ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Μεγάλα Γλωσσικά Μοντέλα για την Αναγνώριση Οικονομικού Συναισθήματος

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Πολυξένη, Θ. Καραμέτου

Επιβλέπουσα : Θεοδώρα, Βαρβαρίγου
Καθηγήτρια ΕΜΠ

Αθήνα, Φεβρουάριος 2025

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Μεγάλα Γλωσσικά Μοντέλα για την Αναγνώριση Οικονομικού Συναισθήματος

Πολυξένη, Θ. Καραμέτου

Επιβλέπουσα : Θεοδώρα, Βαρβαρίγου
Καθηγήτρια ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Φεβρουαρίου 2025.

.....	.....	.....
Θεοδώρα Βαρβαρίγου	Εμμανουήλ Βαρβαρίγος	Δημήτριος Ασκούνης
Καθηγήτρια ΕΜΠ	Καθηγητής ΕΜΠ	Καθηγητής ΕΜΠ

Αθήνα, Φεβρουάριος 2025

.....
Πολυξένη, Θ. Καραμέτου

Διπλωματούχος μεταπτυχιακού προγράμματος Τεχνοοικονομικά Συστήματα της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Περίληψη

Αυτή η διπλωματική εργασία διερευνά τη χρήση Μεγάλων Γλωσσικών Μοντέλων (LLMs) για την ανάλυση συναισθήματος στον οικονομικό τομέα, με εφαρμογή στην ελληνική αγορά. Τα LLMs, όπως το GPT, έχουν την ικανότητα να αναγνωρίζουν και να ερμηνεύουν ανθρώπινα συναισθήματα, κάτι που τα καθιστά ιδανικά για εφαρμογές όπως η πρόβλεψη χρηματιστηριακών τιμών, η ανάλυση μακροοικονομικών δεικτών και η λογιστική. Στην έρευνα, χρησιμοποιήσαμε LLMs για την ανάλυση συναισθήματος σε κριτικές χρηστών του ελληνικού ηλεκτρονικού καταστήματος Skroutz. Εφαρμόσαμε τεχνικές όπως fine-tuning, για την προσαρμογή των μοντέλων στα δεδομένα της ελληνικής γλώσσας, και Prompt Engineering, για τη βελτίωση της ακρίβειας στην αναγνώριση συναισθημάτων. Τα αποτελέσματα έδειξαν ότι τα LLMs μπορούν να αναλύουν αποτελεσματικά συναισθήματα σε ελληνικό κείμενο, προσφέροντας σημαντικές πληροφορίες για τη λήψη αποφάσεων στον οικονομικό τομέα. Η εργασία υπογραμμίζει τη δυνατότητα των LLMs να εφαρμοστούν σε τοπικές αγορές που δεν χρησιμοποιούν την αγγλική γλώσσα ως κύρια, και να συμβάλουν στην ανάλυση δεδομένων και στον στρατηγικό σχεδιασμό.

Λέξεις Κλειδιά: Μεγάλα Γλωσσικά Μοντέλα (LLMs), Ανάλυση συναισθήματος, Ανάλυση δεδομένων

Abstract

This thesis investigates the use of Large Language Models (LLMs) for sentiment analysis in the economic sector, with a focus on the Greek market. LLMs, such as GPT, have the ability to recognize and interpret human emotions, making them ideal for applications such as stock price prediction, macroeconomic indicator analysis, and accounting. In this research, we employed LLMs to analyze sentiment in user reviews from the Greek e-commerce platform Skroutz. Techniques such as fine-tuning, for adapting the models to Greek language data, and Prompt Engineering, for improving the accuracy of sentiment recognition, were applied. The results showed that LLMs can effectively analyze sentiment in Greek text, providing valuable insights for decision-making in the economic sector. The thesis highlights the potential of LLMs to be applied in local markets where English is not the primary language, contributing to data analysis and strategic planning.

Keywords: Large Language Models (LLMs), Sentiment Analysis, Data Analysis

Ευχαριστίες

Η παρούσα εργασία εκπονήθηκε στα πλαίσια του Διατμηματικού Μεταπτυχιακό Προγράμματος Σπουδών «Τεχνο-Οικονομικά Συστήματα» προσφέρεται από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου και από το Τμήμα Βιομηχανικής Διοίκησης και Τεχνολογίας του Πανεπιστημίου Πειραιώς. Θα ήθελα καταρχήν να ευχαριστήσω την επιβλέπουσα καθηγήτριά μου κα.Θεοδώρα Βαρβαρίγου όπως και τον κ. Ευθύμη Χονδρογιάννη για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον αντικείμενο που ανταποκρίνεται απολύτως στα επιστημονικά μου ενδιαφέροντα

Επιπρόσθετα νιώθω την ανάγκη να ευχαριστήσω θερμά όλους τους διδάσκοντες του Διατμηματικού Προγράμματος Μεταπτυχιακών Σπουδών για την στήριξη και την άριστη συνεργασία κατά την διάρκεια των μαθημάτων.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την αμέριστη στήριξή τους και την υπομονή τους κατά τη διάρκεια των σπουδών μου. Η ηθική τους υποστήριξη ήταν ανεκτίμητη.

Με εκτίμηση,

Καραμέτου Πολυξένη

Περιεχόμενα

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Κεφάλαιο 1. Εισαγωγή

1.1 Σημασία του προβλήματος

Η επικοινωνιακή στρατηγική των κεντρικών τραπεζών, όπως αντικατοπτρίζεται στον τόνο και το συναίσθημα που εκφράζουν οι ανακοινώσεις και οι εκθέσεις τους, διαδραματίζει κομβικό ρόλο στη διαμόρφωση των οικονομικών εξελίξεων. Έρευνες έχουν ήδη υπογραμμίσει την καθοριστική επίδραση αυτής της επικοινωνίας στις χρηματοοικονομικές αγορές. Για παράδειγμα, οι Ehrmann και Fratzscher [32] έχουν επισημάνει ότι η επικοινωνία της Ευρωπαϊκής Κεντρικής Τράπεζας (ΕΚΤ) επηρεάζει ουσιαστικά τις προσδοκίες των αγορών και τις τιμές των περιουσιακών στοιχείων. Παράλληλα, οι Gürkaynak et al. [33] προσέφεραν στοιχεία που συνδέουν τον τρόπο επικοινωνίας της Ομοσπονδιακής Τράπεζας των ΗΠΑ με μεταβολές στις αποδόσεις ομολόγων και τις τιμές των μετοχών. Ο τόνος με τον οποίο εκφράζονται οι κεντρικές τράπεζες δύναται να προκαλέσει αντιδράσεις σε επίπεδο τιμών περιουσιακών στοιχείων, συναλλαγματικών ισοτιμιών και επιτοκίων, επηρεάζοντας κατ' επέκταση τη δυναμική των αγορών και τις οικονομικές επιδόσεις [34]. Επιπλέον, η σαφήνεια και η συνέπεια στην επικοινωνία τους μπορούν να βελτιώσουν την προβλεψιμότητα των μελλοντικών νομισματικών πολιτικών, επιτρέποντας στους συμμετέχοντες στις αγορές να λαμβάνουν πιο ενημερωμένες αποφάσεις [35].

Η κατανόηση και η ανάλυση των συναισθημάτων αποτελεί ένα από τα πιο περίπλοκα προβλήματα στην τεχνητή νοημοσύνη, παρά την πρόοδο που έχει σημειωθεί στη μηχανική μάθηση. Η ανίχνευση συναισθημάτων από υπολογιστές δεν αφορά μόνο την κατανόηση της ανθρώπινης γλώσσας, αλλά απαιτεί και την ερμηνεία των συμφραζομένων και των λεπτών συναισθηματικών αντιδράσεων που χαρακτηρίζουν την ανθρώπινη επικοινωνία. Αυτή η πρόκληση αποκτά ακόμη μεγαλύτερη σημασία στον οικονομικό τομέα, όπου οι συναισθηματικές αντιδράσεις επηρεάζουν κρίσιμες αποφάσεις, επενδυτικές συμπεριφορές και την ίδια την αγορά.

Η δυνατότητα αναγνώρισης και κατανόησης οικονομικού συναισθήματος μπορεί να προσφέρει εξαιρετικά χρήσιμα εργαλεία για την παρακολούθηση τάσεων, την πρόβλεψη αγορών και τη βελτιστοποίηση της επικοινωνίας σε χρηματοοικονομικές εφαρμογές. Παρ' όλα αυτά, η πολυπλοκότητα της γλώσσας που χρησιμοποιείται στον οικονομικό τομέα, οι τεχνικοί όροι, καθώς και οι συχνά διφορούμενες εκφράσεις δημιουργούν ένα ιδιαίτερα απαιτητικό πεδίο για την εφαρμογή μοντέλων τεχνητής νοημοσύνης.

Τα τελευταία χρόνια, τα μεγάλα γλωσσικά μοντέλα (LLMs), όπως η σειρά GPT της OpenAI, έχουν σημειώσει αξιοσημείωτη πρόοδο στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP). Τα μοντέλα αυτά αποτελούν σημαντικό ορόσημο στην τεχνολογία τεχνητής νοημοσύνης, ιδιαίτερα όσον αφορά την κατανόηση και τη δημιουργία ανθρώπινης γλώσσας. Η αυξημένη υπολογιστική ισχύς και οι εξελιγμένοι αλγόριθμοι έχουν επιτρέψει στα LLMs να έχουν προχωρημένες δυνατότητες στην κατανόηση σύνθετων κειμένων, στην απάντηση ερωτήσεων και στη δημιουργία περιεχομένου. Η δυναμική τους γίνεται ιδιαίτερα εμφανής στον χρηματοοικονομικό τομέα, όπου οι δυνατότητές τους αναδεικνύονται σταδιακά με ολοένα και μεγαλύτερη επίδραση [1].

1.2 Συνεισφορά της διπλωματικής

Ο βασικός στόχος της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της χρήσης Μεγάλων Γλωσσικών Μοντέλων (LLMs) για την ανάλυση συναισθήματος στον οικονομικό τομέα, με ιδιαίτερη έμφαση στην ελληνική αγορά. Η εργασία επιδιώκει να προσαρμόσει τα LLMs σε ελληνικά δεδομένα, αξιοποιώντας τεχνικές όπως το fine-tuning και το prompt engineering, ώστε να βελτιωθεί η ακρίβεια και η αποτελεσματικότητά τους στην αναγνώριση συναισθημάτων σε χρηματοοικονομικά και καταναλωτικά κείμενα. Μέσω της ανάλυσης συναισθημάτων σε ελληνικά κείμενα, όπως κριτικές χρηστών, αποσκοπείτε η εξαγωγή πληροφοριών σχετικά με τη συμπεριφορά των καταναλωτών.

1.3 Διάρθρωση της διπλωματικής

Στο πρώτο κεφάλαιο, παρουσιάζεται η εισαγωγή στο πρόβλημα, η σημασία της ανάλυσης συναισθήματος για τον οικονομικό τομέα, οι στόχοι της έρευνας, η συνεισφορά της εργασίας, καθώς και η γενική δομή της διπλωματικής.

Το δεύτερο κεφάλαιο αναφέρεται στο θεωρητικό υπόβαθρο της εργασίας. Παρουσιάζεται η θεωρία του συναισθήματος, η σημασία του στις οικονομικές αποφάσεις και οι τεχνικές αναπαράστασης λέξεων. Επιπλέον, γίνεται ανάλυση των γλωσσικών μοντέλων, ξεκινώντας από τις παραδοσιακές προσεγγίσεις και φτάνοντας στα σύγχρονα μοντέλα που βασίζονται σε μετατροπείς, όπως τα Transformers.

Στο τρίτο κεφάλαιο, εξετάζονται τα Μεγάλα Γλωσσικά Μοντέλα, οι εφαρμογές τους στον οικονομικό τομέα, καθώς και οι τεχνικές προσαρμογής τους σε εξειδικευμένα δεδομένα, όπως το fine-tuning και το prompt engineering. Παράλληλα, αναλύονται οι προκλήσεις που προκύπτουν από τη χρήση αυτών των μοντέλων.

Το τέταρτο κεφάλαιο επικεντρώνεται στην πειραματική αξιολόγηση. Περιγράφεται η μεθοδολογία που ακολουθήθηκε για την ανάλυση συναισθήματος σε ελληνικά χρηματοοικονομικά δεδομένα, παρουσιάζονται τα αποτελέσματα των πειραμάτων και σχολιάζεται η απόδοση των χρησιμοποιούμενων μεθόδων.

Τέλος, στο πέμπτο κεφάλαιο, συνοψίζονται τα συμπεράσματα της έρευνας, ενώ προτείνονται πιθανές μελλοντικές επεκτάσεις για τη βελτίωση και την περαιτέρω ανάπτυξη της χρήσης LLMs στην ανάλυση συναισθήματος και σε άλλες εφαρμογές του οικονομικού τομέα.

Κεφάλαιο 2. Θεωρητικό υπόβαθρο

2.1 Θεωρία Συναισθήματος

2.1.1 Εισαγωγή

Η θεωρία των βασικών συναισθημάτων περιγράφει το συναίσθημα ως το θεμέλιο της κοινωνικής αλληλεπίδρασης, λειτουργώντας ως ένα σύστημα επικοινωνίας που διευκολύνει την ανταλλαγή πληροφοριών, τη διατήρηση σχέσεων και την έκφραση ιδεών μεταξύ των ανθρώπων. Παράλληλα, το συναίσθημα αποτελεί αναπόσπαστο ψυχολογικό στοιχείο που συμβάλλει στην επιβίωση του ανθρώπινου είδους, ενώ ταυτόχρονα επηρεάζει βαθιά τη διαμόρφωση κοινωνικών συμπεριφορών και υποστηρίζει την ανάπτυξη της νοητικής σκέψης. Καθώς παίζει καθοριστικό ρόλο σε πλήθος ανθρώπινων δραστηριοτήτων, όπως η αντίληψη, η εκμάθηση, η λήψη αποφάσεων, η συλλογιστική και η κοινωνικοποίηση, το συναίσθημα αναγνωρίζεται ως μια από τις βασικές δυνάμεις που ενισχύουν την ανάπτυξη του ανθρώπινου πολιτισμού [2].

2.1.2 Σημασία των Συναισθημάτων

Η σημασία των συναισθημάτων για τον άνθρωπο μπορεί να συνοψιστεί σε πέντε βασικές διαστάσεις. Πρώτον, η λειτουργία της **επιβίωσης** βασίζεται σε μια μαθημένη φυσιολογική αντίδραση που επιτρέπει στους ανθρώπους να προσαρμόζονται θετικά στο περιβάλλον τους [3]. Τα συναισθήματα ενισχύουν την ικανότητα προσαρμογής μέσω της ρύθμισης της προσοχής, της μνήμης, της αντίληψης και άλλων γνωστικών διεργασιών, εξασφαλίζοντας καλύτερες πιθανότητες επιβίωσης και εξέλιξης κατά τη διάρκεια της ανθρώπινης εξέλιξης.

Δεύτερον, η λειτουργία της **επικοινωνίας** υπογραμμίζει τη σημασία των συναισθημάτων για την ακριβή έκφραση και κατανόηση των ανθρώπινων προθέσεων [4]. Οι ίδιες λέξεις, όταν εκφράζονται με διαφορετικό συναισθηματικό φορτίο,

αποκτούν διαφορετικό νόημα, δείχνοντας πως τα συναισθήματα είναι κρίσιμα για την ερμηνεία των κοινωνικών αλληλεπιδράσεων.

Τρίτον, τα συναισθήματα διαδραματίζουν σημαντικό ρόλο στη **λήψη αποφάσεων**, επηρεάζοντας τόσο τις γρήγορες όσο και τις αργές γνωστικές διαδικασίες. Το ασυνείδητο «Σύστημα 1» βασίζεται κυρίως στα συναισθήματα και τις εμπειρίες, ενώ το συνειδητό «Σύστημα 2» στηρίζεται στη λογική ανάλυση [5]. Συνεπώς, τα συναισθήματα επηρεάζουν βαθιά τη διαδικασία λήψης αποφάσεων, επηρεάζοντας την αποτελεσματικότητα και την ποιότητα των αποτελεσμάτων.

Τέταρτον, τα συναισθήματα έχουν **κινητήρια λειτουργία**, ενισχύοντας και διατηρώντας τις ανθρώπινες συμπεριφορές. Με αυτόν τον τρόπο, επηρεάζουν τον βαθμό δέσμευσης, την επιμονή στις προσπάθειες και την αξιολόγηση των αποτελεσμάτων [6].

Τέλος, τα συναισθήματα λειτουργούν ως **μηχανισμός διατήρησης δεσμών** μεταξύ των μελών εθνοτικών ομάδων, οικογενειών, κοινωνικών κύκλων, κοινωνικών τάξεων και άλλων ομάδων, ενισχύοντας τη συνοχή και τη συνεργασία σε διάφορα κοινωνικά πλαίσια.

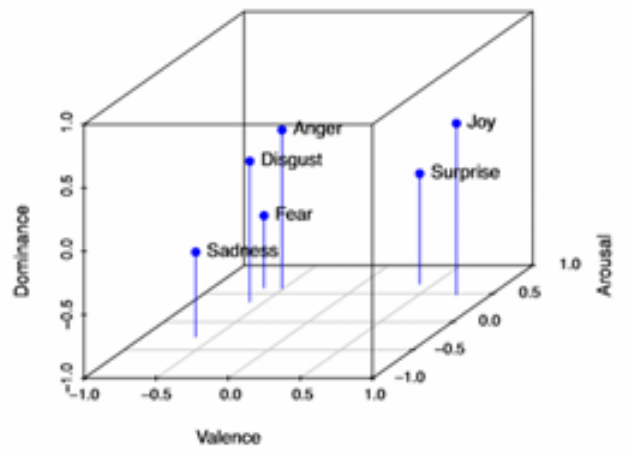
2.1.3 Ταξινόμηση Συναισθημάτων

Η πλέον διαδεδομένη προσέγγιση για την κατηγοριοποίηση των βασικών συναισθημάτων αποδίδεται στον Paul Ekman. Τα συναισθήματα θεωρούνται ξεχωριστές κατηγορίες, όπου κάθε συναίσθημα λειτουργεί ως ανεξάρτητη οντότητα. Μέσα από ανθρωπολογικές έρευνες, ο Ekman κατέληξε σε έξι θεμελιώδη συναισθήματα: θυμός, φόβος, λύπη, χαρά, αηδία και έκπληξη. Αυτά τα συναισθήματα θεωρούνται "οικουμενικά", καθώς εκφράζονται ανεξαρτήτως πολιτιστικής προέλευσης [7]. Μια σημαντική επέκταση σε αυτό το μοντέλο έγινε από τον Plutchik, ο οποίος πρόσθεσε δύο επιπλέον συναισθηματικές καταστάσεις: την "ανυπομονησία" και την "εμπιστοσύνη", διευρύνοντας έτσι το πλαίσιο κατηγοριοποίησης [8].

Σε μία άλλη προσέγγιση, το κάθε συναίσθημα περιγράφεται μέσω δύο ή τριών βασικών χαρακτηριστικών, ενώ οι συναισθηματικές καταστάσεις απεικονίζονται σε έναν πολυδιάστατο χώρο. Αυτό το μοντέλο δίνει τη δυνατότητα να εξεταστούν οι

συνδέσεις και οι συνεχείς μεταβολές μεταξύ διαφορετικών συναισθημάτων, αποτυπώνοντας την πολυπλοκότητα της συναισθηματικής εμπειρίας. Οι Russell και Mehrabian [9], πρότειναν το μοντέλο VAD (Valence, Arousal, Dominance) σύμφωνα με το οποίο, τα συναισθήματα περιγράφονται μέσω τριών ανεξάρτητων διαστάσεων:

1. **Ευχαρίστηση (valence)**: Περιγράφει το εύρος από τη δυσαρέσκεια έως την ευχαρίστηση, αντιπροσωπεύοντας το ευχάριστο ή δυσάρεστο συναίσθημα για κάτι.
2. **Διέγερση (arousal)**: Αναφέρεται στο επίπεδο ενεργοποίησης του συναισθήματος, που κυμαίνεται από την ηρεμία έως τον ενθουσιασμό.
3. **Κυριαρχία (dominance)**: Εξετάζει τον βαθμό ελέγχου που σχετίζεται με τη συναισθηματική κατάσταση, από την υποταγή έως την κυριαρχία.



Εικόνα 1. Το μοντέλο VAD για τα έξι βασικά συναισθήματα.

	Valence	*Arousal*	*Dominance*
Anger	–0.43	0.67	0.34
Joy	0.76	0.48	0.35
Surprise	0.4	0.67	–0.13
Disgust	–0.6	0.35	0.11
Fear	–0.64	0.6	–0.43
Sadness	–0.63	0.27	–0.33

Πίνακας 1. Οι τιμές για τα έξι βασικά συναισθήματα ως συνάρτηση των διαστάσεων[10].

Η **πόλωση (polarity)** στην ανάλυση συναισθημάτων αναφέρεται στη μέτρηση της συναισθηματικής φόρτισης που εκφράζεται σε ένα κείμενο, προσδιορίζοντας αν το συναίσθημα είναι θετικό, αρνητικό ή ουδέτερο. Πρόκειται για έναν βασικό δείκτη που

χρησιμοποιείται για την ποσοτικοποίηση της συναισθηματικής διάθεσης μέσα από λέξεις, φράσεις ή ολόκληρα κείμενα.

Η θετική συναισθηματική φόρτιση (positive polarity) εκφράζει συναισθήματα όπως χαρά, ικανοποίηση ή ενθουσιασμό. Αντίθετα, η αρνητική συναισθηματική φόρτιση (negative polarity) υποδεικνύει συναισθήματα όπως λύπη, απογοήτευση ή θυμό. Η ουδέτερη φόρτιση (neutral polarity) αναφέρεται σε περιπτώσεις όπου το συναίσθημα είναι ισορροπημένο ή αντικειμενικό, χωρίς να εκφράζεται κάποια έντονη διάθεση.

Το polarity υπολογίζεται συχνά με βάση τη συναισθηματική φόρτιση των λέξεων στο κείμενο. Υπολογίζεται από τον τύπο

$$Polarity = \frac{\sum(\theta\varepsilon\tau\iota\kappa\acute{\varepsilon}\varsigma\ \beta\alpha\theta\mu o\lambda o\gamma\acute{\iota}\varepsilon\varsigma)-\sum(\alpha\rho\nu\eta\tau\iota\kappa\acute{\varepsilon}\varsigma\ \beta\alpha\theta\mu o\lambda o\gamma\acute{\iota}\varepsilon\varsigma)}{\sum(\acute{o}\lambda\varepsilon\varsigma\ o\iota\ \beta\alpha\theta\mu o\lambda o\gamma\acute{\iota}\varepsilon\varsigma)} \quad (1)$$

Η **ένταση (intensity)** στην ανάλυση συναισθημάτων αναφέρεται στην ένταση με την οποία εκφράζεται ένα συναίσθημα σε ένα κείμενο, μία πρόταση ή μία λέξη. Αντικατοπτρίζει πόσο ισχυρά ή έντονα εκφράζεται ένα συναίσθημα, ανεξάρτητα από το αν είναι θετικό ή αρνητικό. Οι διαβαθμίσεις της έντασης κυμαίνονται από χαμηλή ένταση, που υποδηλώνει ήπια συναισθήματα, έως υψηλή ένταση, που αντιπροσωπεύει έντονα συναισθήματα.

Για παράδειγμα, φράσεις όπως "είμαι κάπως ευχαριστημένος" εκφράζουν χαμηλή ένταση συναισθήματος, ενώ φράσεις όπως "είμαι απόλυτα ενθουσιασμένος" χαρακτηρίζονται από υψηλή ένταση. Η μέτρηση της έντασης συμπληρώνει τη βασική πληροφορία του polarity, καθώς βοηθά στην κατανόηση όχι μόνο του είδους του συναισθήματος (θετικό, αρνητικό ή ουδέτερο) αλλά και της δύναμής του.

Επίσης έχουμε τα **explicit** και **implicit** συναισθήματα που αναφέρονται στους τρόπους με τους οποίους εκφράζονται οι πληροφορίες, τα συναισθήματα ή οι έννοιες σε ένα κείμενο ή μήνυμα. Το **explicit** αφορά πληροφορίες που εκφράζονται άμεσα και ξεκάθαρα. Το μήνυμα είναι σαφές, χωρίς να αφήνει περιθώρια για παρερμηνείες. Για παράδειγμα, η φράση "Είμαι θυμωμένος" δηλώνει με σαφήνεια το συναίσθημα του θυμού. Αντίθετα, το **implicit** (σιωπηρό) αφορά πληροφορίες που εκφράζονται έμμεσα, μέσω υπονοούμενων ή συμφραζομένων. Τα implicit μηνύματα απαιτούν κατανόηση

του πλαισίου ή της πρόθεσης του ομιλητή για να γίνουν πλήρως αντιληπτά. Για παράδειγμα, η φράση "Δεν αντέχω άλλο" υπονοεί αρνητική συναισθηματική κατάσταση, όπως λύπη ή θυμό, χωρίς να το εκφράζει άμεσα.

Η **temporal διάσταση** στην ανάλυση συναισθημάτων αφορά τη μελέτη των συναισθηματικών αλλαγών που συμβαίνουν με την πάροδο του χρόνου και τη σημασία της χρονικής σειράς στην κατανόηση της συναισθηματικής κατάστασης. Δεν αρκεί μόνο να προσδιοριστεί αν ένα συναίσθημα είναι θετικό, αρνητικό ή ουδέτερο, αλλά και να εξεταστεί πώς εξελίσσεται μέσα σε ένα πλαίσιο, καθώς η σειρά με την οποία εκφράζονται τα συναισθήματα μπορεί να αλλάξει τη συνολική ερμηνεία.

Για παράδειγμα, η φράση *"Πριν ήμουν εντάξει, αλλά τώρα είμαι θυμωμένος"* δείχνει μια αρνητική εξέλιξη, καθώς το συναίσθημα μεταβαίνει από ουδέτερο ή θετικό (εντάξει) σε αρνητικό (θυμό). Αντίθετα, η φράση *"Πριν ήμουν θυμωμένος, αλλά τώρα είμαι εντάξει"* εκφράζει μια θετική εξέλιξη, καθώς το συναίσθημα μεταβαίνει από αρνητικό σε ουδέτερο ή θετικό.

2.1.4 Ανάλυση Συναισθήματος μέσω Χρήσης Υπολογιστή

Η **ανάλυση συναισθήματος κειμένου** εστιάζει στην εξαγωγή, ανάλυση, κατανόηση και δημιουργία συναισθηματικών πληροφοριών από τη φυσική γλώσσα. Στα πρώτα στάδια της αναγνώρισης συναισθημάτων σε κείμενο, οι μέθοδοι βασίζονταν κυρίως σε χειροκίνητα κατασκευασμένα λεξικά συναισθημάτων και κανόνες για τη συναισθηματική ανάλυση. Αυτές οι προσεγγίσεις αξιολογούσαν την πολικότητα των συναισθημάτων αντιστοιχίζοντας λέξεις-κλειδιά με γραμματικούς κανόνες μέσα στο κείμενο. Παρά τα πλεονεκτήματά τους, περιορίζονταν από την κάλυψη του συναισθηματικού λεξικού και τους κανόνες, γεγονός που δυσχέραινε την ανάλυση συναισθημάτων σε πολλαπλούς τομείς.

Με την πρόοδο της μηχανικής μάθησης, αναδείχθηκαν νέες μέθοδοι αναγνώρισης συναισθημάτων, βασισμένες σε στατιστικά μοντέλα και αλγορίθμους μηχανικής μάθησης. Μέσω εκπαίδευσης σε μεγάλης κλίμακας σύνολα δεδομένων κειμένου, αυτά τα μοντέλα μπόρεσαν να μάθουν αυτόματα τις εκφράσεις συναισθημάτων και τα σημασιολογικά χαρακτηριστικά, βελτιώνοντας τόσο την ακρίβεια όσο και τη γενίκευση στις ταξινομήσεις συναισθημάτων.

Τα τελευταία χρόνια, η τεχνολογία βαθιάς μάθησης έχει φέρει σημαντικές αλλαγές στην αναγνώριση συναισθημάτων από κείμενο. Μοντέλα που βασίζονται σε νευρωνικά δίκτυα, όπως τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs), τα συνελικτικά νευρωνικά δίκτυα (CNNs), τα δίκτυα μνήμης μακράς και βραχείας διάρκειας (LSTMs), τους Μετατροπείς (Transformers) όπως το BERT και το GPT, έχουν επιτύχει εξαιρετικά αποτελέσματα σε διάφορες εργασίες συναισθηματικής ανάλυσης [11]. Τα μοντέλα αυτά μπορούν να συλλάβουν συμφραζόμενες πληροφορίες και σημασιολογικές σχέσεις, προσφέροντας βαθύτερη κατανόηση και ανάλυση συναισθημάτων, ενισχύοντας την αποτελεσματικότητα και την ακρίβεια των εφαρμογών ανάλυσης κειμένου.

2.1.5 Σημασία του Συναισθήματος στις Οικονομικές Αποφάσεις

Οι Hansen και McMahon [36] μελέτησαν τον τρόπο με τον οποίο το συναίσθημα που προκύπτει από τις εκθέσεις χρηματοπιστωτικής σταθερότητας των κεντρικών τραπεζών μπορεί να επηρεάσει τις οικονομικές αντιλήψεις και τις προσδοκίες για τη νομισματική πολιτική. Αντίστοιχα, οι Apel και Grimaldi [37] ανέλυσαν τα πρακτικά συνεδριάσεων της Riksbank, χρησιμοποιώντας αυτοματοποιημένες τεχνικές ανάλυσης περιεχομένου. Η προσέγγισή τους έδειξε ότι το συναίσθημα και ο τόνος που εξάγονται από αυτά τα κείμενα μπορούν να συμβάλλουν στην πρόβλεψη μελλοντικών αποφάσεων για τα επιτόκια, αναδεικνύοντας την περίπλοκη αλληλεπίδραση μεταξύ κεντρικής τραπεζικής επικοινωνίας και προσδοκιών των αγορών. Η έρευνα αυτή αποδεικνύει ότι οι προηγμένες τεχνικές ανάλυσης συναισθήματος μπορούν να προσφέρουν βαθύτερη κατανόηση των έμμεσων μηνυμάτων που μεταδίδουν οι κεντρικές τράπεζες.

Παράλληλα, μελέτες όπως αυτές των Hubert και Labondance [38] παρέχουν περαιτέρω ενδείξεις για την αποτελεσματικότητα της ανάλυσης συναισθήματος στον συγκεκριμένο τομέα. Στο έργο τους, εξετάζουν την επίδραση των ανακοινώσεων καθοδήγησης (forward guidance) της Ευρωπαϊκής Κεντρικής Τράπεζας (ΕΚΤ) στα επιτόκια των Overnight Indexed Swaps (OIS). Τα ευρήματά τους αποκαλύπτουν μια συνεπή μείωση των βραχυπρόθεσμων επιτοκίων, ιδιαίτερα σε μεγαλύτερες χρονικότητες, φαινόμενο που παραμένει αισθητό και μετά τον έλεγχο για μακροοικονομικούς παράγοντες.

2.2 Αναπαράσταση λέξεων

Οι παραδοσιακές μέθοδοι αναπαράστασης λέξεων που προηγήθηκαν των σύγχρονων διανυσματικών μοντέλων έδιναν έμφαση σε απλούστερες, συχνά δυαδικές αναπαραστάσεις.

Το One-Hot Encoding αποτελεί μια κατηγορηματική μέθοδο αναπαράστασης λέξεων μέσω διανυσμάτων, κατά την οποία κάθε όρος του λεξιλογίου αποδίδεται σε ένα δυαδικό διάνυσμα μήκους ίσο με το μέγεθος του λεξιλογίου, N. Σε αυτό το διάνυσμα, όλες οι τιμές είναι μηδενικές, εκτός από μία μοναδική θέση, στην οποία εμφανίζεται η τιμή 1 και αντιστοιχεί στη συγκεκριμένη λέξη. Μπορούμε να ορίσουμε τη μέθοδο ως εξής:

Λεξιλόγιο: $V = \{w_1, w_2, \ldots, w_N\}$, όπου N το πλήθος των μοναδικών λέξεων. Διανυσματική Αναπαράσταση: Κάθε λέξη w_i αναπαρίσταται από ένα διάνυσμα e_i μήκους N

$$e_i = [0, 0, \ldots, 1, \ldots, 0]^T \quad (2)$$

όπου το 1 τοποθετείται στη i-οστή θέση, αντιστοιχώντας στον δείκτη της λέξης στο λεξιλόγιο.

Για παράδειγμα, έστω V = {'cat', 'dog', 'hat'}. Οι διανυσματικές αναπαραστάσεις τους θα ήταν:

$$e_{cat} = [1, 0, 0]^T, e_{dog} = [0, 1, 0]^T, e_{hat} = [0, 0, 1]^T. \quad (3)$$

Παρότι ο υπολογισμός των one-hot αναπαραστάσεων είναι ιδιαίτερα αποδοτικός, η προσέγγιση αυτή παρουσιάζει σημαντικά μειονεκτήματα. Ένα από τα βασικά ζητήματα είναι η υψηλή διαστατικότητα των διανυσμάτων, η οποία καθίσταται προβληματική σε μεγάλα λεξιλόγια, αυξάνοντας την υπολογιστική πολυπλοκότητα. Επιπλέον, η μέθοδος αυτή δεν ενσωματώνει καμία σημασιολογική πληροφορία, καθώς δεν προκύπτουν συσχετίσεις μεταξύ των αναπαριστώμενων λέξεων [12].

Μία εξίσου απλή αλλά περιορισμένη μέθοδος είναι η Κωδικοποίηση Ετικέτας (Label Encoding), στην οποία σε κάθε λέξη αποδίδεται απλώς ένας μοναδικός ακέραιος αριθμός. Έστω ένα λεξιλόγιο V = {w_1, w_2, ..., w_N}, ορίζουμε μια συνάρτηση κωδικοποίησης f: V → $\mathbb{Z}$, όπου f(w_i) = i.

Για παράδειγμα, με το ίδιο λεξιλόγιο:

$$f('cat') = 1, f('dog') = 2, f('hat') = 3. \quad (4)$$

Ωστόσο, οι περιορισμοί της κωδικοποίησης ετικέτας είναι ανάλογοι με αυτούς της one-hot αναπαράστασης. Η αυθαίρετη αντιστοίχιση ενός αριθμού σε κάθε λέξη δεν προσφέρει στο μοντέλο τη δυνατότητα να αντλήσει σημασιολογικές συσχετίσεις ή πληροφορίες σχετικά με τη σχετική σημασία των λέξεων μεταξύ τους.

Το μοντέλο Bag of Words (BoW) αποτελεί ένα απλό μοντέλο εξαγωγής χαρακτηριστικών, το οποίο χρησιμοποιείται ευρέως στην επεξεργασία κειμένου από αλγορίθμους ανάκτησης πληροφορίας και μηχανικής μάθησης. Στα μοντέλα BoW, ένα κείμενο (π.χ. μια πρόταση ή ένα έγγραφο) αναπαρίσταται ως μια αδόμητη συλλογή των λέξεων που περιέχονται σε αυτό, αδιαφορώντας για τη σειρά εμφάνισής τους, τη γραμματική δομή και το συμφραζόμενο. Η έμφαση δίνεται αποκλειστικά στη συχνότητα εμφάνισης κάθε λέξης. Στόχος είναι η δημιουργία ενός λεξιλογίου όλων των μοναδικών όρων του δοθέντος corpus και, στη συνέχεια, η αναπαράσταση κάθε κειμένου ως ένα διάνυσμα, του οποίου κάθε συνιστώσα αντανακλά την παρουσία ή την απουσία (αλλά και τη συχνότητα) του αντίστοιχου όρου του λεξιλογίου.

Έστω ότι έχουμε μία συλλογή M εγγράφων, $D = \{d_1, d_2, \ldots, d_m\}$. Το λεξικό είναι ένα σύνολο όλων των μοναδικών λέξεων στο corpus, μεγέθους N,

$$V = \{w_1, w_2, \ldots, w_n\}. \quad (5)$$

Κάθε κείμενο d_j αποτελείται από μια ακολουθία λέξεων (tokens),

$$d_j = [t_{j1}, t_{j2}, \ldots, t_{jLj}]. \quad (6)$$

Για κάθε έγγραφο d_j και για κάθε λέξη w_i στο λεξιλόγιο V, υπολογίζεται η συχνότητα όρου tf_{ij} ως εξής:

$$\mathrm{tf}_{ij} = \mathrm{freq}\left(w_i, d_j\right) = \sum_{k=1}^{L_j} \delta\left(w_i, t_{jk}\right) \quad (7)$$

όπου:

$$\delta\left(w_i, t_{jk}\right) = \begin{cases} 1, & \alpha\nu\ w_i = t_{jk} \\ 0, & \alpha\nu\ w_i \neq t_{jk} \end{cases} \quad (8)$$

Κάθε έγγραφο $\mathrm{d_j}$ αναπαρίσταται ως ένα διάνυσμα $\mathrm{x_j} \in \mathrm{R^n}$

$$\mathbf{x}_j = \left[\mathrm{tf}_{1j}, \mathrm{tf}_{2j}, \mathrm{tf}_{3j}, \ldots, \mathrm{tf}_{Nj}\right]^{\mathsf{T}}. \quad (9)$$

Δημιουργώντας τα διανύσματα για όλα τα έγγραφα, προκύπτει ένα μητρώο εμφάνισης όρων μεγέθους N×M

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_M]^T \quad (10)$$

Για παράδειγμα, έστω ότι έχουμε 3 έγγραφα

- $\mathrm{d_1}$: “the cat sat”
- $\mathrm{d_2}$: “the cat sat in the hat”
- $\mathrm{d_3}$: “the dog barked”.

Το λεξιλόγιο θα είναι

$$V = \{“the”, “cat”, “sat”, “in”, “hat”, “dog”, “barked”\}.$$

Κάθε έγγραφο θα έχει διανυσματική αναπαράσταση

$$\begin{aligned} \mathbf{x}_1 &= [1,1,1,0,0,0,0]^{\mathsf{T}} \\ \mathbf{x}_2 &= [2,1,1,1,1,0,0]^{\mathsf{T}} \\ \mathbf{x}_3 &= [1,0,0,0,0,1,1]^{\mathsf{T}} \end{aligned} \quad (11)$$

ενώ το μητρώο εμφάνισης των όρων θα είναι

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

Το BoW είναι απλό και εύχρηστο, αλλά δημιουργεί πολύ αραιούς πίνακες σε μεγάλα λεξιλόγια και αδυνατεί να διαχωρίσει πολυσημίες. Για παράδειγμα, η πρόταση "Mary is quicker than John"είναι ταυτόσημη με την πρόταση "John is quicker than Mary" [13].

Ένα επαναλαμβανόμενο πρόβλημα στις παραδοσιακές προσεγγίσεις ήταν η αδυναμία τους να αποτυπώσουν τη σημασιολογική πληροφορία που φέρουν τα κείμενα. Η κατανόηση της σημασίας αυτής της έλλειψης, σε συνδυασμό με την ταχεία ανάπτυξη του τομέα της Επεξεργασίας Φυσικής Γλώσσας, οδήγησε στη δημιουργία τεχνικών κατανεμημένης αναπαράστασης λέξεων. Οι μέθοδοι αυτές, εμπνευσμένες από το πεδίο της Κατανεμημένης Σημασιολογίας (Distributional Semantics) [14], στηρίζονται στην αρχή ότι λέξεις οι οποίες εμφανίζονται σε παρόμοια συμφραζόμενα μοιράζονται συναφείς σημασίες. Η ιδέα αυτή υλοποιείται με την ενσωμάτωση των λέξεων σε έναν συνεχόμενο διανυσματικό χώρο, όπου λέξεις με παρεμφερή νόημα βρίσκονται γεωμετρικά κοντά η μία στην άλλη.

Η Λανθάνουσα Σημασιολογική Ανάλυση (**Latent Semantic Analysis–LSA),** γνωστή και ως Latent Semantic Indexing (LSI), αποτελεί μια θεμελιώδη τεχνική στην Επεξεργασία Φυσικής Γλώσσας. Βασιζόμενη σε αρχές γραμμικής άλγεβρας, έχει ως στόχο την κατανόηση λανθανουσών συσχετίσεων μεταξύ λέξεων σε κείμενα μεγάλης κλίμακας. Με τη χρήση της Διάσπασης Ιδιοτιμών Τιμών (Singular Value Decomposition–SVD), το LSA κατορθώνει να εξάγει τη σημασιολογική δομή που κρύβεται στα δεδομένα, μειώνοντας ταυτόχρονα τη διάστασή τους [15].

Ας υποθέσουμε ότι διαθέτουμε ένα σύνολο κειμένων αποτελούμενο από M έγγραφα και ένα λεξιλόγιο N διακριτών όρων. Αρχικά, δημιουργούμε έναν πίνακα όρων-εγγράφων $X \in R^{N \times M}$, στο οποίο κάθε στοιχείο x_{ij} αντιστοιχεί στο «βάρος» του όρου i στο έγγραφο j, όπως για παράδειγμα μια τιμή TF-IDF. Στη συνέχεια,

εφαρμόζουμε SVD στον πίνακα X, προκύπτει ένας επιμερισμός του σε τρεις επιμέρους πίνακες:

$$X = U \Sigma V^T \quad (13)$$

όπου:

- $U \in R^{N \times r}$: Περιλαμβάνει τα μοναδιαία διανύσματα που αντιστοιχούν στους όρους.

- $\Sigma \in R^{r \times r}$: Είναι διαγώνιος πίνακας με τις ιδιοτιμές τιμές.

- $V \in R^{M \times r}$: Αποτελείται από μοναδιαία διανύσματα που αντιστοιχούν στα έγγραφα.

- r: Ο βαθμός του αρχικού πίνακα X, που συνήθως επιλέγεται να είναι πολύ μικρότερος από τις αρχικές διαστάσεις του X.

Στην συνέχεια επιλέγουμε τις k μεγαλύτερες μοναδιαίες τιμές και τα αντίστοιχα διανύσματα

$$\mathbf{X}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^\top \quad (14)$$

Αυτή η επιλογή αποτυπώνει τις πιο σημαντικές λανθάνουσες σημασιολογικές διαστάσεις. Η προσέγγιση της LSA παρέχει τη δυνατότητα να αναδειχθούν οι συνώνυμες λέξεις και οι εννοιολογικές συγγένειες μεταξύ τους, ενώ παράλληλα επιτυγχάνει μείωση του θορύβου μέσω της διάστασης, διατηρώντας έτσι την πληροφοριακή ακεραιότητα. Παρ' όλα αυτά, η υψηλή υπολογιστική πολυπλοκότητα της μεθόδου SVD, όπως και η περιορισμένη ικανότητα αντιμετώπισης της πολυσημίας, συνιστούν δύο σημαντικά μειονεκτήματα. Παρά τις αδυναμίες αυτές, η LSA παραμένει χρήσιμη σε διαδικασίες ανάκτησης πληροφορίας και ταξινόμησης εγγράφων.

2.3 Γλωσσικά Μοντέλα

Γλωσσικό Μοντέλο είναι μία συνάρτηση που αντιστοιχίζει μία πιθανότητα στα αλφαριθμητικά που μπορούν να παραχθούν από ένα λεξιλόγιο [13]. Με άλλα λόγια, ένα γλωσσικό μοντέλο αναλαμβάνει να καθορίσει πόσο πιθανό είναι να εμφανιστεί μια συγκεκριμένη ακολουθία λέξεων ή χαρακτήρων, εφόσον γνωρίζει τις πιθανότητες εμφάνισης των συστατικών της και των πιθανών διαδοχών τους. Έτσι, στηρίζεται σε στατιστικά στοιχεία και μαθηματικές αρχές, αξιοποιώντας συχνότητες, συναρτήσεις

πιθανότητας και άλλες υπολογιστικές τεχνικές, ώστε να προβλέψει ποια λέξη ή φράση μπορεί να ακολουθήσει μια δεδομένη ακολουθία.

Για γλωσσικό μοντέλο M και αλφάβητο Σ έχουμε

$$\sum_{s\in\Sigma^*} P_M(s) = 1 \quad (15)$$

όπου Σ^* αναπαριστά το σύνολο όλων των πιθανών που μπορεί να σχηματιστεί από το αλφάβητο Σ, s είναι μία ακολουθία συμβόλων (μία λέξη ή πρόταση) και $P_M(s)$ είναι η πιθανότητα να εμφανιστεί η ακολουθία s από το μοντέλο.

2.3.1 «Κλασικά» Μοντέλα Γλώσσας

Το απλούστερο είδος γλωσσικού μοντέλου είναι τα **N-grams**. Σε αυτή την προσέγγιση, η πιθανότητα εμφάνισης μιας λέξης εκτιμάται με βάση τις προηγούμενες N−1 λέξεις, λειτουργώντας ως μια «τοπική» προσέγγιση της ιστορίας του κειμένου. Ένα 2-γραμμο (bigram) θεωρείται μοντέλο Markov πρώτης τάξης, αφού στηρίζεται στη μία και μόνο προηγούμενη λέξη, ενώ ένα 3-γραμμο (trigram) αποτελεί μοντέλο Markov δεύτερης τάξης, καθώς λαμβάνει υπόψη τις δύο αμέσως προηγούμενες λέξεις. Γενικεύοντας, ένα N-γραμμο μοντέλο μπορεί να χαρακτηριστεί ως μοντέλο Markov τάξης N−1.

Ουσιαστικά, τα N-γράμματα εφαρμόζουν μια απλοποιητική παραδοχή Markov. Η πιθανότητα της επόμενης λέξης εξαρτάται μόνο από ένα πεπερασμένο σύνολο προηγούμενων λέξεων και όχι από ολόκληρο το ιστορικό. Μαθηματικά, αυτή η προσέγγιση μπορεί να εκφραστεί ως

$$P(w_i \mid w_1, w_2, \dots, w_{i-1}) \approx P(w_i \mid w_{i-N+1}, w_{i-N+2}, \dots, w_{i-1}) \quad (16)$$

Για παράδειγμα, σε ένα 2-γραμμο μοντέλο (bigram)

$$P(w_i \mid w_1, \dots, w_{i-1}) \approx P(w_i \mid w_{i-1}) \quad (17)$$

Η εκτίμηση των πιθανοτήτων γίνεται με την καταμέτρηση συχνοτήτων στο σώμα κειμένων (corpus)

$$P(w_i \mid w_{i-N+1}, \dots, w_{i-1}) = \frac{C(w_{i-N+1},\dots,w_{i-1},w_i)}{C(w_{i-N+1},\dots,w_{i-1})} \quad (18)$$

όπου $C(\cdot)$ η συχνότητα εμφάνισης μιας συγκεκριμένης ακολουθίας λέξεων στο σώμα κειμένων. Ωστόσο, ένα πρόβλημα που συχνά ανακύπτει είναι ότι ορισμένες ακολουθίες

ενδέχεται να μην εμφανιστούν ποτέ στο εκπαιδευτικό σύνολο. Για την αντιμετώπιση αυτού του ζητήματος εφαρμόζονται τεχνικές εξομάλυνσης που μειώνουν την πιθανότητα μηδενικών κατανομών, βελτιώνοντας τη γενικευσιμότητα του μοντέλου [20].

Οι Bengio et al. [21] το 2003, πρότειναν την εισαγωγή των νευρωνικών δικτύων στη δημιουργία γλωσσικών μοντέλων. Τα Feedforward Νευρωνικά Γλωσσικά Μοντέλα (FFNNLM) βασίζονται σε ένα απλό, εμπρός-τροφοδοτούμενο (feedforward) νευρωνικό δίκτυο, το οποίο εκπαιδεύεται με σκοπό να προβλέψει την επόμενη λέξη δεδομένων των προηγούμενων λέξεων. Η βασική ιδέα ήταν να εγκαταλειφθεί η απλή καταμέτρηση συχνοτήτων των Ν-γραμμάτων και να αξιοποιηθεί η ισχύς των νευρωνικών δικτύων για την εκμάθηση σημασιολογικών σχέσεων και συμφραζομένων.

Σε αυτό το πλαίσιο, το νευρωνικό δίκτυο λαμβάνει ως είσοδο τα ενσωματωμένα διανύσματα (embeddings) των προηγούμενων λέξεων και, περνώντας μέσα από ένα ή περισσότερα κρυφά επίπεδα (hidden layers), εξάγει την πιθανότητα για κάθε υποψήφια επόμενη λέξη. Με αυτόν τον τρόπο, το FFNNLM επιτρέπει μια πιο συμπαγή αναπαράσταση της γλωσσικής πληροφορίας, σε αντίθεση με τις μεθόδους Ν-γραμμάτων που αποδίδουν πιθανότητες με βάση συχνότητες και δεν αξιοποιούν πλήρως την σημασιολογική και συντακτική πληροφορία.

Η δυνατότητα του FFNNLM να «μάθει» έμμεσες σχέσεις μεταξύ λέξεων και να προσαρμοστεί σε διαφορετικές γλωσσικές δομές οδήγησε σε βελτίωση των επιδόσεων σε ποικίλα γλωσσικά καθήκοντα. Παρ' όλα αυτά, η συγκεκριμένη αρχιτεκτονική παρουσιάζει δύο βασικά μειονεκτήματα: α) η υπολογιστική πολυπλοκότητα, η οποία μπορεί να είναι σημαντική κατά την εκπαίδευση ή την παραγωγή προβλέψεων για μεγάλα λεξιλόγια, και β) η αδυναμία της να επεξεργάζεται αποτελεσματικά μακροπρόθεσμες εξαρτήσεις μεταξύ λέξεων που βρίσκονται σε μεγάλη απόσταση μέσα στο κείμενο.

Η εισαγωγή των Recurrent Neural Network Language Models (RNNLMs) από τους Mikolov et al. [22] υπήρξε ένα σημαντικό βήμα προς τη βελτίωση της ικανότητας των γλωσσικών μοντέλων να χειρίζονται ακολουθιακά δεδομένα μεγάλου μήκους, όπως προτάσεις και παράγραφοι, χωρίς την ανάγκη προσδιορισμού ενός σταθερού

«παραθύρου» (window) λέξεων. Τα RNNs χρησιμοποιούν έναν κυκλικό βρόχο (recurrent loop) που επιτρέπει την ανάκληση πληροφοριών από προηγούμενες χρονικές στιγμές (ή λέξεις), με αποτέλεσμα να μπορούν να διατηρούν στην «μνήμη» τους στοιχεία σχετιζόμενα με προηγούμενα τμήματα του κειμένου.

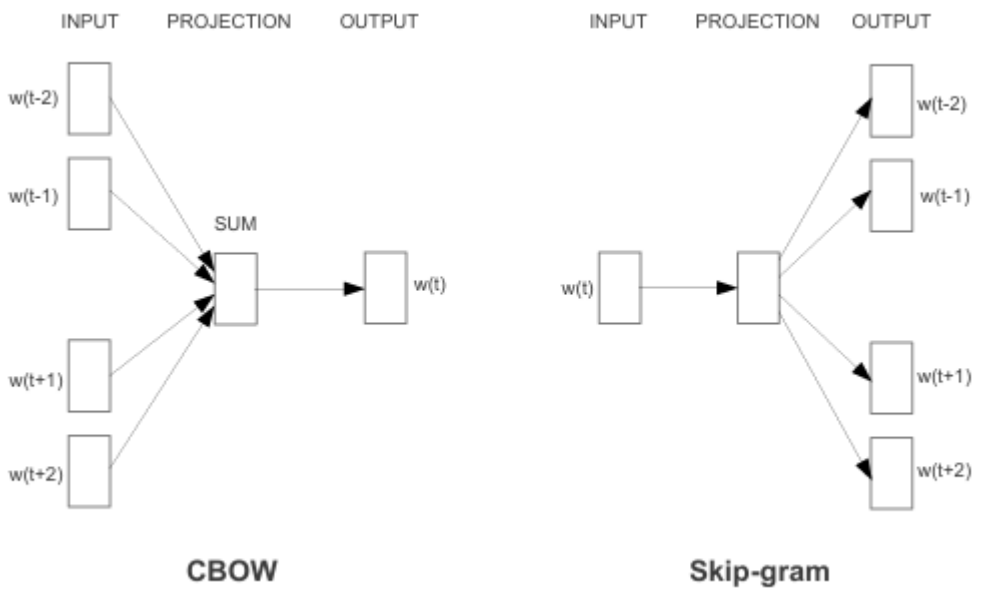
Η βασική ιδέα πίσω από τα RNNLMs είναι ότι σε κάθε χρονικό βήμα t, το μοντέλο λαμβάνει μια λέξη w_t (ή την αναπαράστασή της) και μια κατάσταση κρυφού επιπέδου h_{t-1}, η οποία αντικατοπτρίζει τις πληροφορίες που έχουν ήδη επεξεργαστεί σε προηγούμενα βήματα. Μέσα από μια συνάρτηση μετάβασης, παράγεται μια νέα κατάσταση h_t που συμπυκνώνει τα παλιά και τα νέα δεδομένα, καθώς και μια πρόβλεψη της πιθανότητας για την επόμενη λέξη w_{t+1}

$$\begin{aligned} h_t &= f_\theta(h_{t-1}, w_t) \\ P(w_{t+1} \mid w_1, w_2, \dots, w_t) &= g_\phi(h_t) \end{aligned} \tag{19}$$

Όπου f_θ είναι μια μη γραμμική συνάρτηση που υλοποιείται από τα νευρωνικά επίπεδα του RNN και g_φ είναι μια συνάρτηση εξόδου που δίνει τις πιθανότητες για κάθε υποψήφια λέξη του λεξιλογίου.

Το 2013, οι Mikolov et al. [16] παρουσίασαν ένα σύνολο μοντέλων με το όνομα **Word2Vec** τα οποία παράγουν διανυσματικές αναπαραστάσεις λέξεων (word embeddings) που δύνανται να συλλάβουν σημασιολογικές και συντακτικές ομοιότητες μεταξύ τους. Τα μοντέλα Word2Vec βασίζονται σε δύο αρχιτεκτονικές: το Continuous Bag of Words (CBOW) και το Skip-Gram. Και οι δύο αρχιτεκτονικές εκπαιδεύονται με τη χρήση νευρωνικών δικτύων, αξιοποιώντας ένα μεγάλο σύνολο δεδομένων κειμένου, και αποσκοπούν στη δημιουργία πολυδιάστατων διανυσματικών αναπαραστάσεων που αποτυπώνουν τις σχέσεις μεταξύ των λέξεων.

Η αρχιτεκτονική CBOW προβλέπει την τρέχουσα λέξη βάσει του περιβάλλοντός της, δηλαδή των γειτονικών λέξεων σε ένα συγκεκριμένο παράθυρο λέξεων. Αυτό επιτρέπει στο μοντέλο να μάθει πώς μια λέξη σχετίζεται με τις λέξεις γύρω της. Από την άλλη, το Skip-Gram λειτουργεί αντίστροφα, προβλέπει τις λέξεις του περιβάλλοντος με βάση τη δεδομένη λέξη. Αυτή η μέθοδος είναι ιδιαίτερα αποτελεσματική για την αναγνώριση συντακτικών και σημασιολογικών ομοιοτήτων.



Εικόνα 2. Αρχιτεκτονική των δύο μοντέλων του Word2Vec [16].

Το **GloVe**, το οποίο προτάθηκε από τους Pennington et al. (2014) [17], αποτελεί έναν μη επιβλεπόμενο αλγόριθμο μάθησης που στοχεύει στη δημιουργία διανυσματικών αναπαραστάσεων των λέξεων. Σε αντίθεση με τεχνικές όπως το Word2Vec, οι οποίες στηρίζονται στην εκπαίδευση ενός μοντέλου με βάση ένα τοπικό "παράθυρο" (window) συμφραζομένων, το GloVe αξιοποιεί τις παγκόσμιες στατιστικές συνύπαρξης των λέξεων σε ολόκληρο το σώμα κειμένων.

Η διαδικασία εφαρμόζεται πάνω σε ένα μητρώο συν-εμφάνισης (co-occurrence matrix) $X \in R^{W \times W}$, όπου W είναι το μέγεθος του λεξιλογίου. Κάθε στοιχείο X_{ij} του πίνακα X αναπαριστά το πόσο συχνά η λέξη j εμφανίζεται στο συμφραζόμενο της λέξης i. Μέσω αυτών των συσχετίσεων, το GloVe μαθαίνει να αντιστοιχίζει λέξεις σε διανύσματα τέτοια, ώστε οι γεωμετρικές σχέσεις μεταξύ τους να αντανακλούν τις σημασιολογικές και συντακτικές τους ομοιότητες.

Η εκπαίδευση του GloVe υλοποιείται μέσω της ελαχιστοποίησης μιας συνάρτησης κόστους ελαχίστων τετραγώνων που προσπαθεί να προσεγγίσει τον λογάριθμο της συχνότητας συν-εμφάνισης $logX_{ij}$. Πιο συγκεκριμένα, η συνάρτηση κόστους είναι της μορφής:

$$J = \sum_{i,j=1}^{W} f(X_{ij})\left(w_i^{\top}\tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2 \quad (20)$$

όπου $w_i, \tilde{w}_j$ είναι τα διανύσματα αναπαράστασης για την "κεντρική" λέξη i και τη λέξη j που εμφανίζεται στο περιβάλλον της, $b_i, \tilde{b}_j$ είναι οι παράγοντες πόλωσης (bias terms) που ρυθμίζουν την κεντρική τάση των διανυσμάτων και $f(X_{ij})$ είναι μια συνάρτηση βαρύτητας που ελέγχει τη συμβολή ζευγών λέξεων με διάφορες συχνότητες στην ελαχιστοποίηση του κόστους. Μία τυπική συνάρτηση βαρύτητας είναι η

$$f(X_{ij}) = \begin{cases} \left(\frac{X_{ij}}{x_{max}}\right)^{\alpha}, & \text{if } X_{ij} < x_{max} \\ 1, & \text{else} \end{cases} \quad (21)$$

με παραμέτρους $\alpha = 0.75$ και $x_{\max} = 100$.

Το τελικό αποτέλεσμα είναι ένας διανυσματικός χώρος στον οποίο λέξεις με παρόμοια σημασία τοποθετούνται κοντά η μία στην άλλη. Κατ' αυτόν τον τρόπο, το GloVe επιτυγχάνει την αποτύπωση τόσο τοπικών όσο και παγκόσμιων συσχετίσεων στη γλωσσική πληροφορία, αποτελώντας μια ισχυρή τεχνική για τη δημιουργία πλούσιων σε σημασιολογικό περιεχόμενο word embeddings.

2.3.2 Μοντέλα Γλώσσας βασισμένα σε Μετατροπείς

Οι συμβατικές τεχνικές, όπως τα Word2Vec και GloVe, ενώ παρέχουν χρήσιμη σημασιολογική πληροφορία, δεν καταφέρνουν να αποτυπώσουν την ευελιξία της γλώσσας, όπου μια λέξη μπορεί να έχει πολλαπλές έννοιες ανάλογα με τα συμφραζόμενα. Για να αντιμετωπιστεί αυτό το πρόβλημα, προτάθηκαν οι Contextualized αναπαραστάσεις λέξεων (contextualized word embeddings), οι οποίες προσαρμόζουν δυναμικά την αναπαράσταση μιας λέξης ανάλογα με το συγκεκριμένο περιεχόμενο του κειμένου στο οποίο εμφανίζεται.

Χαρακτηριστικό παράδειγμα αυτής της προσέγγισης αποτελεί το **ELMo** (Embeddings from Language Models) [18], το οποίο αξιοποιεί μοντέλα βασισμένα σε επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) για να παράγει διανυσματικές αναπαραστάσεις που μεταβάλλονται ανάλογα με τα προηγούμενα και επόμενα συμφραζόμενα μιας πρότασης. Με τον τρόπο αυτόν, η λέξη "bank" π.χ. θα έχει διαφορετική αναπαράσταση όταν αναφέρεται σε "τράπεζα" σε οικονομικό συμφραζόμενο, σε σύγκριση με όταν αναφέρεται σε "όχθη ποταμού".

Ακόμα πιο αποτελεσματικές λύσεις προέκυψαν με την αξιοποίηση μοντέλων βασισμένων στους Μετασχηματιστές (Transformers), όπως το BERT (Bidirectional Encoder Representations from Transformers) [19]. Το BERT εκπαιδεύεται με αμφίδρομη προσοχή (bidirectional attention) πάνω σε μεγάλα σώματα κειμένων, επιτρέποντάς του να απορροφά πληροφορίες από ολόκληρη την πρόταση ταυτόχρονα. Αυτό προσφέρει μια προηγμένη κατανόηση του νοήματος και του ρόλου κάθε λέξης, καθώς και πλουσιότερες, πιο ευέλικτες αναπαραστάσεις.

Οι Contextualized αναπαραστάσεις λέξεων έχουν αποδειχθεί ιδιαίτερα αποτελεσματικές σε εφαρμογές επεξεργασίας φυσικής γλώσσας, όπως η Ανάλυση Συναισθήματος, η Απάντηση Ερωτήσεων, η Αναγνώριση Οντοτήτων, η Περίληψη Κειμένου και η Μετάφραση. Επιτρέπουν στα μοντέλα να κατανοούν βαθύτερα την πολυπλοκότητα της φυσικής γλώσσας. Σε αυτή τη βάση, όλο και περισσότερες τεχνικές εστιάζουν στην περαιτέρω βελτίωση των Contextualized αναπαραστάσεων, οδηγώντας σε μοντέλα που προσεγγίζουν την ανθρώπινη ικανότητα κατανόησης και παραγωγής γλώσσας με τρόπους που παλιότερα θεωρούνταν αδύνατοι.

Τα Transformers [23] έχουν αναδειχθεί ως το κυρίαρχο υπόδειγμα για την κατασκευή Μεγάλων Γλωσσικών Μοντέλων. Ο Transformer βασίζεται στη δομή κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) και χρησιμοποιεί τον μηχανισμό προσοχής (attention) για την επεξεργασία ακολουθιών δεδομένων, όπως το κείμενο. Στην αριστερή πλευρά της εικόνας φαίνεται ο κωδικοποιητής (encoder), ο οποίος λαμβάνει την είσοδο και την επεξεργάζεται ώστε να δημιουργήσει της εισόδου. Η είσοδος αρχικά μετατρέπεται σε διανύσματα embeddings σταθερού μεγέθους, ενώ προστίθενται κωδικοποιήσεις θέσης για να εισαχθεί η πληροφορία της σειράς των λέξεων. Ο κωδικοποιητής αποτελείται από επαναλαμβανόμενα επίπεδα, όπου κάθε επίπεδο περιλαμβάνει πολυκέφαλη προσοχή (multi-head attention) για τον υπολογισμό των σχέσεων μεταξύ των λέξεων, ένα πλήρως συνδεδεμένο δίκτυο (για μη γραμμική επεξεργασία και ένα μηχανισμό προσθήκης και κανονικοποίησης.

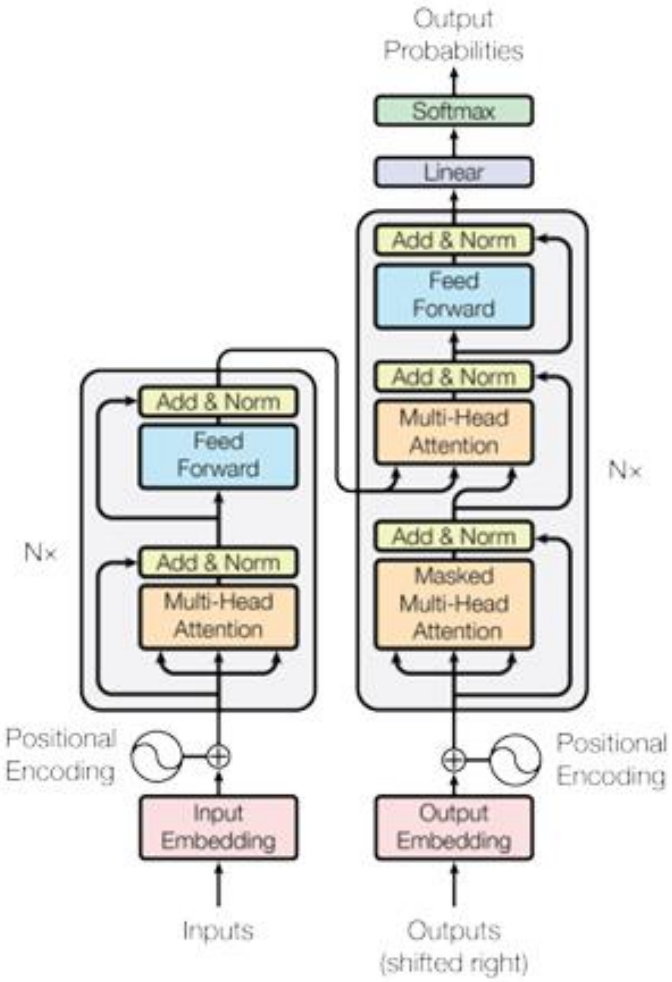
Στη δεξιά πλευρά της εικόνας απεικονίζεται ο αποκωδικοποιητής (decoder), ο οποίος χρησιμοποιεί την έξοδο του κωδικοποιητή για να δημιουργήσει την τελική ακολουθία εξόδου. Η έξοδος ξεκινά με την ενσωμάτωση λέξεων και την προσθήκη κωδικοποιήσεων θέσης, όπως στον κωδικοποιητή. Ο αποκωδικοποιητής περιλαμβάνει επαναλαμβανόμενα επίπεδα που έχουν τρία βασικά μέρη. Τη πολυκέφαλη προσοχή με

μάσκα (masked multi-head attention), η οποία διασφαλίζει ότι οι μελλοντικές λέξεις δεν αποκαλύπτονται κατά την εκπαίδευση, την πολυκέφαλη προσοχή, που συνδέει την έξοδο του κωδικοποιητή με την τρέχουσα ακολουθία εξόδου, και το πλήρως συνδεδεμένο δίκτυο για την επεξεργασία των λέξεων. Στο τέλος, η έξοδος περνά από ένα γραμμικό επίπεδο και ένα softmax, για να παραχθούν οι πιθανότητες για την επόμενη λέξη. Το επίπεδο softmax έχει ως συνάρτηση ενεργοποίησης

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{22}$$

όπου $z = [z_1, z_2, \dots, z_n]$ το διάνυσμα εισόδου.

Το **attention** είναι μια διαδικασία που αντιστοιχίζει ένα ερώτημα και ένα σύνολο ζευγών κλειδιών-τιμών σε μια έξοδο. Στόχος του είναι να δίνει έμφαση στα πιο σημαντικά μέρη της εισόδου όταν το μοντέλο κάνει προβλέψεις ή παράγει αποτελέσματα. Ο μηχανισμός αυτός λειτουργεί μέσω της δυναμικής συσχέτισης μεταξύ διαφορετικών στοιχείων της εισόδου, όπως λέξεις σε μία πρόταση.

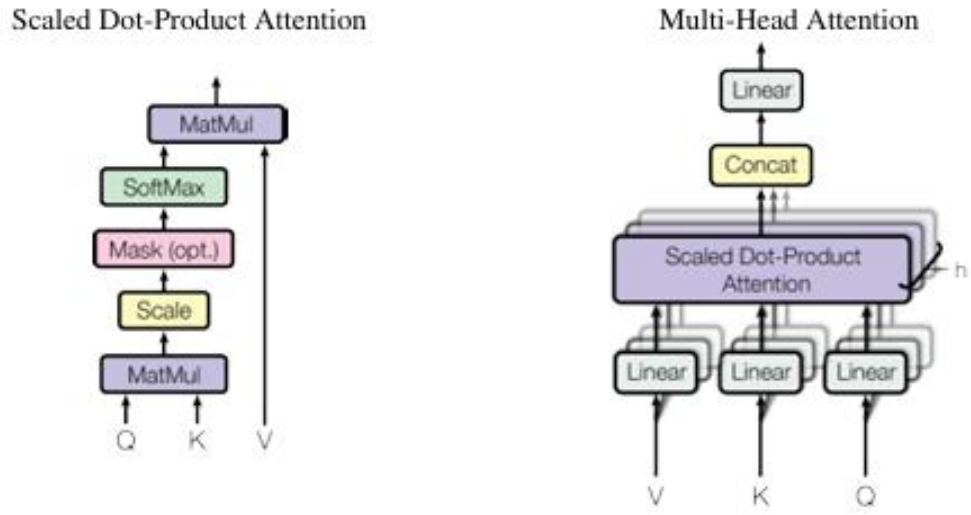


Εικόνα 3. Αρχιτεκτονική του Transformer [23].

Η διαδικασία ξεκινά με τη μετατροπή κάθε στοιχείου εισόδου σε τρία διανύσματα: το **Query (Q),** που αναπαριστά τι "ψάχνει" το μοντέλο, το **Key (K),** που περιγράφει τα κριτήρια για τη σημαντικότητα, και το **Value (V),** που περιέχει την πληροφορία που θα χρησιμοποιηθεί στην έξοδο. Υπολογίζεται η συσχέτιση μεταξύ των Q και K μέσω του εσωτερικού γινομένου τους, και στη συνέχεια κανονικοποιείται με τη συνάρτηση softmax, ώστε να δημιουργηθούν πιθανότητες που δείχνουν τη σχετική σημασία κάθε στοιχείου. Τα αποτελέσματα αυτά εφαρμόζονται στα values, ενισχύοντας έτσι τα πιο σημαντικά μέρη της πληροφορίας.

Μια σημαντική επέκταση του μηχανισμού attention είναι ο multi-headed attention, που εισάγει τη δυνατότητα πολλαπλών παράλληλων "κεφαλών" attention. Κάθε κεφαλή λειτουργεί με διαφορετικά διανύσματα Q, K και V, επιτρέποντας στο μοντέλο να εστιάζει ταυτόχρονα σε διαφορετικά μέρη της εισόδου και να κατανοεί

πολυπλοκότερες συσχετίσεις. Τα αποτελέσματα από τις κεφαλές συνενώνονται και προβάλλονται μέσω ενός γραμμικού στρώματος, επιτρέποντας την αποδοτική ενσωμάτωση των διαφορετικών πληροφοριών.



Εικόνα 4. Τα επίπεδα Attention και Multi-Head Attention [23].

2.4 Προ-εκπαιδευμένα Γλωσσικά Μοντέλα

Το BERT (Bidirectional Encoder Representations from Transformers) [19] αποτελεί ένα γλωσσικό μοντέλο που στηρίζεται στην αρχιτεκτονική των Transformers. Προτάθηκε από την Google το 2018, και υπάρχει σε δύο βασικές διαμορφώσεις, που διαφέρουν στο μέγεθος και την υπολογιστική τους πολυπλοκότητα. Το BERTBASE αποτελείται από 12 επίπεδα (layers) Transformer, με 768 κρυφές μονάδες και 12 κεφαλές προσοχής (attention heads), συνολικά περίπου 110 εκατομμύρια παραμέτρους.

Αυτή η έκδοση έχει σχεδιαστεί ώστε να είναι πιο ελαφριά και ταχύτερη, επιτρέποντας πειραματισμούς με χαμηλότερο υπολογιστικό κόστος. Το BERTLARGE περιλαμβάνει 24 επίπεδα Transformer, με 1024 κρυφές μονάδες και 16 κεφαλές προσοχής, φτάνοντας περίπου τα 340 εκατομμύρια παραμέτρους. Αν και πιο απαιτητικό σε υπολογιστικούς πόρους, αποδίδει καλύτερα και παρέχει υψηλότερη ακρίβεια στις περισσότερες εργασίες NLP.

Ο σχεδιασμός του BERT διαφοροποιείται από προγενέστερα μοντέλα γλώσσας, καθώς βασίζεται σε μια διαδικασία προεκπαίδευσης που αξιοποιεί δύο μη επιβλεπόμενες εργασίες. Στην πρώτη εργασία, Masked Language Model (MLM), ορισμένα τμήματα της εισόδου (tokens) αντικαθίστανται τυχαία από ένα ειδικό σύμβολο [MASK]. Στόχος του μοντέλου είναι να ανακτήσει τις αρχικές λέξεις βάσει του συμφραζομένου τους. Χάρη σε αυτή τη στρατηγική, το BERT λαμβάνει υπόψη ταυτόχρονα πληροφορίες από τα αριστερά και τα δεξιά της λέξης (bidirectional context). Αυτό το χαρακτηριστικό το διαφοροποιεί από προϋπάρχοντα μοντέλα, όπως το GPT, τα οποία επεξεργάζονταν κυρίως πληροφορία προς μία κατεύθυνση. Με το MLM, το BERT μπορεί να προβλέψει μια λέξη που λείπει σε μια πρόταση, αξιοποιώντας πλήρως το περιβάλλον της. Για παράδειγμα, σε μια πρόταση "Το [MASK] είναι ένα καλό εργαλείο.", το μοντέλο μπορεί να ανασυνθέσει τη λέξη "BERT".

Η δεύτερη εργασία, Next Sentence Prediction (NSP), αφορά την ικανότητα του μοντέλου να κατανοεί τη λογική συνάφεια μεταξύ δύο προτάσεων. Συγκεκριμένα, δίνονται στο BERT δύο προτάσεις και το μοντέλο καλείται να προβλέψει αν η δεύτερη πρόταση αποτελεί φυσική συνέχεια της πρώτης ή όχι. Κατά την προεκπαίδευση, το 50% των ζευγών προτάσεων είναι συναφείς και το υπόλοιπο 50% επιλέγεται τυχαία χωρίς σύνδεση μεταξύ τους. Για παράδειγμα, αν η Πρόταση Α είναι "Ο BERT είναι ένα νευρωνικό δίκτυο." και η Πρόταση Β είναι "Το μοντέλο αυτό χρησιμοποιείται στην επεξεργασία φυσικής γλώσσας.", το μοντέλο θα πρέπει να αναγνωρίσει τη λογική συνέχεια μεταξύ τους. Αν όμως η Πρόταση Β είναι "Ο καιρός είναι ηλιόλουστος.", το BERT θα πρέπει να προβλέψει ότι η συνέχεια δεν είναι σχετική.

2.4.1 GPT μοντέλα

Τα μοντέλα GPT (Generative Pre-trained Transformer) είναι μια οικογένεια αρχιτεκτονικών μετασχηματιστών που σχεδιάστηκαν για την παραγωγή φυσικής γλώσσας. Αναπτύχθηκαν από την OpenAI, και έχουν εξελιχθεί μέσα από αρκετές γενιές.

GPT-1

Το GPT-1 [24], το οποίο παρουσιάστηκε το 2018, αποτελεί ένα από τα πρώτα μεγάλα γλωσσικά μοντέλα που στηρίχθηκαν στην αρχιτεκτονική Transformers,

προχωρώντας όμως με μια μονοκατευθυνόμενη προσέγγιση. Δηλαδή, επεξεργαζόταν τις λέξεις από αριστερά προς τα δεξιά, εστιάζοντας κατά κύριο λόγο στην πρόβλεψη της επόμενης λέξης, χωρίς να λαμβάνει ταυτόχρονα υπόψη πληροφορίες από το επόμενο μέρος της πρότασης, όπως συνέβαινε με αμφίδρομα μοντέλα τύπου BERT.

Αυτό το μοντέλο περιλάμβανε 12 επίπεδα (transformer blocks) και συνολικά 117 εκατομμύρια παραμέτρους. Το GPT-1 εκπαιδεύτηκε σε δεδομένα από το BookCorpus, ένα σύνολο δεδομένων με περισσότερα από 7,000 βιβλία μυθοπλασίας, το οποίο περιλαμβάνει περίπου 5 GB καθαρού κειμένου. Η χρήση αυτών των δεδομένων παρείχε στο μοντέλο τη δυνατότητα να κατανοεί δομές φυσικής γλώσσας, καθώς και λογικές και συντακτικές σχέσεις. Μετά την προεκπαίδευση, το GPT-1 προσαρμόζεται (fine-tuning) σε συγκεκριμένες εργασίες NLP, όπως κατηγοριοποίηση κειμένου, απάντηση σε ερωτήσεις και ανάλυση συναισθήματος. Η στρατηγική fine-tuning περιλαμβάνει μικρότερα σύνολα δεδομένων με επιβλεπόμενη εκπαίδευση, προσαρμόζοντας το μοντέλο για εξειδικευμένες χρήσεις. Όμως, το GPT-1 είχε προβλήματα στην αντιμετώπιση πιο πολύπλοκων εργασιών [25].

GPT-2

Παρουσιάστηκε το 2019 από την OpenAI, αποτελεί τη φυσική συνέχεια του GPT-1. Αποτελείται από 1.5 δισεκατομμύρια παραμέτρους, σημαντικά περισσότερες από το GPT-1, γεγονός που του επιτρέπει να διαχειρίζεται πιο σύνθετες γλωσσικές δομές. Η εκπαίδευση πραγματοποιήθηκε σε ένα ευρύ σύνολο δεδομένων, γνωστό ως WebText, το οποίο αποτελείται από περίπου 8 εκατομμύρια ιστοσελίδες που του έδωσε την ικανότητα να παράγει κείμενο υψηλής συνοχής και να ανταποκρίνεται σε ένα ευρύ φάσμα θεμάτων [26].

GPT-3

Το GPT-3 παρουσιάστηκε το 2020 από την OpenAI. Με 175 δισεκατομμύρια παραμέτρους, υπερβαίνει κατά πολύ τα προηγούμενα μοντέλα, επιτρέποντάς του να απορροφήσει μια τεράστια ποσότητα γλωσσικών δεδομένων. Ένα ιδιαίτερο χαρακτηριστικό του GPT-3 είναι ότι, σε αντίθεση με πολλά προηγούμενα μοντέλα, μπορεί να προσαρμοστεί σε νέες εργασίες με ελάχιστη ή και καθόλου πρόσθετη εκπαίδευση μέσω παραδειγμάτων (few-shot learning). Χωρίς να απαιτεί επιπλέον fine-

tuning, το GPT-3 αποκρίνεται σε διαφορετικές απαιτήσεις απλώς αλλάζοντας την προτροπή (prompt) που του δίνεται.

GPT-4

Το GPT-4, παρουσιάστηκε από την OpenAI το 2023, αποτελεί την τελευταία γενιά στη σειρά GPT μοντέλων. Το GPT-4 μπορεί να επεξεργάζεται πολυτροπικές εισόδους (multimodal inputs), δηλαδή όχι μόνο κείμενο αλλά και εικόνες. Επίσης, μπορεί να απαντά σε ερωτήσεις σχετικά με εικόνες, να εξάγει πληροφορίες από γραφήματα ή να περιγράφει περιεχόμενα εικόνων. Υποστηρίζει μεγαλύτερο **context window** (μέχρι 32,000 tokens), επιτρέποντας τη διαχείριση μεγαλύτερων κειμένων, όπως βιβλία, μεγάλα έγγραφα ή σύνθετες συνομιλίες.

2.4.1.1 Claude μοντέλα

Claude 1

Το Claude 1, που παρουσιάστηκε το 2023 από την Anthropic, αποτελεί ένα νέο μεγάλο γλωσσικό μοντέλο το οποίο, όπως και άλλα μοντέλα της γενιάς του, βασίζεται σε τεχνικές βαθιάς μάθησης και αρχιτεκτονικές παρόμοιες με αυτές των Transformers. Αυτό που διαφοροποιεί το Claude 1 είναι η προσήλωσή του σε αρχές ευθυγράμμισης (alignment), με στόχο να παρέχει απαντήσεις που δεν είναι μόνο ορθές και συνεκτικές, αλλά και σύμφωνες με τις αξίες που έχουν τεθεί από τους δημιουργούς του [27].

Claude 2

Συγκριτικά με την προηγούμενη έκδοση, το Claude 2 έχει βελτιώσει σημαντικά τις ικανότητές του στην κατανόηση σύνθετων ερωτημάτων, την αντιμετώπιση δύσκολων θεμάτων και την παραγωγή περισσότερο πληροφοριακών και συνεκτικών κειμένων. Παράλληλα, έχει δοθεί ιδιαίτερη έμφαση στη μείωση της πιθανότητας λανθασμένων ή παραπλανητικών απαντήσεων, ενισχύοντας την αξιοπιστία του μοντέλου.

Claude 3

Η νέα σειρά μοντέλων Claude 3 της Anthropic, που παρουσιάστηκε τον Μάρτιο του 2024, περιλαμβάνει τρία μοντέλα αυξανόμενης πολυπλοκότητας: το Claude 3

Haiku, το Claude 3 Sonnet και το Claude 3 Opus. Το Claude 3 Haiku, με περίπου 15 δισεκατομμύρια παραμέτρους, εστιάζει σε σύντομες μορφές κειμένου, όπως τίτλους και λεζάντες, προσφέροντας ταχύτερη απόκριση και μειώνοντας το ποσοστό λανθασμένων απαντήσεων σε μη εξειδικευμένες εργασίες. Το Claude 3 Sonnet, με περίπου 50 δισεκατομμύρια παραμέτρους, διευρύνει τις δυνατότητες της σειράς, επιτρέποντας την παραγωγή πιο σύνθετων, λογοτεχνικά επεξεργασμένων κειμένων, ενισχύοντας την κατανόηση συμφραζομένων και βελτιώνοντας την ακρίβεια σε τεχνικές θεματολογίες, από την πολυσημία έως τη μετάφραση πολύπλοκων εννοιών. Το Claude 3 Opus, με περίπου 200 δισεκατομμύρια παραμέτρους, αντιπροσωπεύει το ανώτερο επίπεδο της κλίμακας, καλύπτοντας ευρύ φάσμα εργασιών μεγάλης πολυπλοκότητας, όπως η ανάλυση επιστημονικών και νομικών κειμένων, η σύνθεση λεπτομερών αναφορών και η αξιοποίηση πληροφοριών υψηλής εξειδίκευσης, ενώ παράλληλα εμφανίζει ενισχυμένη ευθυγράμμιση με ανθρώπινες αξίες και κατευθυντήριες αρχές.

2.4.2 LLaMA μοντέλα

LLaMA-1

Η σειρά LLaMA (Large Language Model Meta AI), αναπτύχθηκε από τη Meta, με στόχο τη διεύρυνση της πρόσβασης στις προηγμένες δυνατότητες επεξεργασίας φυσικής γλώσσας. Η Meta δημιούργησε τα μοντέλα LLaMA με έμφαση στη μείωση του μεγέθους τους, διατηρώντας όμως την υψηλή απόδοσή τους, ενώ παράλληλα τα κατέστησε διαθέσιμα ως ανοιχτού κώδικα για την ερευνητική κοινότητα. Η εκπαίδευση των μοντέλων LLaMA-1 πραγματοποιήθηκε με τη χρήση δημοσίως διαθέσιμων δεδομένων, όπως το Common Crawl, το C4, το GitHub, και άλλα, αξιοποιώντας περίπου 1 τρισεκατομμύριο τοκενς [28].

LLaMA-2

Η νέα γενιά μοντέλων σχεδιάστηκε ώστε να είναι πιο αποδοτική και κατάλληλη για εξειδικευμένες εργασίες, όπως η δημιουργία κειμένων, η περίληψη και η απάντηση σε ερωτήσεις. Τα διαθέσιμα μεγέθη του LLaMA-2 περιλάμβαναν τα LLaMA-2-7B, LLaMA-2-13B, και LLaMA-2-70B παραμέτρων. Αντιμετωπίζοντας τις ανησυχίες που έχουν εκφραστεί για την ασφάλεια των μεγάλων γλωσσικών μοντέλων [29], η Meta ενσωμάτωσε βελτιωμένες τεχνικές ευθυγράμμισης (alignment), ώστε να περιοριστούν

οι πιθανοί κίνδυνοι από κακή χρήση των μοντέλων. Αυτές οι βελτιώσεις περιλάμβαναν καλύτερη διαχείριση των απαντήσεων και περιορισμό των ανεπιθύμητων συμπεριφορών.

LLaMA-3

Το LLaMA-3 είναι η τρίτη γενιά της σειράς μεγάλων γλωσσικών μοντέλων της Meta. δύο εκδόσεις, 8B και 70B παραμέτρων. Τα μοντέλα αυτά εκπαιδεύτηκαν σε περίπου 15 τρισεκατομμύρια tokens κειμένου, αντλώντας δεδομένα από δημόσια διαθέσιμες πηγές, εξασφαλίζοντας έτσι πλούσιο και ποικιλόμορφο εκπαιδευτικό υλικό. Σύμφωνα με τις αξιολογήσεις της Meta AI που διεξήχθησαν τον Απρίλιο του 2024, το LLaMA-3 70B παρουσίασε ανώτερες επιδόσεις σε σχέση με κορυφαία εμπορικά μοντέλα, όπως το Gemini Pro 1.5 και το Claude 3 Sonnet, ξεπερνώντας τα στις περισσότερες μετρήσεις [31].

	Finetuned	Multilingual	Long context	Tool use	Release
Llama 3 8B	✗	✗[1]	✗	✗	April 2024
Llama 3 8B Instruct	✓	✗	✗	✗	April 2024
Llama 3 70B	✗	✗[1]	✗	✗	April 2024
Llama 3 70B Instruct	✓	✗	✗	✗	April 2024
Llama 3.1 8B	✗	✓	✓	✗	July 2024
Llama 3.1 8B Instruct	✓	✓	✓	✓	July 2024
Llama 3.1 70B	✗	✓	✓	✗	July 2024
Llama 3.1 70B Instruct	✓	✓	✓	✓	July 2024
Llama 3.1 405B	✗	✓	✓	✗	July 2024
Llama 3.1 405B Instruct	✓	✓	✓	✓	July 2024

Εικόνα 5Επισκόπηση των μοντέλων LLaMA 3[31].

Category	Benchmark	Llama 3 8B	Gemma 2 9B	Mistral 7B	Llama 3 70B	Mixtral 8x22B	GPT 3.5 Turbo	Llama 3 405B	Nemotron 4 340B	GPT-4 (0125)	GPT-4o	Claude 3.5 Sonnet
General	MMLU (5-shot)	69.4	**72.3**	61.1	**83.6**	76.9	70.7	87.3	82.6	85.1	89.1	**89.9**
	MMLU (0-shot, CoT)	**73.0**	72.3△	60.5	**86.0**	79.9	69.8	88.6	78.7◁	85.4	**88.7**	88.3
	MMLU-Pro (5-shot, CoT)	**48.3**	–	36.9	**66.4**	56.3	49.2	73.3	62.7	64.8	74.0	**77.0**
	IFEval	**80.4**	73.6	57.6	**87.5**	72.7	69.9	**88.6**	85.1	84.3	85.6	88.0
Code	HumanEval (0-shot)	**72.6**	54.3	40.2	**80.5**	75.6	68.0	89.0	73.2	86.6	90.2	**92.0**
	MBPP EvalPlus (0-shot)	**72.8**	71.7	49.5	**86.0**	78.6	82.0	88.6	72.8	83.6	87.8	**90.5**
Math	GSM8K (8-shot, CoT)	**84.5**	76.7	53.2	**95.1**	88.2	81.6	**96.8**	92.3◇	94.2	96.1	96.4◇
	MATH (0-shot, CoT)	**51.9**	44.3	13.0	**68.0**	54.1	43.1	73.8	41.1	64.5	**76.6**	71.1
Reasoning	ARC Challenge (0-shot)	83.4	**87.6**	74.2	**94.8**	88.7	83.7	**96.9**	94.6	96.4	96.7	96.7
	GPQA (0-shot, CoT)	32.8	–	28.8	**46.7**	33.3	30.8	51.1	–	41.4	53.6	**59.4**
Tool use	BFCL	**76.1**	–	60.4	84.8	–	**85.9**	88.5	86.5	88.3	80.5	**90.2**
	Nexus	**38.5**	30.0	24.7	**56.7**	48.5	37.2	**58.7**	–	50.3	56.1	45.7
Long context	ZeroSCROLLS/QuALITY	81.0	–	–	90.5	–	–	**95.2**	–	**95.2**	90.5	90.5
	InfiniteBench/En.MC	65.1	–	–	78.2	–	–	**83.4**	–	72.1	82.5	–
	NIH/Multi-needle	98.8	–	–	97.5	–	–	98.1	–	**100.0**	**100.0**	90.8
Multilingual	MGSM (0-shot, CoT)	**68.9**	53.2	29.9	**86.9**	71.1	51.4	**91.6**	–	85.9	90.5	**91.6**

Εικόνα 6. Συγκριτική επίδοση fine tuned LLaMA 3 με άλλα ισχυρά γλωσσικά μοντέλα [31].

.

Κεφάλαιο 3. Μεγάλα γλωσσικά μοντέλα

Στον χώρο των χρηματοοικονομικών, εκτός από την παραδοσιακή ανάλυση ποσοτικών δεδομένων, ο τρόπος με τον οποίο παρουσιάζονται οι χρηματοοικονομικές αναφορές, οι αναφορές κερδών, τα ειδησεογραφικά άρθρα και οι αναρτήσεις στα μέσα κοινωνικής δικτύωσης διαδραματίζει καθοριστικό ρόλο στη διαμόρφωση των επενδυτικών αποφάσεων όσων συμμετέχουν στο οικονομικό οικοσύστημα. Η ανάλυση του συναισθήματος και του νοήματος που αποπνέουν αυτά τα κείμενα προσφέρει πολύτιμες πληροφορίες για την κατανόηση της αγοράς και τη λήψη στρατηγικών αποφάσεων [39].

3.1 Εφαρμογές LLMs στην οικονομία

Τα Large Language Models (LLMs), που έχουν εκπαιδευτεί σε μεγάλες και ποικιλόμορφες συλλογές δεδομένων, ξεχωρίζουν για την ικανότητά τους να κατανοούν τη γλώσσα σε βάθος, να ερμηνεύουν το πλαίσιο και να αναγνωρίζουν λεπτές αποχρώσεις της σημασίας. Αυτό τα καθιστά χρήσιμα για εφαρμογές όπως η πρόβλεψη χρηματοοικονομικών τάσεων, η αξιολόγηση της διάθεσης των επενδυτών και η ανάλυση της επίδρασης των χρηματοοικονομικών ειδήσεων στην αγορά. Η χρήση αυτών των μοντέλων προσφέρει στους επαγγελματίες του κλάδου τη δυνατότητα να αξιοποιήσουν τη γλώσσα ως εργαλείο για πιο ακριβείς και αποτελεσματικές επενδυτικές στρατηγικές [40].

3.1.1 Πρόβλεψη χρηματιστηριακών τιμών

Τα γεγονότα "Μαύρου Κύκνου" όπως η παγκόσμια χρηματοοικονομική κρίση του 2007-2008, αναφέρονται σε δραματικές μεταβολές ή αλλαγές στις τιμές των μετοχών, οι οποίες φαίνεται να αψηφούν κάθε λογική εξήγηση. Το τυπικό χρηματοοικονομικό μοντέλο, το οποίο βασίζεται στην υπόθεση ότι οι επενδυτές δρουν με ψυχρή λογική και διασφαλίζουν ότι οι τιμές στις αγορές κεφαλαίων ισούνται με την ορθολογική παρούσα αξία των αναμενόμενων μελλοντικών ταμειακών ροών, συχνά αποτυγχάνει να εξηγήσει τέτοια φαινόμενα [41].

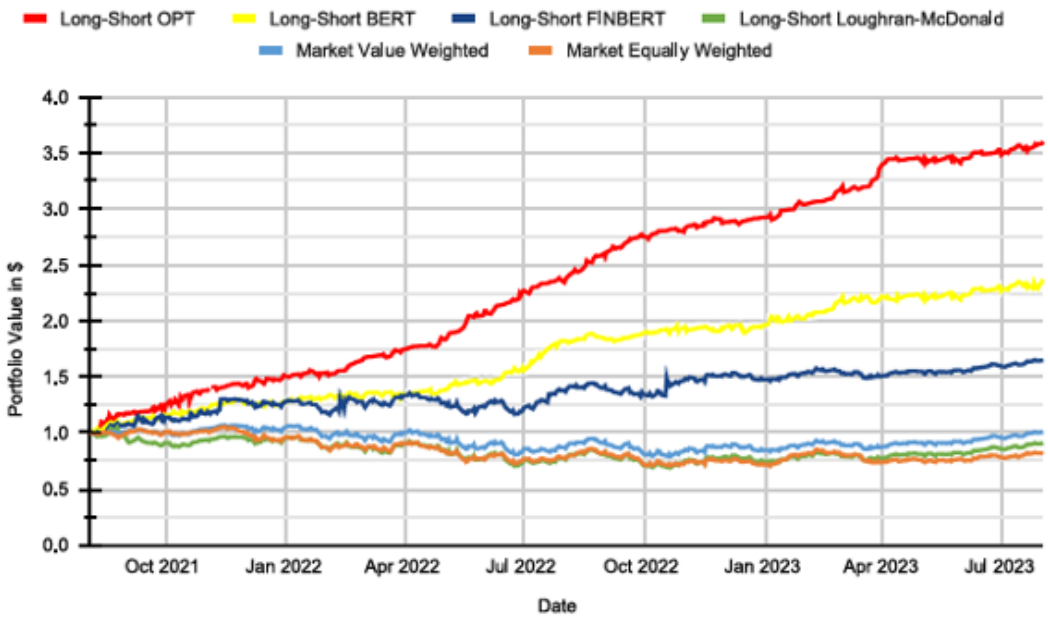
Για να καλυφθεί αυτό το κενό, οι ερευνητές στον τομέα των συμπεριφορικών χρηματοοικονομικών εργάζονται πάνω σε ένα εναλλακτικό μοντέλο, το οποίο στηρίζεται σε δύο βασικές υποθέσεις:

1. Την επενδυτική ψυχολογία**.** Σύμφωνα με την εργασία των Delong, Shleifer, Summers και Waldmann (1990) [42], οι επενδυτές υπόκεινται σε συναισθήματα και προκαταλήψεις. Η επενδυτική ψυχολογία, ορισμένη ευρέως, περιγράφει την πίστη των επενδυτών για τις μελλοντικές ταμειακές ροές και τους κινδύνους που σχετίζονται με τις επενδύσεις, η οποία όμως δεν δικαιολογείται από τα πραγματικά δεδομένα. Αυτή η διαστρέβλωση οδηγεί τους επενδυτές να υπερεκτιμούν ή να υποτιμούν τις προοπτικές μιας αγοράς, με αποτέλεσμα ακραίες αποκλίσεις τιμών.
2. Το κόστος και ο κίνδυνος αντιστροφής των τιμών. Όπως τόνισαν οι Shleifer και Vishny (1997) [43], η αντίδραση απέναντι στους επενδυτές που δρουν υπό το κράτος του συναισθήματος δεν είναι απλή ούτε οικονομικά αποδοτική. Οι ορθολογικοί επενδυτές αντιμετωπίζουν υψηλό κόστος και σημαντικό κίνδυνο όταν επιχειρούν να διορθώσουν αυτές τις στρεβλώσεις και να επαναφέρουν τις τιμές σε ορθολογικά επίπεδα. Για τον λόγο αυτό, η "αγορά" αργεί να προσαρμοστεί στα θεμελιώδη μεγέθη και παραμένει ευάλωτη σε συναισθηματικά φορτισμένες μεταβολές.

Οι Kirtac και Germano [44] εξετάζουν την αποτελεσματικότητα LLMs, όπως του FinBERT, στην ανάλυση συναισθήματος οικονομικών ειδήσεων και στην πρόβλεψη των αποδόσεων της αγοράς. Η ανάλυση πραγματοποιήθηκε σε σύνολα δεδομένων που περιλάμβαναν περισσότερα από 965.000 άρθρα ειδήσεων από το Refinitiv από την 1^{η} Ιανουαρίου, 2010 έως την 30^{η} Ιουνίου, 2023.

Οι ερευνητές προχώρησαν στην ανάπτυξη στρατηγικών συναλλαγών βασισμένων στις βαθμολογίες συναισθήματος που εξάγονται από τα LLMs. Οι στρατηγικές αυτές περιλαμβάνουν τη δημιουργία long, short και long-short χαρτοφυλακίων, λαμβάνοντας υπόψη τις βαθμολογίες που προκύπτουν από μοντέλα όπως το OPT, το BERT, το FinBERT και το λεξικό Loughran-McDonald το οποίο είναι ένα εξειδικευμένο λεξικό για οικονομικά κείμενα και χρησιμοποιείται στην παραδοσιακή ανάλυση κειμένου όπου δεν χρησιμοποιείται μηχανική μάθηση. Στην μεθοδολογία ενσωμάτωσαν ρεαλιστικές συνθήκες συναλλαγών, όπως το κόστος συναλλαγών και ο συγχρονισμός

με τις δημοσιεύσεις ειδήσεων, προσφέροντας έτσι μια πιο ακριβή εικόνα για την αποτελεσματικότητα των στρατηγικών σε πραγματικά περιβάλλοντα.



Εικόνα 7. Σωρευτικές αποδόσεις από την επένδυση $1 με σταθμισμένα ως προς την αξία, μηδενικού κόστους long-short χαρτοφυλάκια βασισμένα σε OPT (κόκκινο), BERT (κίτρινο), FinBERT (σκούρο μπλε) και το λεξικό Loughran-McDonald (πράσινο), με καθημερινό κόστος συναλλαγής 10 μονάδων βάσης. Επίσης, παρουσιάζονται ένα σταθμισμένο ως προς την αξία χαρτοφυλάκιο της αγοράς (γαλάζιο) και ένα ισοσταθμισμένο χαρτοφυλάκιο αγοράς (πορτοκαλί), και τα δύο χωρίς κόστος συναλλαγής [44].

	OPT			BERT			FinBERT		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Sharpe ratio	1.81	1.42	3.05	1.59	1.28	2.11	1.51	1.19	2.07
MDR (%)	0.32	0.25	0.55	0.25	0.21	0.45	0.22	0.18	0.39
StdDev (%)	2.18	2.91	2.49	2.49	3.19	2.68	2.59	3.31	2.81
MDD (%)	−14.76	−24.69	−18.57	−17.89	−27.95	−21.95	−19.71	−29.94	−23.82
	LM dictionary			EW			VW		
	Long	Short	L-S	Long	Short	L-S	Long	Short	L-S
Sharpe ratio	0.87	0.66	1.23	1.25	1.05	1.40	1.28	1.08	1.45
MDR (%)	0.12	0.13	0.22	0.18	0.15	0.33	0.19	0.16	0.35
StdDev (%)	3.54	4.13	3.74	2.90	3.70	3.20	2.95	3.75	3.25
MDD (%)	−35.47	−45.39	−38.29	−31.13	−42.21	−32.87	−28.76	−38.95	−31.87

Εικόνα 8. Περιγραφικά στατιστικά στοιχεία συναλλαγών για διαφορετικά μοντέλα [44].

Οι Γ. Ιακωβίδης et.al. [45] ανέπτυξαν το **FinLlama**, ένα LLM βασισμένο στο Llama 2 7B και σκοπό είχε να βοηθήσει στις αλγοριθμικές συναλλαγές. Οι ερευνητές χρησιμοποίησαν τέσσερα επισημασμένα σύνολα δεδομένων οικονομικού κειμένου για την προσαρμογή του μοντέλου Llama 2, με σκοπό το μοντέλο να έχει την ικανότητα κατανόησης της γλώσσας που συναντάται στον χρηματοοικονομικό τομέα. Η ειδική

αυτή εκπαίδευση δίνει στο μοντέλο τη δυνατότητα να επεξεργάζεται και να αναλύει δεδομένα με τρόπο που αντικατοπτρίζει τις ιδιαιτερότητες του συγκεκριμένου τομέα.

Για την αλλαγή της βασικής λειτουργίας του **Llama 2** από τη δημιουργία κειμένου στην ταξινόμηση συναισθήματος, εφαρμόζουν ένα επίπεδο **SoftMax** ταξινόμησης τριών κατηγοριών στην έξοδο του μοντέλου. Αυτή η προσέγγιση κάνει το προτεινόμενο μοντέλο έναν συνδυασμό γεννήτριας-διακριτή, ο οποίος παράγει αποφάσεις συναισθήματος για τρεις κατηγορίες: θετικό, αρνητικό ή ουδέτερο.

Επιπλέον, εφαρμόζονται πέντε μέθοδοι ανάλυσης συναισθήματος. Για τις μεθόδους που βασίζονται σε λεξικά, χρησιμοποιούνται τα **Loughran-McDonald (LMD)** και **HIV-4**, καθώς και ο αλγόριθμος **VADER**. Όσον αφορά τις μεθόδους βαθιάς μάθησης, χρησιμοποιούν τα μοντέλα **FinBERT** και **FinLlama**. Η ανάλυση συναισθήματος που παράγεται από κάθε μέθοδο χρησιμοποιείται για την κατασκευή μακροπρόθεσμων-βραχυπρόθεσμων χαρτοφυλακίων (long-short portfolios).

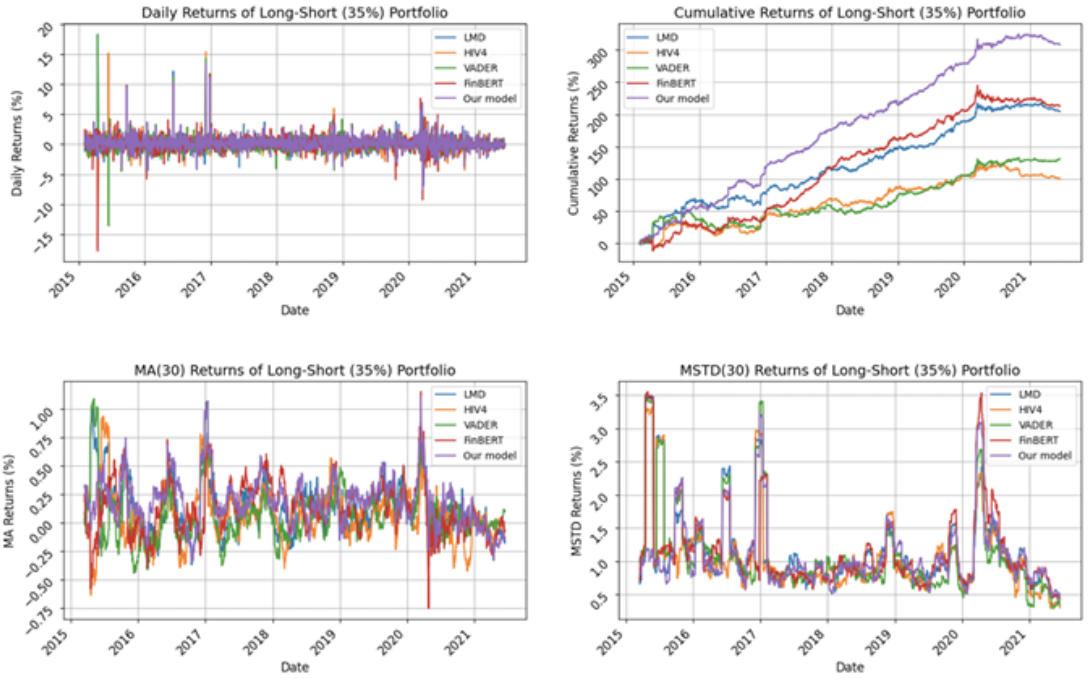
Οι ερευνητές κατατάσσουν καθημερινά τις εταιρείες σύμφωνα με το συναίσθημα που παράγεται από κάθε μέθοδο. Για τις εταιρείες που δεν υπήρχαν δεδομένα συναισθήματος για μια συγκεκριμένη ημέρα, εξαιρούνται από την κατάταξη. Οι εταιρείες με τις υψηλότερες θετικές τιμές κατατάσσονται σε μακροπρόθεσμες θέσεις, ενώ εκείνες με τις ισχυρότερες αρνητικές τιμές κατατάσσονται σε βραχυπρόθεσμες θέσεις.

Στην συνέχεια οι ερευνητές χρησιμοποιούν μια ισοσταθμισμένη στρατηγική χαρτοφυλακίου, μια τακτική που χρησιμοποιείται ευρέως από hedge funds. Καθορίζουν το 35% των εταιρειών με τις καλύτερες επιδόσεις για μακροπρόθεσμες θέσεις, ενώ το 35% με τις χαμηλότερες επιδόσεις κατατάσσεται σε βραχυπρόθεσμες θέσεις. Στόχος της προσέγγισης είναι η μεγιστοποίηση των αποδόσεων μέσω της σωστής επιλογής θέσεων βασισμένων σε συναισθηματικά δεδομένα.

Οι μέθοδοι βαθιάς μάθησης υπερείχαν σημαντικά των μεθόδων που βασίζονται σε λεξικά, όσον αφορά τις αποδόσεις. Ιδιαίτερα, οι μέθοδοι που στηρίζονται σε γενικού σκοπού λεξικά, όπως τα **HIV-4** και **VADER**, παρουσίασαν χαμηλότερες επιδόσεις. Αυτό ήταν αναμενόμενο, καθώς οι μέθοδοι που βασίζονται σε λεξικά συχνά αδυνατούν να αποδώσουν το συμφραστικό νόημα των προτάσεων, ενώ η εξειδικευμένη φύση των

οικονομικών κειμένων μειώνει ακόμη περισσότερο την ακρίβεια των γενικού σκοπού λεξικών.

Η διαφορά στις σωρευτικές αποδόσεις μεταξύ του προτεινόμενου μοντέλου **FinLlama** και της καλύτερης μεθόδου από αυτές που εξετάστηκαν αυξήθηκε με την πάροδο του χρόνου. Η σημαντική υπεροχή του **FinLlama** μετά το 2019 μπορεί να εξηγηθεί από την αύξηση του ημερήσιου μέσου αριθμού εταιρειών που διαπραγματεύονταν, κάτι που οδήγησε σε μεγαλύτερο αριθμό άρθρων κατά τη διάρκεια των ετών.



Εικόνα 9. Σύγκριση της απόδοσης των χαρτοφυλακίων long-short που κατασκευάστηκαν χρησιμοποιώντας πέντε μοντέλα για την ανάλυση συναισθημάτων για το χρονικό διάστημα από τον Φεβρουάριο του 2015 έως τον Ιούνιο του 2021. MA(30) και MSTD(30) αντιπροσωπεύουν τον κινητό μέσο όρο και την κινούμενη τυπική απόκλιση, αντίστοιχα. Οι αποδόσεις υπολογίζονται σε κυλιόμενο παράθυρο 30 ημερών [45].

	LMD	HIV-4	VADER	FinBERT	FinLlama (Ours)	S&P 500
Cumulative Returns (%)	204.6	100.4	130.6	213.0	**308.2**	83.1
Annualized Return (%)	29.1	13.5	17.9	30.3	**45.0**	11.3
Sharpe Ratio	1.5	0.7	0.9	1.5	**2.4**	0.62
Annualized Volatility (%)	19.5	18.9	19.6	20.3	18.6	**18.5**

Εικόνα 10. Σύγκριση μεταξύ των πέντε μεθόδων ανάλυσης συναισθήματος χρησιμοποιώντας ένα χαρτοφυλάκιο long-short 35%. Για τις σωρευτικές αποδόσεις, την ετήσια απόδοση και την αναλογία Sharpe, το υψηλότερο είναι καλύτερο. Για την ετήσια μεταβλητότητα, το χαμηλότερο [45].

3.1.2 Λογιστική

Στον τομέα της λογιστικής και των χρηματοοικονομικών αναφορών, τα LLMs έχουν τη δυνατότητα να φέρουν αλλαγές στην ανάλυση δεδομένων, την προετοιμασία οικονομικών καταστάσεων και τη συμμόρφωση με κανονισμούς [46]. Ο S. Lehner [47] μελέτησε την εφαρμογή LLMs στις ενότητες Management Discussion and Analysis (MD&A) στις λογιστικές αναφορές 10-K και 10-Q που ζητάνε οι αρχές των ΗΠΑ. Αυτές οι ενότητες αποτελούν κρίσιμα στοιχεία της εταιρικής αναφοράς, καθώς παρέχουν τη διοικητική άποψη σχετικά με την οικονομική κατάσταση, τους κινδύνους και τις προοπτικές της εταιρείας. Αυτές οι ενότητες που έχουν και αφηγηματικά στοιχεία, συχνά διαδραματίζουν καθοριστικό ρόλο στη διαμόρφωση των αντιλήψεων των επενδυτών, καθιστώντας την ανάλυση συναισθήματος αναγκαίο εργαλείο για την αξιολόγηση του τόνου αυτών των γνωστοποιήσεων. Θετικό συναίσθημα σε αυτές τις ενότητες μπορεί να σηματοδοτεί εμπιστοσύνη και ανάπτυξη, ενώ αρνητικό συναίσθημα μπορεί να εγείρει ανησυχίες για κινδύνους ή προκλήσεις απόδοσης.

Στην εργασία, εξετάζεται η εφαρμογή των LLMs για την ενίσχυση του συναισθήματος στις ενότητες MD&A των καταθέσεων της Επιτροπής Κεφαλαιαγοράς (SEC). Μέσω των LLMs, τα κείμενα αναδιατυπώνονται με πιο θετικό τόνο. Η έρευνα διερευνά πώς αυτή η μεταβολή συναισθήματος επηρεάζει τη σχέση μεταξύ συναισθήματος και αποδόσεων μετοχών. Η κεντρική υπόθεση είναι ότι τα LLMs μπορούν να χρησιμοποιηθούν στρατηγικά για να ενισχύσουν την επίδραση των οικονομικών καταθέσεων στις τιμές των μετοχών.

Τα αποτελέσματα δείχνουν ότι τα LLMs είναι ιδιαίτερα αποτελεσματικά στην ενίσχυση του θετικού συναισθήματος, το οποίο συσχετίζεται θετικά με μεγαλύτερες αποδόσεις μετοχών. Αυτό υπογραμμίζει τη δυνατότητα των LLMs να χρησιμοποιηθούν ως εργαλεία για τη βελτιστοποίηση της παρουσίασης οικονομικών γνωστοποιήσεων με σκοπό την επιρροή της συμπεριφοράς των επενδυτών.

Τα αποτελέσματα αποκαλύπτουν σημαντικές αλλαγές στο συναίσθημα μεταξύ των αρχικών και των αναδιατυπωμένων οικονομικών αναφορών. Στις αρχικές καταθέσεις 10-K, το συναίσθημα παραμένει γενικά ουδέτερο, με σχετικά χαμηλά έως μέτρια ποσοστά θετικών λέξεων, προτάσεων και αναφορών, καθώς και μέτριες βαθμολογίες μέσου συναισθήματος. Αυτός ο ουδέτερος τόνος φαίνεται να συνάδει με

τη φύση των επίσημων χρηματοοικονομικών γνωστοποιήσεων, οι οποίες συνήθως επιδιώκουν να παρουσιάσουν πληροφορίες με αντικειμενικό τρόπο.

Ωστόσο, μετά τη διαδικασία αναδιατύπωσης, σημειώνεται σημαντική αύξηση στις μετρήσεις συναισθήματος, ιδιαίτερα στις εκδοχές που δημιουργήθηκαν από το **GPT-4o** και, σε μικρότερο βαθμό, από το **GPT-4o-mini**.

Measure	Approach	Original	GPT-4o	GPT-4o-mini
Positive Words	L&M Dictionary	38.3%	67.9%	64.4%
Positive Sentences	SVM (Embedding)	49.8%	81.7%	75.9%
	XGBoost (Embedding)	60.0%	84.2%	80.2%
	Ensemble (Embedding)	54.8%	84.4%	79.3%
Positive Filings	FinBERT	16.4%	57.7%	45.7%
	LSTM	16.0%	32.7%	39.3%
	SVM (TF-IDF)	43.3%	87.9%	75.6%
	XGBoost (TF-IDF)	61.2%	96.1%	94.9%
	Ensemble (TF-IDF)	23.6%	82.7%	67.7%
Average Rating	GPT-4o	4.72	7.06	6.80
	GPT-4o-mini	5.49	7.55	7.36

Εικόνα 11. Σύγκριση συναισθήματος διαφορετικών μοντέλων σε πρωτότυπα και ξαναγραμμένα 10-K αναφορών [47].

Από τα παραπάνω εγείρονται ανησυχίες σχετικά με την ακρίβεια και την ακεραιότητα της αυτοματοποιημένης αναδιατύπωσης σε οικονομικές αναφορές. Η πιθανότητα τα LLMs να χρησιμοποιηθούν για την παρουσίαση υπερβολικά αισιόδοξων ή παραπλανητικών αναλύσεων δημιουργεί προβληματισμούς για τον ρόλο τους στη διασφάλιση της διαφάνειας και της αντικειμενικότητας των αναφορών. Επιπλέον, καθώς οι επενδυτές βασίζονται όλο και περισσότερο στις αναφορές αυτές για τη λήψη αποφάσεων, οποιαδήποτε μορφή προκατάληψης ή υπερβολικής θετικότητας μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις κινδύνου ή απόδοσης. Οπότε κρίνεται αναγκαία η εισαγωγή κανονιστικών πλαισίων που θα ορίζουν σαφή όρια και κατευθυντήριες γραμμές για τη χρήση τεχνητής νοημοσύνης σε οικονομικές αναφορές. Επίσης, πριν τη χρήση τους σε οικονομικές αναφορές, τα LLMs θα πρέπει να υποβάλλονται σε ενδελεχείς ελέγχους για την ανίχνευση πιθανής προκατάληψης και την επιβεβαίωση ότι αποδίδουν το περιεχόμενο με ακρίβεια.

3.1.3 Πρόβλεψη μακροοικονομικών δεικτών

Η εκτίμηση μακροοικονομικών δεικτών αποτελεί μια ιδιαίτερα απαιτητική διαδικασία, καθώς προϋποθέτει την κατανόηση της πολύπλοκης αλληλεπίδρασης πλήθους παραγόντων. Δείκτες όπως το Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ), οι ρυθμοί πληθωρισμού και τα ποσοστά ανεργίας επηρεάζονται από ποικίλες μεταβλητές, όπως οι παγκόσμιες και εθνικές οικονομικές τάσεις, οι πολιτικές αποφάσεις, η πολιτική σταθερότητα, αλλά και αστάθμητοι παράγοντες, όπως πανδημίες ή φυσικές καταστροφές. Επιπλέον, η απρόβλεπτη φύση της ανθρώπινης συμπεριφοράς μπορεί να προκαλέσει αιφνίδιες αλλαγές στις αγορές ή στη γενική εμπιστοσύνη του πληθυσμού [48].

Η ανάλυση συναισθήματος στις οικονομικές ειδήσεις είναι μία προσέγγιση για την αντιμετώπιση αυτών των προκλήσεων. Μέσω αυτής, παρέχονται άμεσα πληροφορίες για τη διάθεση του κοινού και τις τάσεις της αγοράς, αποτυπώνοντας ψυχολογικές και συμπεριφορικές πτυχές που συχνά διαφεύγουν από τα παραδοσιακά οικονομικά μοντέλα. Με τον τρόπο αυτό, η ανάλυση συναισθήματος προσφέρει μια πιο ολοκληρωμένη εικόνα των οικονομικών δυναμικών, συμβάλλοντας στην κατανόηση της εμπιστοσύνης των καταναλωτών, των επενδυτικών τάσεων και της γενικότερης πορείας των αγορών.

Οι Carriero et al. [49] πραγματοποίησαν μια ανάλυση της χρήσης LLMs για την πρόβλεψη μακροοικονομικών χρονικών σειρών, με έμφαση στη δυνατότητα εφαρμογής τους για το nowcasting μακροοικονομικών δεικτών. Τα αποτελέσματά τους ανέδειξαν ότι συγκεκριμένα μοντέλα, όπως το **Moirai** της Salesforce και το **TimesFM** της Google, μπορούν να ανταγωνιστούν τις παραδοσιακές οικονομετρικές μεθόδους όσον αφορά την ακρίβεια. Ωστόσο, αυτά τα μοντέλα παρουσίασαν μεγαλύτερη μεταβλητότητα στις προβλέψεις τους, γεγονός που αποτελεί κρίσιμο παράγοντα κατά την εφαρμογή τους σε πρακτικά σενάρια.

Η μεταβλητότητα αυτή αποκαλύπτει μια πρόκληση στη χρήση των LLMs για προβλέψεις, όπου η συνέπεια και η αξιοπιστία είναι απαραίτητες. Σε πρακτικές εφαρμογές, όπως η πρόβλεψη οικονομικών δεικτών, η σταθερότητα των μοντέλων είναι μεγάλης σημασίας για τη λήψη τεκμηριωμένων αποφάσεων. Παρά την πρόοδο που έχουν σημειώσει τα LLMs, τα ευρήματα της έρευνας υποδεικνύουν ότι η

ενσωμάτωση αυτών των μοντέλων σε κρίσιμες οικονομικές εφαρμογές απαιτεί περαιτέρω βελτιστοποίηση, ώστε να μειωθεί η αστάθεια και να ενισχυθεί η εμπιστοσύνη στη χρήση τους.

Η εργασία των Όnozó, Arthur, και Gyires-Tóth (2024) εστιάζει στη χρήση LLMs για την ανάλυση χρηματοοικονομικών ειδήσεων και την πρόβλεψη μακροοικονομικών δεικτών, όπως το Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ), ο Δείκτης Υπευθύνων Προμηθειών (PMI) και το ποσοστό ανεργίας. Ο στόχος της μελέτης έγκειται στην εφαρμογή τεχνικών ανάλυσης συναισθήματος, οι οποίες επιτρέπουν την εξαγωγή έγκαιρων και χρήσιμων πληροφοριών από οικονομικά άρθρα. Με τη χρήση λεξικογραφικών μεθόδων και μοντέλων Transformer, όπως το BERT, οι ερευνητές ανέπτυξαν εργαλεία που συσχετίζουν συναισθηματικούς δείκτες με μακροοικονομικά δεδομένα. Παράλληλα, αξιοποίησαν ενεργή μάθηση (active learning) σε συνδυασμό με LLMs, όπως το GPT-4, για την αυτοματοποίηση της κατηγοριοποίησης δεδομένων, επιτυγχάνοντας έτσι μεγαλύτερη ακρίβεια και συνέπεια στα αποτελέσματα.

Η εργασία δείχνει ότι η ανάλυση συναισθήματος μπορεί να αποτελέσει ένα χρήσιμο εργαλείο για την άμεση εκτίμηση της οικονομικής κατάστασης, παρέχοντας πληροφορίες σε πραγματικό χρόνο που συχνά προηγούνται των επίσημων στατιστικών. Μέσω της σύγκρισης των παραγόμενων δεικτών συναισθήματος με επίσημους οικονομικούς δείκτες, οι ερευνητές επιβεβαίωσαν την αξία αυτής της προσέγγισης ως εργαλείου πρόβλεψης (nowcasting). Οπότε τα LLMs προσφέρουν σημαντικές δυνατότητες για κεντρικές τράπεζες, οικονομικούς φορείς και επενδυτές, διευκολύνοντας τη λήψη τεκμηριωμένων αποφάσεων.

3.2 Τεχνικές προσαρμογής

Τα LLMs βασίζονται σε εκτενή σύνολα δεδομένων για την εκπαίδευσή τους απαιτώντας τεράστιους υπολογιστικούς πόρους. Αυτή η απαίτηση καθιστά την εκ νέου εκπαίδευση τους ιδιαίτερα απαιτητική. Παρά τις προκλήσεις αυτές, έχουν αναπτυχθεί μέθοδοι που επιτρέπουν την προσαρμογή προ-εκπαιδευμένων μοντέλων, αξιοποιώντας την προηγούμενη εκπαίδευση τους για την κατανόηση γλώσσας και εξειδικεύοντας τις δυνατότητές τους σε συγκεκριμένα προβλήματα.

Η προσαρμογή των LLMs σε δύσκολες εργασίες, όπως η ανάλυση συναισθήματος ή η κατηγοριοποίηση κειμένων, επιτυγχάνεται κυρίως με δύο βασικές στρατηγικές. Η πρώτη αφορά τη χρήση fine-tuning, όπου το μοντέλο εκπαιδεύεται περαιτέρω σε εξειδικευμένα δεδομένα, βελτιστοποιώντας τις παραμέτρους του για συγκεκριμένη χρήση. Η δεύτερη στρατηγική περιλαμβάνει τεχνικές βασισμένες σε προτροπές (prompts), οι οποίες διακρίνονται σε prompt engineering, όπου σχεδιάζονται εξειδικευμένες οδηγίες για την προσαρμογή της λειτουργίας του μοντέλου, και prompt-tuning, όπου εκπαιδεύονται συγκεκριμένες παραμέτροι για τη βελτίωση της απόδοσης του μοντέλου σε προκαθορισμένες εργασίες.

3.2.1 Fine-tuning

Η μέθοδος Universal Language Model Fine-tuning (**ULMFiT**), αναπτύχθηκε από τους Jeremy Howard και Sebastian Ruder [50], και αποτελεί μια μεθοδολογία για τη μεταφορά μάθησης στην NLP. η ULMFiT προσαρμόζει της μεταφοράς μάθησης στην υπολογιστική όραση στις ανάγκες του NLP, προσφέροντας μια καθολική, αποδοτική και ευέλικτη μέθοδο που μπορεί να εφαρμοστεί σε πλήθος εργασιών χωρίς να απαιτείται εκπαίδευση από την αρχή.

Η ULMFiT βασίζεται στη χρήση ενός γλωσσικού μοντέλου που έχει προεκπαιδευτεί σε ένα μεγάλο γενικού τομέα σύνολο δεδομένων, όπως η Wikipedia, για να καταγράψει γενικά χαρακτηριστικά της γλώσσας. Η μέθοδος ενσωματώνει τρία βασικά στάδια για την προσαρμογή του μοντέλου σε συγκεκριμένες εργασίες. Πρώτον Το LM εκπαιδεύεται σε δεδομένα γενικού τομέα για να καταγράψει ευρύτερες γλωσσικές δομές και σχέσεις. Στην συνέχεια το LM προσαρμόζεται στα δεδομένα της εκάστοτε εργασίας, ώστε να ενσωματώσει τις ιδιαιτερότητες της εργασίας. Τέλος Προστίθενται επιπλέον γραμμικά επίπεδα στο μοντέλο, τα οποία εκπαιδεύονται ειδικά για την εργασία στόχο, διατηρώντας ταυτόχρονα τις γενικές γλωσσικές γνώσεις.

Για την αποφυγή της καταστροφικής λήθης (catastrophic forgetting) και τη βελτιστοποίηση της απόδοσης, η ULMFiT χρησιμοποιεί τρεις καινούργιες μεθόδους. Η διακριτική προσαρμογή (Discriminative Fine-Tuning) επιτρέπει τη χρήση διαφορετικών ρυθμών εκμάθησης σε κάθε επίπεδο του νευρωνικού δικτύου. Η μέθοδος Slanted Triangular Learning Rates (STLR) χρησιμοποιεί δυναμική αλλαγή των ρυθμών εκμάθησης, επιταχύνοντας τη σύγκλιση και αποτρέποντας την υπερπροσαρμογή

(overfitting). Επιπλέον, η σταδιακή αποδέσμευση (Gradual Unfreezing) προσαρμόζει το μοντέλο ξεκινώντας από τα τελευταία επίπεδα, που είναι τα λιγότερο γενικά, διασφαλίζοντας τη σταδιακή προσαρμογή χωρίς απώλεια γνώσης.

Η ULMFiT αποδείχθηκε αποτελεσματική σε διάφορες εργασίες NLP, όπως ανάλυση συναισθήματος, κατηγοριοποίηση θεμάτων και ερωτήσεων. Σε σύγκριση με τις υπάρχουσες μεθόδους, επιτυγχάνει σημαντική μείωση του σφάλματος (18-24%) σε ευρέως χρησιμοποιούμενα σύνολα δεδομένων, όπως τα IMDb, Yelp και AG News. Επίσης, η μέθοδος επιτυγχάνει υψηλή απόδοση ακόμη και σε περιπτώσεις με περιορισμένα δεδομένα ετικετών, και δίνει την δυνατότητα για αποτελεσματική μάθηση με 100 φορές λιγότερα δεδομένα από ό,τι απαιτούν οι παραδοσιακές μέθοδοι.

Οι C. Sun et.al. [51] εξετάζουν τη βελτίωση των επιδόσεων του BERT σε εργασίες ταξινόμησης κειμένων μέσω διαφόρων στρατηγικών προσαρμογής και προεκπαίδευσης, επιτυγχάνοντας νέα state-of-the-art αποτελέσματα. Το BERT εφαρμόστηκε σε οκτώ σύνολα δεδομένων, όπως τα IMDb, Yelp, TREC, Yahoo! Answers, AG's News, DBPedia και Sogou News, καλύπτοντας εργασίες ανάλυσης συναισθήματος, ταξινόμησης θεμάτων και κατηγοριοποίησης ερωτήσεων.

Η περαιτέρω προεκπαίδευση σε δεδομένα του ίδιου πεδίου ή συγκεκριμένης εργασίας είχε σημαντικά αποτελέσματα. Ειδικότερα, η προεκπαίδευση εντός εργασίας (Within-task Pre-training) βελτίωσε την απόδοση σε μικρότερες βάσεις δεδομένων, ενώ η προεκπαίδευση σε δεδομένα του ίδιου πεδίου (In-domain Pre-training) παρείχε γενικότερες βελτιώσεις. Αντίθετα, η διατομεακή προεκπαίδευση (Cross-domain Pre-training) δεν παρουσίασε σημαντικά οφέλη, υποδεικνύοντας ότι το BERT είναι ήδη επαρκώς εκπαιδευμένο σε γενικά δεδομένα.

Προτάθηκαν μέθοδοι διαχείρισης μακροσκελών κειμένων, επιλογής επιπέδων και εφαρμογής διαφορετικών ρυθμών εκμάθησης για κάθε επίπεδο του BERT, που συνέβαλαν στη βελτίωση της απόδοσης. Το φαινόμενο της καταστροφικής λήθης αντιμετωπίστηκε επιτυχώς με τη χρήση χαμηλών ρυθμών εκμάθησης και σταδιακής προσαρμογής των επιπέδων. Επιπλέον, η πολυεργασία (Multi-task Fine-Tuning), δηλαδή η ταυτόχρονη προσαρμογή σε πολλαπλές σχετικές εργασίες, παρείχε περαιτέρω βελτιώσεις, αν και η επίδραση ήταν μικρότερη όταν είχε προηγηθεί προεκπαίδευση σε δεδομένα του ίδιου πεδίου.

Εντυπωσιακή ήταν η απόδοση του BERT σε περιπτώσεις μικρών βάσεων δεδομένων. Για παράδειγμα, στο IMDb, η χρήση μόνο του 0,4% των δεδομένων εκπαίδευσης οδήγησε σε μείωση του ποσοστού σφάλματος από 17,26% σε 9,23%. Συγκριτικά με άλλα μοντέλα, όπως τα CNNs, RNNs και ULMFiT, το BERT υπερείχε σε όλες τις εργασίες. Συγκεκριμένα, οι παραλλαγές του BERT με περαιτέρω προεκπαίδευση σημείωσαν μείωση των σφαλμάτων κατά μέσο όρο 18,57%.

Η εφαρμογή των ίδιων τεχνικών στο BERT Large απέδειξε περαιτέρω τη δυναμική του μοντέλου, επιτυγχάνοντας ακόμη καλύτερες επιδόσεις σε σύγκριση με προηγμένα μοντέλα, όπως το ULMFiT. Τα αποτελέσματα της εργασίας αναδεικνύουν τη σημασία της σωστής προσαρμογής και προεκπαίδευσης των Μεγάλων Γλωσσικών Μοντέλων για τη μεγιστοποίηση των επιδόσεών τους τόσο σε μεγάλα όσο και σε μικρά σύνολα δεδομένων.

Οι M. Zhao et. al. [52] πρότειναν μια μέθοδο μείωσης του υπολογιστικού κόστους του fine-tuning, εισάγοντας τη χρήση δυαδικών μασκών. Αντί να προσαρμόζονται όλες οι παράμετροι ενός προεκπαιδευμένου μοντέλου, όπως το BERT, RoBERTa ή DistilBERT, η προτεινόμενη μέθοδος επιλέγει κρίσιμα βάρη μέσω δυαδικών μασκών που εκπαιδεύονται ειδικά για κάθε εργασία. Αυτό μειώνει δραστικά την απαίτηση μνήμης, καθώς οι μάσκες αποθηκεύουν μόνο τα απαραίτητα στοιχεία, καταλαμβάνοντας περίπου το 3% του χώρου που απαιτούν οι πλήρως βελτιστοποιημένες παράμετροι.

3.2.2 Μάθηση με προτροπές

Η μάθηση μέσω προτροπών (prompt tuning) αποτελεί μια προσέγγιση για την προσαρμογή προεκπαιδευμένων γλωσσικών μοντέλων σε εξειδικευμένες εργασίες, χωρίς την ανάγκη τροποποίησης όλων των παραμέτρων τους. Σε αντίθεση με το fine-tuning, όπου απαιτείται η αποθήκευση ενός νέου αντιγράφου του μοντέλου για κάθε εργασία, η μάθηση μέσω προτροπών χρησιμοποιεί οδηγίες προστίθενται στην είσοδο της εκάστοτε εργασίας, επιτρέποντας την επαναχρησιμοποίηση του ίδιου προεκπαιδευμένου μοντέλου για πολλαπλές εφαρμογές [53].

Μαθηματικά, ένα γλωσσικό μοντέλο υπολογίζει την πιθανότητα $P(Y|X)$, όπου X είναι η είσοδος (π.χ., ένα κείμενο ή ερώτηση) και Y η επιθυμητή έξοδος (π.χ., μια

απάντηση ή κατηγορία). Με την εισαγωγή ενός prompt P, η πιθανότητα τροποποιείται σε P(Y|[P;X]), όπου [P;X] είναι η συνένωση του prompt με την είσοδο. Το prompt P είναι διατυπωμένο με λέξεις, π.χ. "Classify the sentiment of the following text as positive or negative." και λειτουργεί ως συνθήκη που καθοδηγεί το μοντέλο να παράγει αποτελέσματα προσαρμοσμένα στην εκάστοτε εργασία. Οι προτροπές μπορούν να δημιουργηθούν με βάση τη γνώση του ανθρώπου για τη δομή της εργασίας. Επειδή η αποτελεσματικότητα των prompts εξαρτάται από την ακρίβεια και την πληρότητα της διατύπωσης και επειδή μπαίνει και ο ανθρώπινος παράγοντας για την δημιουργία τους, είναι συχνά δύσκολο να δημιουργηθούν βέλτιστες προτροπές για σύνθετες εργασίες, καθώς μπορεί να απαιτείται πειραματισμός ή δοκιμή πολλών παραλλαγών και ένα επίπεδο εξειδίκευσης από τον άνθρωπο που κάνει την εργασία.

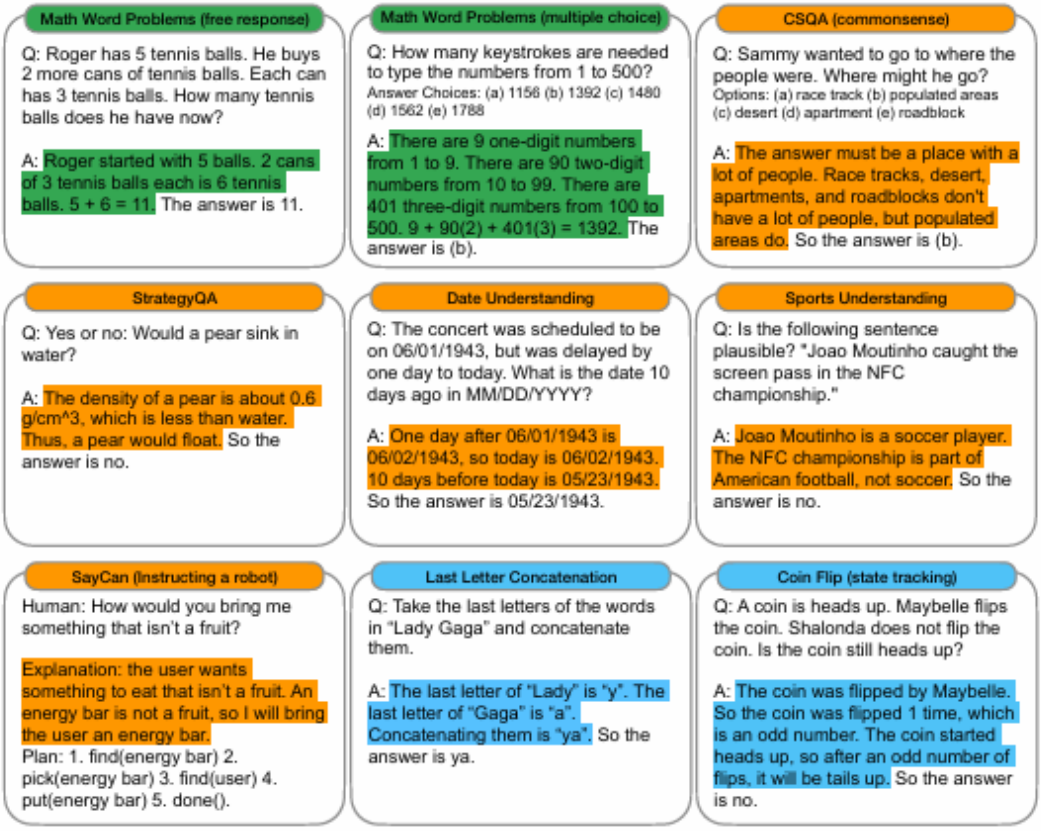
Στην περίπτωση των μαλακών προτροπών (soft prompts), το prompt P δεν είναι απλά μια στατική ακολουθία λέξεων, αλλά παραμετροποιείται μέσω embeddings θ_P, τα οποία εκπαιδεύονται ειδικά για κάθε εργασία. Η διαδικασία εκπαίδευσης στοχεύει στη μεγιστοποίηση της πιθανότητας $P_\theta(Y|[P;X])$, διατηρώντας σταθερά τα βάρη του προεκπαιδευμένου μοντέλου θ και βελτιστοποιώντας μόνο τις παραμέτρους του prompt θ_p [53].

Το **Chain-of-Thought Prompting (CoT)** αποτελεί μια προσέγγιση που ενισχύει την ικανότητα λογικής σκέψης των LLMs μέσω της εισαγωγής ενδιάμεσων βημάτων λογικής σε προβλήματα. Οι J. Wei et.al. [54] έδειξαν ότι η χρήση του CoT βελτίωσε δραματικά την απόδοση σε προβλήματα μαθηματικών, κοινής λογικής και συμβολικής λογικής, καθιστώντας το μια από τις πιο αποδοτικές μεθόδους ενίσχυσης της επίδοσης αυτών των μοντέλων.

Συγκεκριμένα, στην κατηγορία της μαθηματικής λογικής, η εφαρμογή του CoT στο benchmark GSM8K οδήγησε το PaLM 540B να πετύχει ποσοστό επίλυσης 74%, ξεπερνώντας προηγούμενα μοντέλα που είχαν εκπαιδευτεί ειδικά για αυτό το σκοπό. Η βελτίωση ήταν ιδιαίτερα εμφανής σε προβλήματα πολλαπλών βημάτων, όπου το CoT επέτρεψε την ανάλυση σε διαδοχικά λογικά βήματα. Παρόμοια αποτελέσματα παρατηρήθηκαν και σε άλλα benchmarks, όπως τα SVAMP, ASDiv, AQuA, και MAWPS, όπου η χρήση του CoT αύξησε την ακρίβεια, ιδίως σε προβλήματα υψηλότερης πολυπλοκότητας.

Στα προβλήματα κοινής λογικής, το CoT ενίσχυσε την ικανότητα των μοντέλων να κατανοούν και να απαντούν σε ερωτήματα που βασίζονται σε γλωσσικά συμφραζόμενα. Για παράδειγμα, στο StrategyQA, το PaLM 540B πέτυχε ακρίβεια 75.6%, ξεπερνώντας το προηγούμενο state-of-the-art ποσοστό του 69.4%. Στο Sports Understanding, το μοντέλο όχι μόνο ξεπέρασε τα υπόλοιπα μοντέλα αλλά και ανθρώπινες επιδόσεις, με ακρίβεια 95.4%. Αντίστοιχα, στο CSQA (Commonsense Question Answering), το CoT παρείχε μια μικρή αλλά σημαντική βελτίωση.

Στην κατηγορία της συμβολικής λογικής, το CoT αποδείχθηκε ιδιαίτερα αποτελεσματικό στη γενίκευση σε πιο σύνθετα προβλήματα. Στη συνένωση τελευταίων γραμμάτων λέξεων, το CoT πέτυχε σχεδόν 100% επιτυχία σε εντός domain προβλήματα, ενώ επέτρεψε στο μοντέλο να αντιμετωπίσει επαρκώς και εκτός domain προβλήματα μεγαλύτερης πολυπλοκότητας. Παρόμοια αποτελέσματα παρατηρήθηκαν και σε προβλήματα αναστροφής νομισμάτων, όπου το CoT βελτίωσε την κατανόηση των καταστάσεων και την ικανότητα του μοντέλου να επεξεργάζεται μεγαλύτερες ακολουθίες.



Εικόνα 12. Παραδείγματα εισόδου, αλυσίδας συλλογισμών και εξόδου για δοκιμαστικά σύνολα αριθμητικής, κοινής λογικής και συμβολικής συλλογιστικής. Οι αλυσίδες συλλογισμών υπογραμμίζονται [54].

Ένα σημαντικό εύρημα ήταν ότι η αποτελεσματικότητα του CoT εξαρτάται από την κλίμακα του μοντέλου. Η μέθοδος απέδωσε εντυπωσιακά καλά σε μοντέλα με περισσότερες από 100 δισεκατομμύρια παραμέτρους, ενώ σε μικρότερα μοντέλα οι παραγόμενες αλυσίδες σκέψης ήταν λιγότερο συνεπείς και συχνά φλύαρες. Επιπλέον, η ανάλυση σφαλμάτων έδειξε ότι όταν το CoT απέτυχε, περίπου στο 46% των περιπτώσεων τα σφάλματα ήταν μικρά και διορθώσιμα, ενώ στο 54% υπήρχαν μεγαλύτερες ασυνέπειες.

Εικόνα 13. Παραδείγματα κατανόησης σημασιολογίας και σφαλμάτων ενός βήματος που διορθώθηκαν με την κλιμάκωση του PaLM από 62B σε 540B [54].

3.3 Προκλήσεις

Η εκπαίδευση LLMs για ανάλυση συναισθήματος στον χρηματοοικονομικό τομέα παρουσιάζει σημαντικές προκλήσεις λόγω της πολυπλοκότητας και της εξειδίκευσης της χρηματοοικονομικής γλώσσας. Οι τεχνικοί όροι και η ιδιαίτερη σημασία πολλών λέξεων δημιουργούν εμπόδια στην αποτελεσματική εκπαίδευση και βελτίωση αυτών των μοντέλων, καθιστώντας τη διαδικασία απαιτητική και σύνθετη.

Μία από τις μεγαλύτερες δυσκολίες έγκειται στο γεγονός ότι οι όροι στη χρηματοοικονομική γλώσσα έχουν συχνά ειδική σημασία που διαφέρει από τη χρήση τους στην καθημερινή γλώσσα. Για παράδειγμα, λέξεις όπως «bull» και «bear», οι οποίες σε γενικές χρήσεις είναι ουδέτερες, στον χρηματοοικονομικό τομέα δηλώνουν έντονες καταστάσεις αγοράς. Ο «bull» υποδηλώνει μια ανερχόμενη αγορά, ενώ ο «bear» μια πτωτική αγορά. Αυτή η μεταφορική χρήση μπορεί να είναι δύσκολο να κατανοηθεί από γλωσσικά μοντέλα που έχουν εκπαιδευτεί σε γενικά δεδομένα [44].

Επιπλέον, η πολυπλοκότητα και η δομή των χρηματοοικονομικών κειμένων αυξάνουν το επίπεδο δυσκολίας. Κείμενα όπως οικονομικές εκθέσεις, δελτία τύπου και αναλύσεις είναι συχνά μεγάλης έκτασης, γεμάτα με τεχνικούς όρους και σύνθετες προτάσεις. Τα γλωσσικά μοντέλα αντιμετωπίζουν περιορισμούς στο μέγεθος των εισόδων που μπορούν να επεξεργαστούν, γεγονός που απαιτεί στρατηγικές όπως ο διαχωρισμός των κειμένων σε τμήματα ή η χρήση ιεραρχικών μηχανισμών προσοχής [44].

Η έλλειψη μεγάλων και κατάλληλα επισημασμένων συνόλων δεδομένων για τη χρηματοοικονομική ανάλυση συναισθήματος αποτελεί επίσης σημαντική πρόκληση. Ενώ υπάρχουν κάποιες δημόσιες βάσεις δεδομένων, όπως το **Financial PhraseBank** [55], τα περισσότερα σύνολα δεδομένων είναι ιδιωτικά ή απαιτούν ακριβά δικαιώματα πρόσβασης. Αυτό περιορίζει τις δυνατότητες εκπαίδευσης μοντέλων με επαρκή δεδομένα για τη σωστή κατανόηση της πολυπλοκότητας του τομέα.

Μια άλλη πρόκληση είναι η ευαισθησία των χρηματοοικονομικών κειμένων σε μικρές διαφοροποιήσεις στο ύφος ή τη διατύπωση. Τα συναισθήματα σε τέτοια κείμενα είναι συχνά μεικτά, γεγονός που απαιτεί μοντέλα ικανά να κατανοούν αποχρώσεις. Για παράδειγμα, μια πρόταση όπως «Τα κέρδη μειώθηκαν, αλλά η εταιρεία προβλέπει

σημαντική ανάπτυξη το επόμενο τρίμηνο» απαιτεί την αναγνώριση ενός θετικού υπονοούμενου παρά τη φαινομενικά αρνητική έναρξη.

Παράλληλα, τα μικρότερα γλωσσικά μοντέλα συχνά αποτυγχάνουν να δημιουργήσουν ακριβείς αναλύσεις, καθώς η εξειδικευμένη φύση του τομέα απαιτεί μεγάλη κλίμακα παραμέτρων και προσαρμοσμένη εκπαίδευση. Ωστόσο, ακόμα και τα μεγαλύτερα μοντέλα ενδέχεται να υποφέρουν από υπερπροσαρμογή (overfitting), όταν εκπαιδεύονται με μικρά σύνολα δεδομένων, ή να παρουσιάζουν μεροληψία λόγω περιορισμένης ποικιλίας στα δεδομένα εκπαίδευσης.

Κεφάλαιο 4. Πειραματική αξιολόγηση

Η μεθοδολογία που ακολουθήθηκε έχει ως στόχο την παρουσίαση διαφορετικών τρόπων χρήσης των LLMs για την ανάλυση συναισθήματος. Ξεκινάμε με ανάλυση του dataset που θα χρησιμοποιήσουμε και τα επόμενα βήματα περιλαμβάνουν τη χρήση απλών αλγορίθμων μηχανικής μάθησης ως σύγκριση, την αξιοποίηση προεκπαιδευμένων μοντέλων, τη διαδικασία fine-tuning, και την εφαρμογή τεχνικών Prompt Engineering.

4.1 Σύνολο Δεδομένων

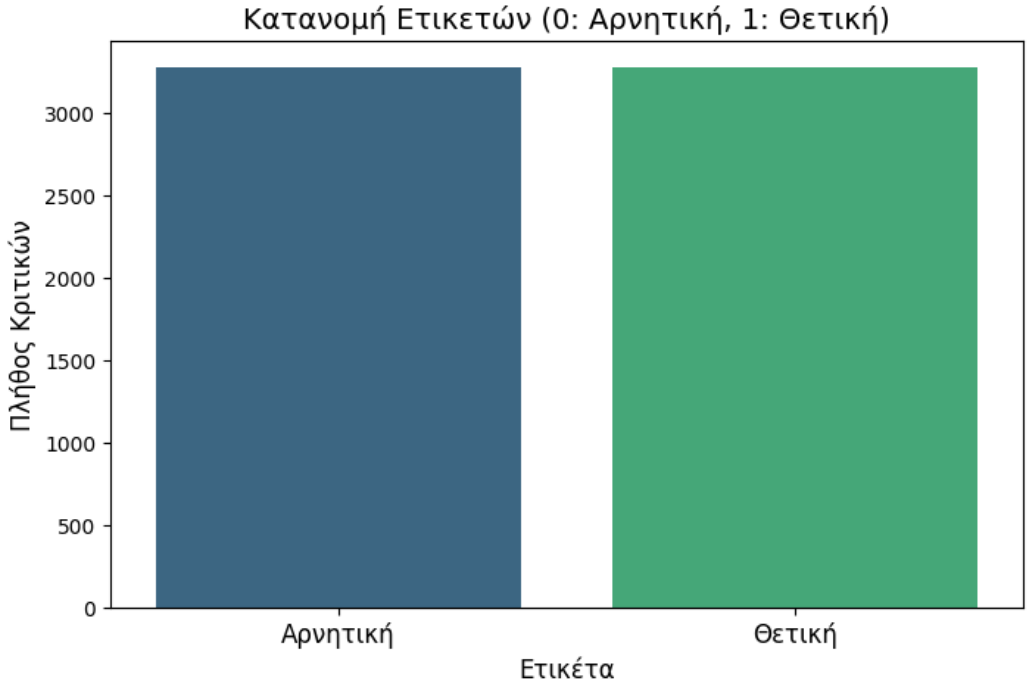
Σε αυτή την διπλωματική εργασία χρησιμοποιήθηκε το *Skroutz Shop Reviews Sentiment Analysis Dataset [56]*, το οποίο βρίσκεται στην πλατφόρμα Kaggle. Το dataset περιέχει κριτικές χρηστών από το ελληνικό ηλεκτρονικό κατάστημα Skroutz, με σκοπό τη μελέτη της ανάλυσης συναισθήματος (sentiment analysis).

Το dataset αποτελείται από 6552 κριτικές γραμμένες στα ελληνικά και έχουν ταξινομηθεί ως θετικές ή αρνητικές. Πέντε τυχαίες εγγραφές δίνονται στην Εικόνα 14

id	review	label
1183	Ολα καλα! Μου εκανε εντυπωση πως το καταστημα ειχε τοσο πολυ δουλεια Τεταρτη μεσημερι που πηγα να παραλαβω. Κατι κανουν καλα εκει...Ευχαριστω	Positive
756	Ταχύτατη εξυπηρέτηση ευγενικό και φιλικό προσωπικό. Το συστήνω ανεπιφύλακτα.	Positive
3880	Έχω παραγγείλει απο 27 Μαρτίου Σήμερα είναι 7 Απριλίου Δεν σηκώνουν τηλέφωνα Δεν απαντούν σε mail Έχω την εφαρμογή τους με κάρτα μέλους και δεν δίνουν καμία ενημέρωση για την πορεία της παραγγελίας Δεν μπορώ ούτε να την ακυρώσω Τα λεφτά τα τράβηξαν από την αρχή από την κάρτα	Negative
763	Πολύ άμεση αποστολή.. πολλά συγχαρητήρια.	Positive
5893	Μια παραγγελία δύο προϊόντα. Έτυχε να χαλάσουν και τα δύο φέτος εντός εγγύησης. Το πρώτο ήταν ένας Technisat Digipal2 DVBT. Πέραν του ότι λανθασμένα μου είπαν αρχικά ότι ήταν εκτός εγγύησης υποστηρίζοντας ότι έχει μόνο ένα χρόνο εγγύησης **ΔΕΝ** αναλάμβαναν την υποστήριξή του. Έπρεπε να ψάξω να βρω την αντιπροσωπία του μόνος μου μιας και το τηλέφωνο που μου έδωσαν δεν ίσχυε. Και εντάξει εγώ είμαι Αθήνα ο άλλος από την επαρχία; Απλά πετάει το προϊόν στον κοντινότερο κάδο υποθέτω. Το δεύτερο ένα Labtec media wireless desktop 800 είχε πρόβλημα το τελευταίο διάστημα με το ροδάκι του ποντικιού. Ένα λεπτό είναι πρακτικά ο έλεγχος αλλά επειδή είναι και περιφερειακό όφειλα να κάτσω σε λίστα αναμονής για 7 μέρες. Δεν είχα πρόβλημα. Με ενημέρωσαν να πάω να το πάρω διότι αντικαταστάθηκε. Πήγα την επόμενη το μεσημέρι στα κεντρικά τους και **ΔΕΝ** το είχαν ετοιμοπαράδοτο. 3 ώρες τζάμπα διαδρομή. Είπαν ότι θα μου το έφερναν σήμερα με courier δωρεάν. Είναι 6 το απόγευμα και κανείς και τίποτα **ΔΕΝ** ήρθε. Αν στα παραπάνω προσθέσω και το γεγονός ότι για οτιδήποτε αγοραστεί online η μοναδική λύση είναι η μετάβαση στα κεντρικά είναι προφανές ότι η υποστήριξη είναι μόνο για όσους μένουν δίπλα στα κεντρικά και μόνο για όσα προϊόντα αναλαμβάνει το RMA τους το κατάστημα και δεν σε στέλνει σε quest για τον αντιπρόσωπο. Δεν νομίζω να ξανά αγοράσω ποτέ κάτι από αυτούς.	Negative

Εικόνα 14. Πέντε τυχαίες εγγραφές του συνόλου δεδομένων.

Στην Εικόνα 15 έχουμε το πλήθος των κριτικών που είναι θετικές και αρνητικές. Βλέπουμε ότι έχουμε ίσο αριθμό θετικών και αρνητικών κριτικών ήτοι 3276 θετικές και 3276 αρνητικές. Οπότε το σύνολο δεδομένων μας είναι ισορροπημένο.



Εικόνα 15Κατανομή των ετικετών.

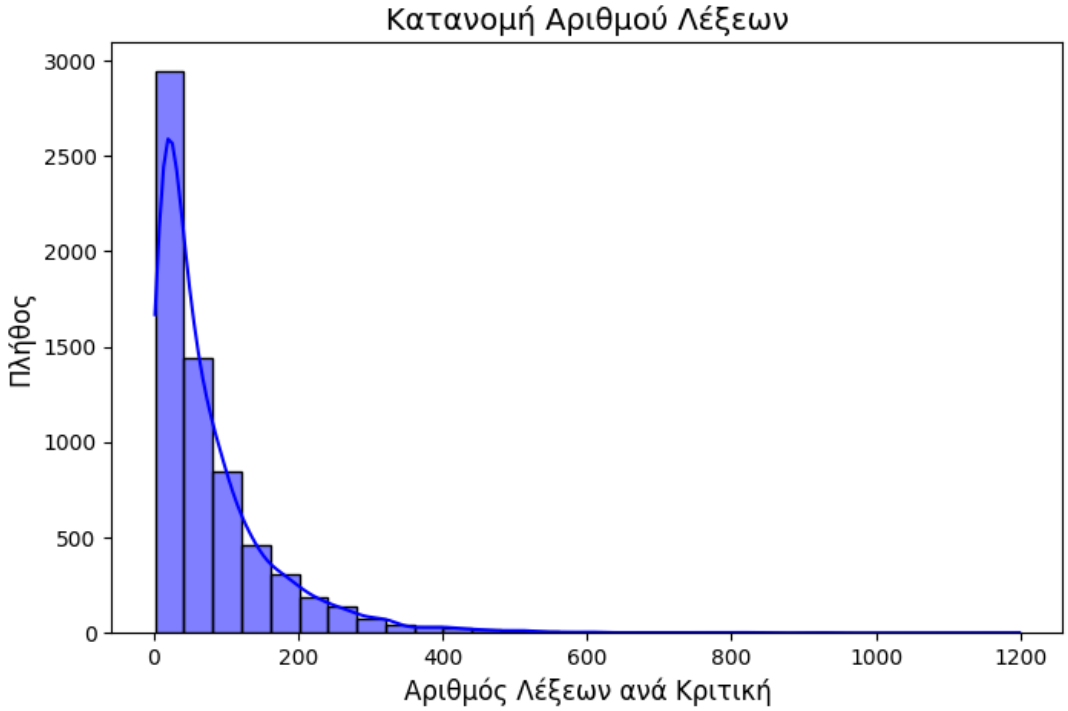
Στην συνέχεια μελετάμε το πλήθος λέξεων σε κάθε κριτική και εξετάζουμε αν το πλήθος των λέξεων και το συναίσθημα της κριτικής είναι ανεξάρτητα. Στον πίνακα 2 έχουμε τα μέτρα θέσης για τον αριθμό λέξεων ανά κριτική. Ο μέσος αριθμός λέξεων ανά κριτική είναι περίπου 78.27, κάτι που υποδεικνύει ότι οι κριτικές είναι γενικά σύντομες. Η σχετικά υψηλή τιμή της τυπικής απόκλισης (88.36) δείχνει ότι υπάρχει σημαντική ποικιλία στο μήκος των κριτικών. Κάποιες κριτικές είναι εξαιρετικά σύντομες, ενώ άλλες είναι αρκετά εκτενείς. Το 25% των κριτικών έχουν 21 λέξεις ή λιγότερες, κάτι που φανερώνει ότι υπάρχει μεγάλος αριθμός πολύ σύντομων κριτικών και υπάρχει και μία απάντηση με μόνο δύο λέξεις. Το 50% των κριτικών περιλαμβάνουν περίπου 48.5 λέξεις, ενώ το 75% φτάνει τις 103 λέξεις. Αυτές οι τιμές δείχνουν ότι οι περισσότερες κριτικές έχουν μήκος μεταξύ 21 και 103 λέξεων.

Mean	78.27
Std	88.36
Min	2
25%	21
50%	48.5
75%	103
Max	1,199

Πίνακας 2. Στατιστικά μέτρα για τον αριθμό των λέξεων ανά κριτική.

Στην Εικόνα 16 έχουμε και το ιστόγραμμα των αριθμών λέξεων ανά κριτική. Βλέπουμε ότι η κατανομή είναι δεξιά-ασύμμετρη που επιβεβαιώνει ότι η πλειονότητα των κριτικών έχουν μικρό μήκος. Επιπλέον, βλέπουμε μία μακριά ουρά που δείχνει ότι υπάρχουν μερικές κριτικές που είναι πολύ μεγάλες.

Η ποικιλία στον αριθμό λέξεων δείχνει ότι το dataset περιλαμβάνει τόσο σύντομες όσο και εκτενείς κριτικές. Αυτό μπορεί να επηρεάσει την ανάλυση συναισθήματος, καθώς οι σύντομες κριτικές ενδέχεται να περιέχουν λιγότερες πληροφορίες για την εξαγωγή συναισθημάτων, ενώ οι εκτενείς κριτικές παρέχουν περισσότερο περιεχόμενο αλλά αυξάνουν την υπολογιστική πολυπλοκότητα.



Εικόνα 16. Κατανομή του αριθμού λέξεων.

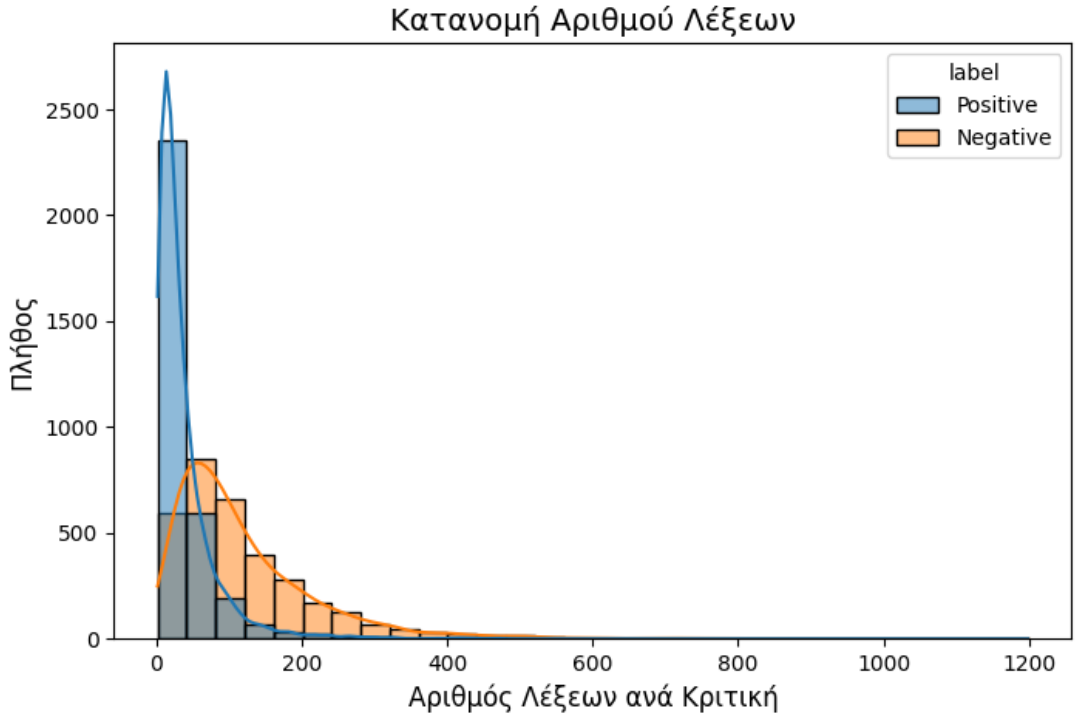
Στην συνέχεια, μελετάμε την κατανομή του αριθμού λέξεων ανά ετικέτα για να δούμε αν υπάρχει διαφορά μεταξύ των αρνητικών και θετικών κριτικών. Στον Πίνακα 4 έχουμε τα μέτρα θέσης ανάλογα με την ετικέτα της κριτικής. Για τις αρνητικές κριτικές έχουμε μέσο όρο λέξεων 119.44 που δείχνει ότι οι χρήστες αναλύουν περισσότερο τις αρνητικές εμπειρίες. Επίσης, οι αρνητικές κριτικές έχουν υψηλή τυπική απόκλιση που δείχνει μεγάλη ποικιλία στο μήκος των κριτικών. Οι θετικές κριτικές έχουν μέσο όρο λέξεων 37 που δείχνει ότι οι θετικές κριτικές είναι συνήθως

πιο σύντομες. Επιπλέον, έχουν χαμηλότερη τυπική απόκλιση που δείχνει ότι έχουν λιγότερη ποικιλία στο μήκος των κριτικών.

label	count	mean	std	min	25%	50%	75%	max
Negative	3276	119.44	101.59	2	52	91	157	1199
Positive	3276	37.09	43.62	2	12	23	46	468

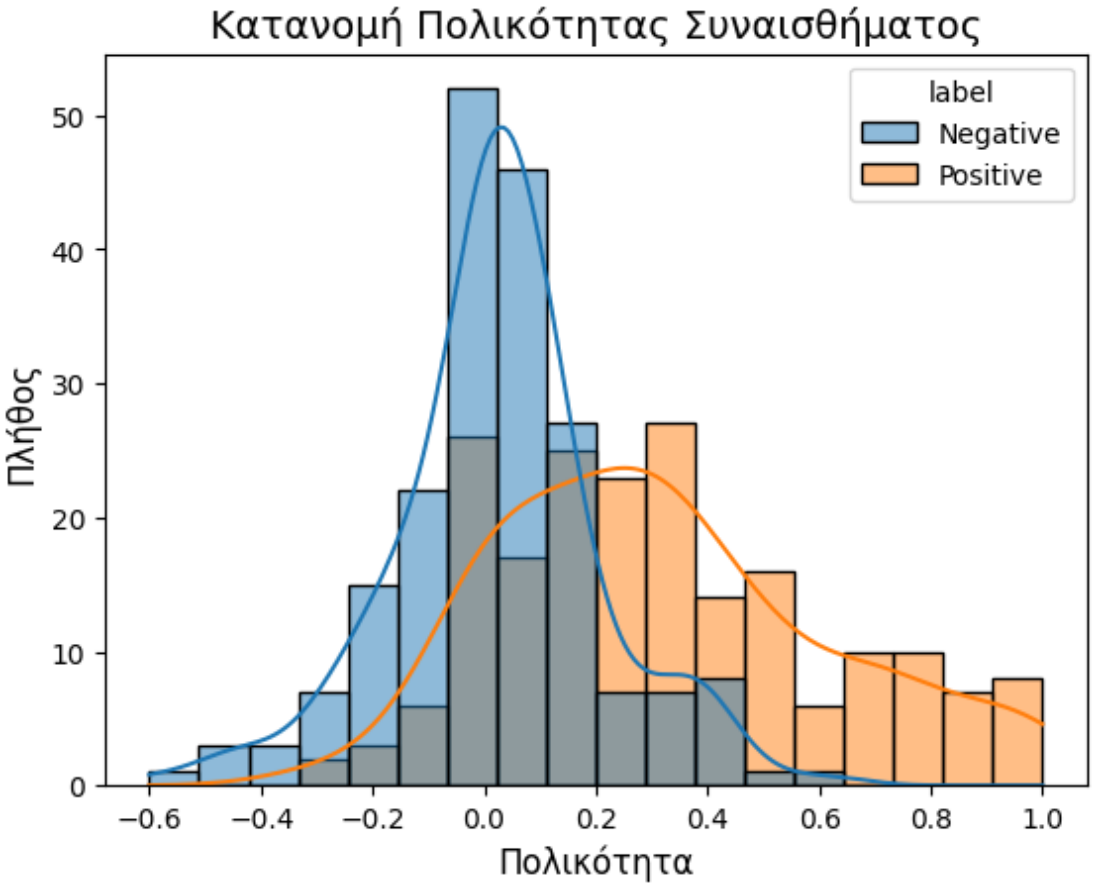
Πίνακας 2. Στατιστικά μέτρα του πλήθους λέξεων ανά ετικέτα.

Στην Εικόνα 17 έχουμε το ιστόγραμμα των παραπάνω. Παρατηρούμε ότι οι θετικές κριτικές είναι αρκετά πιο σύντομες από τις αρνητικές. Κάνουμε και έναν έλεγχο t-test με μηδενική υπόθεση ότι το πλήθος των λέξεων είναι ίδια για τις δύο ετικέτες έναντι της εναλλακτικής υπόθεσης ότι υπάρχει διαφορά. Παίρνουμε $p = 0.0$ που σημαίνει ότι απορρίπτουμε την μηδενική υπόθεση με επίπεσο σημαντικότητας 5%, άρα αποδεχόμαστε ότι υπάρχει διαφορά στο πλήθος των λέξεων. Στην συνέχεια, κάνουμε την μηδενική υπόθεση ότι το μέσο πλήθος των λέξεων της θετικής ετικέτας είναι μεγαλύτερο από το μέσο πλήθος για την αρνητική ετικέτα έναντι της εναλλακτικής υπόθεσης η αρνητική ετικέτα να έχει μεγαλύτερο μέσο πλήθος λέξεων. Πάλι έχουμε p=0.0 που σημαίνει ότι απορρίπτουμε την μηδενική υπόθεση με επίπεδο σημαντικότητας 5%, οπότε δεχόμαστε ότι η αρνητική ετικέτα έχει μεγαλύτερο πλήθος λέξεων.



Εικόνα 17. Ιστόγραμμα του αριθμού λέξεων ανά κριτική και συναίσθημα.

Στην συνέχεια μελετάμε την πολικότητα των κριτικών. Επειδή για πολικότητα χρειαζόμαστε βιβλιοθήκη η οποία χρησιμοποιείται για αγγλικό κείμενο, πρέπει να μεταφράζουμε κάθε κριτική. Η μετάφραση γίνεται μέσω της βιβλιοθήκης deep_translator σε Python και χρησιμοποιούμε τον GoogleTranslator που μας επιτρέπει να χρησιμοποιούμε το Google Translation μέσω κώδικα. Όμως, επειδή είναι χρονοβόρο, παίρνουμε ένα δείγμα 200 κριτικών από την κάθε κατηγορία. Στην Εικόνα 18 έχουμε την πολικότητα των κριτικών όπου οι τιμές πιο κοντά στο ένα δείχνουν θετικό συναίσθημα και οι τιμές κοντά στο -1 δείχνουν αρνητικό συναίσθημα. Όπως θα περιμέναμε οι αρνητικές κριτικές είναι μετατοπισμένες πιο αριστερά από τις θετικές. Αυτό το επιβεβαιώνουμε και με έναν έλεγχο t-test όπου παίρνουμε p=1.4e-27, που είναι στατιστικά σημαντικό σε επίπεδο σημαντικότητας 5%.



Εικόνα 18. Πολικότητα κριτικών της κάθε κατηγορίας.

Στην συνέχεια σχηματίζουμε και word clouds για τις θετικές και αρνητικές κριτικές. Στις θετικές κριτικές βλέπουμε να εμφανίζονται λέξεις όπως «κανένα πρόβλημα», «γρήγορη αποστολή», «Μπράβο» κ.α. Για τις αρνητικές κριτικές έχουμε λέξεις όπως «επιστροφή χρημάτων», «μέχρι σήμερα», «εξέλιξη παραγγελίας» κ.α. Επίσης, βλέπουμε ότι οι πιο κοινές λέξεις όπως «κατάστημα», «παραγγελία», «προϊόν»

είναι και στις δύο ετικέτες, κάτι που σημαίνει ότι δεν είναι αρκετές για να κάνουμε την ταξινόμηση, αλλά χρειαζόμαστε και τα συμφραζόμενα.

Εικόνα 19. Word cloud για θετικές κριτικές.

Εικόνα 20. Word cloud για τις αρνητικές κριτικές.

4.2 Μεθοδολογία

Για την ανάπτυξη του κατάλληλου κώδικα χρησιμοποιήσαμε την γλώσσα προγραμματισμού Python, η οποία είναι μία διερμηνευόμενη, γενικού σκοπού και υψηλού επιπέδου γλώσσα προγραμματισμού. Η Python, είναι γνωστή για την

υιοθέτηση μίας προγραμματιστικής ιδεολογίας που υποστηρίζει τη σημασία του καθαρού, απλού και κατανοητού κώδικα όπως περιγράφεται στο μανιφέστο «The Zen of Python». Αυτό την έχει κάνει την πιο δημοφιλή γλώσσα προγραμματισμού για την ανάπτυξη μοντέλων μηχανικής μάθησης, βαθιάς μάθησης και για την ανάπτυξη των LLMs. Η βασική βιβλιοθήκη που χρησιμοποιήσαμε είναι η **HuggingFace.**

Η **HuggingFace** είναι μια πλατφόρμα που αναπτύχθηκε με στόχο να κάνει τα LLMs και τις τεχνολογίες μηχανικής μάθησης πιο προσιτές και εύχρηστες για ερευνητές και επαγγελματίες. Δημιουργήθηκε για να διευκολύνει τη χρήση και την ανάπτυξη state-of-the-art μοντέλων όπως τα BERT, GPT, RoBERTa, και T5, παρέχοντας εργαλεία για την χρησιμοποίηση και την προσαρμογή τους σε ποικίλες εφαρμογές. Η HuggingFace προσφέρει τη βιβλιοθήκη **Transformers**,, η οποία παρέχει έτοιμα προεκπαιδευμένα μοντέλα καθώς και δυνατότητα για fine-tuning. Η σημασία της έγκειται στο γεγονός ότι μειώνει δραστικά το χρόνο και τους πόρους που απαιτούνται για την ανάπτυξη μοντέλων LLMs. Επιπλέον, η κοινότητά της και το αποθετήριο μοντέλων της βοηθούν στην ελεύθερη χρήση προηγμένων μοντέλων, και έτσι προωθείται η καινοτομία στην τεχνητή νοημοσύνη.

Η ανάλυση ξεκίνησε με τη χρήση του αλγορίθμου **LogisticRegression** ως σημείο αναφοράς χρησιμοποιώντας την βιβλιοθήκη **sklearn**. Στη συνέχεια, χρησιμοποιήθηκε μια έκδοση του **XLM-RoBERTa**, σχεδιασμένη για εργασίες Natural Language Inference (NLI) σε πολυγλωσσικά περιβάλλοντα. Το επόμενο βήμα ήταν η προσαρμογή ενός μοντέλου BERT που ήταν ήδη εκπαιδευμένο σε ελληνικό σύνολο δεδομένων και εμείς προχωρήσμε με το συγκεκριμένο dataset σε fine-tuning για να μάθει να κάνει κατηγοριοποίηση συναισθήματος. Η διαδικασία περιλάμβανε την χρήση της βιβλιοθήκης HuggingFace για φόρτωση του προεκπαιδευμένου BERT μοντέλου., προσαρμογή των παραμέτρων του μοντέλου όπως των αριθμό των εποχών εκπαίδευσης και του ρυθμού μάθησης και τέλος εκπαίδευση του μοντέλου με το ελληνικό dataset. Το τελικό βήμα περιλάμβανε την εφαρμογή τεχνικών **Prompt Engineering** για την αξιοποίηση δυνατοτήτων των LLMs χωρίς την ανάγκη εκτεταμένης εκπαίδευσης.

Για την αξιολόγηση των μοντέλων χρησιμοποιούμε μετρικές που είναι κατάλληλες για προβλήματα ταξινόμησης. Η ακρίβεια (precision) ορίζεται ως το ποσοστό των σωστών προβλέψεων του μοντέλου επί του συνόλου των παρατηρήσεων που θεώρησε ως θετικές και υπολογίζεται με τον τύπο

$$Precision = \frac{Tp}{Tp+Fp} \tag{23}$$

όπου Tp είναι οι παρατηρήσεις που ορθώς βρέθηκαν θετικά και Fp είναι οι παρατηρήσεις που ταξινομήθηκαν λανθασμένα ως θετικές. Η ακρίβεια εκφράζει το ποσοστό των θετικών προβλέψεων που είναι σωστές και είναι σημαντική σε εφαρμογές όπου οι λανθασμένες θετικές προβλέψεις έχουν μεγάλο κόστος. Για παράδειγμα, σε ένα μοντέλο που ανιχνεύει ανεπιθύμητα email, μια υψηλή τιμή precision σημαίνει ότι από τα μηνύματα που ταξινομήθηκαν ως spam, η πλειοψηφία είναι όντως spam.

Η ανάκληση (recall) χρησιμοποιείται για να μετρήσει πόσο καλά το μοντέλο εντοπίζει τις πραγματικά θετικές περιπτώσεις. Ορίζεται από τον τύπο

$$Recall = \frac{Tp}{Tp+Fn} \tag{24}$$

όπου Fn οι παρατηρήσεις που ταξινομήθηκαν ως αρνητικές ενώ ήταν θετικές. Η ανάκληση εκφράζει το ποσοστό των θετικών περιπτώσεων που το μοντέλο εντόπισε σωστά και χρησιμοποιείται σε εφαρμογές όπου το κόστος των λανθασμένων αρνητικών προβλέψεων είναι υψηλό. Για παράδειγμα, σε ένα μοντέλο ανίχνευσης καρκίνου, είναι σημαντικό να έχει υψηλή ανάκληση, καθώς αν το μοντέλο αποτύχει να εντοπίσει μια καρκινική περίπτωση, το κόστος είναι πολύ μεγάλο.

Τέλος, το F1-score είναι ένας συνδυασμός των παραπάνω μετρικών και ορίζεται ως ο αρμονικός μέσος

$$F1 = 2\frac{Precision*Recall}{Precision+Recall} \tag{25}$$

4.3 Πειραματική αξιολόγηση

4.3.1 Logistic Regression

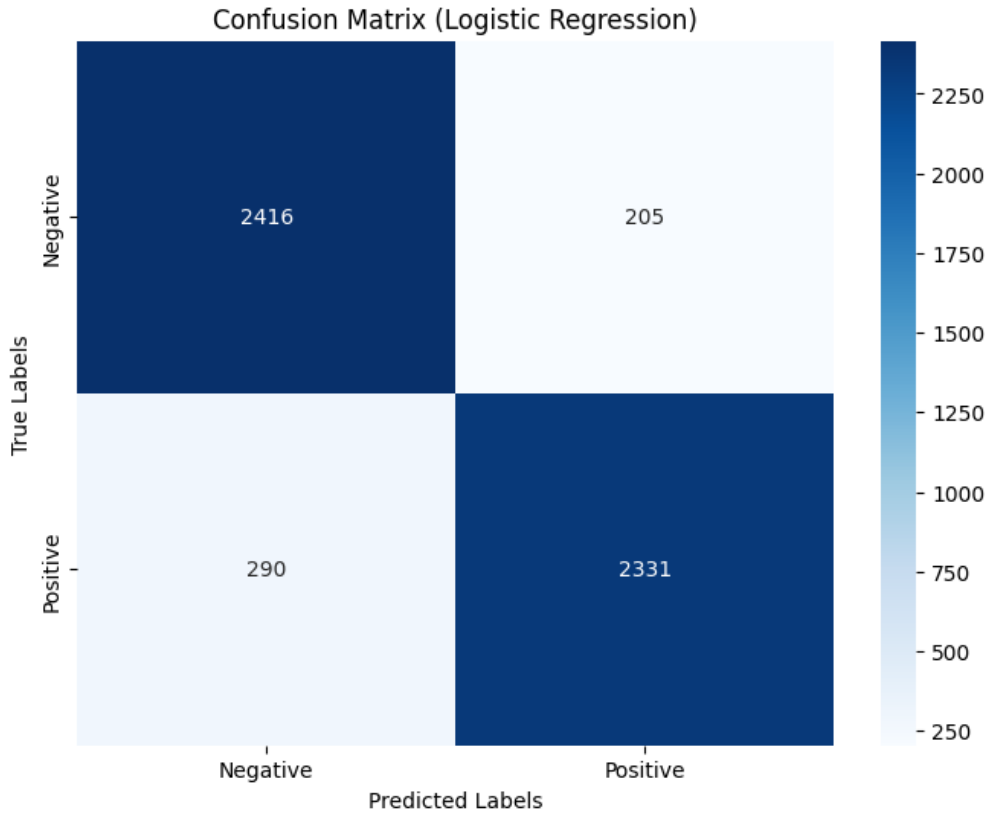
Αρχικά εφαρμόζουμε έναν κλασσικό αλγόριθμο της μηχανικής μάθησης, τον λογιστική παλινδρόμηση για να την συγκρίνουμε με τα LLMs που θα δούμε στην συνέχεια. Στην Εικόνα 21 βλέπουμε την ακρίβεια, ανάκληση και F1-score για το μοντέλο μας. Βλέπουμε γενικά πολύ υψηλά ποσοστά, οπότε το μοντέλο μας είναι αποτελεσματικό.

	precision	recall	f1-score	support
Negative	0.89	0.92	0.91	2621
Positive	0.92	0.89	0.90	2621
accuracy			0.91	5242
macro avg	0.91	0.91	0.91	5242
weighted avg	0.91	0.91	0.91	5242

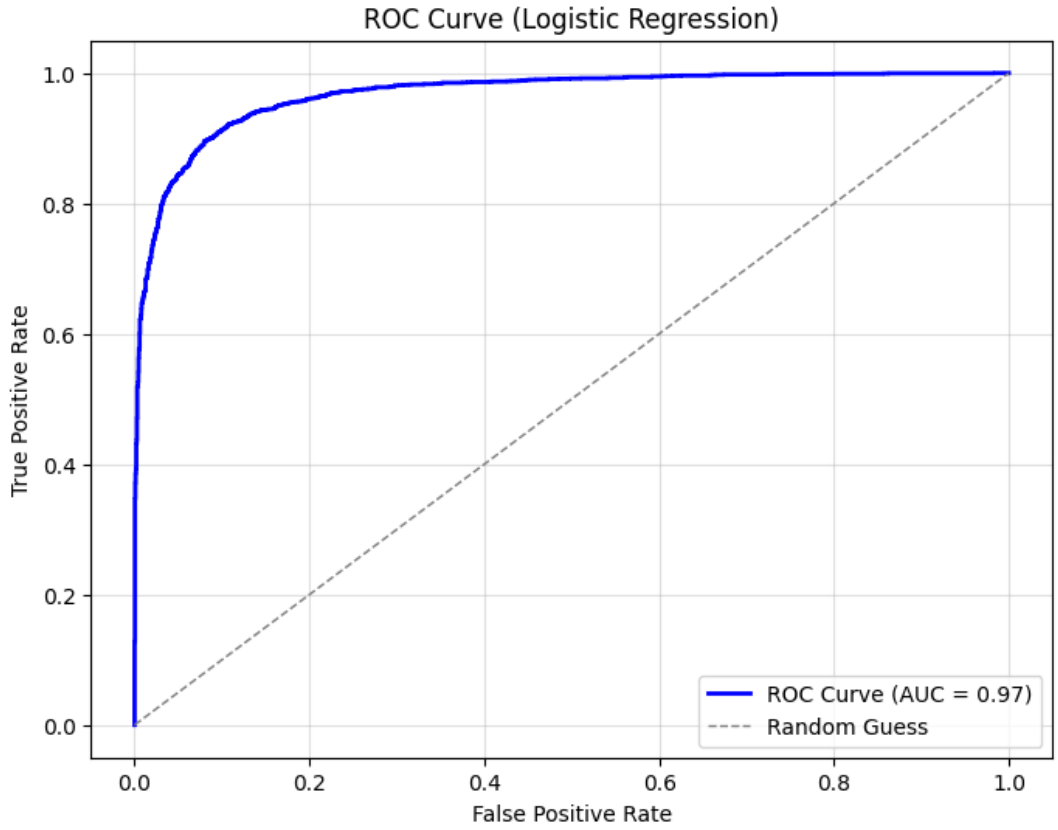
Εικόνα 21. Αναφορά ταξινόμησης για την λογιστική παλινδρόμηση.

Στην Εικόνα 22 έχουμε και τον πίνακα σύγχυσης για την λογιστική παλινδρόμηση. Βλέπουμε στην διαγώνιο ότι το μοντέλο καταφέρνει πολύ καλά να βρει τις πραγματικά αρνητικές και θετικές κριτικές. Επίσης, η τιμή για τα False Negatives (290) είναι υψηλότερη από τα False Positives (205), δείχνοντας ότι το μοντέλο δυσκολεύεται περισσότερο να εντοπίσει θετικές κριτικές.

Στην Εικόνα 23 έχουμε την καμπύλη ROC για το μοντέλο. Το εμβαδόν κάτω από την καμπύλη είναι 0.97 που είναι πολύ κοντά στην μονάδα που δείχνει ότι έχουμε ένα πολύ καλό μοντέλο.



Εικόνα 22. Πίνακας σύγχυσης για λογιστική παλινδρόμηση.



Εικόνα 23. Καμπύλη ROC για λογιστική παλινδρόμηση.

4.3.2 RoBERTa

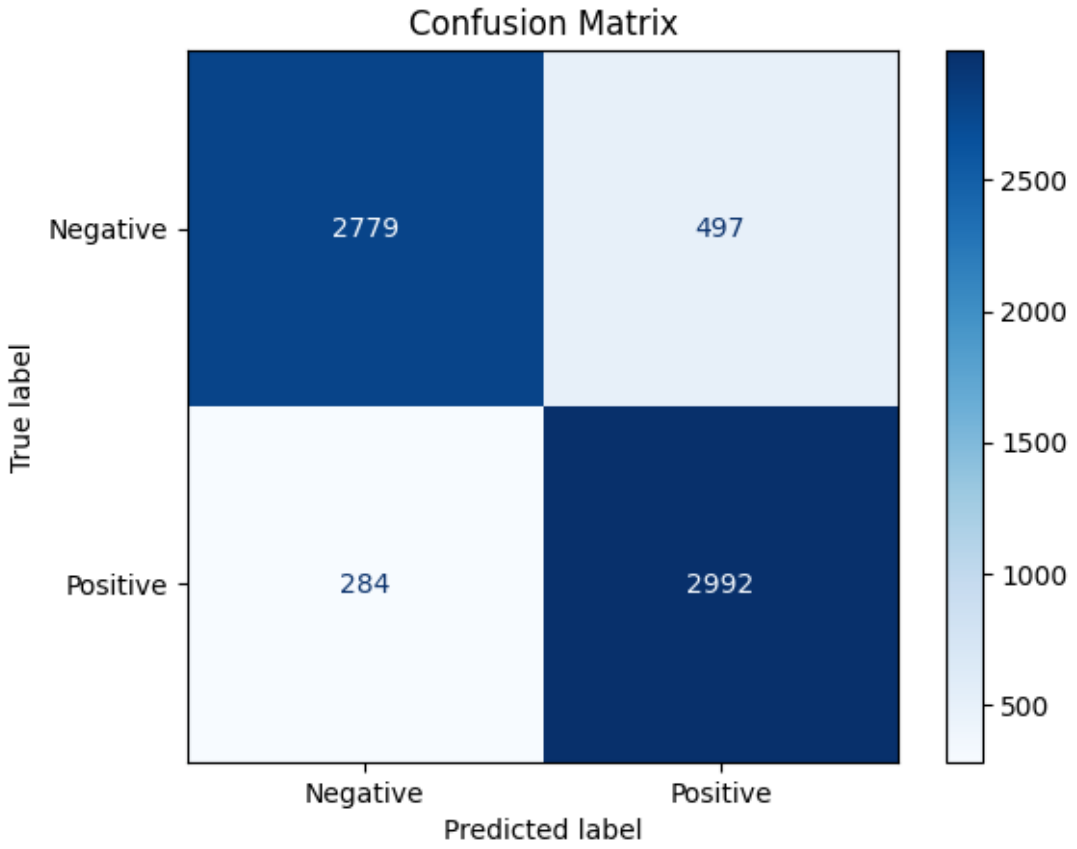
Μετά χρησιμοποιούμε ένα μοντέλο της οικογένειας του RoBERTa το οποίο έχει εκπαιδευτεί σε ένα σύνολο δεδομένο με 15 διαφορετικές γλώσσες συμπεριλαμβανομένου και των ελληνικών για να κάνει zero-shot classification. Μέσω του HuggingFace χρησιμοποιούμε το μοντέλο για το σύνολο δεδομένων μας χωρίς περαιτέρω εκπαίδευση.

Το RoBERTa έχει καλύτερη ακρίβεια για τα αρνητικά, αλλά χειρότερη για τα θετικά που σημαίνει ότι περισσότερες από τις αρνητικές προβλέψεις ήταν σωστές. Για την ανάκληση, το RoBERTa είναι καλύτερο για τις θετικές κριτικές, αλλά υστερεί στις αρνητικές. Αυτό σημαίνει ότι εντοπίζει καλύτερα τις θετικές κριτικές. Τέλος, το F1-score είναι χειρότερο για το RoBERTa.

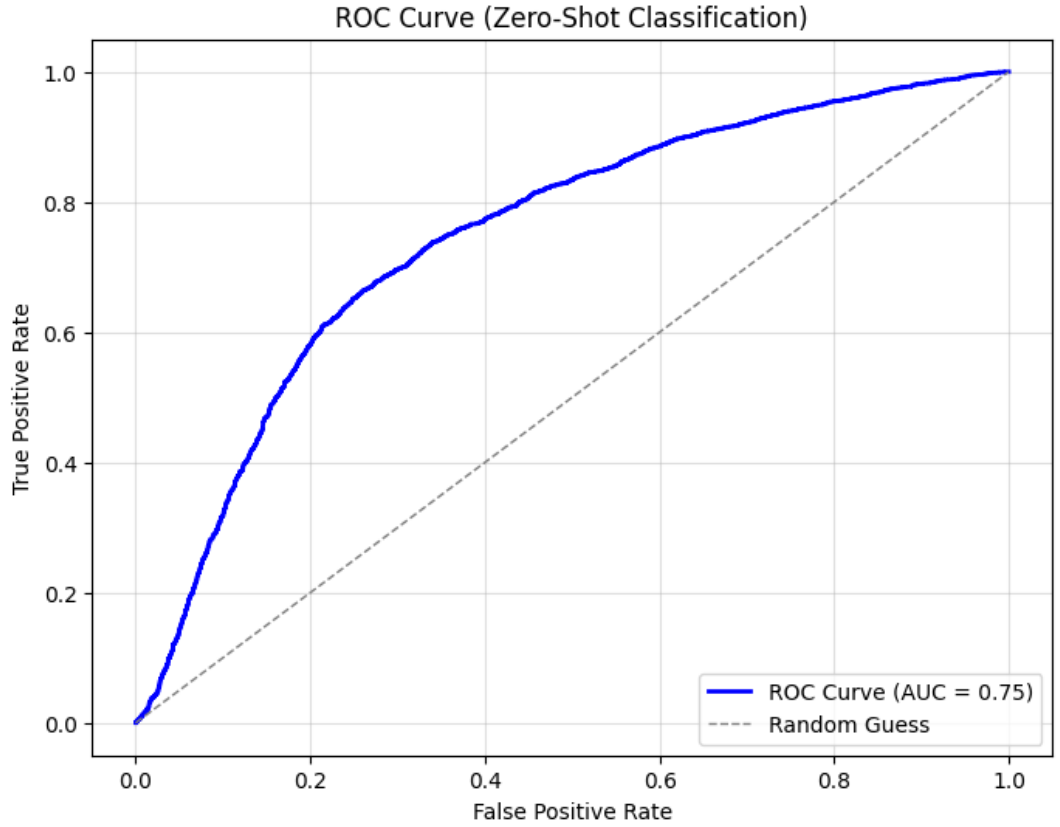
	precision	recall	f1-score	support
Negative	0.91	0.85	0.88	3276
Positive	0.86	0.91	0.88	3276
accuracy			0.88	6552
macro avg	0.88	0.88	0.88	6552
weighted avg	0.88	0.88	0.88	6552

Εικόνα 24, Αναφορά ταξινόμησης για το RoBERTa.

Στην Εικόνα 25 έχουμε τον πίνακα σύγχυσης και στην Εικόνα 26 την καμπύλη ROC. Βλέπουμε ότι το εμβαδόν κάτω από την καμπύλη είναι μικρότερο από το μοντέλο του Logistic Regression.



Εικόνα 25. Πίνακας σύγχυσης για το RoBERTa.



Εικόνα 26. Καμπύλη ROC για το RoBERTa.

4.3.3 BERT

Στην συνέχεια χρησιμοποιούμε ένα μοντέλο BERT που έχει ήδη γίνει fine-tuned στα ελληνικά. Αρχικά, τα δεδομένα προετοιμάστηκαν για εισαγωγή στο μοντέλο. Κάθε κείμενο μετατράπηκε σε ακολουθίες tokens χρησιμοποιώντας τον tokenizer του BERT από το HuggingFace. Για την εκπαίδευση χρησιμοποιήθηκαν οι υπερπαράμετροι του Πίνακα 3. Επίσης, για να αποφευχθεί η υπερεκπαίδευση, χρησιμοποιήθηκε η τεχνική **Early Stopping** όπου η εκπαίδευση θα σταματήσει αν για δύο συνεχόμενες εποχές δεν υπάρχει βελτίωση των μετρικών.

Υπερπαράμετρος	Τιμή	Περιγραφή
learning_rate	2e-5	Ρυθμός μάθησης για τη βελτιστοποίηση. Χαμηλή τιμή για σταθερή προσαρμογή.
per_device_train_batch_size	8	Μέγεθος batch ανά συσκευή (GPU/CPU).
num_train_epochs	10	Μέγιστος αριθμός εποχών για εκπαίδευση.
weight_decay	0.01	Ρυθμιστικό βάρος για αποφυγή υπερεκπαίδευσης.
evaluation_strategy	epoch	Αξιολόγηση του μοντέλου μετά από κάθε εποχή.
logging_steps	10	Καταγραφή δεδομένων μετά από 10 βήματα.
save_strategy	epoch	Αποθήκευση του μοντέλου στο τέλος κάθε εποχής.
load_best_model_at_end	True	Φόρτωση του καλύτερου μοντέλου στο τέλος της εκπαίδευσης.
metric_for_best_model	f1	Κριτήριο επιλογής του καλύτερου μοντέλου (F1-score).

Πίνακας 3Υπερπαράμετροι του BERT.

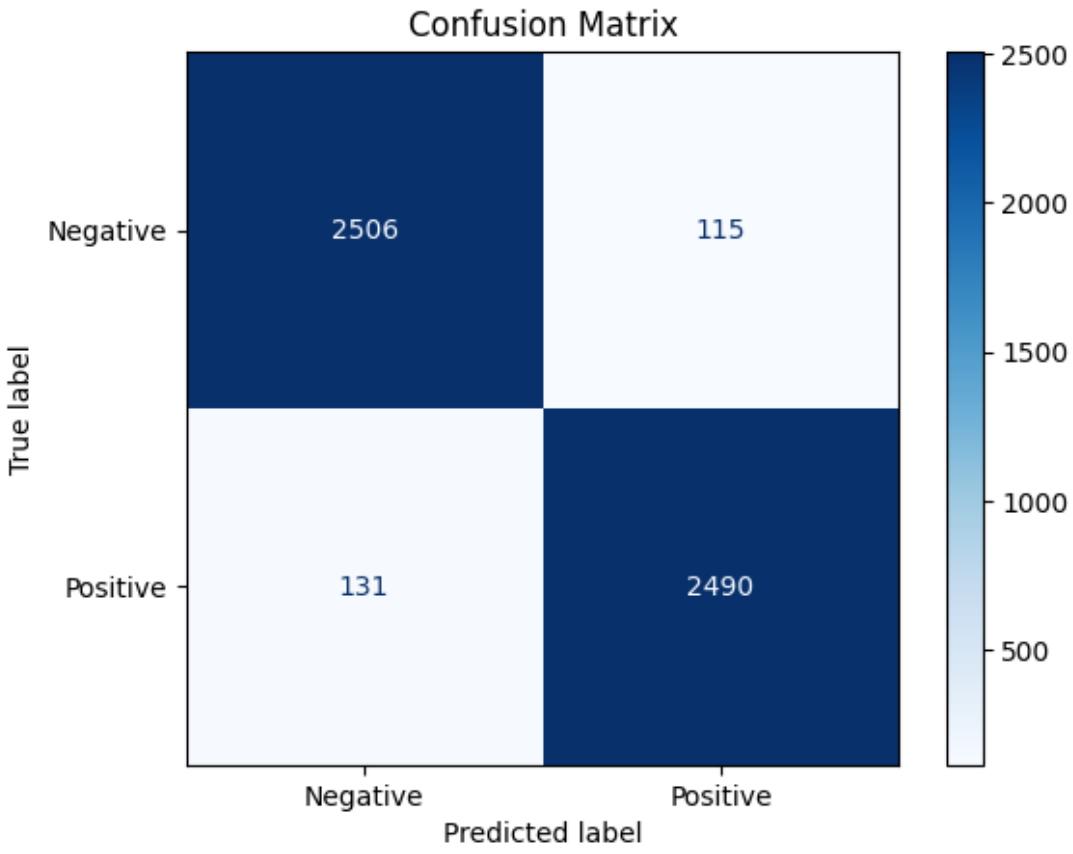
Πριν το εκπαιδεύσουμε, το δοκιμάζουμε στο σύνολο δεδομένων και πετυχαίνει ακρίβεια 0.50 και ανάκληση 0.02. Αυτό δεν είναι παράξενο αφού δεν έχει γίνει fine-tuned για την εργασία που το προορίζουμε.

Στην Εικόνα 27 έχουμε την αναφορά ταξινόμησης για τον BERT. Βλέπουμε ότι σε κάθε μετρική υπερτερεί του Logistic Regression.

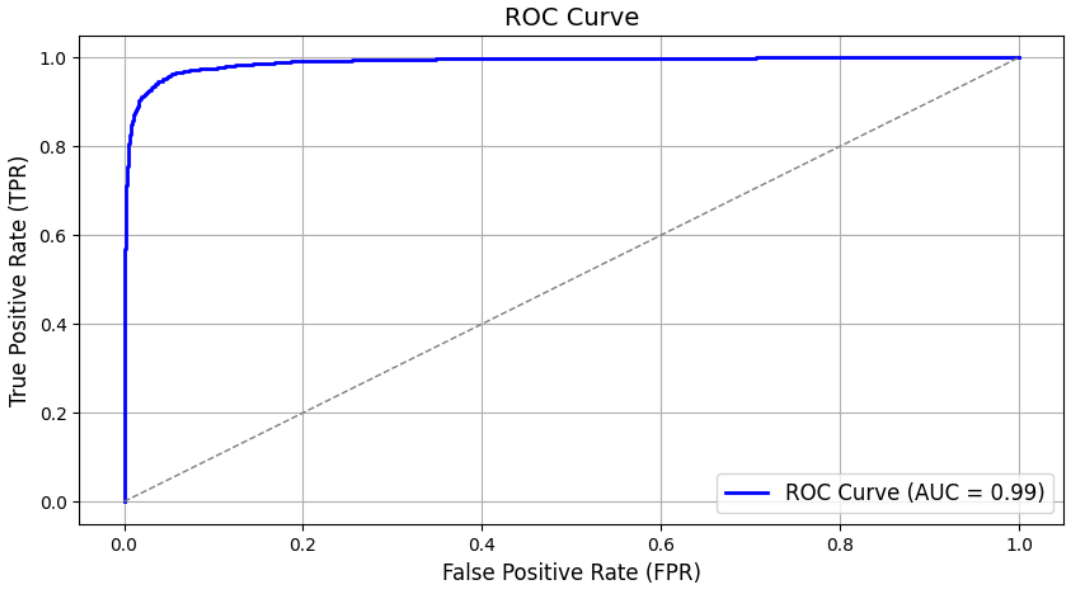
	precision	recall	f1-score	support
Negative	0.95	0.96	0.95	2621
Positive	0.96	0.95	0.95	2621
accuracy			0.95	5242
macro avg	0.95	0.95	0.95	5242
weighted avg	0.95	0.95	0.95	5242

Εικόνα 27Αναφορά ταξινόμησης για το BERT.

Στην Εικόνα 28 έχουμε τον πίνακα συγχύσεως όπου βλέπουμε ότι σχεδόν όλες τις κριτικές έχει καταφέρει να τις ταξινομήσει σωστά. Τέλος, στην Εικόνα 29 έχουμε την καμπύλη ROC για τον BERT. Το εμβαδόν κάτω από την καμπύλη είναι 0.99 που είναι καλύτερο από τον Logistic Regression.



Εικόνα 28. Πίνακας συγχύσεως για το BERT.



Εικόνα 29 Καμπύλη ROC για τον BERT.

4.3.4 Prompt Engineering

Χρησιμοποιήθηκε το μοντέλο **FLAN-T5 Base**, μια έκδοση του **T5** της Google που έχει εκπαιδευτεί σε ένα ευρύ φάσμα εργασιών μέσω τεχνικών fine-tuning και δέχεται prompts. Δεδομένου ότι το dataset αποτελείται από ελληνικές κριτικές προϊόντων, το μοντέλο FLAN-T5 δεν μπορούσε να επεξεργαστεί απευθείας τα δεδομένα, καθώς είναι προεκπαιδευμένο σε αγγλικά δεδομένα. Για να ξεπεραστεί αυτό το ζήτημα, όλες οι κριτικές μεταφράστηκαν στα αγγλικά πριν εισαχθούν στο μοντέλο.

Η μετάφραση έγινε με τη χρήση της βιβλιοθήκης **deep_translator**, όπου κάθε κριτική μετατράπηκε από ελληνικά σε αγγλικά και στη συνέχεια προσαρμόστηκε μέσα στα prompts που χρησιμοποιήθηκαν για την ταξινόμηση συναισθήματος. Ωστόσο, η διαδικασία της μετάφρασης ενδέχεται να εισάγει σημασιολογικές αλλοιώσεις ή να επηρεάσει το συναίσθημα του κειμένου, καθώς η σημασία λέξεων και εκφράσεων μπορεί να αλλάξει κατά τη μετάφραση. Επίσης οι ελληνικές κριτικές περιέχουν ιδιωματισμούς που μπορεί να μην μεταφράζονται σωστά και οι συναισθηματικοί τόνοι μιας φράσης μπορεί να χαθούν στη μετάφραση, επηρεάζοντας την ακρίβεια του μοντέλου. Επίσης, λόγω υπολογιστικού κόστους επιλέχθηκε τυχαίο δείγμα από 100 θετικές και 100 αρνητικές κριτικές.

4.3.4.1 Απλό prompt

Το απλό prompt ήταν:

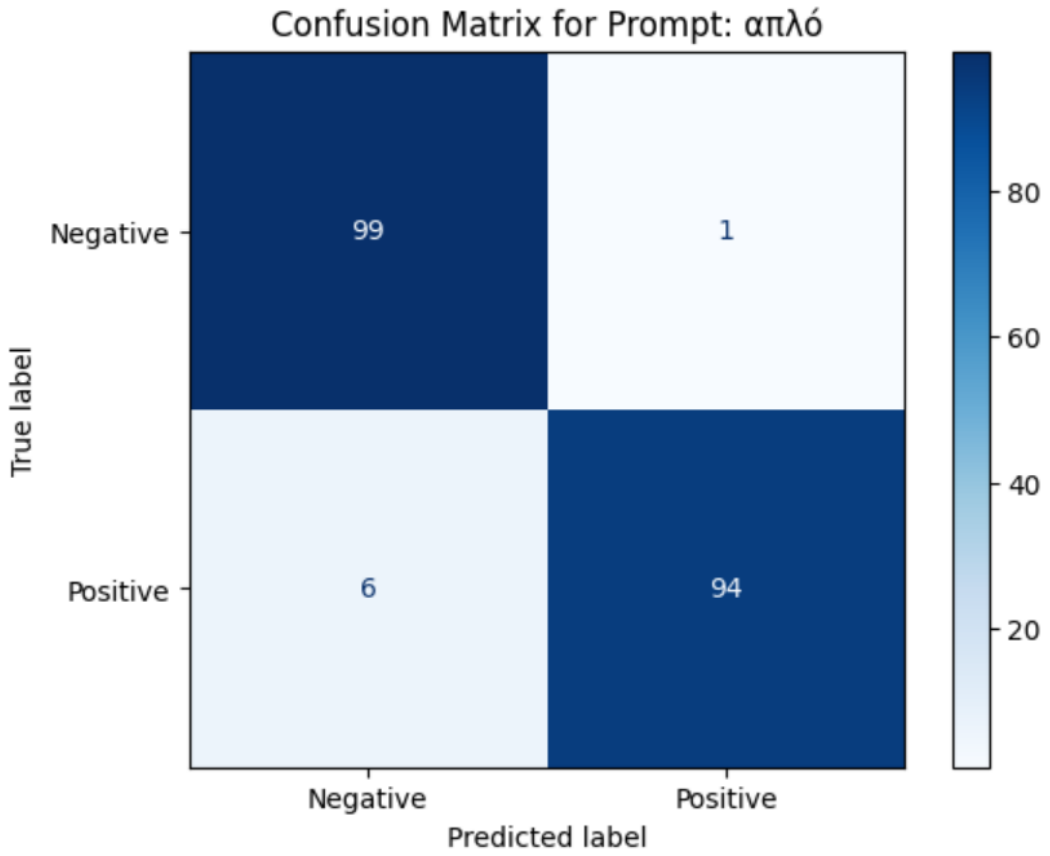
«Να απαντήσεις αν η παρακάτω κριτική είναι 'Positive' ή 'Negative': '{}'»

Στην Εικόνα 30 βλέπουμε πολύ καλά αποτελέσματα και για τις αρνητικές και για τις θετικές κριτικές. Τα αποτελέσματα είναι καλύτερα και από το logistic regression και από το BERT.

	precision	recall	f1-score	support
Negative	0.94	0.99	0.97	100
Positive	0.99	0.94	0.96	100
accuracy			0.96	200
macro avg	0.97	0.96	0.96	200
weighted avg	0.97	0.96	0.96	200

Εικόνα 30. Αναφορά ταξινόμησης για το απλό prompt.

Στην Εικόνα 31 βλέπουμε ότι μόνο μία αρνητική κριτική δεν κατηγοριοποίησε σωστά και μόλις 6 θετικές κατηγοριοποίησε λάθος.



Εικόνα 31Πίνακας σύγχυσης για το απλό prompt.

4.3.4.2 Λεπτομερές

Το λεπτομερές prompt ήταν:

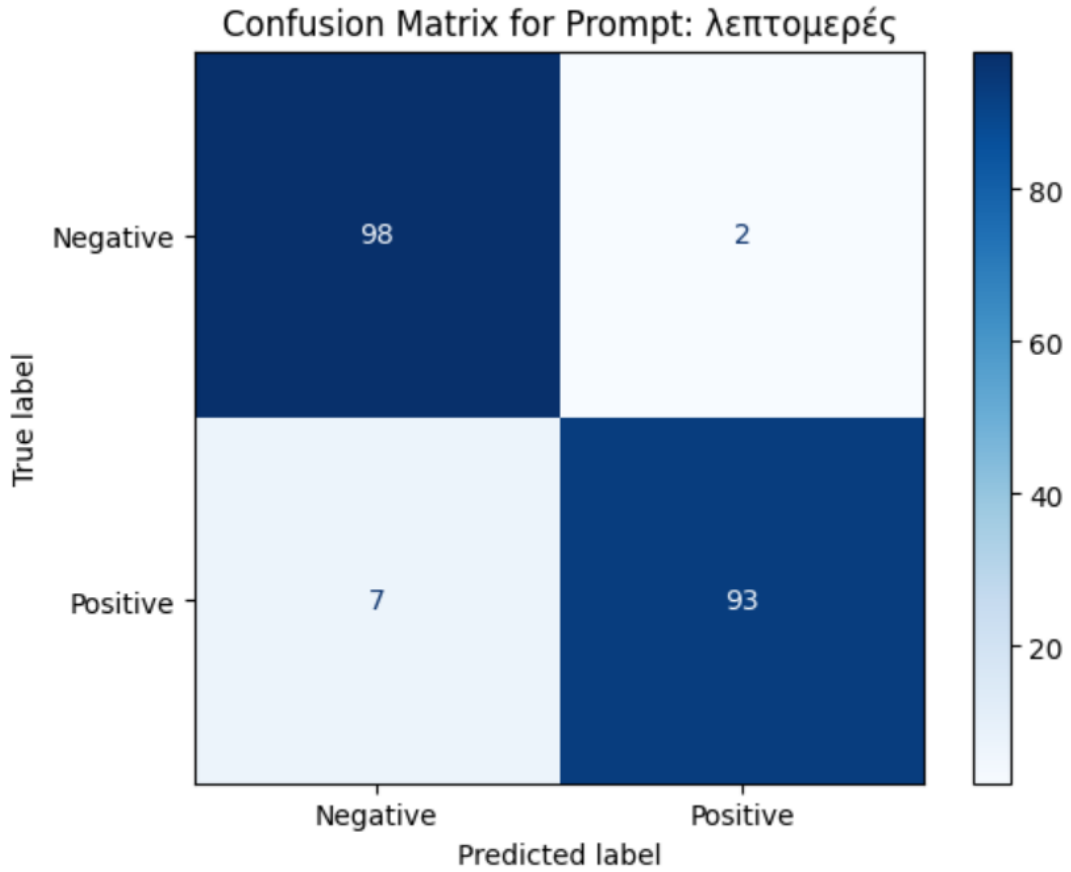
«Διάβασε προσεκτικά την παρακάτω κριτική. Απάντησε αν είναι 'Positive' ή 'Negative'. Εδώ είναι η κριτική: '{}'»

Στην Εικόνα 32 έχουμε επίσης πολύ καλά αποτελέσματα, αν και ελαφρώς χειρότερα από το απλό prompt.

	precision	recall	f1-score	support
Negative	0.93	0.98	0.96	100
Positive	0.98	0.93	0.95	100
accuracy			0.95	200
macro avg	0.96	0.96	0.95	200
weighted avg	0.96	0.95	0.95	200

Εικόνα 32. Αναφορά ταξινόμησης για το λεπτομερές prompt.

Στην Εικόνα 33 έχουμε ότι το μοντέλο άλλαξε λανθασμένα την απόφασή του για μία θετική και μία αρνητική κριτική.



Εικόνα 33. Πίνακας σύγχυσης για το λεπτομερές prompt.

4.3.4.3 Εξειδικευμένο prompt

Στο εξειδικευμένο prompt αναφερόμαστε στο μοντέλο ως ειδικό επί του θέματος:

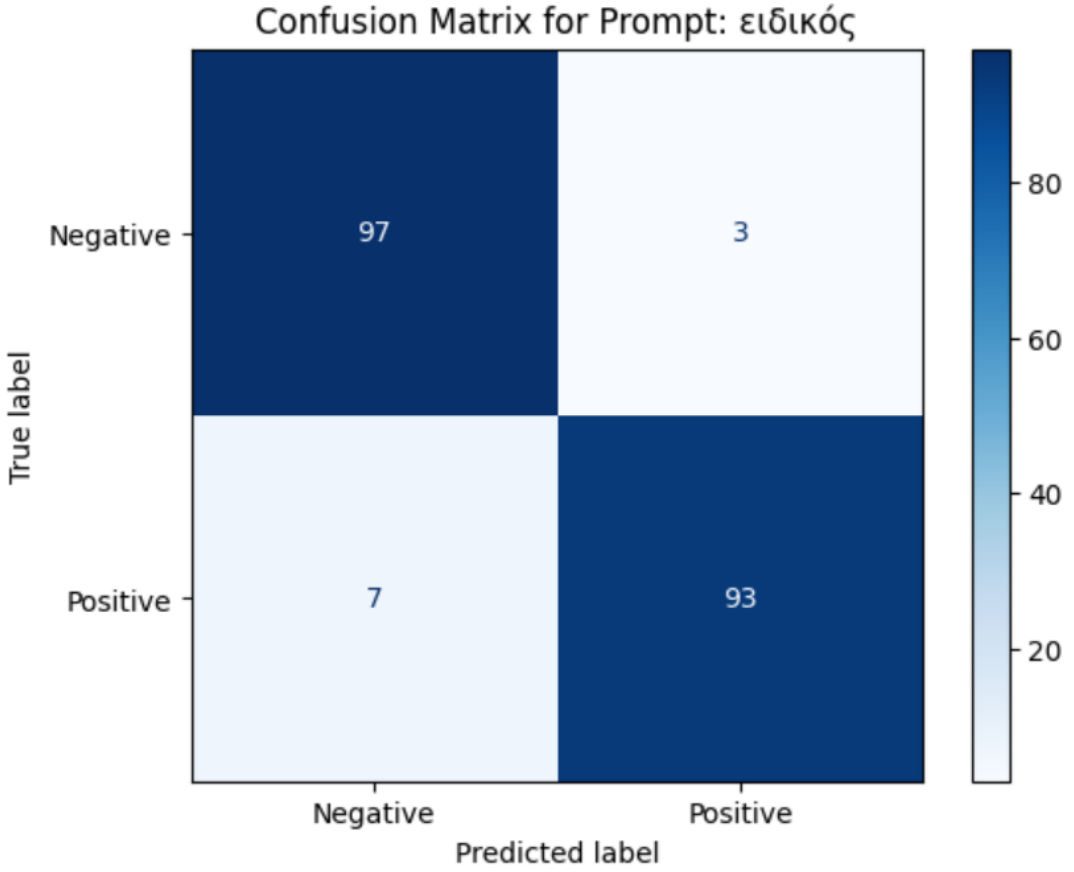
«Είσαι ειδικός στην ανάλυση συναισθήματος. Ανέλυσε το συναίσθημα της παρακάτω κριτικής. Απάντησε αν είναι 'Positive' ή 'Negative': '{}'»

Στην Εικόνα 34 βλέπουμε ότι το μοντέλο παραμένει καλύτερο από το Logistic Regression, αλλά είναι κοντά στο BERT.

	precision	recall	f1-score	support
Negative	0.93	0.97	0.95	100
Positive	0.97	0.93	0.95	100
accuracy			0.95	200
macro avg	0.95	0.95	0.95	200
weighted avg	0.95	0.95	0.95	200

Εικόνα 34. Αναφορά ταξινόμησης για το prompt του ειδικού.

Στην Εικόνα 35 βλέπουμε ότι άλλαξε λανθασμένα άλλη μία αρνητική κριτική.



Εικόνα 35. Πίνακας σύγχυσης για το prompt του ειδικού.

4.3.4.4 Few-shot

Στο few-shot του δίνουμε κάποια παραδείγματα

«Παράδειγμα 1: Κριτική: 'Η εξυπηρέτηση ήταν εξαιρετική και η παράδοση άμεση.'
Ταξινόμηση: Positive
Παράδειγμα 2: Κριτική: 'Το προϊόν ήταν ελαττωματικό και η υποστήριξη ανύπαρκτη.'
Ταξινόμηση: Negative

Κατηγοριοποίησε την παρακάτω κριτική ως 'Positive' ή 'Negative': '{}'»

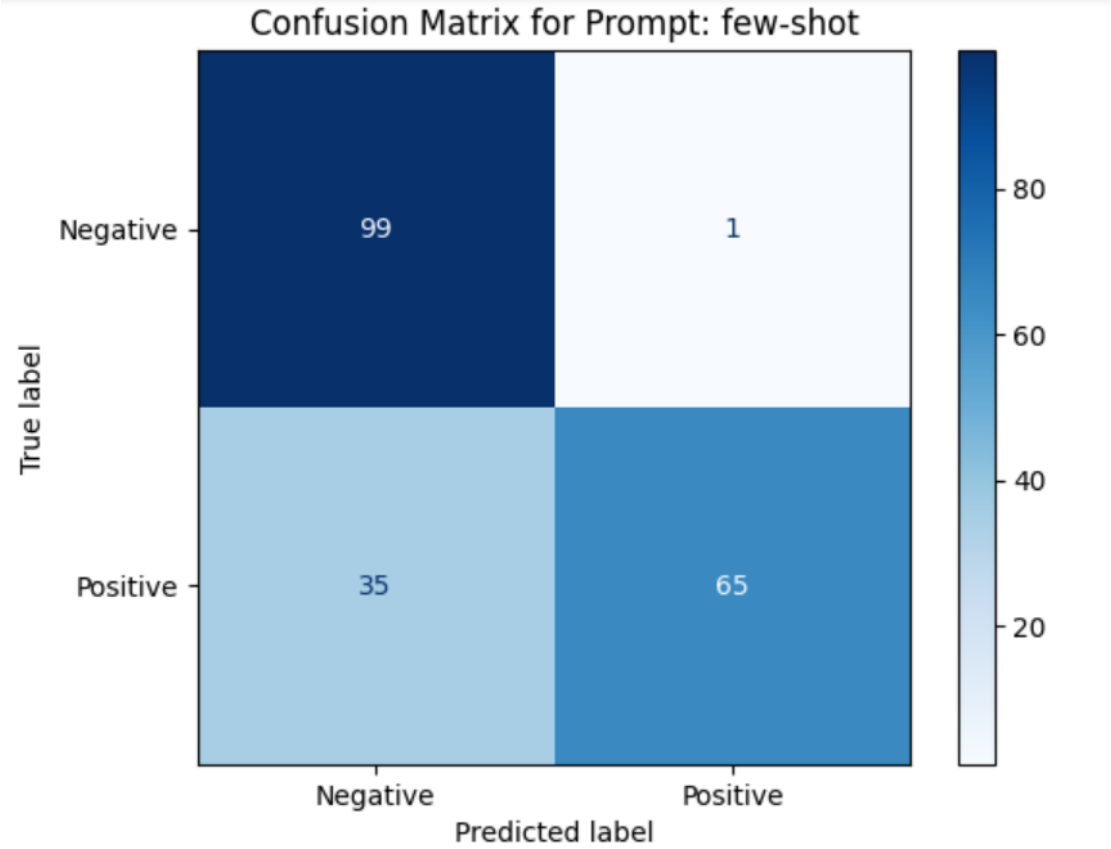
Στην Εικόνα 36 βλέπουμε ότι η ακρίβεια του μοντέλου έπεσε αισθητά.

Αναφορά Ταξινόμησης για το ύφος few-shot:

	precision	recall	f1-score	support
Negative	0.74	0.99	0.85	100
Positive	0.98	0.65	0.78	100
accuracy			0.82	200
macro avg	0.86	0.82	0.81	200
weighted avg	0.86	0.82	0.81	200

Εικόνα 36. Αναφορά ταξινόμησης για το few-shot prompt.

Στην Εικόνα 37 βλέπουμε ότι έχει πρόβλημα να κατηγοριοποιήσει τις θετικές κριτικές. Συγκεκριμένα άλλαξε λανθασμένα την απόφασή του για άλλες 28 θετικές κριτικές.



Εικόνα 37. Πίνακας σύγχυσης για το few-shot prompt.

4.3.4.5 Many-shot

Στο many-shot prompt δίνουμε στο μοντέλο περισσότερα παραδείγματα:

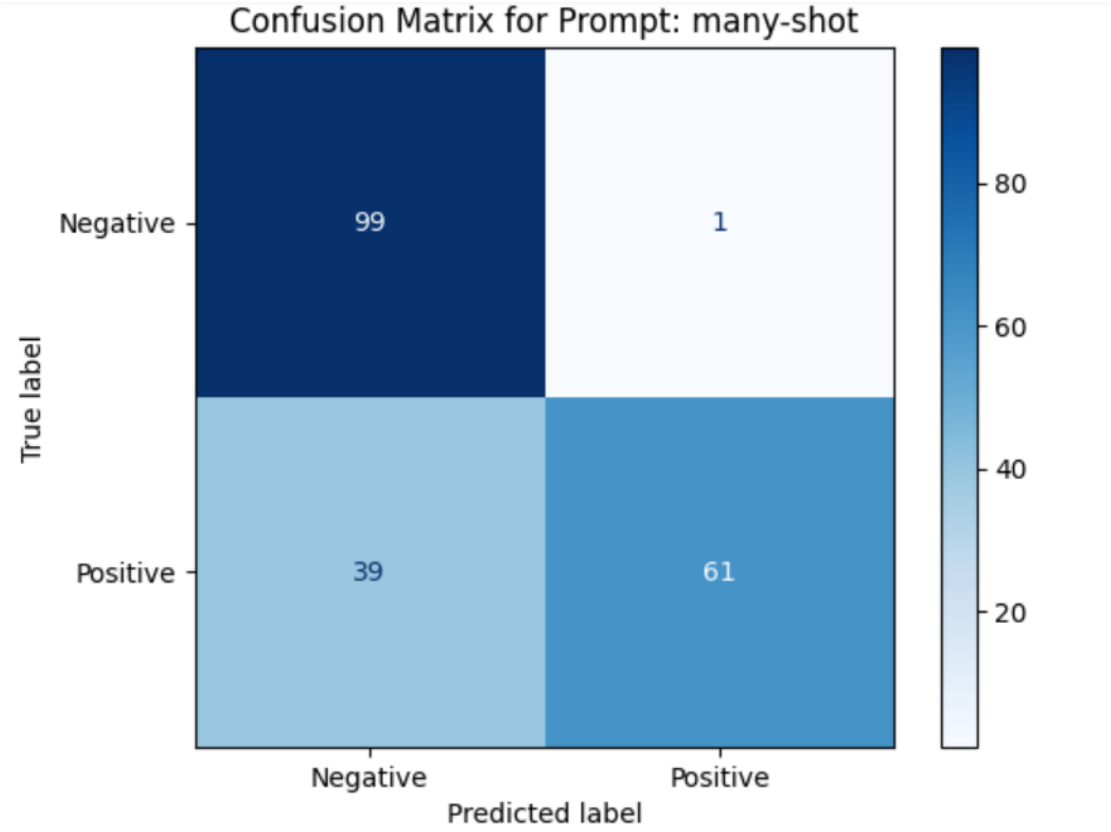
«Παράδειγμα 1: Κριτική: 'Η εξυπηρέτηση ήταν εξαιρετική και η παράδοση έγινε εγκαίρως.' Ταξινόμηση: Positive
Παράδειγμα 2: Κριτική: 'Το προϊόν δεν λειτουργούσε όπως περιγράφεται.' Ταξινόμηση: Negative
Παράδειγμα 3: Κριτική: 'Η παραγγελία έφτασε σε τέλεια κατάσταση και πολύ γρήγορα.' Ταξινόμηση: Positive
Παράδειγμα 4: Κριτική: 'Πολύ κακή εμπειρία με την εξυπηρέτηση πελατών.' Ταξινόμηση: Negative
Κατηγοριοποίησε την παρακάτω κριτική ως 'Positive' ή 'Negative': '{}'»

Στην Εικόνα 38 βλέπουμε ότι μοντέλο δεν βελτιώθηκε.

	precision	recall	f1-score	support
Negative	0.71	0.99	0.83	100
Positive	0.98	0.60	0.75	100
accuracy			0.80	200
macro avg	0.85	0.79	0.79	200
weighted avg	0.85	0.80	0.79	200

Εικόνα 38. Αναφορά ταξινόμησης για το many-shot prompt.

Στην Εικόνα 39 βλέπουμε ότι άλλαξε λανθασμένα την απόφασή του για άλλες τέσσερις θετικές κριτικές.



Εικόνα 39. Πίνακας σύγχυσης για το many-shot prompt.

Κεφάλαιο 5. Επίλογος

5.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία έδειξε τη σημασία και τις δυνατότητες των LLMs στην ανάλυση συναισθήματος, με ιδιαίτερη έμφαση στον οικονομικό τομέα και τα δεδομένα της ελληνικής αγοράς. Η χρήση τεχνικών προσαρμογής, όπως το fine-tuning και το prompt engineering, κατέστησε δυνατή τη βελτίωση της απόδοσης των LLMs σε ελληνικά δεδομένα.

Το μοντέλο Logistic Regression έδειξε πολύ υψηλή απόδοση, με ακρίβεια, ανάκληση και F1-score στο 0.91. Ο πίνακας σύγχυσης έδειξε ότι το μοντέλο ταξινομεί σωστά τις περισσότερες κριτικές. Η καμπύλη ROC επιβεβαίωσε την υψηλή απόδοση του μοντέλου, με εμβαδόν κάτω από την καμπύλη (AUC) 0.97.

Το μοντέλο RoBERTa, χωρίς περαιτέρω εκπαίδευση, έδειξε καλύτερη ακρίβεια για τις αρνητικές κριτικές, αλλά χειρότερη για τις θετικές. Η ανάκληση ήταν καλύτερη για τις θετικές κριτικές, αλλά υστερεί στις αρνητικές. Το F1-score ήταν χαμηλότερο σε σύγκριση με το Logistic Regression. Η καμπύλη ROC έδειξε εμβαδόν 0.94, κάτι που υποδηλώνει καλή απόδοση, αλλά όχι τόσο υψηλή όσο του Logistic Regression.

Το μοντέλο BERT, μετά από fine-tuning, έδειξε καλύτερη απόδοση μεταξύ των προηγούμενων μοντέλων. Η ακρίβεια, η ανάκληση και το F1-score ήταν όλα πάνω από 0.95, υποδηλώνοντας ότι το μοντέλο ταξινομεί πολύ καλά τις κριτικές. Ο πίνακας σύγχυσης έδειξε ότι το μοντέλο έκανε ελάχιστα λάθη, ενώ η καμπύλη ROC έδειξε εμβαδόν 0.99.

Χρησιμοποιήθηκε το μοντέλο FLAN-T5 Base με διάφορα prompts για την ταξινόμηση συναισθήματος. Τα αποτελέσματα ήταν πολύ, ιδιαίτερα με το απλό prompt, όπου το μοντέλο έδειξε ακρίβεια και ανάκληση πάνω από 0.98. Ο πίνακας σύγχυσης έδειξε ότι το μοντέλο έκανε ελάχιστα λάθη, με μόνο μία αρνητική και έξι θετικές κριτικές να ταξινομούνται λανθασμένα. Τα λεπτομερή και εξειδικευμένα prompts έδειξαν επίσης υψηλή απόδοση, αν και ελαφρώς χαμηλότερη από το απλό

prompt. Τα few-shot και many-shot prompts έδειξαν μειωμένη απόδοση, με το μοντέλο να δυσκολεύεται περισσότερο στην ταξινόμηση των θετικών κριτικών.

Συνολικά, τα αποτελέσματα δείχνουν ότι και τα παραδοσιακά μοντέλα μηχανικής μάθησης (όπως το Logistic Regression) και τα προηγμένα μοντέλα γλωσσικής επεξεργασίας (όπως το BERT και το FLAN-T5) μπορούν να επιτύχουν υψηλή απόδοση στην ανάλυση συναισθήματος. Ωστόσο, η χρήση τεχνικών Prompt Engineering με μοντέλα όπως το FLAN-T5 μπορεί να προσφέρει ανταγωνιστικά αποτελέσματα χωρίς την ανάγκη εκτεταμένης εκπαίδευσης, ιδιαίτερα όταν χρησιμοποιούνται απλά και λεπτομερή prompts.

Τα αποτελέσματα της πειραματικής αξιολόγησης έδειξαν ότι τα LLMs μπορούν να αναγνωρίσουν με υψηλή ακρίβεια συναισθήματα σε ελληνικά κείμενα, ακόμη και σε τομείς όπου η χρήση της ελληνικής γλώσσας παρουσιάζει προκλήσεις λόγω της περιορισμένης διαθεσιμότητας δεδομένων.

5.2 **Μελλοντικές Επεκτάσεις**

Ένα σημαντικό πεδίο μελλοντικής έρευνας είναι η δημιουργία μεγαλύτερων και πιο αντιπροσωπευτικών συνόλων δεδομένων στην ελληνική γλώσσα, ώστε να εμπλουτιστούν τα μοντέλα και να επιτευχθεί καλύτερη γενίκευση. Δεδομένης της έλλειψης επαρκών ελληνικών δεδομένων για εξειδικευμένες εφαρμογές, η συνεργασία με εταιρείες, οργανισμούς και φορείς θα μπορούσε να ενισχύσει την πρόσβαση σε ποιοτικά δεδομένα.

Επιπλέον, η εργασία μπορεί να επεκταθεί προς την κατεύθυνση της πολυτροπικής ανάλυσης (multimodal analysis), ενσωματώνοντας δεδομένα από διαφορετικές πηγές, όπως εικόνες, γραφήματα ή ηχητικά δεδομένα, για την ενίσχυση της ανάλυσης συναισθήματος.

Μια επιπλέον κατεύθυνση είναι η εφαρμογή των LLMs σε πραγματικό χρόνο για την ανάλυση οικονομικών ειδήσεων, κοινωνικών δικτύων ή αναφορών των μέσων μαζικής ενημέρωσης, ώστε να παρέχουν δυναμική πληροφόρηση για την πρόβλεψη αγορών

Βιβλιογραφία

[1] Zhao, Huaqin, et al. “Revolutionizing Finance with LLMs: An Overview of Applications and Insights.” ArXiv (Cornell University), 21 Jan. 2024, https://doi.org/10.48550/arxiv.2401.11641. Accessed 18 Mar. 2024.

[2] Pei, Guanxiong, et al. “Affective Computing: Recent Advances, Challenges, and Future Trends.” Intelligent Computing, vol. 3, 1 Jan. 2024, https://doi.org/10.34133/icomputing.0076.

[3] Bach, Dominik R., and Peter Dayan. “Algorithms for Survival: A Comparative Perspective on Emotions.” Nature Reviews Neuroscience, vol. 18, no. 5, 1 May 2017, pp. 311–319, www.nature.com/articles/nrn.2017.35, https://doi.org/10.1038/nrn.2017.35. Accessed 23 Feb. 2022.

[4] Chen, Luefeng, et al. “Three-Layer Weighted Fuzzy Support Vector Regression for Emotional Intention Understanding in Human–Robot Interaction.” IEEE Transactions on Fuzzy Systems, vol. 26, no. 5, Oct. 2018, pp. 2524–2538, https://doi.org/10.1109/tfuzz.2018.2809691. Accessed 27 Sept. 2020.

[5] McGlynn, N.F. (2014). Thinking fast and slow. *Australian veterinary journal*, *92 12*, N21 .

[6] Fanselow, Michael S. “Emotion, Motivation and Function.” Current Opinion in Behavioral Sciences, vol. 19, Feb. 2018, pp. 105–109, https://doi.org/10.1016/j.cobeha.2017.12.013.

[7] Ekman, Paul. “An Argument for Basic Emotions.” Cognition and Emotion, vol. 6, no. 3-4, May 1992, pp. 169–200, https://doi.org/10.1080/02699939208411068.

[8] Plutchik, Robert. Human Emotions Have Deep Evolutionary Roots, a Fact That May Explain Their Complexity and Provide Tools for Clinical Practice. 1 Jan. 2016. Accessed 18 Feb. 2025.

[9] Russell, James A, and Albert Mehrabian. “Evidence for a Three-Factor Theory of Emotions.” Journal of Research in Personality, vol. 11, no. 3, Sept. 1977, pp. 273–294, https://doi.org/10.1016/0092-6566(77)90037-x.

[10] Bălan, Oana, et al. “Emotion Classification Based on Biophysical Signals and Machine Learning Techniques.” Symmetry, vol. 12, no. 1, 20 Dec. 2019, p. 21, https://doi.org/10.3390/sym12010021.

[11] Mao, Rui, et al. “The Biases of Pre-Trained Language Models: An Empirical Study on Prompt-Based Sentiment Analysis and Emotion Detection.” IEEE Transactions on Affective Computing, vol. 14, no. 3, 2022, pp. 1–11, https://doi.org/10.1109/taffc.2022.3204972.

[12] Hancock, John T., and Taghi M. Khoshgoftaar. “Survey on Categorical Data for Neural Networks.” Journal of Big Data, vol. 7, no. 1, 10 Apr. 2020, https://doi.org/10.1186/s40537-020-00305-w.

[13] Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., & Quarteroni, S. (2013). An Introduction to Information Retrieval.

[14] Ferrone, Lorenzo, and Fabio Massimo Zanzotto. “Symbolic, Distributed, and Distributional Representations for Natural Language Processing in the Era of Deep Learning: A Survey.” Frontiers in Robotics and AI, vol. 6, 21 Jan. 2020, https://doi.org/10.3389/frobt.2019.00153. Accessed 11 Apr. 2020.

[15] Deerwester, Scott, et al. “Indexing by Latent Semantic Analysis.” Journal of the American Society for Information Science, vol. 41, no. 6, Sept. 1990, pp. 391–407, onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9, https://doi.org/10.1002/(sici)1097-4571(199009)41:6%3C391::aid-asi1%3E3.0.co;2-9. Accessed 9 July 2019.

[16] Mikolov, Tomas, et al. “Efficient Estimation of Word Representations in Vector Space.” ArXiv.org, 6 Sept. 2013, arxiv.org/abs/1301.3781. Accessed 21 May 2023.

[17] Pennington, Jeffrey, et al. “Glove: Global Vectors for Word Representation.” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543, www.aclweb.org/anthology/D14-1162/, https://doi.org/10.3115/v1/d14-1162.

[18] Peters, Matthew, et al. “Deep Contextualized Word Representations.” Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), vol. 1, 2018, aclweb.org/anthology/N18-1202/, https://doi.org/10.18653/v1/n18-1202.

[19] Devlin, Jacob, et al. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” ArXiv (Cornell University), 11 Oct. 2018.

[20] Clark, Alexander, et al. Statistical Representation of Grammaticality Judgements: The Limits of N-Gram Models. 1 Aug. 2013, pp. 28–36.

[21] Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research, 3*, 1137-1155.

[22] Mikolov, T., Karafiát, M., Burget, L., Černocký, J.H., & Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech*.

[23] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*.

[24] Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training.

[25] Thakkar, Krishna Yatin, and Nimit Jagdishbhai. “Exploring the Capabilities and Limitations of GPT and Chat GPT in Natural Language Processing.” Journal of Management Research and Analysis, vol. 10, no. 1, 15 Apr. 2023, pp. 18–20, https://doi.org/10.18231/j.jmra.2023.004.

[26] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.

[27] Priyanshu, A., Maurya, Y., & Hong, Z. (2024). AI Governance and Accountability: An Analysis of Anthropic's Claude. *ArXiv, abs/2407.01557*.

[28] Touvron, Hugo, et al. “LLaMA: Open and Efficient Foundation Language Models.” ArXiv (Cornell University), 27 Feb. 2023, https://doi.org/10.48550/arxiv.2302.13971.

[29] Yao, Yifan, et al. “A Survey on Large Language Model (LLM) Security and Privacy: The Good,the Bad, and the Ugly.” ArXiv (Cornell University), 4 Dec. 2023, https://doi.org/10.48550/arxiv.2312.02003.

[30] Touvron, H., Martin, L., Stone, K.R., et.al.(2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. *ArXiv, abs/2307.09288*.

[31] Dubey, A., Jauhri, A., Pandey, A., et.al. (2024). The Llama 3 Herd of Models. *ArXiv, abs/2407.21783*.

[32] Ehrmann, M., & Fratzscher, M. (2005). Central bank communication: Evidence from a survey of international central banks. *Journal of Economic Policy Reform, 37*(3), 591–516

[33] Gürkaynak, R. S., Sack, B., & Swanson, E. T. (2005). The bond yield conundrum from a macro-finance perspective. *Federal Reserve Bank of New York Economic Policy Review, 11*(2), 151–165.

[34] Blinder, Alan S, et al. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." Journal of Economic Literature, vol. 46, no. 4, 1 Nov. 2008, pp. 910–945, https://doi.org/10.1257/jel.46.4.910. Accessed 13 Oct. 2020.

[35] Swanson, Eric T. "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets." Journal of Monetary Economics, vol. 118, Mar. 2021, pp. 32–53, https://doi.org/10.1016/j.jmoneco.2020.09.003.

[36] Hansen, Stephen, and Michael McMahon. "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication." Journal of International Economics, vol. 99, Mar. 2016, pp. S114–S133, https://doi.org/10.1016/j.jinteco.2015.12.008. Accessed 22 Sept. 2019.

[37] Apel, M., & Grimaldi, M. B. (2014). How informative are central bank minutes? *International Journal of Central Banking, 65*(1), 53–76.

[38] Hubert, P., & Labondance, F. (2018). The effect of ECB forward guidance on the term structure of interest rates. *International Journal of Central Banking, 14*(5), 193–222.

[39] Fatemi, Sorouralsadat, et al. "A Comparative Analysis of Instruction Fine-Tuning Large Language Models for Financial Text Classification." ACM Transactions on Management Information Systems, 4 Dec. 2024, https://doi.org/10.1145/3706119. Accessed 12 Jan. 2025.

[40] Liu, Chenghao, et al. "Large Language Models and Sentiment Analysis in Financial Markets: A Review, Datasets, and Case Study." IEEE Access, vol. 12, 2024, pp. 134041–134061, ieeexplore.ieee.org/document/10638546, https://doi.org/10.1109/access.2024.3445413. Accessed 11 Oct. 2024.

[41] Salur, B.V. (2013). Investor sentiment in the stock market.

[42] de Long, J., Shleifer, A.V., Summers, L., & Waldmann, R.J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy, 98*, 703 - 738.

[43] Shleifer, A., & Vishny, R.W. (1995). The Limits of Arbitrage. *NBER Working Paper Series*.

[44] Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters*.

[45] Giorgos Iacovides, et al. FinLlama: LLM-Based Financial Sentiment Analysis for Algorithmic Trading. 14 Nov. 2024, pp. 134–141, https://doi.org/10.1145/3677052.3698696. Accessed 18 Feb. 2025.

[46] Föhr, Tassilo Lars, et al. "Assuring Sustainable Futures: Auditing Sustainability Reports Using AI Foundation Models." Social Science Research Network, 4 July 2023, papers.ssrn.com/sol3/papers.cfm?abstract_id=4502549. Accessed 27 Oct. 2023.

[47] Lehner, S. (2024). *The future of accounting: Sentiment analysis in AI-rewritten SEC filings*. SSRN. https://doi.org/10.2139/ssrn.4984337

[50] Lívia Réka Ónozó, et al. "Leveraging LLMs for Financial News Analysis and Macroeconomic Indicator Nowcasting." IEEE Access, vol. PP, no. 99, 1 Jan. 2024, p. 1,

www.researchgate.net/publication/385412244_Leveraging_LLMs_for_Financial_News_Analysis_and_Macroeconomic_Indicator_Nowcasting, https://doi.org/10.1109/ACCESS.2024.3488363.

[51] Sun, Chi-Kuang, et al. “How to Fine-Tune BERT for Text Classification?” ArXiv (Cornell University), 14 May 2019, https://doi.org/10.48550/arxiv.1905.05583. Accessed 26 Apr. 2023.

[52] Zhao, M., Lin, T., Jaggi, M., & Schütze, H. (2020). Masking as an Efficient Alternative to Finetuning for Pretrained Language Models. *Conference on Empirical Methods in Natural Language Processing*.

[53] Lester, Brian, et al. “The Power of Scale for Parameter-Efficient Prompt Tuning.” Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, https://doi.org/10.18653/v1/2021.emnlp-main.243.

[54] Jason Zhanshun Wei, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. 27 Jan. 2022, https://doi.org/10.48550/arxiv.2201.11903.

[55] Malo, Pekka, et al. “Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts.” Journal of the Association for Information Science and Technology, vol. 65, no. 4, 26 Nov. 2013, pp. 782–796, https://doi.org/10.1002/asi.23062.

[56] Fragkis, N. (n.d.). Skroutz Shop Reviews - Sentiment Analysis [Dataset]. Kaggle. Retrieved January 28, 2025, from https://www.kaggle.com/datasets/nikosfragkis/skroutz-shop-reviews-sentiment-analysis/data