



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ  
ΠΟΛΥΤΕΧΝΕΙΟ**

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ  
ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ**

*Ακαδημαϊκό έτος 2022-2023*

## **Επεξεργασία Φωνής και Φυσικής Γλώσσας**

Αναφορά της προπαρασκευής της 3ης  
εργαστηριακής άσκησης

Γκοτζιάς Γεώργιος - 03119047

Δημήτρης Δαμιανός - 03119825

## Εισαγωγή:

Για τις ανάγκες αυτής της άσκησης καλούμαστε να αναπτύξουμε ένα βαθύ νευρωνικό δίκτυο, το οποίο, με τη βοήθεια προ-εκπαιδευμένων LLMs, θα μπορεί να κατηγοριοποιήσει προτάσεις ανάλογα με το συναίσθημα (sentiment analysis).

### 1. Προ-επεξεργασία Δεδομένων:

Σε αυτό το βήμα θα μετατρέψουμε τα δεδομένα μας σε μορφή κατάλληλη ώστε να μπορεί να τα χρησιμοποιήσει το δίκτυό μας και να κάνει χρήση των ήδη γνωστών embeddings. Συγκεκριμένα:

**1.1)** Με την βοήθεια της `LabelEncoder()` απεικονίζουμε τα “positive” σε 1 και τα “negative” σε 0, για το MR dataset:

```
Before label encoding: ['positive', 'positive', 'positive', 'positive', 'positive']  
After label encoding: [1 1 1 1 1]
```

Στο Semeval2017A η κωδικοποίηση είναι: ‘negative’ = 0, ‘neutral’=1, ‘positive’=2.

**1.2)** Με την βοήθεια απλού διαχωρισμού στο κενό ως κριτήριο ανάλυσης, έχουμε τα παρακάτω tokens για το MR:

```
Tokenized input: ['the', 'rock', 'is', 'destined', 'to',  
'be', 'the', '21st', 'century's', 'new', 'conan',  
'and', 'that', 'he's', 'going', 'to', 'make', 'a', 'sp  
lash', 'even', 'greater', 'than', 'arnold', 'schwarzenegg  
er', 'jean-claud', 'van', 'damme', 'or', 'steven', 'se  
gal', '.']
```

Για το Semeval2017A χρησιμοποιήσαμε τον `TweetTokenizer()`.

**1.3)** Με την βοήθεια του word2idx πετυχαίνουμε την παρακάτω κωδικοποίηση των προτάσεων:

```

Item 0:
example=[32611 84 37 24030 2 118 1590 5636 78 17802 5677 111
7004 8 31149 5848 11 1814 3 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0]
label=1
length=19

Item 1:
example=[ 8 320 4 8132 34204 11216 3 28223 908 96
1717 20 27 1396 40238 5 1 5563 13 5
31 474 15 1182 5 34 5 3009 3 400001
8 4882 5 33093 3 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0]
label=1
length=35

Item 2:
example=[400001 30 3218 2 16455 400001 18 8 1728 1699
6 78 7873 4040 6 3469 3 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0]
label=1
length=17

Item 3:
example=[ 4038 26 135 4545 400001 22 8 2008 2 7675
12535 3 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0]
label=1
length=12

Item 4:
example=[ 1 2114 5805 349 370 32 956 8 320 13 49 87
13409 4467 20 863 2 11559 6 1482 2519 118 85 69
320 7 398 300 3 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0
0 0]
label=1
length=29

```

## 2. Μοντέλο:

Σε αυτό το βήμα θα δημιουργήσουμε το Βαθύ Νευρωνικό Δίκτυο που θα βρίσκει τα representations της πρότασης εισόδου με την βοήθεια ενός embeddings layer, και στην συνέχεια θα κατηγοριοποιεί την είσοδο ανάλογα με το sentiment analysis.

### 2.1) Embedding Layer

Το DNN μας έχει 4 επίπεδα:

- input
- embedding layer
- hidden layer
- output

Τα βάρη του δικτύου μας αρχικοποιούνται σε τιμές από προ-εκπαιδευμένα word embeddings, ενώ δίνει την δυνατότητα στον χρήστη να “παγώνει” και να “ξεπαγώνει” το embedding layer, ώστε τα βάρη του να αλλάζουν στην εκπαίδευση ή όχι.

### **Ερώτηση 1:**

Χρησιμοποιούμε τα προ-εκπαιδευμένα word embeddings διότι θέλουμε να διευκολύνουμε την εκπαίδευση του νευρωνικού. Συγκεκριμένα, η τυχαία αρχικοποίηση των βαρών μπορεί να οδηγήσει σε σύγκλιση σε τοπικό και όχι ολικό ελάχιστο, το οποίο αποφεύγεται όταν χρησιμοποιούμε προ-εκπαιδευμένα βάρη.

Ο λόγος που συμβαίνει αυτό είναι ότι χάρη στα προ-εκπαιδευμένα embeddings, έχουμε ήδη μια σημασιολογική αναπαράσταση των λέξεων, με αποτέλεσμα η σύγκλιση να είναι ευκολότερη.

### **Ερώτηση 2:**

Παγώνουμε το embeddings layer και δεν επιτρέπουμε την εκπαίδευση του για να αποφύγουμε το overfitting, ενώ παράλληλα κερδίζουμε πολύ σε υπολογιστικό χρόνο.

## **2.2) Output Layer**

Στο τελικό αυτό κομμάτι του νευρωνικού γίνεται η κατηγοριοποίηση της εισόδου. Συγκεκριμένα, χρησιμοποιήσαμε ένα κρυφό layer με ReLU ενεργοποίηση, και στο τέλος το τελευταίο layer που προβάλλει τις επεξεργασμένες αναπαραστάσεις στον χώρο των κλάσεων.

### **Ερώτηση 3:**

Χρησιμοποιούμε μη-γραμμικές συναρτήσεις διότι με αυτό τον τρόπο καταφέρνουμε να πετύχουμε έναν μη-γραμμικό διαχωρισμό των δεδομένων μετά την εκπαίδευση του νευρωνικού.

Αντίθετα, ο διαχωρισμός που θα πετύχουμε με την χρήση γραμμικών μετασχηματισμών είναι γραμμικός. Κάτι τέτοιο δεν είναι ωφέλιμο, καθώς συνήθως τα δεδομένα μας δεν διαχωρίζονται γραμμικά.

## **2.3) Forward Pass**

Στο σημείο αυτό υλοποιούμε το forward propagation.

Η είσοδος του μοντέλου είναι mini batches διανυσμάτων εισόδων, και η έξοδος προκύπτει ως εξής:

- Βρίσκουμε το word representation της λέξης μέσα από το embedding layer.
- Υπολογίζουμε το συνολικό representation της πρότασης ως το μέσο όρο των embeddings κάθε λέξης της.
- Εφαρμόζουμε μη-γραμμικό μετασχηματισμό και κάνουμε την κατηγοριοποίηση, προβάλλοντας τις αναπαραστάσεις στον χώρο των κλάσεων.

### **Ερώτημα 4-5:**

Οι αναπαραστάσεις των προτάσεων προκύπτουν ως μέσοι όροι/ κέντρα βάρους των αναπαραστάσεων των λέξεων.

Αυτή η αναπαράσταση ουσιαστικά δίνει ισόβαρη προσοχή σε όλες τις λέξεις, θεωρώντας ότι το συνολικό νόημα είναι ο μέσος όρος των επιμέρους εννοιών.

Η συγκεκριμένη προσέγγιση δεν λαμβάνει υπόψη την σειρά των λέξεων, ούτε συντακτικούς κανόνες και σημεία στίξης, τα οποία συνήθως είναι αναγκαία για την κατανόηση του νοήματος μια πρότασης.

Συνεπώς είναι μια ατελής προσέγγιση.

### **3. Εκπαίδευση**

#### **3.1) Φόρτωση Παραδειγμάτων:**

Σε αυτό το βήμα χρησιμοποιούμε την συνάρτηση `Dataloader()` ώστε να χωρίσουμε τα δεδομένα μας σε mini-batches για την εκπαίδευση.

#### **Ερώτηση 6:**

Με τα μικρά mini-batches έχουμε γρηγορότερη σύγκλιση, καθώς έχουμε ενημέρωση των βαρών πιο συχνά. Όμως πολύ μικρά μεγέθη των mini-batches μπορεί να προσθέτουν θόρυβο στο τελικό αποτέλεσμα, και δεν είναι επιθυμητά.

Με τα μεγάλα mini-batches αποφεύγουμε τον θόρυβο αλλά υπάρχει ο κίνδυνος να συγκλίνουμε σε τοπικό αντί για ολικό ελάχιστο, συνεπώς να έχουμε ένα κακά εκπαιδευμένο δίκτυο.

#### **Ερώτηση 7:**

Το shuffling πριν την εκπαίδευση είναι χρήσιμο καθώς εξασφαλίζει ότι το δίκτυο δεν θα μάθει την σειρά εμφάνισης των δεδομένων, ενώ βοηθάει στην σύγκλιση στο τοπικό ελάχιστο.

#### **3.2) Βελτιστοποίηση:**

Σε αυτό το σημείο θα υλοποιήσουμε την βελτιστοποίηση του μοντέλου, σύμφωνα με τις εξής παραμέτρους:

- Το κριτήριο σφάλματος θα είναι `BCEWithLogitsLoss` για κατηγοριοποίηση 2 κλάσεων, αλλιώς θα χρησιμοποιήσουμε `CrossEntropyLoss`.
- Ως παραμέτρους παίρνουμε τις παραμέτρους του δικτύου.
- Ως optimizer επιλέξαμε των Adam.

### 3.3) Εκπαίδευση

Σε αυτό το σημείο εκπαιδεύουμε το νευρωνικό μας για 10 εποχές. Παρακάτω φαίνεται το loss που έχουμε στο τέλος κάθε εποχής:

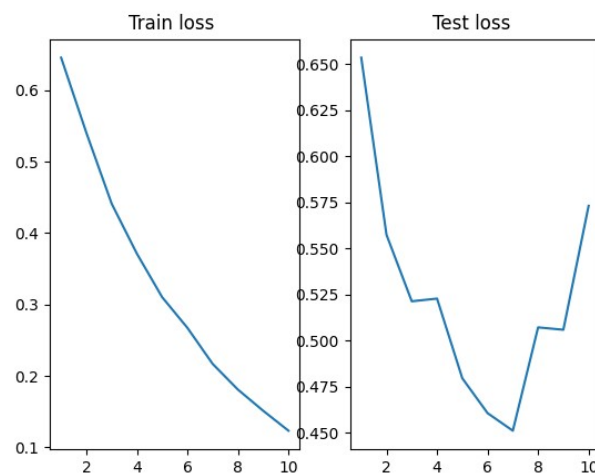
```
[=====] ...Epoch 1, Loss: 0.6286  
[=====] ...Epoch 2, Loss: 0.5937  
[=====] ...Epoch 3, Loss: 0.5610  
[=====] ...Epoch 4, Loss: 0.1760  
[=====] ...Epoch 5, Loss: 0.2679  
[=====] ...Epoch 6, Loss: 0.2977  
[=====] ...Epoch 7, Loss: 0.1532  
[=====] ...Epoch 8, Loss: 0.1436  
[=====] ...Epoch 9, Loss: 0.2037  
[=====] ...Epoch 10, Loss: 0.2204
```

### 3.4) Αξιολόγηση

Παρακάτω βλέπουμε τα scores των μετρικών accuracy, F1 score και recall για το MR dataset:

```
Training set  
accuracy: 0.7499  
recall: 0.7499  
f1-score: 0.7488421204999671  
  
Test set  
accuracy: 0.6963746223564955  
recall: 0.6963746223564955  
f1-score: 0.6944159288986875
```

Παρακάτω βλέπουμε τις καμπύλες του loss για τα δύο datasets συναρτήσει των εποχών εκπαίδευσης:



Παρατηρούμε ότι το σφάλμα μειώνεται με το πέρασμα των εποχών, όμως στο test set έχουμε ένα ξαφνικό spike μετά την 7η εποχή.

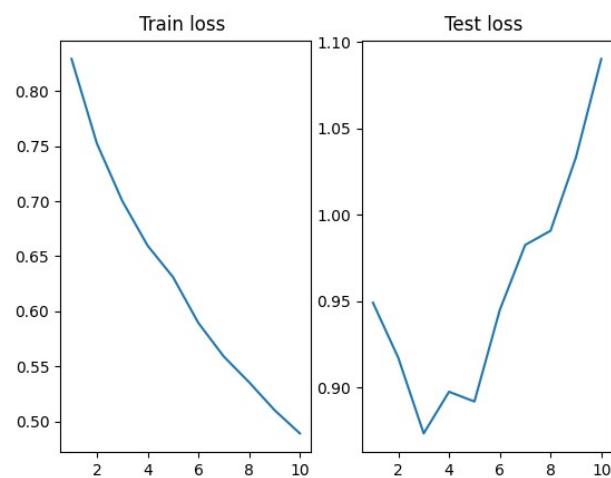
Αυτό ίσως είναι αποτέλεσμα overfitting, καθώς εδώ δεν κρατήσαμε παγωμένα τα βάρη.

Για το Semeval2017A έχουμε τα εξής αποτελέσματα:

```
Training set
accuracy: 0.6202945329836594
recall: 0.5469830964168646
f1-score: 0.5590673799073064

Test set
accuracy: 0.561055030934549
recall: 0.5427016457917943
f1-score: 0.5369813146793825
```

Και οι καμπύλες:



Παρατηρούμε ότι σε αυτό το dataset έχουμε πάρα πολύ μεγάλο test loss, το οποίο πιθανό να είναι αποτέλεσμα overfit.



## 4. Κατηγοριοποίηση με χρήση LLMs

Σε αυτό το μέρος θα χρησιμοποιήσουμε ένα LLM (στην περίπτωση μας το chatGPT) για να κάνουμε αναγνώριση συναισθήματος.

Συγκεκριμένα θα δοκιμάσουμε 20 προτάσεις από κάθε κατηγορία του MR dataset και θα μετρήσουμε τα αποτελέσματα του.

Για τις πρώτες 20 της κατηγορίας θετικού συναισθήματος έχουμε τα παρακάτω αποτελέσματα:

1. Positive sentiment: The phrase expresses excitement and anticipation for "the rock" as the new "Conan" and predicts that he will surpass the achievements of Arnold Schwarzenegger, Jean-Claude Van Damme, and Steven Seagal.
  2. Positive sentiment: The phrase praises the "gorgeously elaborate continuation" of "The Lord of the Rings" trilogy and emphasizes that words cannot adequately describe the expanded vision of director Peter Jackson.
  3. Mixed sentiment: The phrase suggests that the biopic is effective but tepid, indicating a combination of positive and negative sentiment.
  4. Positive sentiment: The phrase suggests that "Wasabi" is a good movie to have fun and enjoy.
  5. Positive sentiment: The phrase describes the film as rare and honest, providing great insight into the neurotic mindset of comics.
  6. Positive sentiment: The phrase suggests that the film offers a rare combination of entertainment and education.
  7. Negative sentiment: The phrase implies that the picture shows the negative consequences of good intentions.
  8. Mixed sentiment: The phrase acknowledges the cleverness of the screenplay but also expresses a desire to hate it, though ultimately recognizing that it succeeds.
  9. Positive sentiment: The phrase describes "Take Care of My Cat" as a refreshingly different slice of Asian cinema.
  10. Positive sentiment: The phrase suggests that the film is worth seeing and discussing, praising its uniqueness.
  11. Positive sentiment: The phrase expresses surprise at the low-key quality and genuine tenderness of "Wisegirls."
  12. Positive sentiment: The phrase suggests that "Wendigo" fulfills the reasons why people go to the cinema, feeding the eye, heart, and mind.
  13. Positive sentiment: The phrase describes the movie as one of the greatest family-oriented, fantasy-adventure movies ever.
  14. Positive sentiment: The phrase suggests that the movie ponders the importance of stories and why we need them.
  15. Positive sentiment: The phrase describes the documentary as illuminating, despite being overly talkative.
  16. Positive sentiment: The phrase describes the movie as a masterpiece that took four years to make.
  17. Positive sentiment: The phrase suggests that the movie's beauty is ripe and enrapturing, tempting those who want to explore its mysteries.
  18. Positive sentiment: The phrase suggests that the movie offers a breath of fresh air and true sophistication.
- Overall, most of the phrases express positive sentiment, with a few indicating mixed sentiment and one phrase with a negative sentiment.


Βλέπουμε ότι πετυχαίνει 15/18 ενώ δίνει 2 mixed sentiment και 1 negative.

Για την 3η πρόταση, το chatGPT μας αναφέρει ότι λόγω της λέξης “terid” έδωσε αυτή την αξιολόγηση, συνεπώς έδωσε μεγάλη βαρύτητα σε αυτή την λέξη.

Για την 7η πρόταση, μας αναφέρει ότι έδωσε αρνητική αξιολόγηση επειδή “μιλάει για τις αρνητικές συνέπειες των καλών προθέσεων”. Διαβάζοντας την φράση καταλαβαίνουμε ότι ο χρήστης που έκανε αυτό το σχόλιο θεωρεί ως κάτι καλό το γεγονός ότι η ταινία μπορεί να απεικονίσει κάτι τέτοιο. Φαίνεται ότι το chatGPT δίνει σημασία στις λέξεις καθεαυτές και όχι στο συνολικό νόημα της πρότασης σε αυτό το παράδειγμα.

Για την 8η πρόταση, μας αναφέρει ότι η πρόταση ναι μεν έχει θετικά σχόλια (π.χ. cleverness) όμως αναφέρει και αρνητικά στοιχεία (π.χ. η διάθεση να την μισήσει). Εδώ βλέπουμε ότι πάλι ότι δίνεται έμφαση στις λέξεις ξεχωριστά αντί στην συνολική πρόταση.

Θα δοκιμάσουμε τώρα τις 20 προτάσεις της κατηγορίας αρνητικού αισθήματος:

-  Here's the sentiment classification for each phrase:
1. Negative sentiment: The phrase describes something as simplistic, silly, and tedious.
  2. Negative sentiment: The phrase suggests that the content is laddish, juvenile, and only enjoyable for teenage boys.
  3. Negative sentiment: The phrase criticizes the work as exploitative and lacking depth or sophistication, making it difficult to watch due to its graphic treatment of crimes.
  4. Negative sentiment: The phrase implies that the filmmaker missed the opportunity for a pathological study and instead focuses on skewed melodrama and circumstantial situations.
  5. Negative sentiment: The phrase describes the work as visually flashy but lacking narrative clarity and emotional depth, emphasizing style and mystification over substance.
  6. Negative sentiment: The phrase criticizes the story as unoriginal and recycled multiple times, lacking originality or freshness.
  7. Negative sentiment: The phrase acknowledges the movie's bravado in revisiting a stale concept but criticizes it for lacking any other noteworthy aspects.
  8. Negative sentiment: The phrase characterizes the work as sour rather than farcical, suggesting a lack of humor or comedic quality.
  9. Negative sentiment: The phrase criticizes the script as hackneyed, which does a disservice to both the story and the actors.
  10. Negative sentiment: The phrase suggests that the work fails to humanize its subject and falls short despite its relatively gore-free approach to serial murders.
  11. Negative sentiment: The phrase describes the work as a sentimental mess that lacks authenticity.
  12. Neutral sentiment: The phrase acknowledges that the performances are engaging but suggests that the work would have been better suited for a one-hour TV documentary.
  13. Neutral sentiment: The phrase expresses that the work is interesting but not compelling.
  14. Negative sentiment: The phrase suggests that the work falls short of being a trenchant, ironic cultural satire and instead feels like a frustrating misfire.
  15. Negative sentiment: The phrase suggests that the actor's first starring vehicle is unlikely to make him a household name, despite his previous notable roles in other films.
  16. Negative sentiment: The phrase implies that there is a distinction between movies that have the courage to be over the top and those that don't care about being stupid, suggesting a negative evaluation of the latter.
  17. Negative sentiment: The phrase suggests that the humor in the movie, including Joe Viterelli's performance, is not as funny as it was in "Analyze This."
  18. Negative sentiment: The phrase criticizes the screenwriting of John Pogue, suggesting that his previous works are not impressive, and concludes with a negative evaluation of the film.
  19. Negative sentiment: The phrase criticizes the lack of common sense in the movie and the unrealistic nature of the action scenes.
  20. Negative sentiment: The phrase implies that the movie has only a small portion of decent

Παρατηρούμε ότι σε αυτή την κατηγορία σε είχε καμία δυσκολία, έκανε σωστή αναγνώριση.

Σε αυτό θα οφείλεται στη χρήση καθαρά αρνητικών επιθέτων από τους χρήστες, σε αντίθεση με την προηγούμενη κατηγορία που είχαμε μίξη θετικών και αρνητικών επιθέτων.

## Θα εξετάσουμε τώρα τις 20 πρώτες προτάσεις του Semeval2017A dataset

1. "Won the match #getin": Positive sentiment (expressing excitement and satisfaction).
2. "Plus, tomorrow is a very busy day, with Awareness Day's and debates. Gulp. Debates...": Negative sentiment (expressing anxiety or apprehension).
3. "Some areas of New England could see the first flakes of the season Tuesday.": Neutral sentiment (providing factual information).
4. "@francesco\_con40 2nd worst QB. DEFINITELY Tony Romo. The man who likes to share the ball with everyone. Including the other team.": Negative sentiment (expressing criticism towards Tony Romo's skills).
5. "#Thailand Washington - US President Barack Obama vowed Wednesday as he visited storm-ravaged New Jersey shore to...": Neutral sentiment (providing factual information about Barack Obama's visit).
6. "Did y'all hear what Tony Romo dressed up as for Halloween? A Giants quarterback! Cause that's all he could throw to Sunday night.": Negative sentiment (expressing criticism towards Tony Romo's performance).
7. "Tim Tebow may be available! Wow Jerry, what the heck you waiting for!": Positive sentiment (expressing excitement or eagerness).
8. "@mariakaykay aga tayo tomorrow ah. :) Good night, Ces. Love you! >:<": Positive sentiment (expressing affection and well-wishes).
9. "Tina Fey & Amy Poehler are hosting the Golden Globe awards on January 13. What do you think?": Neutral sentiment (asking for an opinion).
10. "Lunch from my new Lil spot... THE COTTON BOWL... pretty good#1st#time#will be going back#": Positive sentiment (expressing satisfaction with the lunch experience).
10. "Lunch from my new Lil spot... THE COTTON BOWL... pretty good#1st#time#will be going back#": Positive sentiment (expressing satisfaction with the lunch experience).
11. "SNC Halloween Pr. Pumped. Let's work it for Sunday... Packers vs.... who knows or caresn. #SNC #cheerpracticeonhalloween": Neutral sentiment (expressing anticipation for a cheerleading practice).
12. "@jacquelinemegan I'm sorry, I Heart Paris is no longer available at the Rockwell branch! You may call 8587000 to get a copy transferred! :)": Neutral sentiment (providing information).
13. "Manchester United will try to return to winning ways when they face Arsenal in the Premier League at Old Trafford on Saturday.": Neutral sentiment (providing information about a football match).
14. "Going to a bulls game with Aaliyah & hope next Thursday": Positive sentiment (expressing excitement about attending a Bulls game).
15. "Any Toon Fans with a spare ticket for Anfield on Sunday? Willing to pay extra #NUFC": Neutral sentiment (making a request).
16. "Louis inspired outfit on Monday and Zayn inspired outfit today... 4/5 done just need Harry :)": Positive sentiment (expressing excitement about outfits inspired by different individuals).
17. "Going to bed now... Rose parade then game tomorrow": Neutral sentiment (informing about personal plans).
18. "@\_Nenaah oh cause my friend got something from China, and they said it will take at least 6 to 8 weeks, and it came in the 2nd week :P": Neutral sentiment (discussing delivery times).
19. "I love the banner that was unfurled in the United end last night. It read: 'Chelsea - Standing up against racism since Sunday'": Positive sentiment (expressing admiration for the banner's message).
20. "#Repost Chris Bosh may be ugly. But he has a gorgeous wife and adorbs baby. I want to be happy like them one": Positive sentiment (expressing admiration for Chris Bosh's family).

Συγκρίνοντας με τα labels που έχουμε, βλέπουμε ότι πέτυχε μόνο 7/20. Θα πρέπει να σημειώσουμε ότι στο τέλος μας γράφει:

Please note that sentiment analysis can be subjective and may vary depending on the context and the individual's interpretation.

Πράγματι, λόγω του σύντομου μήκους των tweets, είναι δύσκολο να κάνουμε sentiment analysis,, καθώς δίνεται αποκλειστικά έμφαση στις λέξεις καθεαυτές και όχι στο context.

Παράδειγμα το "won the match", που το κατηγοριοποιεί ως positive, ενώ είναι neutral, καθώς αναφέρει απλά το γεγονός της νίκης.