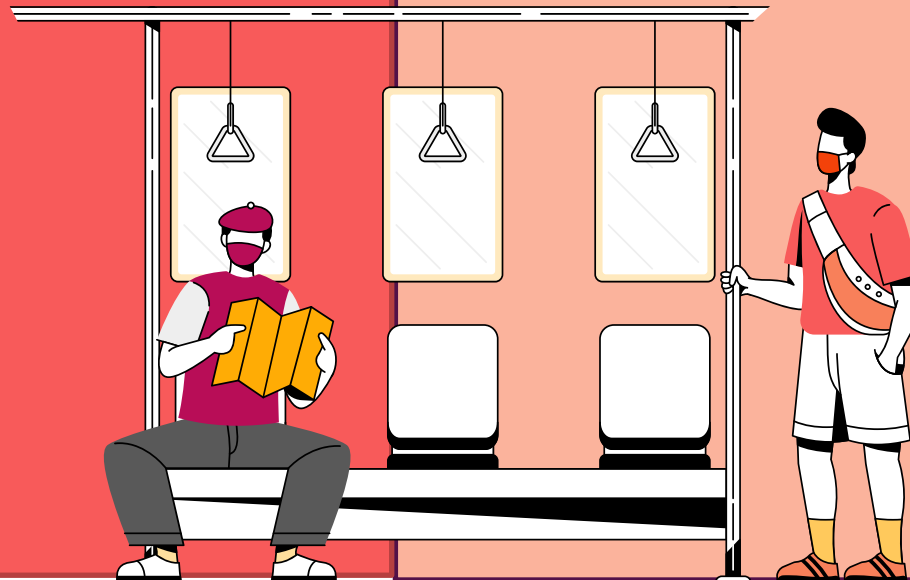
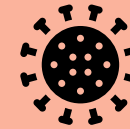


中國時報

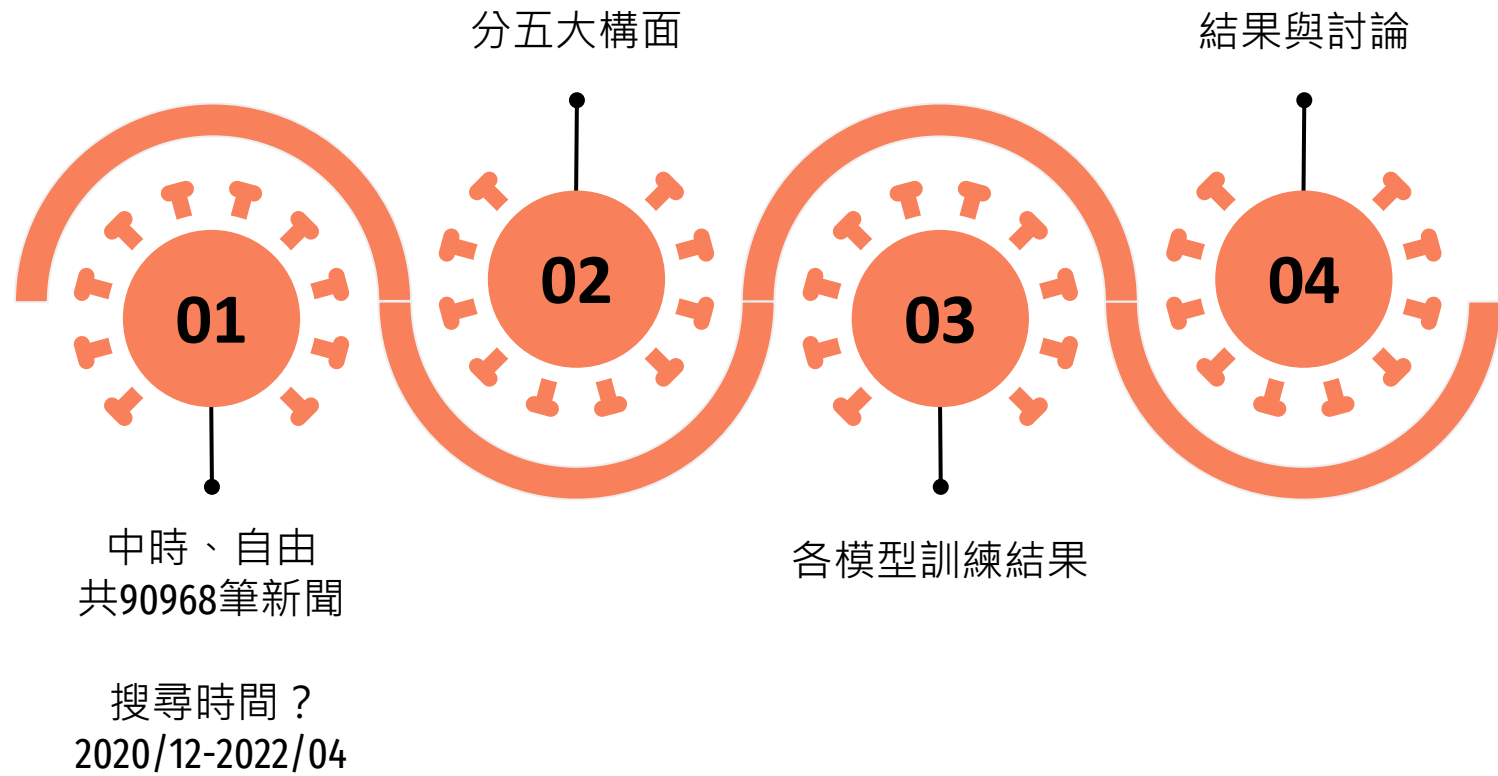
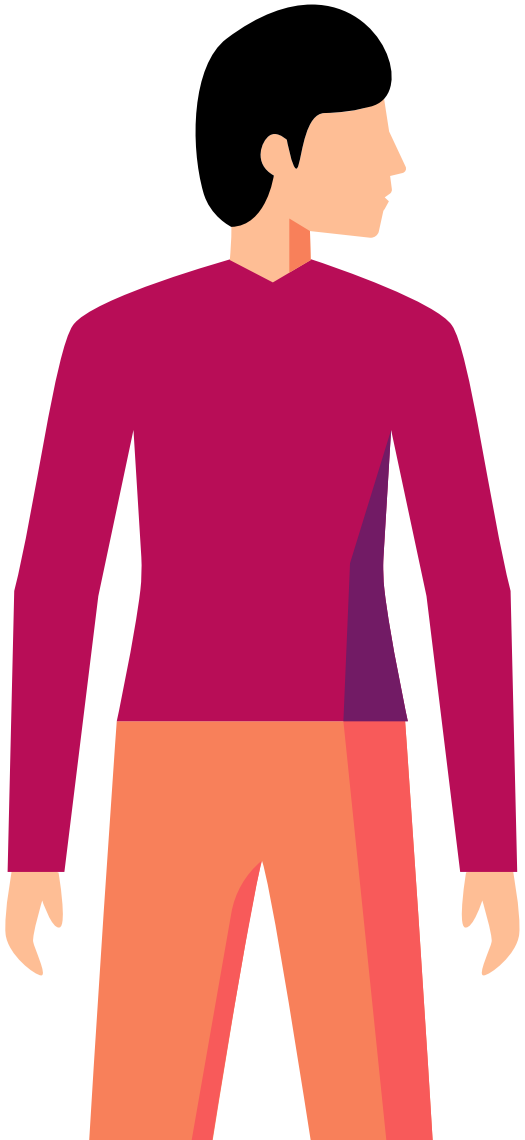
搜尋時間：2020/12-至今

自由時報

搜尋時間：2020/12-至今



CONCEPTUAL FRAMEWORK



keywords

- **Politician**

- 陳時中(阿中)、蔣萬安、黃珊珊

- **Adaptive**[自主檢測]

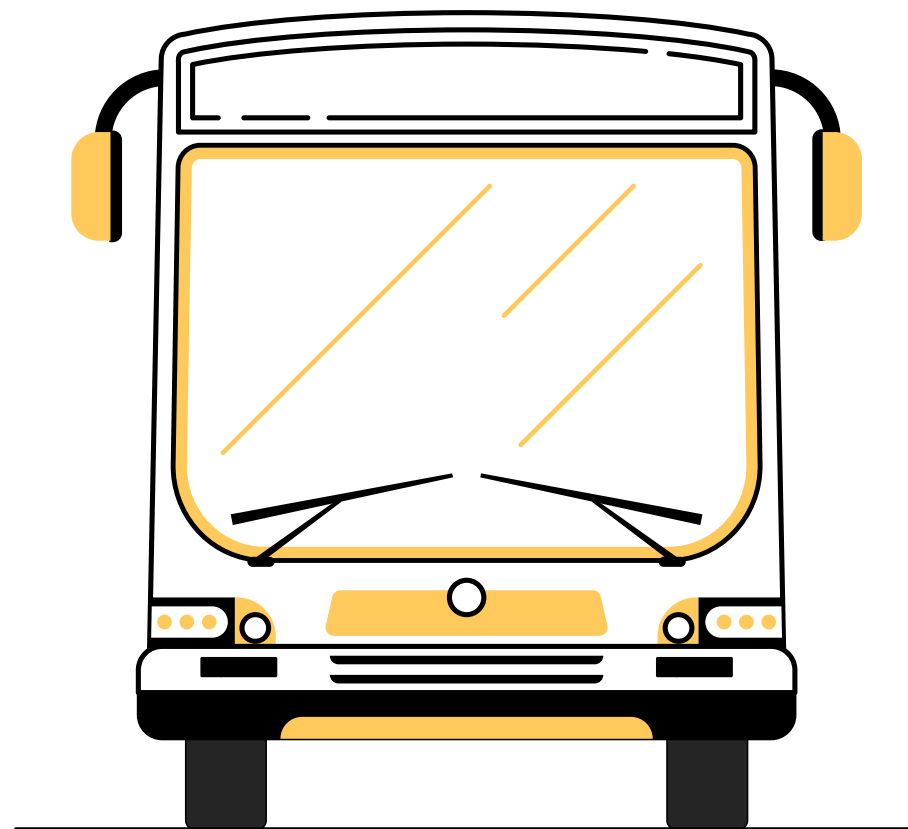
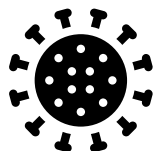
- 居家快篩、快篩試劑

- **Contact tracing**[感染追蹤]

- 居家照顧、居家隔離（居隔）、居家檢疫（居檢）、社區感染、自我健康監測、自主健康管理

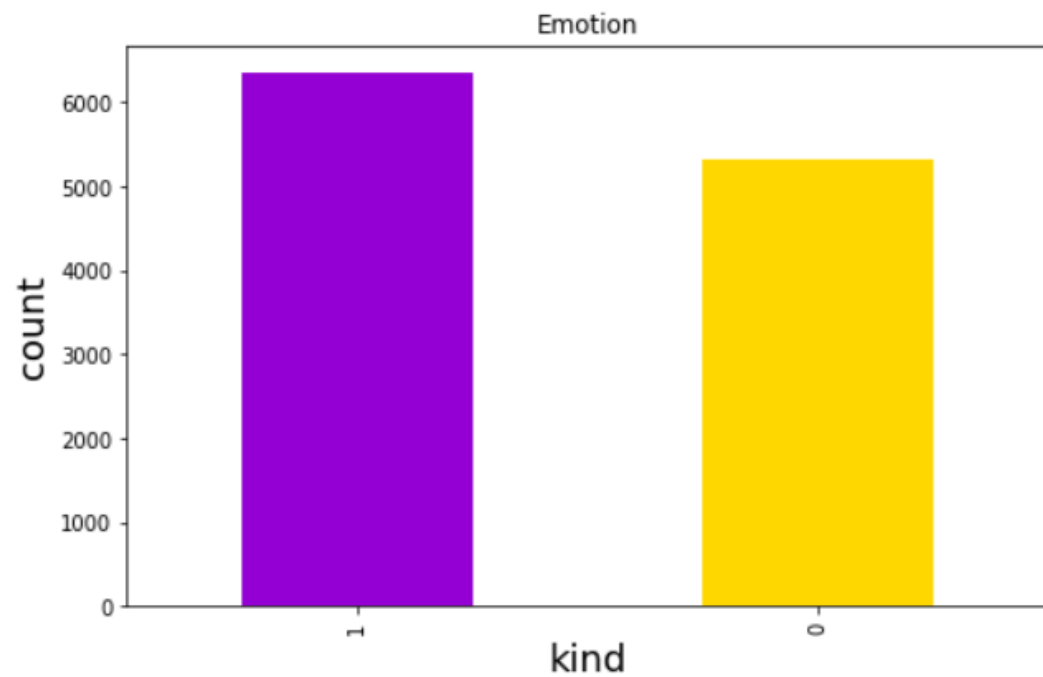
- **Vaccination**[疫苗接種]

- 疫苗、莫德納（moderna）、AZ（AstraZeneca、阿斯特捷利康）、BNT（Pfizer–BioNTech、輝瑞）、高端（MVC）



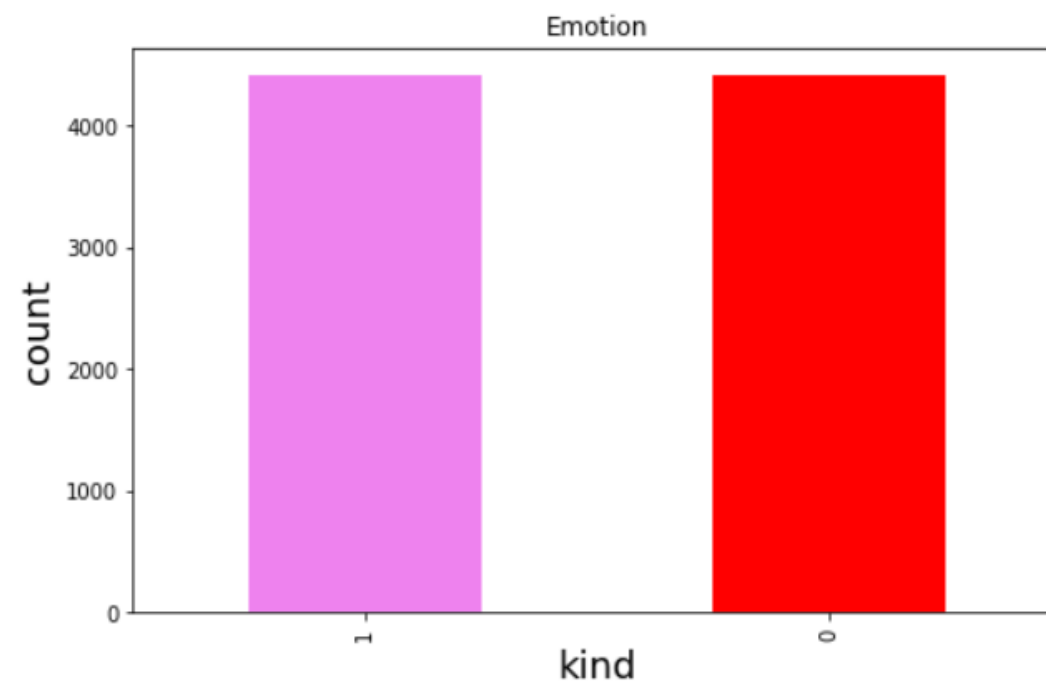
Proportion of information

Text(0.5, 0, 'kind')

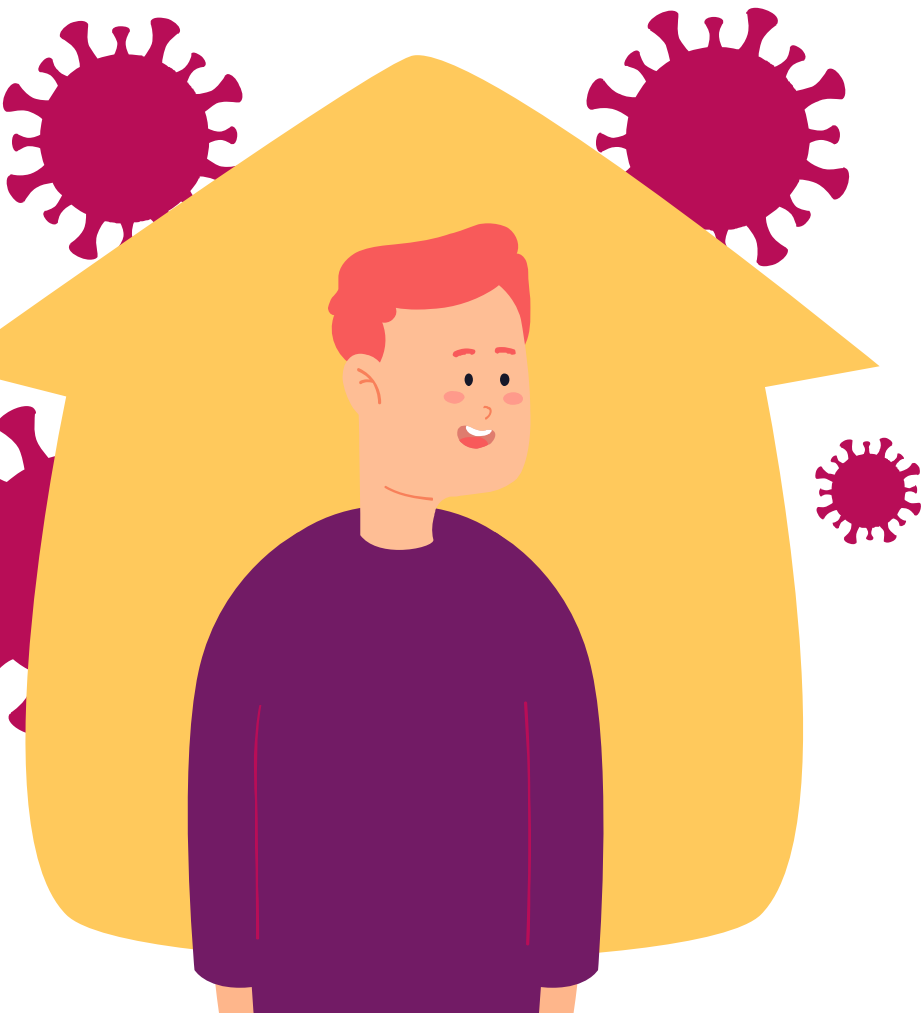


Original : 資料不平衡

Text(0.5, 0, 'kind')



SMOTE : 經由取樣達到樣本數量一致



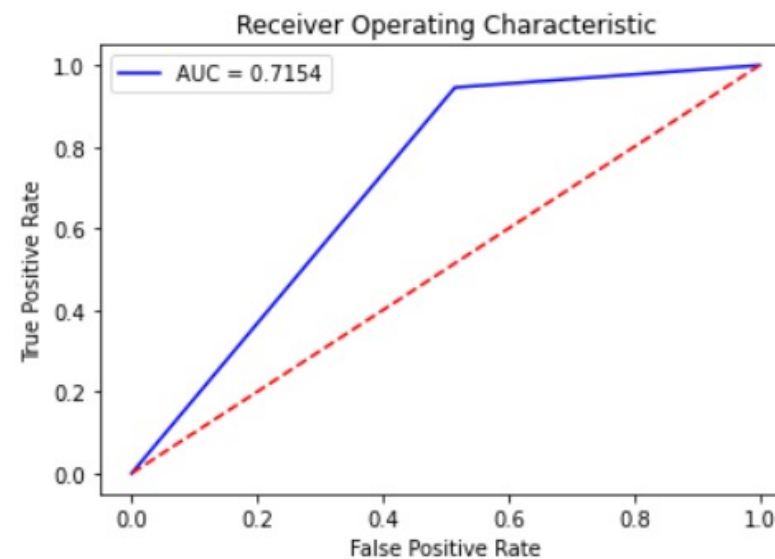
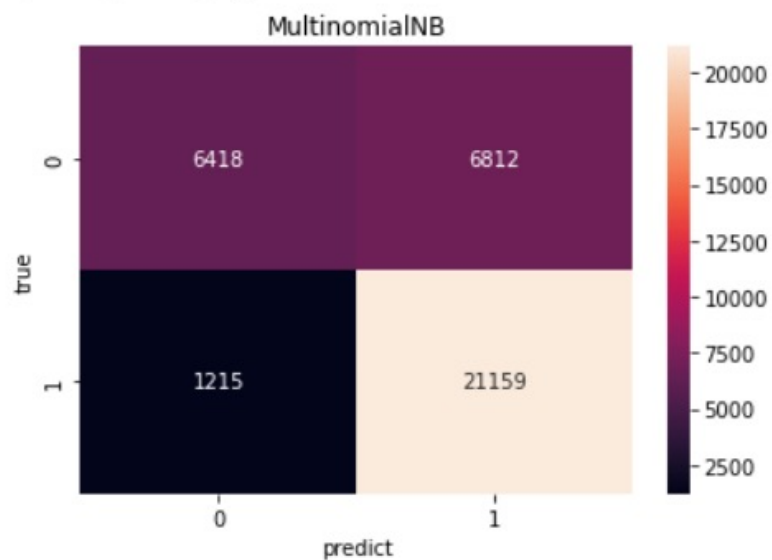
Multinomial NB

Accuracy: 77.45%

Confusion matrix:

```
[[ 6418  6812]
```

```
 [ 1215 21159]]
```



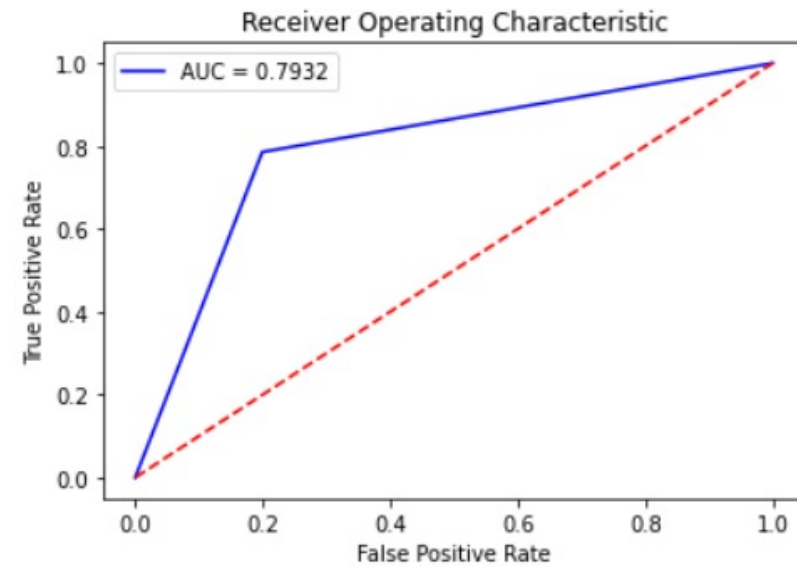
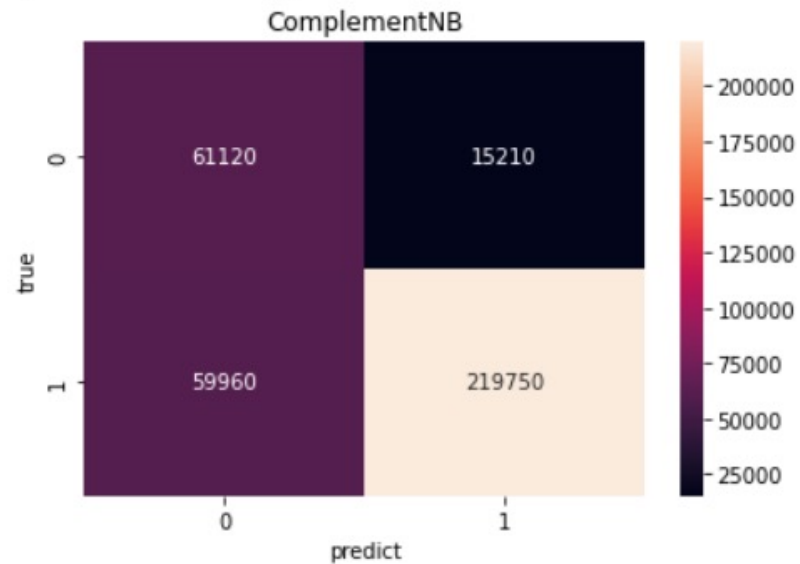
Complement NB

Accuracy: 78.89%

confusion matrix:

```
[[ 61120  15210]
```

```
 [ 59960 219750]]
```



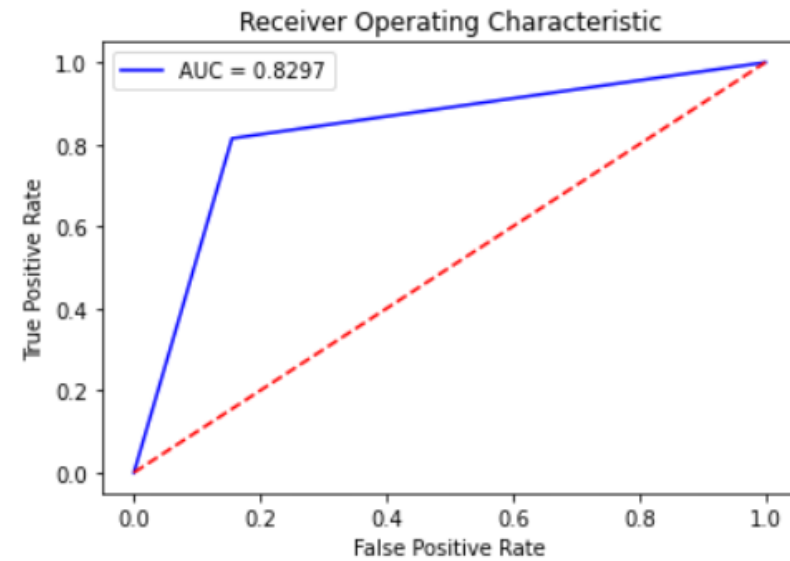
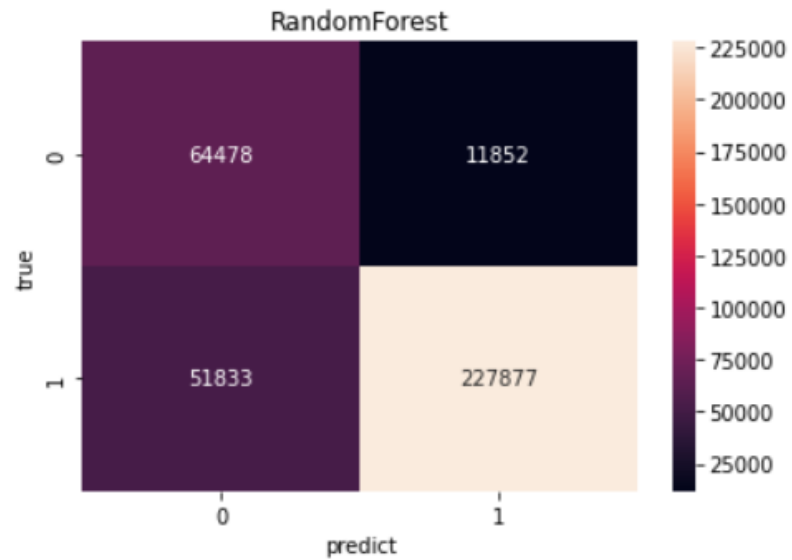
Random Forest

Accuracy: 82.11%

confusion matrix:

```
[[ 64478 11852]
```

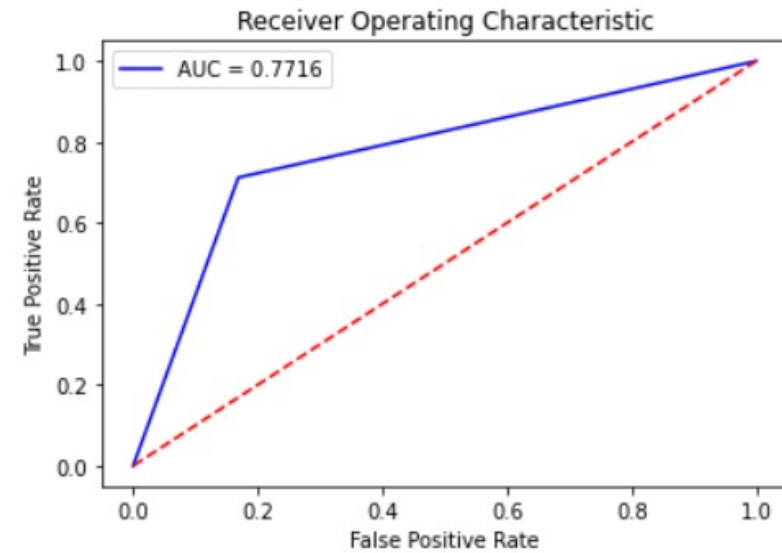
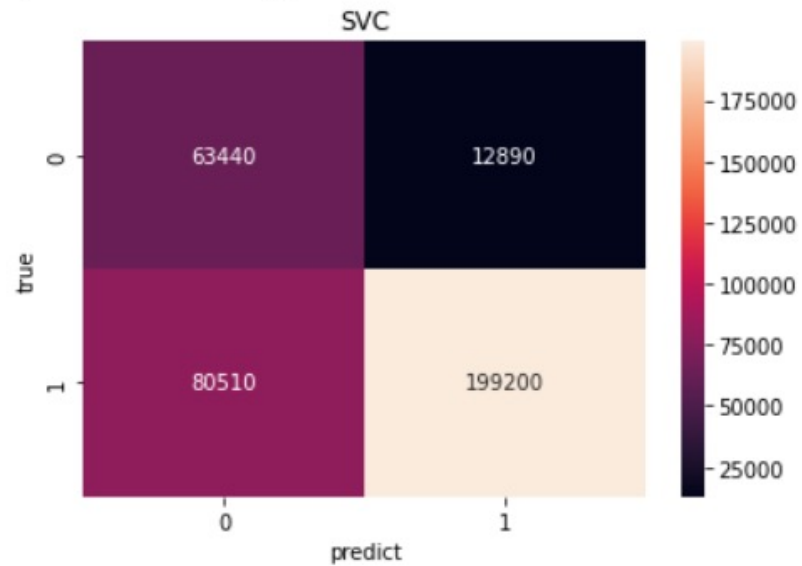
```
[ 51833 227877]]
```



Support Vector Machine - SVC

Accuracy: 73.77%

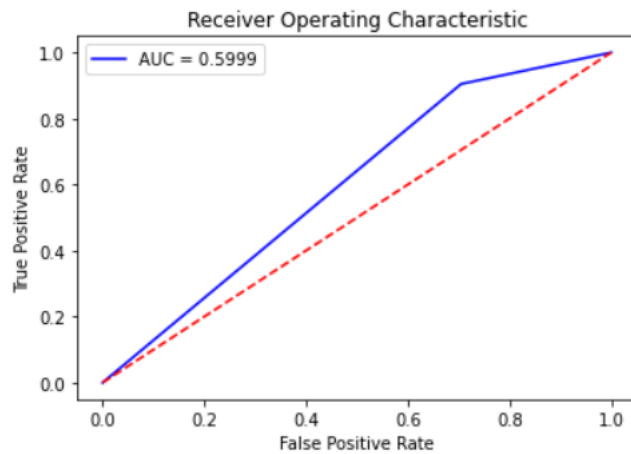
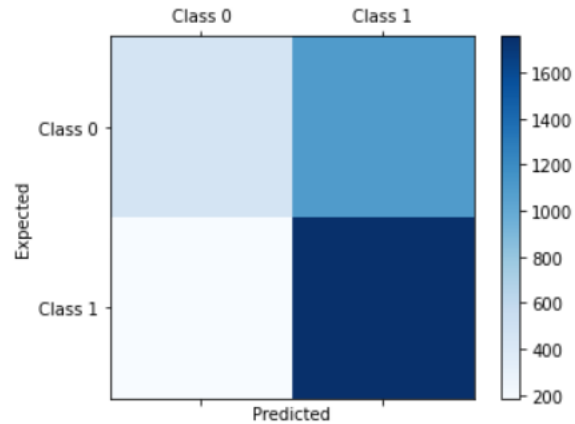
confusion matrix:
[[63440 12890]
 [80510 199200]]



XGBClassifier

Confusion matrix:

```
[[ 460 1099]
 [ 185 1758]]
```

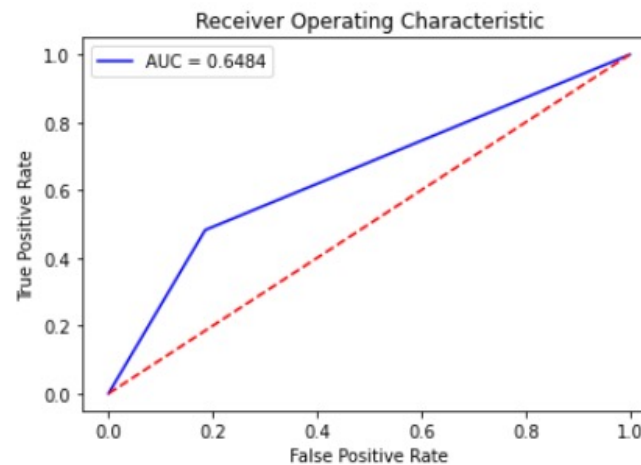
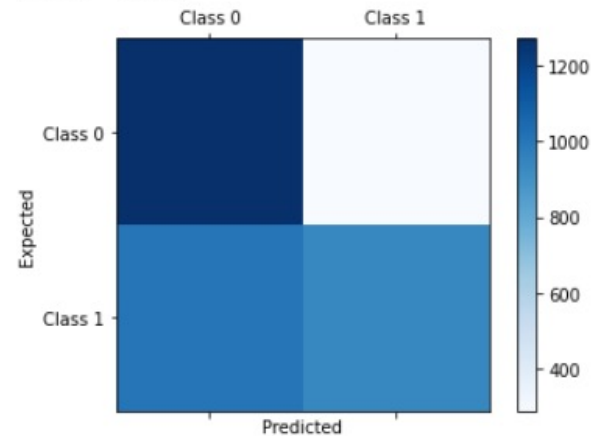


Accuracy: 63.34%

SMOTE+XGBClassifier

Confusion matrix:

```
[[1270  289]
 [1006  937]]
```

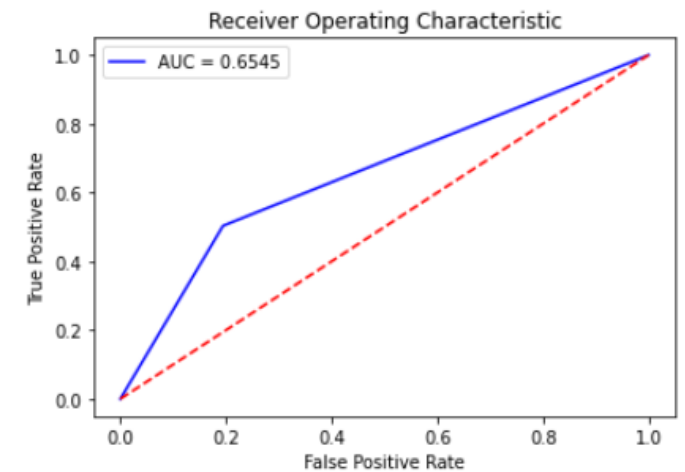
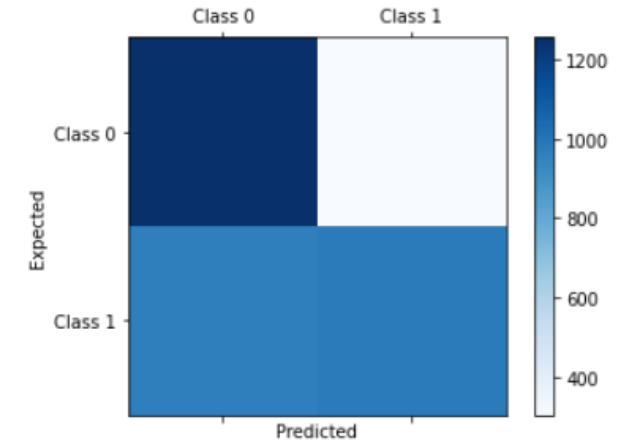


Accuracy: 63.02%

BL-SMOTE+XGBClassifier

Confusion matrix:

```
[[1256  303]
 [ 965  978]]
```

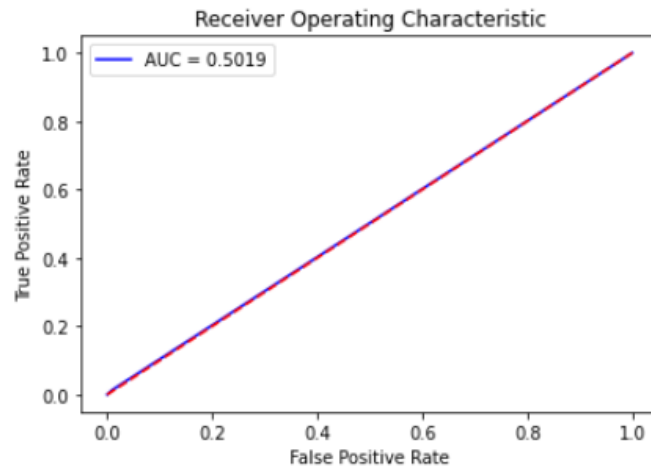
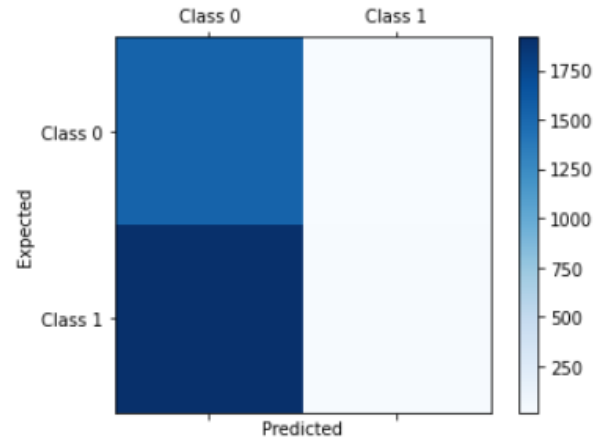


Accuracy: 63.79%

GaussianNB()

Confusion matrix:

```
[[1545  14]  
 [1918  25]]
```

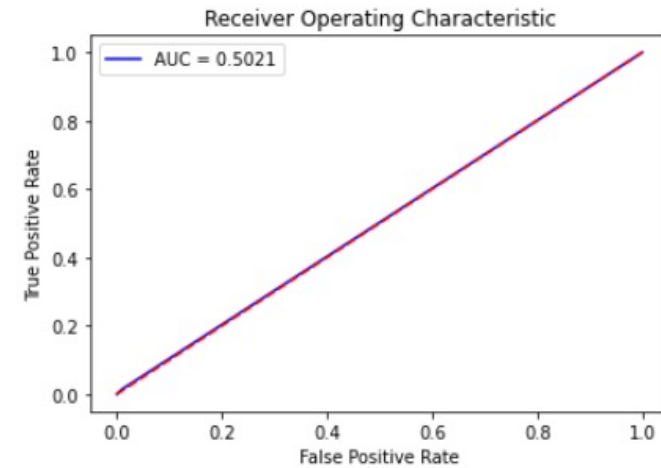
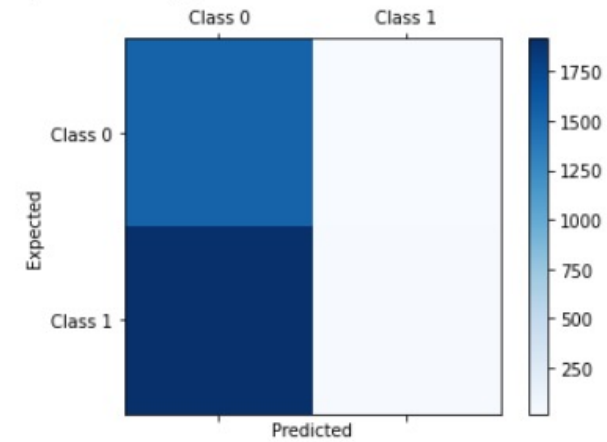


Accuracy: 44.83%

SMOTE+GaussianNB()

Confusion matrix:

```
[[1544  15]  
 [1916  27]]
```

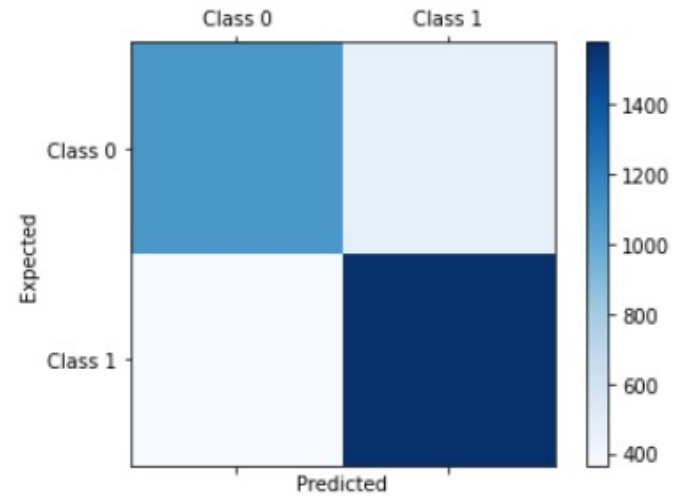


Accuracy: 44.86%

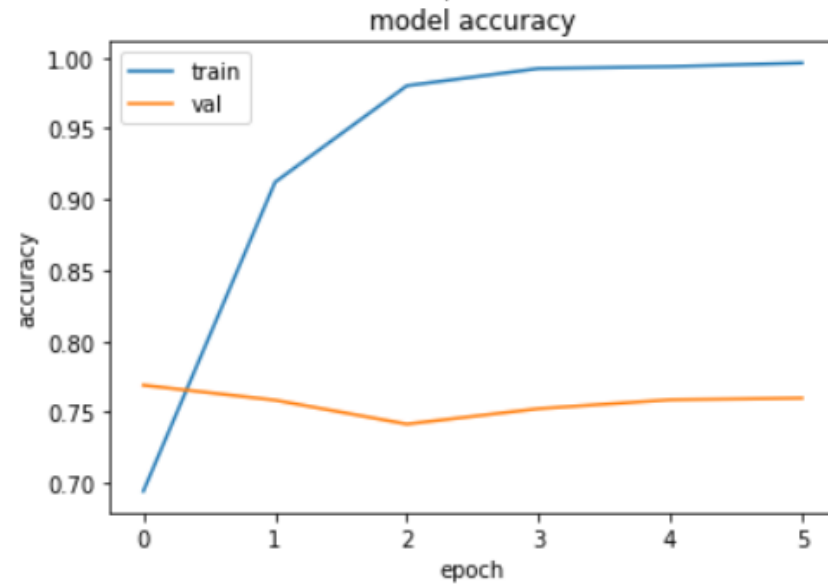
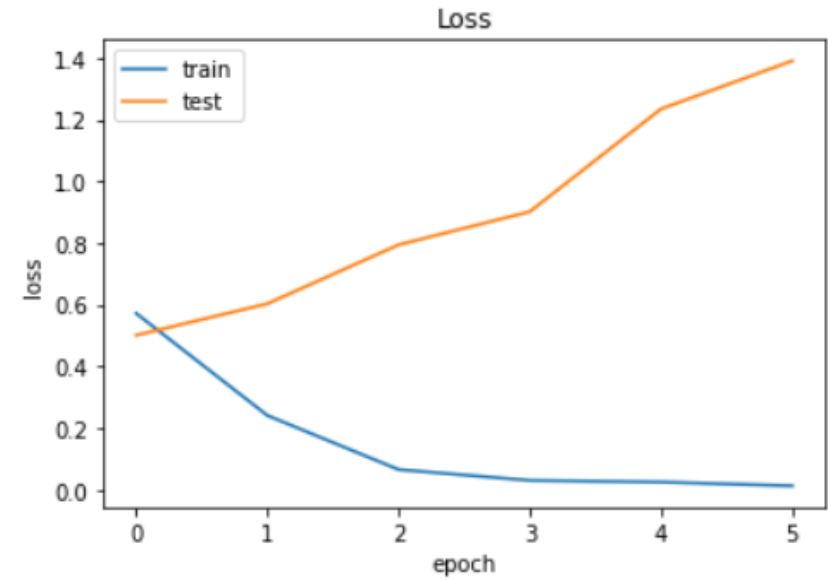
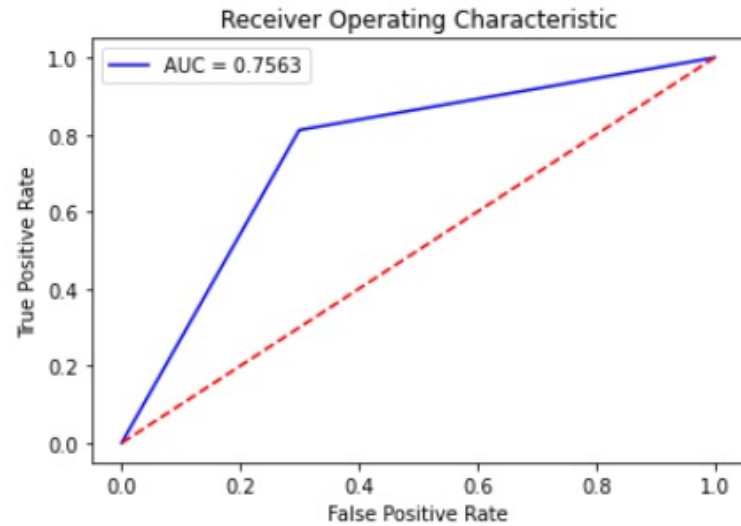
Bi-LSTM

Confusion matrix:

```
[[1092  467]
 [ 365 1578]]
```



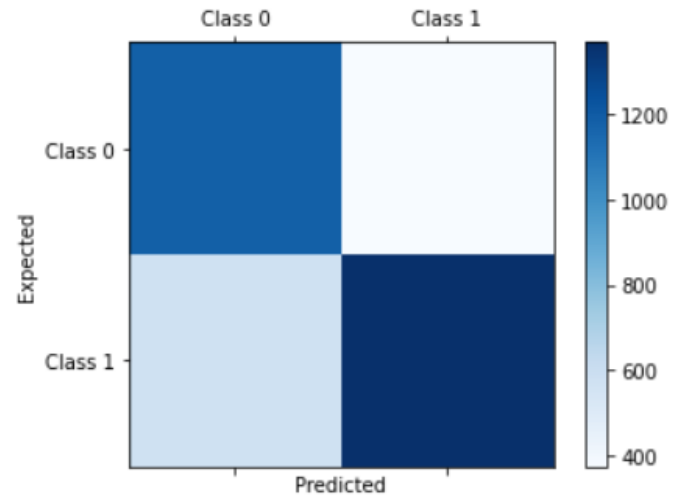
Accuracy: 75.99%



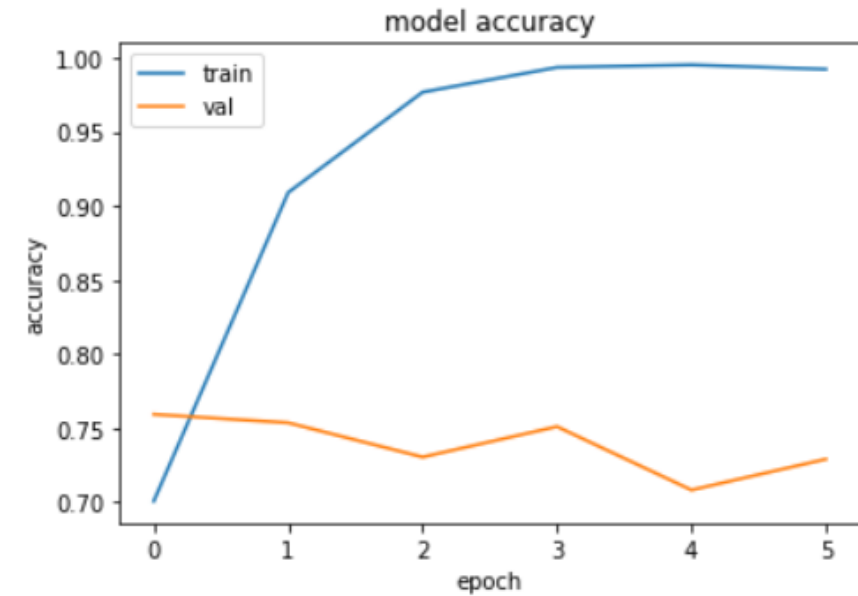
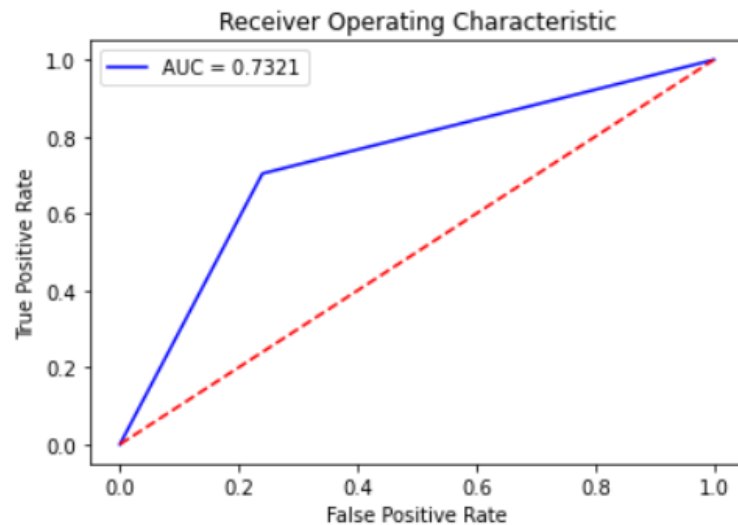
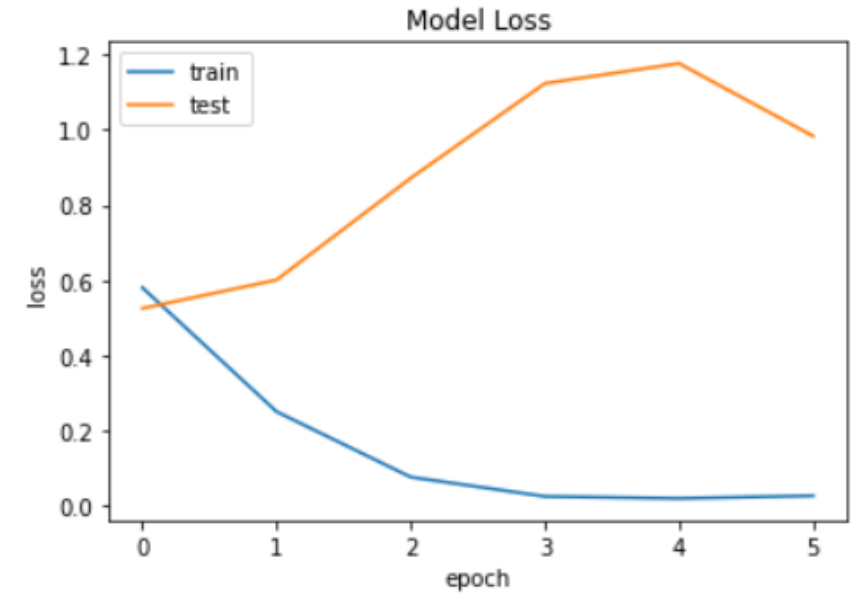
SMOTE+Bi-LSTM

Confusion matrix:

```
[[1185  374]
 [ 575 1368]]
```



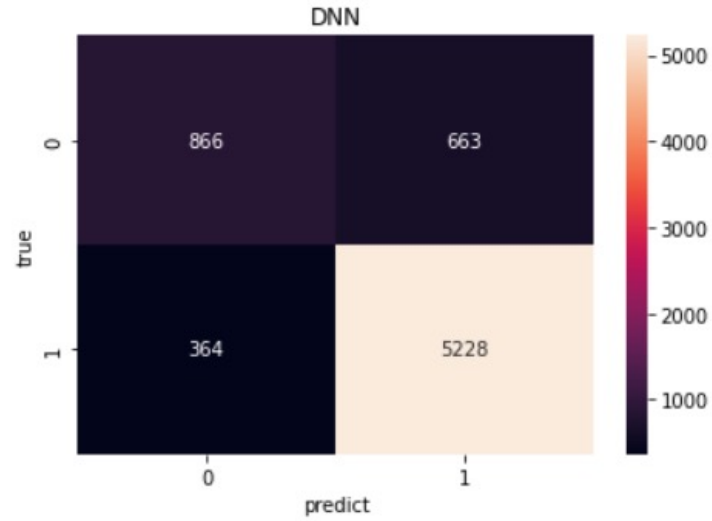
Accuracy: 72.90%



DNN

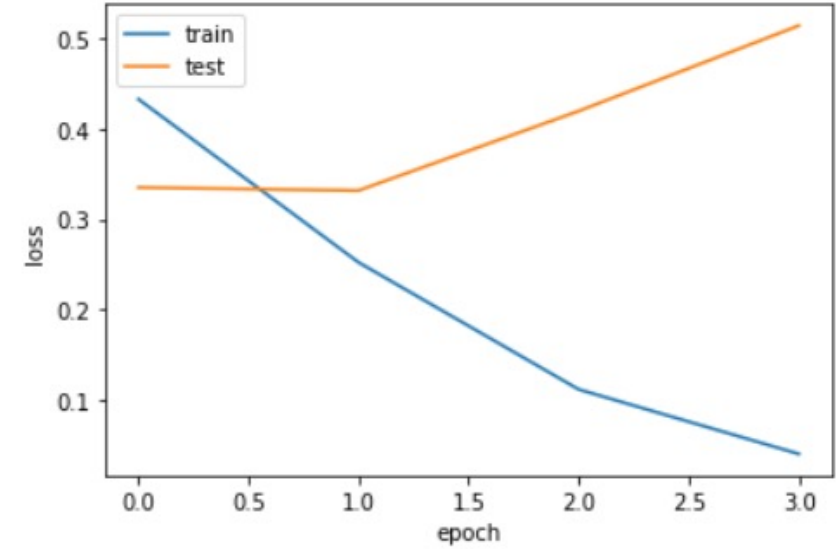
Confusion Matrix

```
[[ 866  663]
 [ 364 5228]]
```

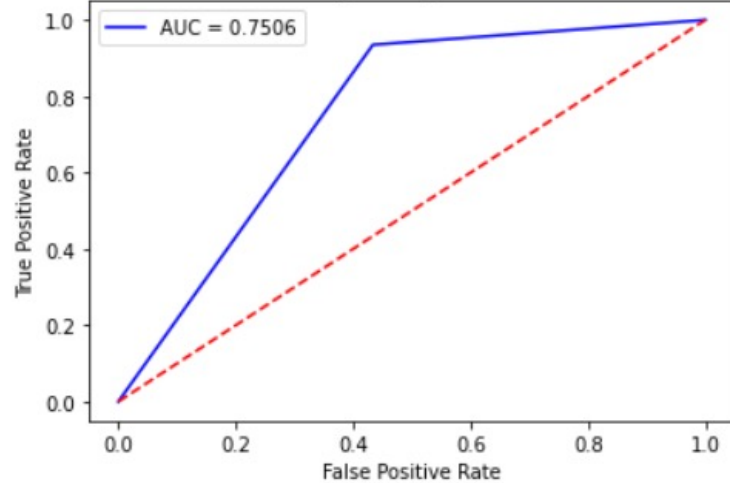


Accuracy: 85.58%

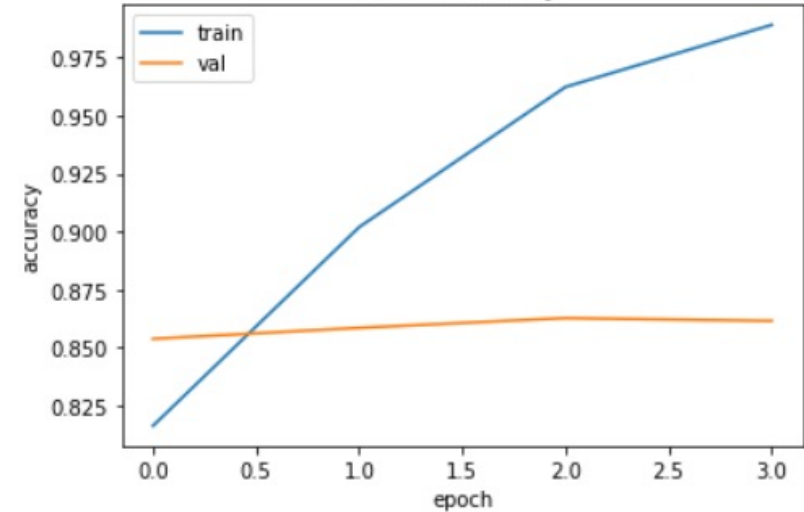
Model Loss

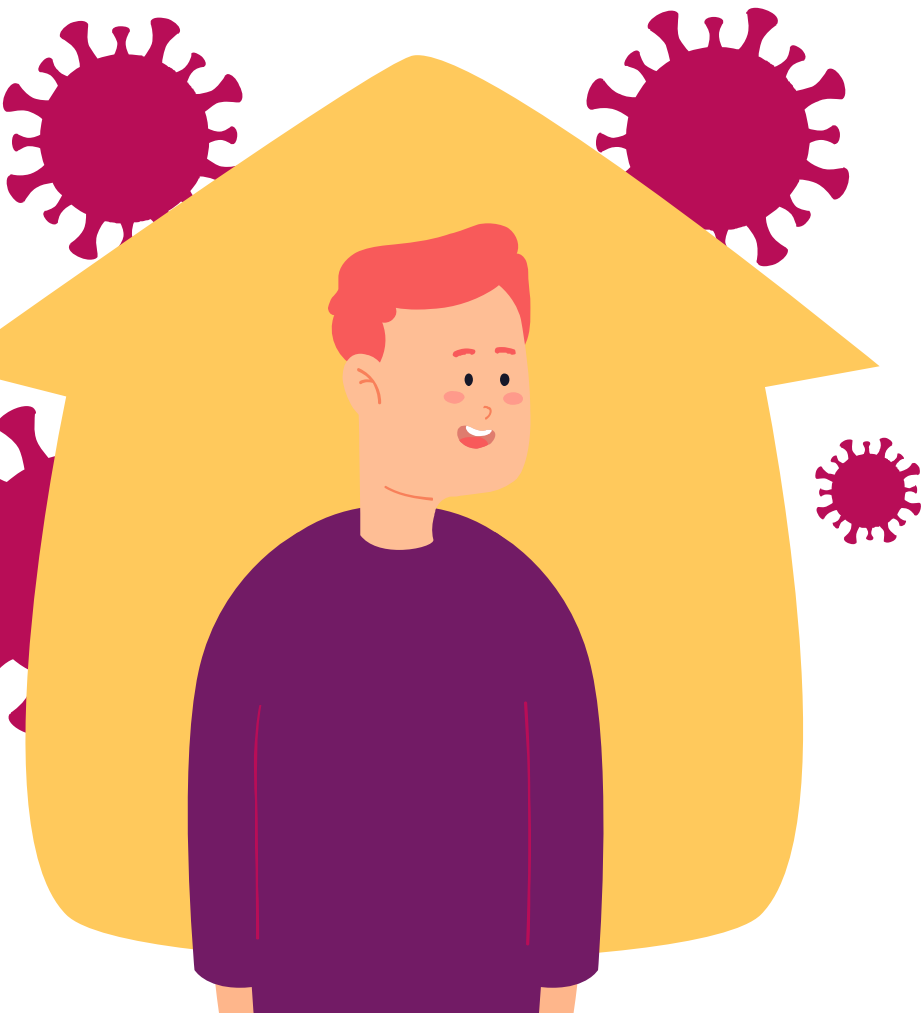


Receiver Operating Characteristic



model accuracy





METHODS

MODEL	label	precision	recall	f1-score
MNB	neg	0.84	0.49	0.62
	pos	0.76	0.95	0.84
CNB	neg	0.50	0.80	0.62
	pos	0.94	0.79	0.85
RandomForest	neg	0.55	0.84	0.67
	pos	0.95	0.81	0.88
SVC linear	neg	0.44	0.83	0.58
	pos	0.94	0.71	0.81
XGBClassifier	neg	0.71	0.30	0.42
	pos	0.62	0.90	0.73
SMOTE+XGBClassifier	neg	0.56	0.81	0.66
	pos	0.76	0.48	0.59
BL-SMOTE+XGBClassifier	neg	0.54	0.81	0.66
	pos	0.76	0.50	0.61
GaussianNB	neg	0.45	0.99	0.62
	pos	0.64	0.01	0.03
Bi-LSTM	neg	0.70	0.74	0.72
	pos	0.80	0.77	0.79
SMOTE + Bi-LSTM	neg	0.67	0.76	0.71
	pos	0.79	0.70	0.74
DNN	neg	0.70	0.57	0.63
	pos	0.89	0.93	0.91

neg means negative emotion / pos means positive emotion

METHODS - MNB

01

辨識文字類型資料快速



02

具一定程度 precision 、 recall 、 f1



03

有標籤資料預測情感後，再用 Predict_proba 算該新聞被預測0或1機率
ex :若新聞預測為 [0.56651809 0.43348191] → 非「負面」而是中性情緒

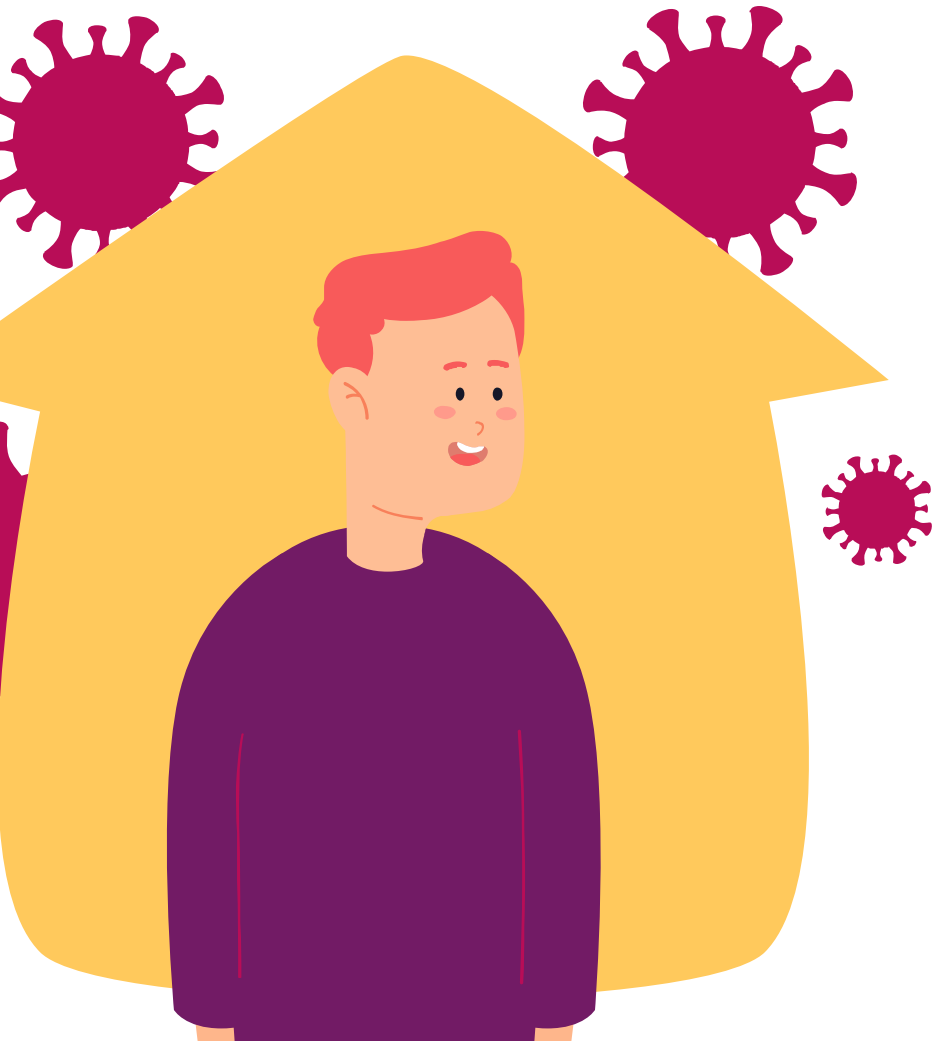


04

已研發DNN、Bi-LSTM



Discussion



加入SMOTE算法資料集訓練結果，大多Accuracy會下降或持平。

欲優化該資料集訓練的準確度，目前有2大方向可以進行優化:

- 1.考慮其他模型(如：CNN、BERT)
- 2.加入borderline-2之SMOTE優化該模型

接下來這2大方向為我們優化模型的方向。