Sentiment analysis based on Greek Bert and MT5

George Kountouris, Minas Alifragkis, Christos Morfopoulos

3 July 2022

Contents

1 Abstract

Sentiment Analysis is a significant task in Natural Language Processing (NLP). It uses to understand the sentiments of the customer/people for products, movies, and other such things, whether they feel positive, negative, or neutral about it. It is often performed on textual data, helps companies and other related entities know about their products/services, and helps them work on the feedback to improve it further. Sentiment analysis is one of the most challenging tasks in NLP because even humans struggle to analyze sentiments accurately. Some immediate challenges derived from languages are subjectivity and tone, context, polarity, irony, and sarcasm. In the current work, we are comparing different approaches performed on Greek movie reviews. More specifically, on a dataset named Athinorama Greek Movie Reviews, Sentiment Analysis2 from Kaggle. So, we choose the Sentiment Analysis task because it is difficult and because we think it has vast applications in industry and research.

2 Data exploration

The selected dataset includes data from movie reviews in the Greek language. More specifically, it is downloaded from the Kaggle platform and contains a large amount of data to train our models. Data extracted from Athinorama, an informative Greek site concerning many categories of entertainment. It has 148.795 reviews for 6.481 unique movies. The provided information is the following: id number, greek title, original title, category, director, movie length, movie date, author, review date, review, stars, label, mean of star ratings, number of reviews, and URL.

	id number	greek title	original title	category	director	movie lenght	movie date	author	review date	review	stars	label	mean of stars	number of reviews	url
0	0	Και Τώρα Κάτι Τελείως Διαφορετικό	AND NOW FOR SOMETHING COMPLETELY DIFFERENT	Κωμωδία	Ίαν Μακ Νότον	88	1971	Marvin	2002	Φοβερή η σύλληψη του χιούμορ από τους Python"...	4 αστεράκΑ	4.0	4.5	2	https://www.athinorama.gr/cinema/movieratings....
1	1	Και Τώρα Κάτι Τελείως Διαφορετικό	AND NOW FOR SOMETHING COMPLETELY DIFFERENT	Κωμωδία	Ίαν Μακ Νότον	88	1971	Χριστόφορος Ζώνας	2002	Από τις καλύτερς στιγμές των Μ.Π., ισάξιο μόνο...	5 αστεράκΑ	5.0	4.5	2	https://www.athinorama.gr/cinema/movieratings....
2	2	Το Αδελφάτο των Ιπποτών της Ελεεινής Τραπέζης	Monty Python and the Holy Grail	Κωμωδία	Τέρι Γκίλιαμ, Τέρι Τζόουνς	91	1975	Vandim	2020	Κλασικό! Από τις καλύτερες και ανατρεπτικότερε...	4 αστεράκΑ	4.0	4.0	32	https://www.athinorama.gr/cinema/movieratings....
3	3	Το Αδελφάτο των Ιπποτών της Ελεεινής Τραπέζης	Monty Python and the Holy Grail	Κωμωδία	Τέρι Γκίλιαμ, Τέρι Τζόουνς	91	1975	dH	2015	Μου θυμιζει τα κόμικ του Αστερίξ που όσο μεγάλ...	5 αστεράκΑ	5.0	4.0	32	https://www.athinorama.gr/cinema/movieratings....
4	4	Το Αδελφάτο των Ιπποτών της Ελεεινής Τραπέζης	Monty Python and the Holy Grail	Κωμωδία	Τέρι Γκίλιαμ, Τέρι Τζόουνς	91	1975	Orestis	2014	Το κάτι άλλο!	5 αστεράκΑ	5.0	4.0	32	https://www.athinorama.gr/cinema/movieratings....

Figure 1: Data in dataframe

We also calculated the minimum, maximum and average number of reviews for each movie grouped by sentiment as you may see below

```
Negative Reviews
Minimum reviews: 1 / Maximum reviews: 160 / Mean reviews: 7.52

Neutral Reviews
Minimum reviews: 1 / Maximum reviews: 123 / Mean reviews: 8.68

Positive Reviews
Minimum reviews: 1 / Maximum reviews: 505 / Mean reviews: 12.22
```

Figure 2: Statistics or reviews concerning sentiment

3 Preprocessing

First and foremost, we checked if the data had any missing values in columns we were interested in (review and label). As it has no missing values dataset looks to be well created. Then we noticed that sentiment is in the form of stars, so we decided to transform these labels to form 3 classes, positive, neutral, and negative. So all comments with 0 to 1.5 stars are classified as negative, 2 to 3.5 as neutral, and 4 to five as positive.

	review	label	review_cleansed	sentiment
131878	Ε Ξ Α Ι Ρ Ε Τ Ι Κ Η !!	2.0	ε ξ α ι ρ ε τ ι κ η	positive
31607	ΤΑΙΝΙΑΡΑ!!!!!Η ΤΣΟΛΗ ΦΟΒΕΡΗ Ο ΙΘΑΝ ΜΑΝΑΡΙ ΔΕΝ ...	2.0	ταινιαραη τσολη φοβερη ο ιθαν μαναρι δεν θα επ...	positive
121574	Εξαιρετικές ερμηνείες και από τους δυο πρωταγω...	1.0	εξαιρετικες ερμηνειες και απο τους δυο πρωταγω...	neutral
65312	Κάτι οι κάποιες κριτικές που διάβασα (LIFO),κά...	0.0	κατι οι καποιες κριτικες που διαβασα lifoκατι ...	negative
38318	Αρκετά καλη ταινία,το σενάριο περίπλοκο,η ιστο...	1.0	αρκετα καλη ταινιατο σεναριο περιπλοκοη ιστορι...	neutral
...	...	...	...	...
88510	εξαιρετικό	2.0	εξαιρετικο	positive
27856	Αρκετά καλή, αξίζει τα ευρώ σας, θα περάσετε π...	2.0	αρκετα καλη αξιζει τα ευρω σας θα περασετε πολ...	positive
116970	Εντυπωσιακή τεχνικά και κινηματογραφικά, εξαντ...	1.0	εντυπωσιακη τεχνικα και κινηματογραφικα εξαντλ...	neutral
17262	Πρέπει να επιλέξετε αστεράκια ή να γράψετε παρ...	2.0	πρεπει να επιλεξετε αστερακια η να γραψετε παρ...	positive
23084	Εντυπωσιακή ατμόσφαιρα & αγωνία!	2.0	εντυπωσιακη ατμοσφαιρα αγωνια	positive

Figure 3: Transformation from stars to sentiments classes

Concerning the reviews, we transformed all words into lower case, removing accents and punctuation.

4 Data visualizations

In following Bar Plot, you may see the labels (negative, neutral, and positive) distribution over movies. Translated to exact numbers, we have 34.495 reviews with negative sentiment, 46.011 reviews with the neutral sentiment, and 68.289 reviews with positive sentiment.

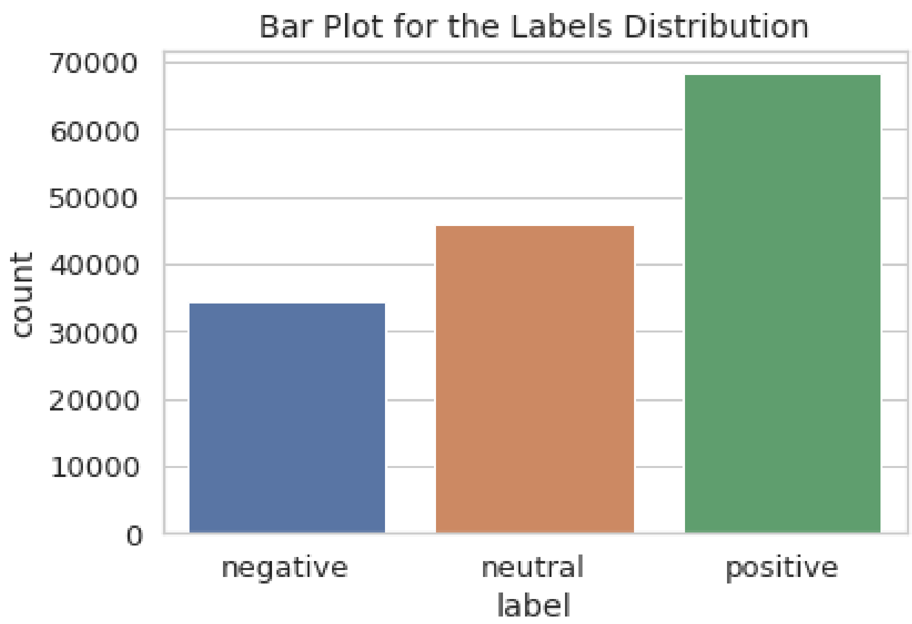

Figure 4: Sentiment distribution over reviews

The graph below is a plot that describes the different sentiment probability of the number of reviews through the movies axis.

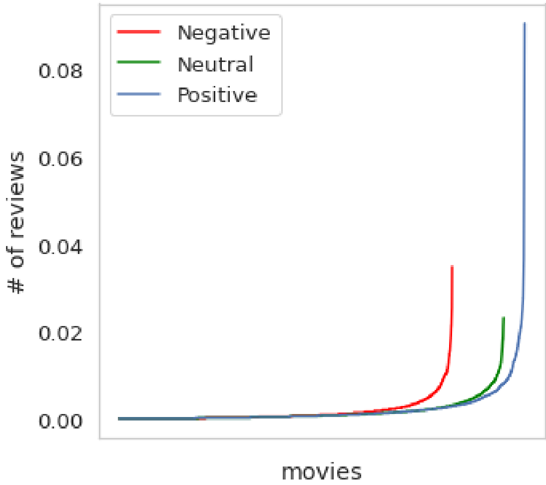

Figure 5: Review probability distribution throughout movies

We used a Word Cloud representation for each sentiment based on reviews that contain some indicative words. For instance Word Cloud for negative reviews contains the words “κρίμα”, “δυστυχώς”, “χάλια” etc, for neutral reviews contains the words “βέβαια”, “όμως”, “σενάριο” etc, for positive reviews contains the words “καλή”, “αριστούργημα”, “εξαιρετική” etc.

Figure 6: Word cloud representation

Finally, in following plot we display some of the top 3-grams for each sentiment. This visualization provide as with confidence that generally the dataset contains good quality data.

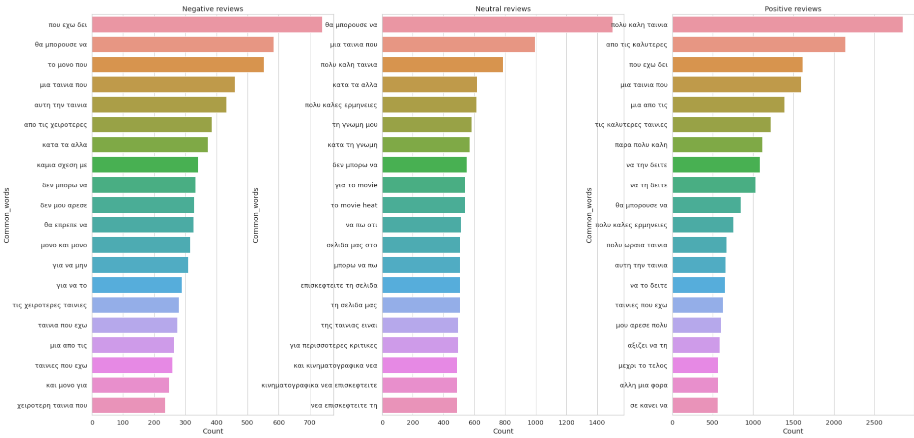

Figure 7: 3-grams for each sentiment

In negative reviews some of the most indicative n-grams are "από τις χειροτερες","δεν μου αρεσε" and "χειροτερη ταινια που". In neutral reviews some of the most indicative n-grams are "κατ τα αλλα","κατα τη γνωμη" and "να πω οτι". In positive reviews some of the most indicative n-grams are "πολυ καλη ταινια","από τις καλυτερες" and "πολυ καλες ερμηνειες".

5 Implementation of Greek BERT

Our initial approach was to implement sentiment analysis along with the BERT architecture. The dataset, as explained before, consists of Greek movie reviews from Athinorama, with their labels ranging from 0 to 5 stars. As we mentioned in the pre-processing section, we grouped the labels into negative, neutral, and positive sentiments. Therefore, our sentiment analysis problem is nothing more than a multi-class classification problem with three classes. As our dataset is mainly in Greek, we should apply transfer learning from the Greek version of the BERT model, which has robust embeddings with Greek vocabulary and has trained in the Greek Wikipedia, Greek legislation corpus, and many more.

Google developed and published the famous BERT in 2018 to better understand user searches. BERT uses the Transformer architecture as its basic structure. The Transformer is an architecture that relies solely on the attention mechanism and removes the need for recurrence by processing the entire input string at once instead of in a word-by-word fashion. However, as its acronym implies, BERT uses only Encoders from the Transformer architecture, and in our case, the Greek BERT is the base model with 12 encoders with 12 bidirectional self-attention heads. Finally, BERT has many applications such as speech recognition, question answering, topic modeling, summarization, language understanding, and many more.

A critical section is the role of the tokenizer. Apart from the pre-trained Greek model, we have loaded and also utilized the pre-trained Greek tokenizer. The tokenizer modifies the input text data into encoded tokens with a respective id and creates an attention mask for each representation to finally have the correct format for the model. Therefore, the role of the tokenizer is crucial. The embeddings of the BERT have a max threshold of 512 tokens, and with respect to the main training corpus, the user should determine the max length of the tokens. We should also mention that the max length of the tokens significantly impacts computational cost. Therefore, we have reduced training time if we keep the max length parameter low. In our case, we plotted the tokens' length to visualize their distribution.

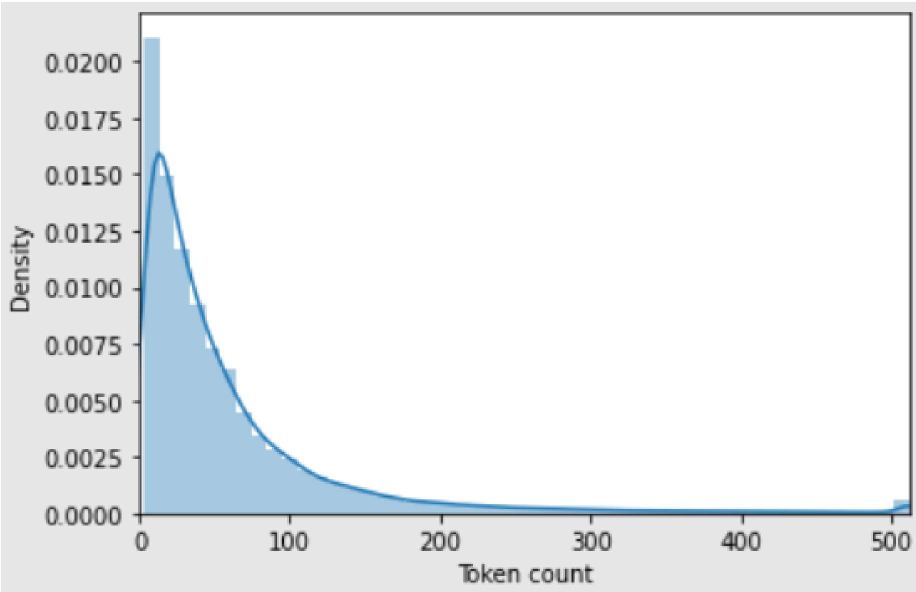

Figure 8: Distribution of length of the Tokens

From the above plot, we can represent most tokens with a length of 120 to 150. Therefore we applied on the tokenizer a max length equal to 128. In the figure below, we can see an illustration of the tokenizer for giving a text along with its padding to reach the proper max length.

```
αριστουργημα του κορυφαιου μπαστερ κιτον
[CLS] αριστουργημα του κορυφαιο ##υ μπασ ##τερ κιτ ##ον [SEP] [PAD] [PAD] [PAD] [PAD] [PAD] [PAD] [PAD] [PAD]
max length : 128
```

Figure 9: Illustration of the tokenizer

In our dataset, we split the data into 80% training (roughly 120 thousand rows), 10% testing (15 thousand rows), and 10% validation (15 thousand rows), and also we applied a Dataloader with a batch size equal to 128. In addition, we have integrated the tokenizer in the Dataloader, so we can have in every batch the encoded input ids and the attention mask to fit our model.

As we mentioned before, the main architecture on which we based our model is the pre-trained Greek Bert, but as our task is sentiment analysis a classification with three classes, we adjust on top of our pre-trained model a Dropout Layer (with probability equal to 0.3) and a Dense Layer with three outputs. In this way, we finetune the pre-trained Greek model with our training corpus and finally classify a review in a range of 3 classes. We should mention that we base our implementation on PyTorch framework. We set the loss function to Cross-Entropy loss and the optimizer to Adam. Finally, we trained the model for 10 Epochs, and after each epoch, we monitored the training accuracy and loss and the validation accuracy and loss. We save the best model with respect to validation accuracy.

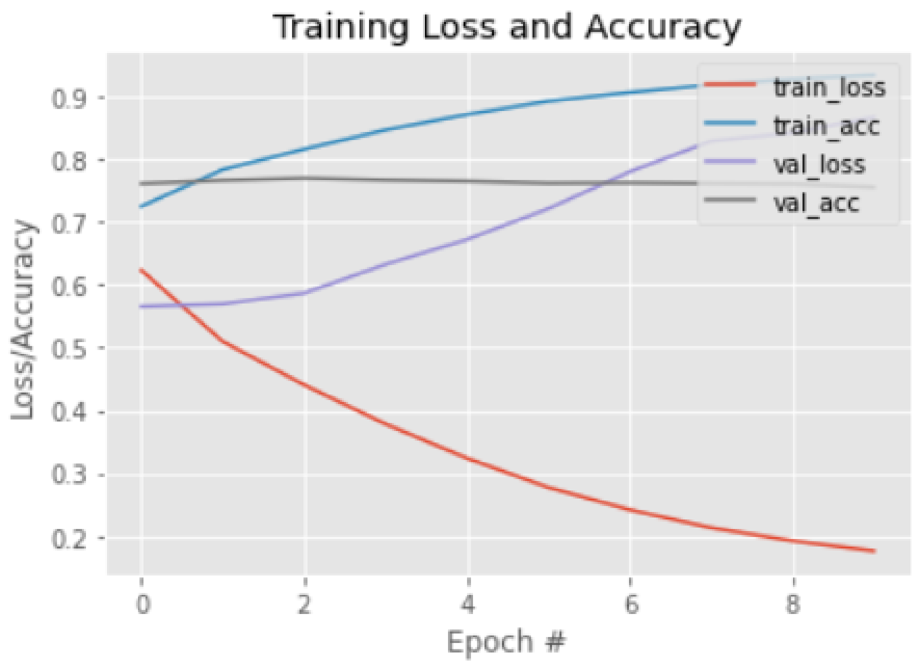

Figure 10: Evaluation metrics History of the model

In the above graph, we can notice the training loss is continuously decreasing achieving a training accuracy of 93% in the $10th$ epoch, but the validation loss is fluctuating between 0.55 and 0.85, achieving the best validation accuracy of 77% in the $2nd$ epoch.

6 Implementation of MT5

Regarding our sentiment analysis problem, we tried to integrate the MT5 model as a second approach. Google has also developed the MT5 model, a powerful multilingual model with a vocabulary of 101 languages. The MT5 is a transformer-based sequence-to-sequence model, which "models" every problem in text-to-text format, meaning that the input has a text format, and the output has a text format. The applications of the MT5 model have a great variety, such as classification, summarization, and translation. The main difference between BERT and MT5 is that MT5 uses Encodes and Decoders as the main Transformer architecture. Finally, we should mention that we utilized the smallest version of the MT5 series (google/mt5-small) as these models have very high computational costs due to their size. MT5 follows Bert's preprocessing steps, except for not using labels. Because of the text-to-text format, we should apply the sentiment values (negative, neutral, positive) to the model, not their numeric labels.

To finetune the MT5 model, which contains the Greek language, to our training corpus, we have loaded the model and its respective tokenizer. The MT5 tokenizer has a different technique to encode the tokens in contrast to the BERT tokenizer, but the procedure is the same. First, we have to set the max length of the tokens, as we have elaborated with Bert. In the figures bellow, we notice the distribution of the length of the tokens according to the MT5 tokenization technique and an illustrative example of the MT5 tokenizer.

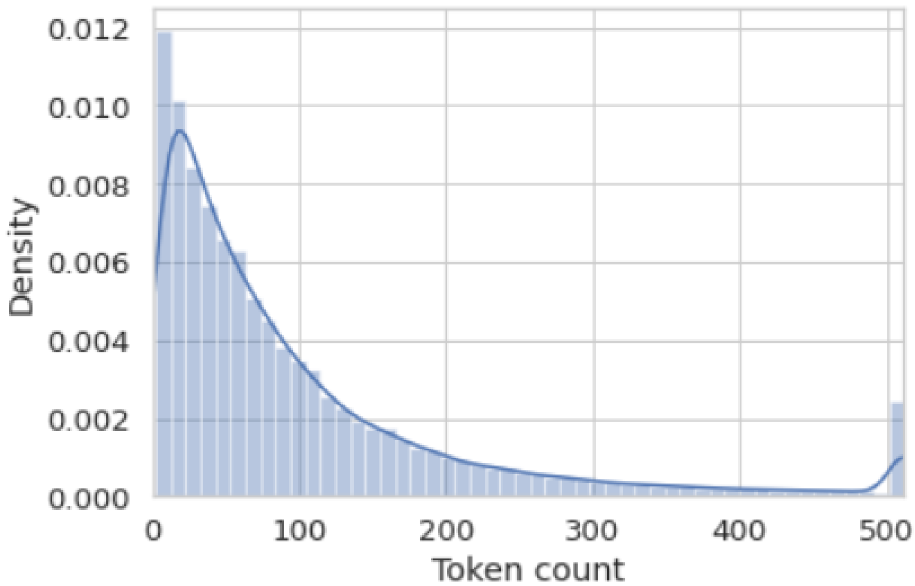

Figure 11: Distribution of length of the Tokens with MT5 tokenizer

```
ως τωρα την εχω δει δυο φορες και θα ξαναπηγαινα ευχαριστως ειναι εξαιρετικη με εκπληκτικα εφε δυνατες συγκινησεις και μια μαγικη ιστορια να διηγηθει  ιδανικη για οπο
_ ως _τ ωρα _ την _ε χω _ δει _δυ ο _φορ ες _και _θα _ξανα πη γα ινα _ευχαριστ ως _ ει ναι _ε ξαιρετικ η _με _εκ πληκτ ικα _εφ ε _ δυνατ ες _συγκ ινη σεις _και _
max length : 128

positive
_positive </s> <pad>
max length : 3
```

Figure 12: Illustration of the MT5 tokenizer

We have set the max length of tokens equal to 128 for the MT5 tokenizer, the same as the Bert tokenizer. However, a significant remark is that the MT5 tokenizer, in contrast to the BERT tokenizer, has to encode not only the input ids and attention mask but also the target input id and its attention, as shown in the example above. That occurs because the MT5 uses decoders, whereas the BERT only uses encoders. Therefore, we have set the max length of "target" tokens equal to 3.

MT5 is very agile that we do not have to modify the architecture of the pre-trained model as we have done with BERT. This way, we can apply classification or sentiment analysis and finetune the pre-trained model with our corpus. The only requirement is to tokenize the data so we can give four inputs to the model.

In the training section of the MT5 model, we have trained the model for 12 epochs and used the Cross-Entropy Loss as the loss function and Adam as the optimizer. In every epoch, we monitor the training loss and validation loss-accuracy. A very interesting part is that to have predictions, we should apply the general method of the MT5 model concerning some input ids and attention masks. Then we have to convert the ids to tokens through the MT5 tokenizer and apply the evaluation metrics. In the figure below, we visualized the training history of the model.

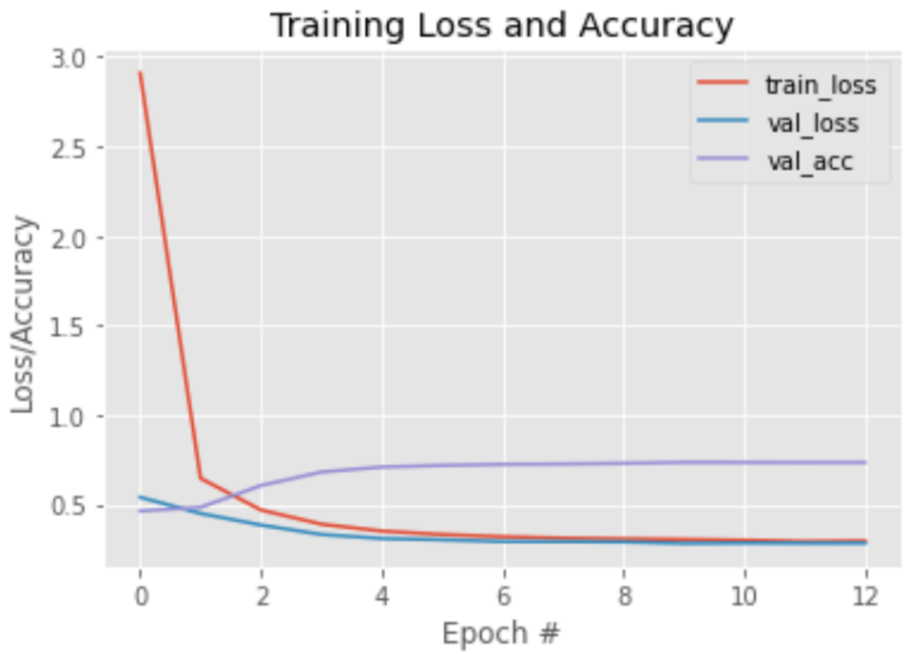

Figure 13: Evaluation metrics & History of the model

From the plot above, we should notice that the training loss as the validation loss were decreasing and also that the validation accuracy metrics was increasing across the epochs. The best validation accuracy was roughly at 73% at 9th epoch.

7 Implementation of GPT 2

Generative Pre-trained Transformer 2 (GPT-2) is an open-source artificial intelligence model created by OpenAI in February 2019. GPT-2 translates text, answers questions, summarizes passages, and can generate text output on a level that, while sometimes indistinguishable from that of humans. It is a general-purpose learner and was not trained to do specific tasks. We faced a number of problems with the tokenizer because GPT-2 based on decoders. At the end we achieve good results similar with the Bert and T5 pretrained models. One other problem is that when we decode the tokenized words many of them lose their meaning.

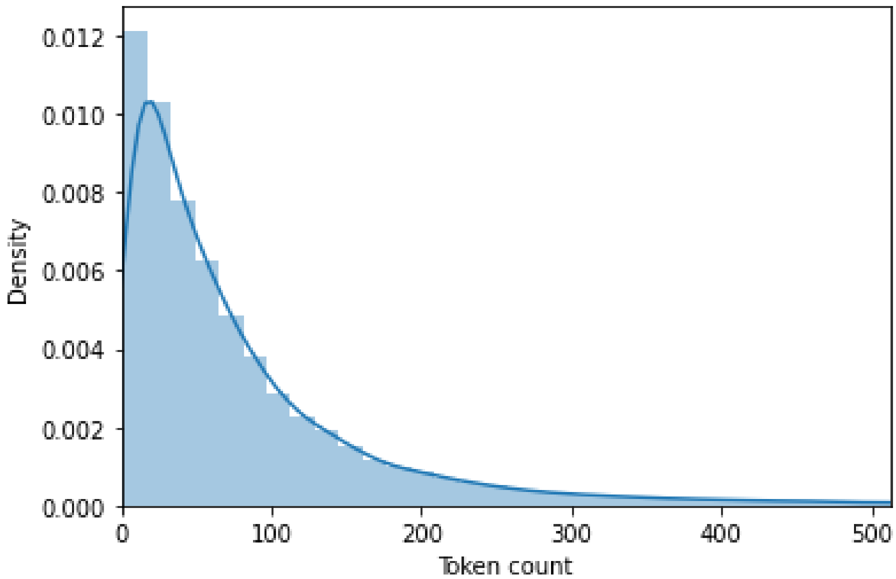

Figure 14: Distribution of length of the Tokens with MT5 tokenizer

```
ως τωρα την εχω δει δυο φορες και θα ξαναπηγαινα ευχαριστως ειναι εξαιρετικη με εκπληκτικα εφε δυνατες συγκινησεις και μια μαγικη ιστορια να διηγηθει  ιδανικη για οπο
_ ως _τ ωρα _ την _ε χω _ δει _δυ ο _φορ ες _και _θα _ξανα πη γα ινα _ευχαριστ ως _ ει ναι _ε ξαιρετικ η _με _εκ πληκτ ικα _εφ ε _ δυνατ ες _συγκ ινη σεις _και _
max length : 128

positive
_positive </s> <pad>
max length : 3
```

Figure 15: Illustration of the MT5 tokenizer

We should mention that we found a Greek paper from Aigean university that they have implement GPT-2 as a sentiment classifier on covid-19 tweets dataset and they have achieve the best results in comparison with the other transformers.

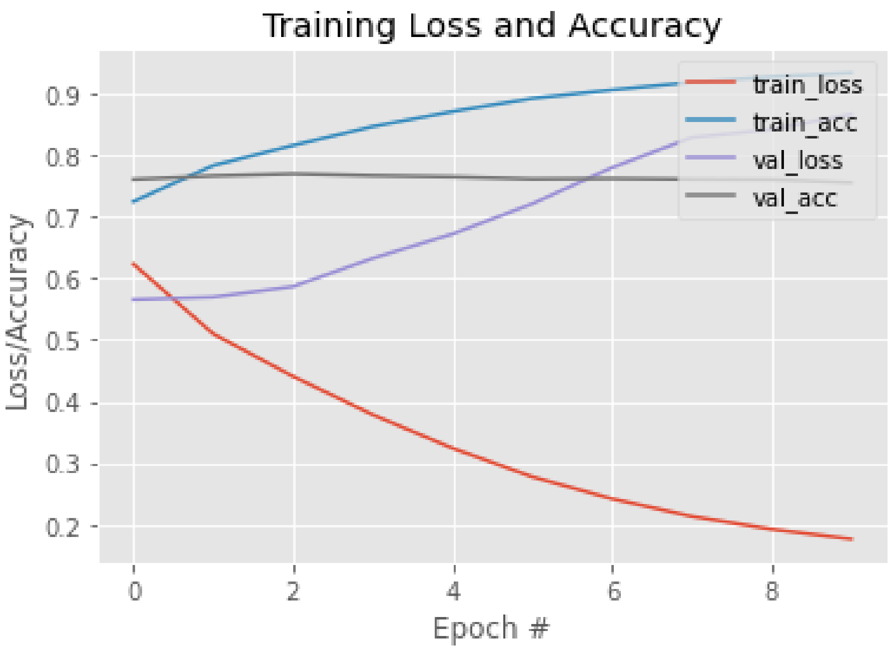

Figure 16: Evaluation metrics & History of the model

8 Implementation of our architecture

In our previous attempts for sentiment classification using the transformer models, we use the embeddings of each sentence of our dataset. This time we try to use the word embeddings of each of the 128 tokens in our sentences. We will use these embeddings as an input in our model, which has the classifier role. The model consists of five parts: the Pre-train layer Greek Bert transformer, a BiLSTM layer, an Attention layer, a CNN layer, and a Full Connected layer. The figure below shows the structure of the model. Specifically, we use the pre-trained Bert model for tokenizing our data. Each sentence has 128 tokens. As an input, we use the tokenized sentences of our dataset using special tokens like [SEP] to mark the end of a sentence or the separation between two sentences and [CLS] at the beginning of our text. This token is used for classification tasks. We sum the output of the last 4 layers of the Bert transformer to represent each word's embeddings and use them as an input in the BiLSTM. At this time, the text sentence starts to classify. The BiLSTM network can process a long text sequence and apply it after the pre-train layer, extracting the text's dependencies in the forward and backward directions. Different words have different effects on judging the sentiment polarity of the entire text. Therefore, the model introduces an attention mechanism to assign different sentiment weights to different words. Convolution is extracting the most meaningful local feature information in the input text information through a filter, reducing the data dimension, and the model can get the position invariant. After the convolutional operation is over, we perform the pooling operation. The pooling layer's role is to compress the convolutional feature vector further, which can reduce the vector dimension and computational complexity. We used the max-pooling method to keep the sentence's most important features. Finally, we connect our pooled feature vectors to a fully connected layer which can improve the robustness of the feature vector while retaining important features as much as possible. To achieve our classification task, we add a drop-out layer with 0.3 probability and a Dense layer with 3 outputs representing the probability for each class. As an optimizer, we use AdamW, and for loss function, we used cross-entropy loss. To fully complete each Epoch, the model needed more than 5 hours of training. Unfortunately, time did not permit the adoption of different architectures or the use of Lexikon.

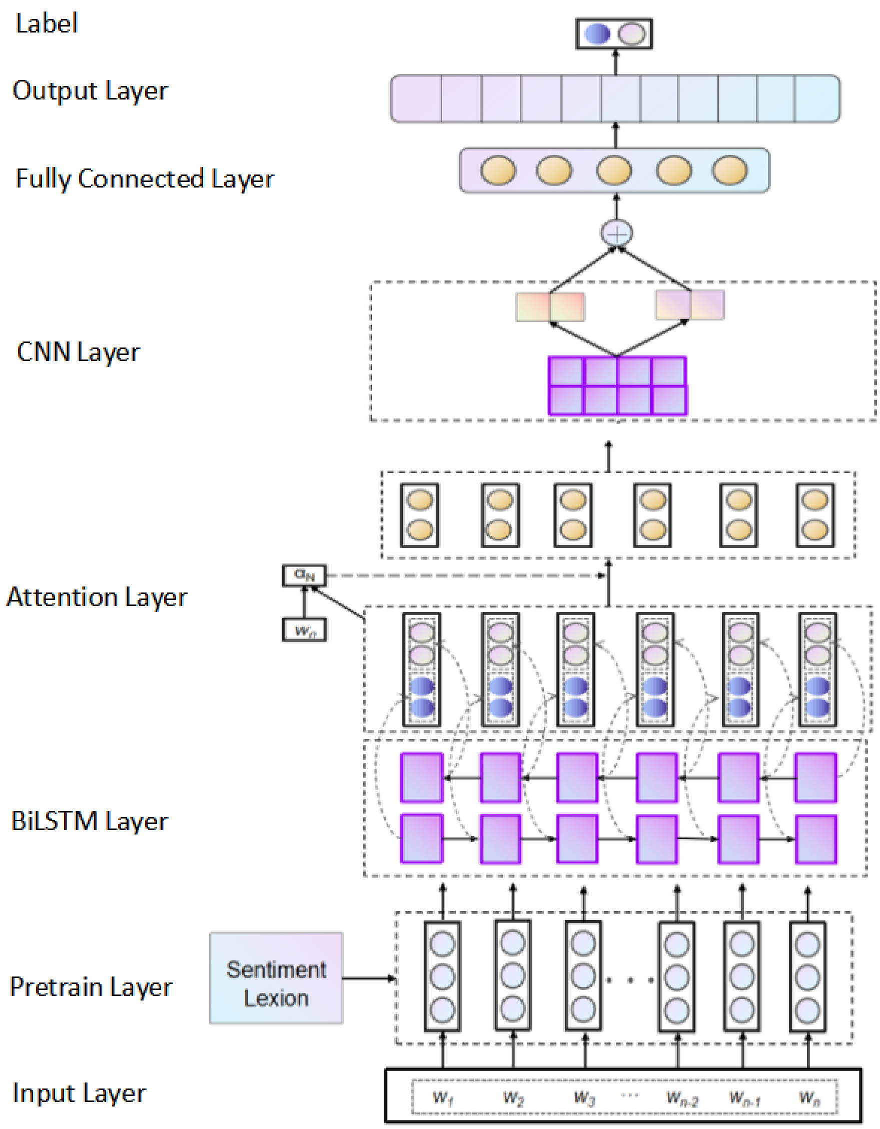

Figure 17: Our model architecture based on Weibo sentiment analysis paper

In the figure bellow we have the training results on accuracy and the output of the loss function for the training and the validation set. We observe that our model did not achieved good results. The training loss haven't change so there is a possibility that with more epochs we could see a difference.

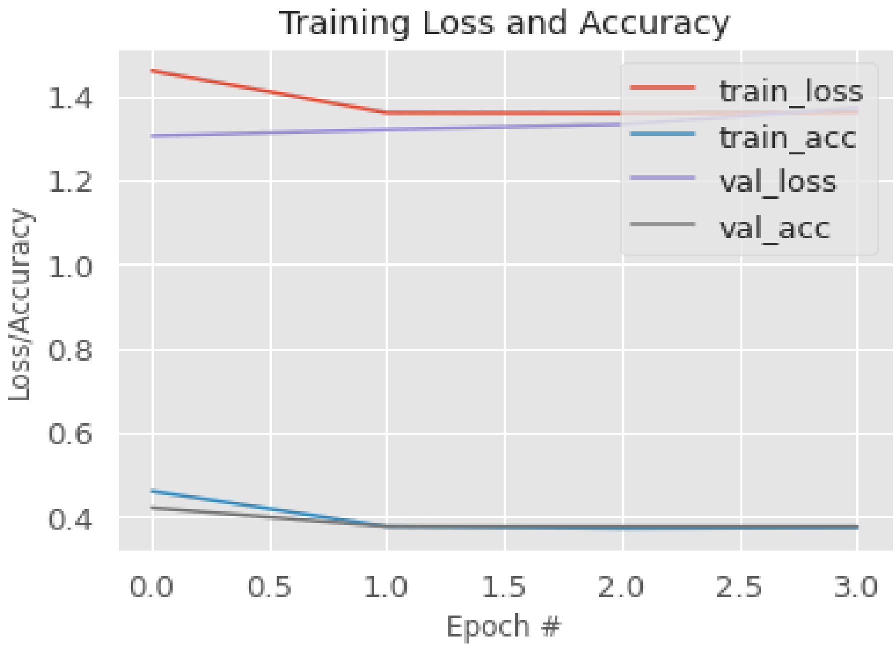

Figure 18: Evaluation metrics & History of the model

9 Evaluation of the results

To evaluate the performance of the different transformers and our proposed model, we randomly divide the dataset into ten parts, using eight of them as the training set, one as the validation set, and the remaining one part as the test set, taking the results as the evaluation result of the models.

As we all know, the model's performance is affected by the number of epochs. Therefore, we train Bert and T5 models for 10 epochs and our model for 4 epochs saving each time the results. We kept the weights that give the best validation accuracy for each model. Having picked the best, we evaluate our results on the test set, calculating the accuracy, precision, recall, and F1 score. Finally, we visualize our results with the confusion matrix. Also, we create a table that aggregates the results for each model. From the metrics below, we can easily claim that most mistakes happen between positive-neutral and negative-neutral comments with no high polarity. (We had better metrics but we lost them)

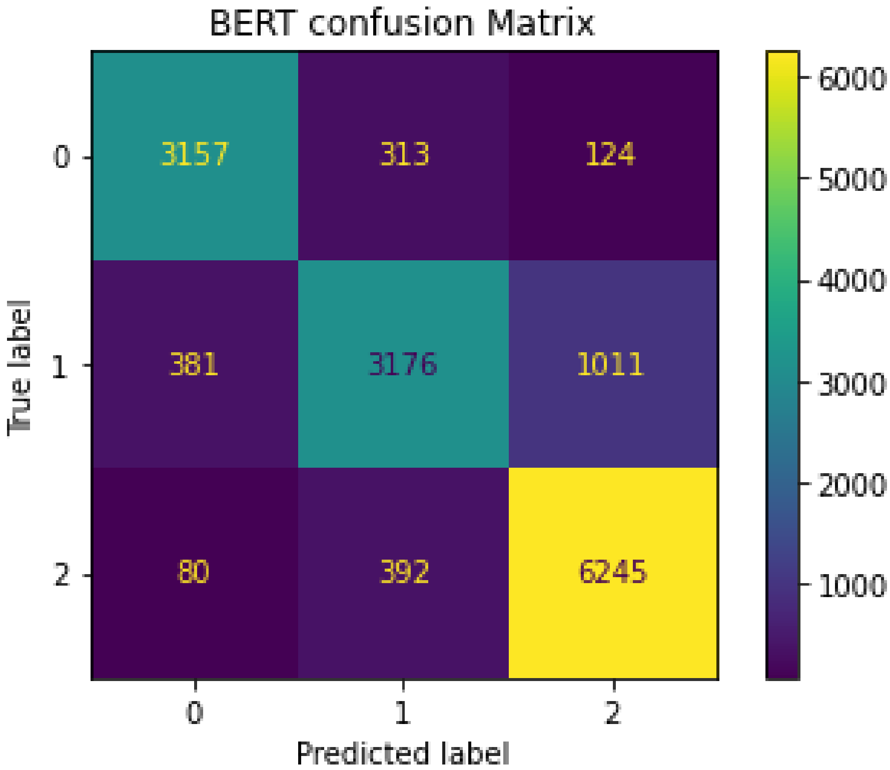

Figure 19: Confusion matrix for bert model

In the figure bellow we represent some metrics for the Bert based model.

	precision	recall	f1-score	support
negative	0.79	0.75	0.77	20
neutral	0.84	0.68	0.75	38
possitive	0.82	0.98	0.89	42
accuracy			0.82	100
macro avg	0.82	0.80	0.80	100
weighted avg	0.82	0.82	0.81	100

Figure 20: Metrics for the Bert model

In order to have a deeper understanding of our models, we create dataframes that visualize in the same frame, the sentences, the tokenization, the true and the predicted labels.

	Review	True Label	Predicted Label	Model Name	BERT_Prediction
0	πολυ καλη γρηγορη με αποψη και ρυθμο πολυ καλο καστ νομιζω οτι ο ανθρωπος πισω απο την ταινια εχει πολλα να δωσει στο συγχρονο ελληνικο κινηματογραφο νοημα χωρις δηθεν κουλτουρα	2	2	BERT	True
1	οταν ειμαστε μια φορα αυστηροι και απαιτητικοι με τις ελληνικες ταινιεστοτε πρεπει να ειμαστε 100 φορες με τις ξενες υπερπαραγωγεσαλλα η υπερχειλιζουσα αγαπη για το ξενο που ανεκαθεν μας διακατεχειμας ωθει να υπερτιμουμε τις ξενες και να ριχνουμε στον καιαδα αβιαστα και ακριτα τις ελληνικεσπρεπει ολοι να δειξουμε επιεικια σε αυτες τις προσπαθειες και οχι να εξαντλουμε την αυστηροτητα μας	1	2	BERT	False
2	μερικες φορες παραγεται το αντιθετο αποτελεσμα απ αυτο που επιθυμουμε ετσι και εδω αντι για ταινια θριλερ βλεπουμε μια προχειρα φτιαγμενη ταινια με γελοιο σεναριο που μονο γελια μπορει να προκαλεσει εγω τουλαχιστον γελασα με την ψυχη μου την συνιστω ανεπιφυλακτα αλλα για κωμωδια οχι για θριλερ	2	2	BERT	True
3	περιμενα περισσοτερα ισως μετα το οι εραστες μενουν μονο ζωντανοι και πιο πισω τα σπασμενα λουλουδια	1	1	BERT	True
4	σταδιακη κορυφωση με νοτες τρυφεροτητας που κοβει την ανασα	2	2	BERT	True
5	αισχος	0	0	BERT	True
6	ειναι μια συμπαθητικη ταινια και ειδικα για αυτο το καλοκαιρι πιστευω οτι αξιζει αρκετα περνας ευχαριστα ενα απο τα ατου τις ταινιας ειναι το δυνατο cast και οι πετυχημενες δοσεις χιουμορ νομιζω οτι ξεφευγεις λιγο απο την καθημερινοτητα και αν και ξερω οτι θα εχω αντιδραση απο το γυναικιο κοινο ωφειλω να ομολογησω οτι καθε ανδρας ονειρευεται μια τετοια γυναικα ρομποτ και μαλιστα αν γινεται να κανω και εγω μια παραγγελια σε νικολ κιντμαν please god do something τωρα θα ψαξω και θα βρω την πρωτοτυπη ταινια ειμαι περιεργος και θελω να κανω συγκρισεις να ενημερωσω τελος οτι η ταινια ειναι μαυρη κωμωδιασας αφηνω τωρα γιατι η γυναικα μου θελει να παμε super market kai δυστυχως δεν εχουμε ρομποτκλαψ	1	1	BERT	True
7	ειχα καιρο ηταν να απολαυσω ενα εργο σαν αυτο οποιος θελειναδει καλες ερμηνειεςφωτογραφιαμουσικη και να την θυματε για καιρο αςπαει να την δει ¶σε που θα νοιωσει ενταση και αγωνια στο επακροφιλοι των αμερικανικων θριλερ μακρια	2	2	BERT	True
8	ωραια ταινια αν και τη βρηκα καπως πιο παιδιαστικη απο αλλες του ιδιου ειδους οπως ο πχ harry potter κα παρ ολ αυτα πρεπει να παραδεχτω πως η φωτογραφια της ειναι μαγευτικη και οι ηθοποιοι πολυ καλοι	1	1	BERT	True
9	εξυπνη διασκεδαστικη ταινιουλα προσφερεται για χαβαλε	1	1	BERT	True

Figure 21: Bert Dataframe predictions

32	ποσο σπουδαιος ειναι μαζι με τον fincher και τον singer οι καλυτεροι συγχρονοι σκηνοθετες περιμενω και το illusionist που επισης πρεπει να αξιζει	positive	positive	positive	MT5	True
33	εξαιρετικη φωτογραφια	positive	positive	positive	MT5	True
34	ριξε στο μπλεντερ μια ιστορια με το χολιγουντ να βλογαει τα γενια του ακατασχετα πρωταγωνιστη μιουζικαλ εναν παντελως αφωνο και λαβ στορι της ξεπετας και βουαλα το αριστουργημα μιλαμε για τρελο προβληματισμο εγιναν και οι δυο πλουσιοι και διασημοι και αυτο δεν το αντεξε η σχεση τους	neutral	negative	negative	MT5	False
35	απιστευτα βαρετη	negative	negative	negative	MT5	True
36	ολη η σκληροτητα η τρυφεροτητα η θλιψη και η ποιηση που συναντουμε στον κοσμο ειλικρινες σπαρακτικο και βαθια ρομαντικο παραμυθι ενηλικων με ηρωες που ειναι λιγοτερο ακραιοι απο οσο δειχνουν η μοναξια η αδυναμια να προσεγγισουμε να εκφραστουμε ακομη και να φωναξουμε ποσο πραγματικα αναγκη εχουμε απο καποιον αλλον ανθρωπο για να μας β	positive	positive	positive	MT5	True
37	μετα τις εμπορικα επιτυχημενες αλλα και καταπληκτικες μεταφορες των xmen και spiderman απο τους bryan singer και sam raimi αντιστοιχα η marvel και η κινηματογραφικη βιομηχανια του χολυγουντ φαινεται αποφασισαν οτι βρηκαν την κοτα με τα χρυσα αυγα αποτελεσμα ενας καταλογος απογοητευτικων μην πω αισχρων κινηματογραφικων εκδοσεων μερικων απο τους πλεον cult τιτλους κομικ οπως dare	negative	neutral	neutral	MT5	False
38	η ταινια συνιστα μια συγκλονιστικη αποκαλυψη ως προς το που μπορει να φτασει ο ανθρωπος οταν ερχεται αντιμετωπος με τον ιδιο του τον εαυτοτα ανθρωπινα ορια ειναι παντοτε επεκτασιμα στο παιχνιδι της επιβιωσης και ο τομ χανκς το αποδιδει αυτο με περισση υποκριτικη μαεστριατο τριτο οσκαρ του αξιζει	positive	positive	positive	MT5	True
39	χρρρρρρρρρρρ	negative	negative	negative	MT5	True
40	με το 1ο trailer τοτε ειχα ενθουσιαστει η ταινια ομως ηταν απλοικη παιδικη και αδιαφορη η 1η ταινια ηθελα πολυ να μου αρεσει γιατι μου αρεσε ο κοσμος του θορ ηταν ομως μετρια γιατι ελειπε η δραση η 2η ταινια ηταν ναι μεν λιγο απλοικη αλλα συνολικα καλυτερη και ευχαριστη να την βλεπεις στην 3η ομως το παρακαναν με την αυτοκριτικ	negative	neutral	neutral	MT5	False
41	δεν μπορω να ειμαι αντικειμενικος με την jennifer lawrence την λατρευω ως ηθοποιο και ως γυναικα αλλα δεν ηταν μονο αυτη ο λογος που μου αρεσε πολυ η ταινια η διαφορετικη προσεγγιση σε θεματα μανιοκαταθλιψης και νευρωτικης συμπεριφορας και το συνολο των ερμηνειων στα του οσκαρ τωρα οι 8 υποψηφιοτητες ειναι υπερβολικες καθως υπηρχαν αξιολογοτερες επιλογες ας	positive	positive	positive	MT5	True

Figure 22: T5 Dataframe predictions

0	μερικες φορες παραγεται το αντιθετο αποτελεσμα...	[CLS] μερικες φορες παραγεται το αντιθετο αποτ...	2	2	BERT	True	μερικες φορες παραγεται το αντιθετο αποτελεσμα...	positive	negative	<pad> negative</s>	MT5	False
1	η κωμωδια της εβδομαδας θα πεθανεις στα γελια ...	[CLS] η κωμωδια της εβδομαδας θα πεθανεις στα ...	0	0	BERT	True	η κωμωδια της εβδομαδας θα πεθανεις στα γελια ...	negative	positive	<pad> positive</s> <pad>	MT5	False
2	καμια σχεση με το νουμερο 1 ισως και αλλα ηθη ...	[CLS] καμια σχεση με το νουμερο 1 ισως και αλλ...	0	1	BERT	False	καμια σχεση με το νουμερο 1 ισως και αλλα ηθη ...	negative	negative	<pad> negative</s>	MT5	True
3	μια αιρετικη αλληγορια για το χριστιανισμο ειδ...	[CLS] μια αιρετικη αλληγορια για το χριστιανισ...	1	1	BERT	True	μια αιρετικη αλληγορια για το χριστιανισμο ειδ...	neutral	negative	<pad> negative</s>	MT5	False
4	οπως παντατο ζευγαρι μπαρτον ντεπ απογοητευει	[CLS] οπως παντατο ζευγαρι μπαρτον ντεπ απογοη...	0	1	BERT	False	οπως παντατο ζευγαρι μπαρτον ντεπ απογοητευει<...	negative	negative	<pad> negative</s>	MT5	True
5	δεν γνωριζω αν ακολουθει τα γεγονοτα με ιστορι...	[CLS] δεν γνωριζω αν ακολουθει τα γεγονοτα με ...	1	2	BERT	False	δεν γνωριζω αν ακολουθει τα γεγονοτα με ιστορι...	neutral	neutral	<pad> neutral</s> <pad>	MT5	True
6	μετα τις εμπορικα επιτυχημενες αλλα και καταπλ...	[CLS] μετα τις εμπορικα επιτυχημενες αλλα και ...	0	0	BERT	True	μετα τις εμπορικα επιτυχημενες αλλα και καταπλ...	negative	neutral	<pad> neutral</s> <pad>	MT5	False
7	με το 1ο trailer τοτε ειχα ενθουσιαστει η ταιν...	[CLS] με το 1ο trailer τοτε ειχα ενθουσιαστει ...	0	0	BERT	True	με το 1ο trailer τοτε ειχα ενθουσιαστει η ταιν...	negative	neutral	<pad> neutral</s> <pad>	MT5	False
8	για να σκανε οι εκσυγχρονιστεςο σοσιαλιστικος ...	[CLS] για να σκανε οι εκσυγχρονιστεςο σοσιαλισ...	1	1	BERT	True	για να σκανε οι εκσυγχρονιστεςο σοσιαλιστικος ...	neutral	positive	<pad> positive</s> <pad>	MT5	False
9	ο ηρωας του γιμου εχει πολυ καλυτερες σκηνες ...	[CLS] ο ηρωας του γιμου εχει πολυ καλυτερες σκ...	1	1	BERT	True	ο ηρωας του γιμου εχει πολυ καλυτερες σκηνες δ...	neutral	positive	<pad> positive</s> <pad>	MT5	False
10	τη βρηκα πολυ ενδιαφερουσακαταδεικνυει το ποσο...	[CLS] τη βρηκα πολυ ενδιαφερουσακαταδεικνυει τ...	2	1	BERT	False	τη βρηκα πολυ ενδιαφερουσακαταδεικνυει το ποσο...	positive	positive	<pad> positive</s> <pad>	MT5	True
11	πρωτη φορα γραφω κριτικη καποιας ταινιας στο α...	[CLS] πρωτη φορα γραφω κριτικη καποιας ταινιας...	0	0	BERT	True	πρωτη φορα γραφω κριτικη καποιας ταινιας στο α...	negative	positive	<pad> positive</s> <pad>	MT5	False
12	απλη και ειλικρινης ταινια θυμιζει εκεινα τα β...	[CLS] απλη και ειλικρινης ταινια θυμιζει εκειν...	1	1	BERT	True	απλη και ειλικρινης ταινια θυμιζει εκεινα τα β...	neutral	positive	<pad> positive</s> <pad>	MT5	False
13	δε μπορω να καταλαβω τις κριτικες των ειδικων ...	[CLS] δε μπορω να καταλαβω τις κριτικες των ει...	0	0	BERT	True	δε μπορω να καταλαβω τις κριτικες των ειδικων ...	negative	neutral	<pad> neutral</s> <pad>	MT5	False
14	ντροπη στη κριτικη του αθηνοραματος αισχος	[CLS] ντροπη στη κριτικη του αθηνοραματος αισχ...	2	2	BERT	True	ντροπη στη κριτικη του αθηνοραματος αισχος</s>...	positive	negative	<pad> negative</s>	MT5	False
15	αρκετα καλη ταινια1 αρχισε λιγοαθλιατι γινοταν...	[CLS] αρκετα καλη ταινια1 αρχισε λιγοαθλιατι γ...	1	2	BERT	False	αρκετα καλη ταινια1 αρχισε λιγοαθλιατι γινοταν...	neutral	neutral	<pad> neutral</s> <pad>	MT5	True
16	καλα εφφε πολυ καλος κακος αυτα απο κει και πε...	[CLS] καλα εφφε πολυ καλος κακος αυτα απο κει ...	1	1	BERT	True	καλα εφφε πολυ καλος κακος αυτα απο κει και πε...	neutral	negative	<pad> negative</s>	MT5	False

Figure 23: Bert and T5 combined

We can observe that the models often classify the sentence incorrectly because the commentator made some typos. We think those typos create tokens that the tokenizers cannot "understand". As a result, the transformer creates random embeddings. That randomness makes the models make wrong decisions, especially when there is no high polarity; for example, positive sentences are treated as neutral.

10 Conclusion

At present, deep learning models are popular in the field of sentiment analysis. Transformers like Bert, GPT-2, or GPT-3 and T5 are the base of the models with higher accuracy. However, for the Greek language, there is not any public available sentiment analysis model. In this project, we tried to implement different pre-trained transformers and use them to classify the Athinorama dataset and propose our model. The highest accuracy we achieved using the pre-trained Bert transformers was close to 77%. Those results are not very good if we compare them with other models in other languages. However, we must understand that a much larger corpus has been used for training in languages like English. In our research, we have also found a paper that has trained a GPT-2 transformer to perform sentiment analysis classification for a greek tweets dataset, but despite our efforts. One problem we faced was tokenizing the words from the pre-trained transformers. The Bert tokenizer produces the most sensible results. We used Bert embeddings to train our model, but we didn't manage to have great results. Bert model consumes enormous hardware resources and requires a considerable corpus for training. In the given time, we could experiment with many models and cherry-pick the best. For future work, we will try improve GPT-2 model.