UNIVERSITY OF ATHENS
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

Deep Learning for NLP

Student name: <Αγγελος Κοντος>
sdi: <*sdi2000089*>

Course: *Artificial Intelligence II (M138, M226, M262, M325)*
Semester: *Fall Semester 2023*

Contents

1. Abstract

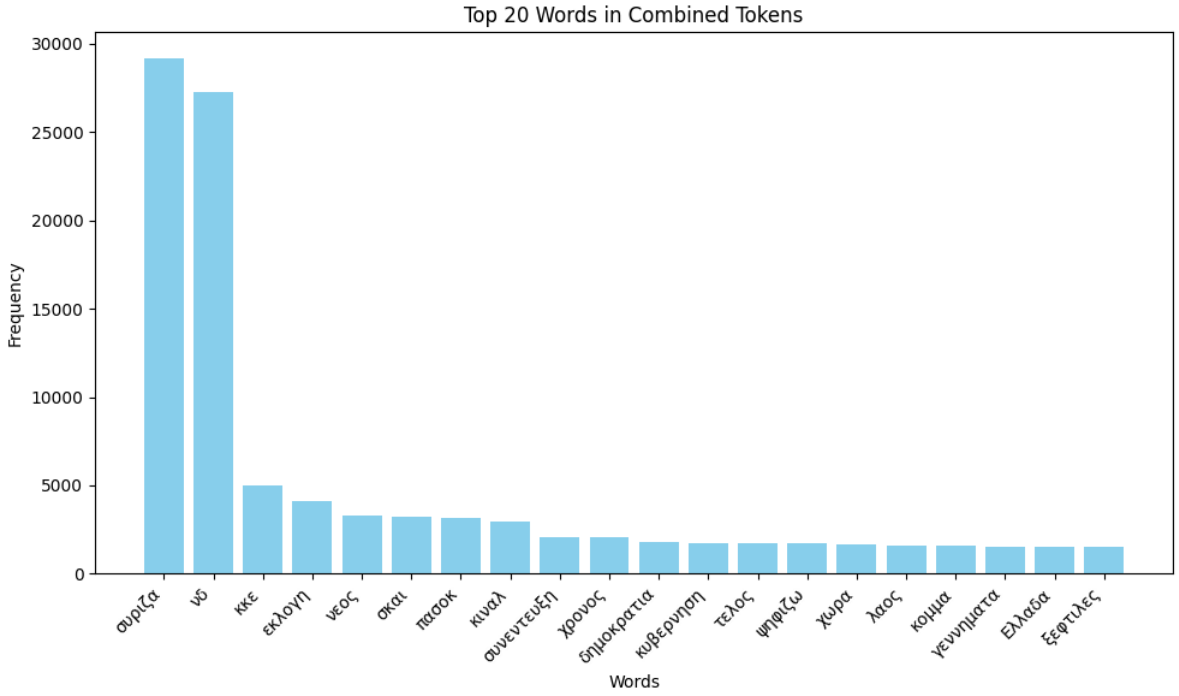
The task we have to do is exactly the same as the previous three tasks, specifically to find the sentimental value from Greek sentences (Tweets) using bert and distilbert, using the same text editor and instead of word2vec I will use bert's tokenazer

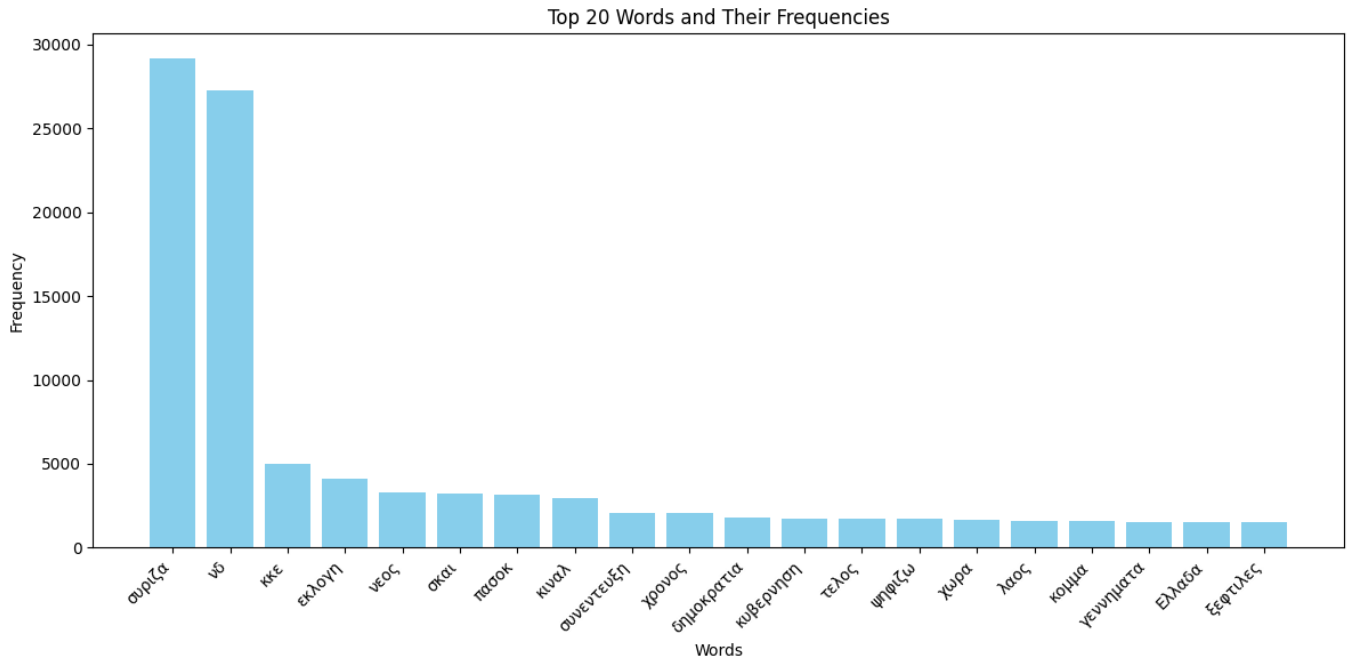
2. Data processing and analysis

2.1. Pre-processing

This part is exactly the same as the 3rd task, specifically in the pre process I put lemetazation, I took out links, English letters, accents, I also joined as much as I could the words of parties, syriza tsipras, etc., I also took out the #, and other symbols

2.2. Analysis





2.3. Data partitioning for train, test and validation

Because I use optuna and dataloader every time part of the total material goes through, but I saw that the best is between batch_size expa from 10 to 1000 according to my experiments

2.4. Vectorization

I use bert's tokenizer=vectorizer and then concatenate the words together and pass the average

3. Algorithms and Experiments

3.1. Experiments

The experiments I did initially large epoch (500-1000) small $\text{batch}_size(10-50), small epoch(0-50) small batch_size(10-50), medium epoch(100-500) small batch_size(10-50) I saw that the epochs didn't t 100) and large batch_size(100-500) which didn't change either, I also did with 1 and 2 h levels didn't change an$

1) large epoch (500-1000) small batch_size (10-50)
2) small epoch (0-50) small batch_size (10-50)
3) medium epoch (100-500) small batch_size (10-50)
4) small epoch (0-50) medium batch_size (50-100)
5) small epoch (0-50) large batch_size (100-500)
6) small epoch (0-50) small batch_size (10-50) 2 h level
7) small epoch (0-50) small batch_size (10-50) 1 h level micro lr
8) small epoch (0-50) small batch_size (10-50) 1 h level mirko lr again with more experiments

Trial				Score
1	0%	34%	0%	34%
2	0%	0%	34%	34%
3	0%	34%	0%	34%
4	34%	0%	0%	34%
5	0%	0%	34%	34%
6	0%	0%	34%	34%
7	34%	0%	0%	34%
8	0%	0%	34%	34%

Table 1: Trials

3.1.1. Table of trials.

3.2. Hyper-parameter tuning

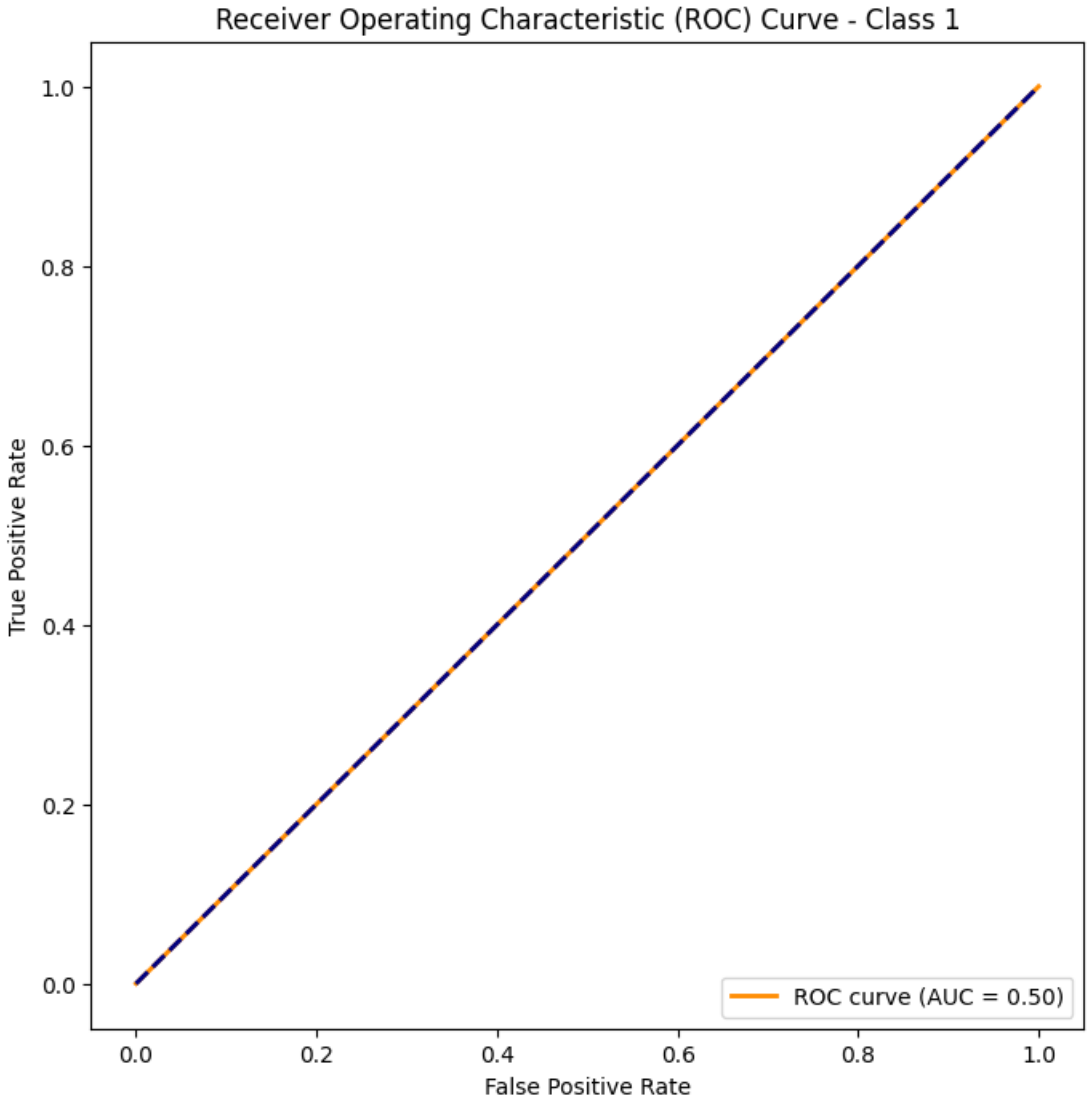
I would say it is under fitting because it is never to blame for all three qualities, whether neutral, positive or negative, this from what I see is due to the dataloader, but even after extra effort to make sure that it will have objects from all its categories, it still produced this, and also the not so good dataset, i.e. even worse results than the previous three on this part and I played with all the intermediate values from epoch large to small, respectively lr and batch_size etc. etc. all the changes I didn't do anything important it's more a matter of luck what it will produce than parameters

3.3. Optimization techniques

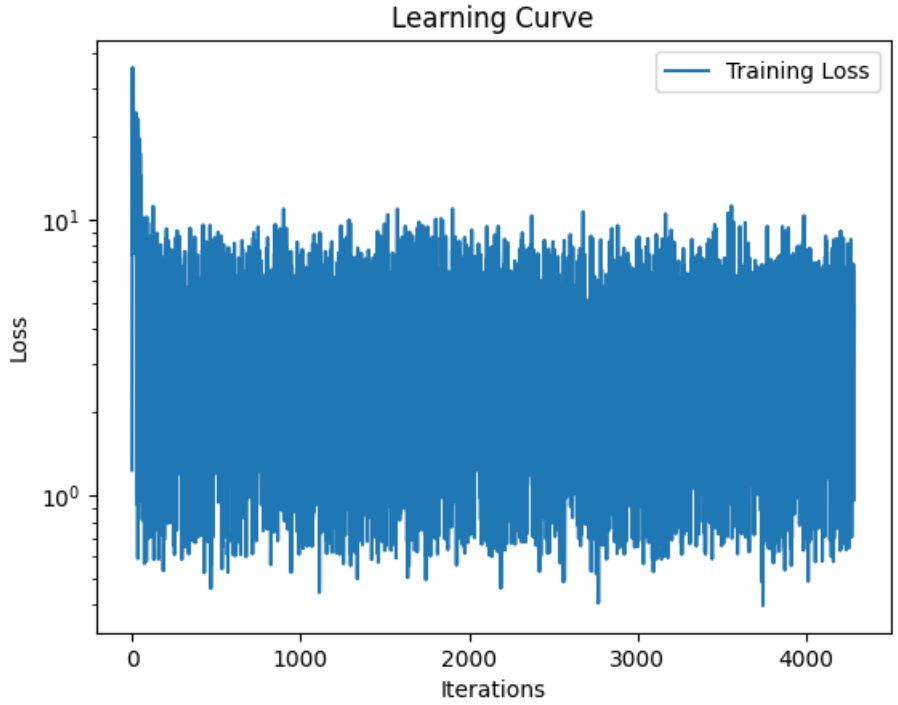
I used optuna and bert, both distilbert, bert also the same with work 3 lr, batch_size, epoch master, layer etc.

3.4. Evaluation

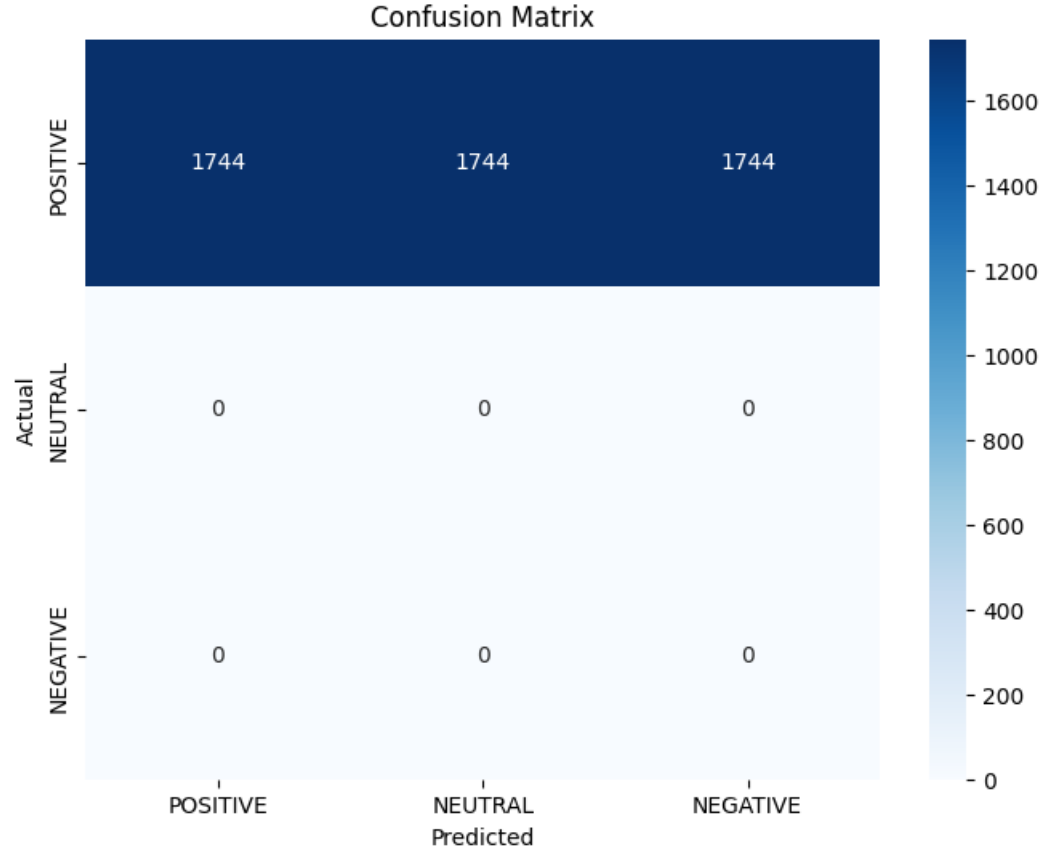
3.4.1. ROC curve. xa



3.4.2. Learning Curve. xa



3.4.3. Confusion matrix. xa



4. Results and Overall Analysis

4.1. Results Analysis

The results are quite negative and because I noticed that everything is played in a game of chance instead of some logic that gets better or worse so whatever experiment is done will not provide a basis, and I noticed after the 4 tasks that the reason for the bad dataset is the more underdeveloped the learning system that produces good results

<Provide and comment diagrams and curves> Done on roc, matrix, loss cut the charts

Trial	αρνητικα	ουδετερα	θετικα	Score
8	0%	0%	34%	34%

Table 2: Trials

4.1.1. Best trial. <Showcase best trial>

4.2. Comparison with the first project

The first project with logical regression is the experiment with the best results, with enough significance in the parameter changes

4.3. Comparison with the second project

The second project with simple layers does not have such big differences in the creation, just here we use bert (except for the results I produce) but in the process for their creation there are small differences in the bert part

4.4. Comparison with the third project

The third one has various things like in the third one we have to make a relaxed and small attention and memory rnn piece by ourselves while here there is between bert and they have the same one with a little worse various ones found the right sentiments

5. Bibliography

References

Apart from the pdf and the information on the exercises, I only used verses

pdf εργασιας