Rapport TP3

- From the previous lab (lab2) we have obtained:

	Model	Vectorizer	Accuracy	Precision	Recall	F1-score
0	logistic	CountVectorizer	0.413687	0.602092	0.413687	0.247276
1	tanh	CountVectorizer	0.779707	0.780978	0.779707	0.779796
2	logistic	TF-IDF	0.411985	0.169732	0.411985	0.240416
3	tanh	TF-IDF	0.828056	0.828061	0.828056	0.828023

	Model	Vectorizer	Accuracy	Precision	Recall	F1-score
0	logistic	w2v	0.411985	0.169732	0.411985	0.240416
1	logistic	ft	0.394961	0.257566	0.394961	0.280625
2	logistic	gl	0.595506	0.595056	0.595506	0.595247
3	tanh	w2v	0.352400	0.270071	0.352400	0.291409
4	tanh	ft	0.411645	0.275096	0.411645	0.247362
5	tanh	gl	0.614573	0.614209	0.614573	0.614334

  According to the figures we can clearly see that the model which uses **TF-IDF** and **tanh** activation function has the best results across all metrics, this is going to be the model that we'll try to improve.

Improving the model

- The model was trained on a simple architecture of 2 hidden layers of sizes [32, 64, 128]neurons with tanh activation function, a learning rate of 1e-5, 1e-3 and Adam Optimizer.
- The trivial thing to do is to try **Random Search** to try and improve the architecture of the neural network, therefore Random Search was realized with

3 target variables:

- The number of neurons in the first layer .
- The number of neurons in the second layer .
- The learning rate.

- Running the Random Search for 20 trials, with Adam optimizer and 20 epochs each yields the following results :

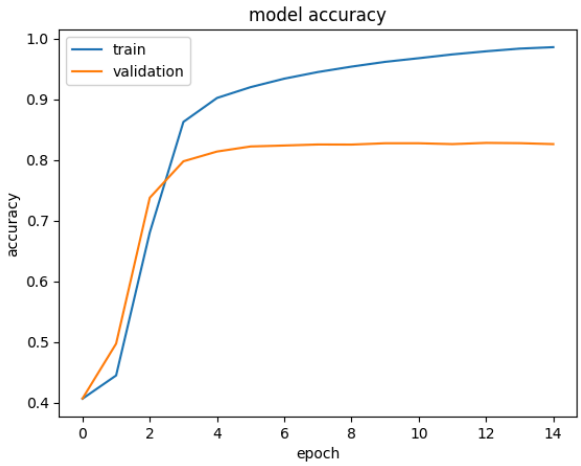
```
Trial 20 Complete [00h 02m 03s]
val_accuracy: 0.8276582658290863

Best val_accuracy So Far: 0.8294381499290466
Total elapsed time: 00h 41m 35s
```

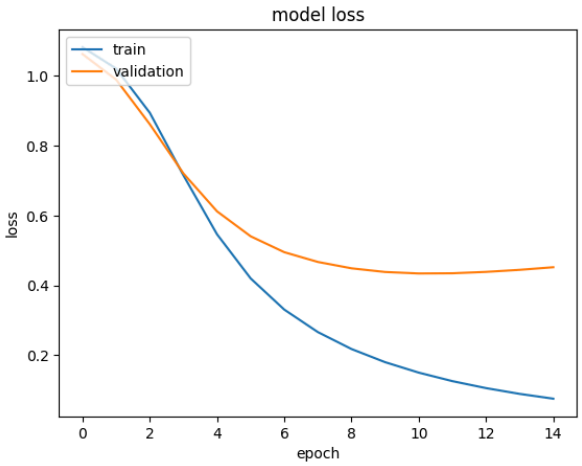
```
{'layer_1': 128, 'layer_2': 64, 'learning_rate': 3.161391335908762e-05}
```

- The best obtained Hyper-parameters were:
  - 128 neurons for the first hidden layer
  - 64 neurons for the second hidden layer
  - A learning rate of 0.0000031613

Final results and discussion

After running the model with the best Hyper-parameters for 15 epochs the following results were obtained:





- The model achieves a validation accuracy of 82%.

Metrics results:

```
Validation accuracy: 0.8278903961181641
Validation precision: 0.8252391616503397
Validation recall: 0.810555636882782
Validation f1 score: 0.817831496276939
```