Lab 03 (NLP Non Deep)

Question 4 :

The 2 models differ on the way they were treated and their results.

TREATMENT

On question 2, we trained a FastText classifier with default parameters on the training data and evaluated it on the test data using accuracy.

On question 3 we trained a FastText classifier with hyperparameters search functionality.
First, we split our training set into a training and a validation set.
Then we trained the model, and in order to activate hyperparameter optimization, we provided a validation file with the autotune validation argument. Then, fastText searches the hyperparameters that give the best f1-score on 'imdb_valid.txt'. Finaly we tested our obtained model.

RESULTS

The default model yielded values for precision at one as 0.5 and recall of one as 0.5 as well.

For the model with attributes found with hyperparameters search, we used the test method to evaluate the classifier on the training set that we split into a training and a validation set, which yielded values for precision at one as 0.7142 and recall of one as 0.7142 as well.

The precision is the number of the correct labels predicted by the classifier among all the labels and the recall is the number of labels successfully predicted among the real labels.

The precision and the recall of the model with attributes found with hyperparameters are 0.2142 superior than the default model. Therefore, the former model performed much better than the latter.

Question 5:

According to FastText's documentation, training a FastText classifier takes a text file as input and every line corresponds to a sample and must have the following format:

__label__<your_label> <corresponding text>

Using the tuned model, we took 2 wrongly classified examples from the test set using the incorrect following formats for the model to fail:

__label__<corresponding text> <your_label> (switched <your_label> and <corresponding text>)

<your_label> <corresponding text> (removed __label__)