Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση

(2024)

8ο Πανελλήνιο Επιστημονικό Συνέδριο «Ένταξη και Χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία»

Μια συγκριτική αξιολόγηση πολυ-γλωσσικών μοντέλων ενσωματώσεων κειμένου στο πλαίσιο Αναλυτικής της Μάθησης

Βασιλική Ραγάζου, Χαράλαμπος Παπαδήμας, Ηλίας Καρασαββίδης, Βασίλειος Κόλλιας

Βιβλιογραφική αναφορά:

Ραγάζου Β., Παπαδήμας Χ., Καρασαββίδης Η., & Κόλλιας Β. (2025). Μια συγκριτική αξιολόγηση πολυ-γλωσσικών μοντέλων ενσωματώσεων κειμένου στο πλαίσιο Αναλυτικής της Μάθησης. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 480–493. ανακτήθηκε από https://eproceedings.epublishing.ekt.gr/index.php/cetpe/article/view/8465

Μια συγκριτική αξιολόγηση πολυ-γλωσσικών μοντέλων ενσωματώσεων κειμένου στο πλαίσιο Αναλυτικής της Μάθησης

Βασιλική Ραγάζου[1], Χαράλαμπος Παπαδήμας[1], Ηλίας Καρασαββίδης[1], Βασίλειος Κόλλιας[2]

ragazou@uth.gr,papadimas@uth.gr,ikaras@uth.gr,vkollias@uth.gr

[1] Παιδαγωγικό Τμήμα Προσχολικής Εκπαίδευσης, Πανεπιστήμιο Θεσσαλίας

[2] Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης, Πανεπιστήμιο Θεσσαλίας

Περίληψη

Στο πλαίσιο της Αναλυτικής της Μάθησης (ΑΜ) η αξιοποίηση των κειμένων που δημιουργούν φοιτητές είναι περιορισμένη. Η ραγδαία πρόοδος που έχει συντελεστεί στο πεδίο της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) έχει οδηγήσει στην ανάπτυξη διαφόρων μοντέλων ενσωμάτωσης κειμένου, διανοίγοντας νέους δρόμους αξιοποίησης για την ΑΜ. Η παρούσα εργασία επιχειρεί να αξιολογήσει συγκριτικά διάφορα μοντέλα ενσωμάτωσης κειμένου που υποστηρίζουν την Ελληνική γλώσσα. Στην έρευνα που περιγράφεται συμμετείχαν 254 προπτυχιακοί φοιτητές ενός περιφερειακού πανεπιστημίου, οι οποίοι παρακολούθησαν μια σειρά έξι βιντεοδιαλέξεων και κλήθηκαν να συντάξουν μια σύντομη περίληψη για την κάθε μια. Τα ανύσματα που δημιουργήθηκαν από τα μοντέλα ενσωμάτωσης κειμένου χρησιμοποιήθηκαν ως είσοδοι σε οκτώ αλγορίθμους Μηχανικής Μάθησης (ΜΜ) για την πρόβλεψη της επίδοσης. Τα αποτελέσματα έδειξαν ότι η απόδοση των μοντέλων ενσωμάτωσης κειμένου είναι αρκετά ικανοποιητική για την Ελληνική γλώσσα, αλλά υπάρχουν μεγάλα περιθώρια βελτίωσης.

Λέξεις κλειδιά: Πολυ-γλωσσικά μοντέλα ενσωμάτωσης κειμένου, Αναλυτική της μάθησης, Ηλεκτρονική μάθηση, Πρόβλεψη επίδοσης

Εισαγωγή

Η Αναλυτική της Μάθησης (ΑΜ) (Learning Analytics) ορίζεται ως η "μέτρηση, συλλογή, ανάλυση και αναφορά δεδομένων για μαθητές και τα πλαίσια τους με σκοπό την κατανόηση και βελτιστοποίηση της μάθησης και των περιβαλλόντων εντός των οποίων συντελείται" (Long & Siemens, 2011, σ. 34). Παρόλο που τα κείμενα που δημιουργούν οι φοιτητές αποτελούν έναν από τους τύπους δεδομένων που μπορούν δυνητικά να αξιοποιηθούν για τη βελτίωση της μάθησης, η έρευνα δείχνει ότι η χρήση τους στο πλαίσιο της ΑΜ δεν είναι διαδεδομένη (Mangaroska & Giannakos, 2018; Banihashem et al., 2022). Δεδομένων των καταιγιστικών εξελίξεων στο πεδίο της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) κατά την τελευταία δεκαετία, πολλές μελέτες έχουν εστιάσει στην ανάλυση κειμένων που δημιουργούν φοιτητές για διάφορα έργα, όπως π.χ. επιτυχημένη ολοκλήρωση μαθήματος (Robinson et al., 2016), εντοπισμός ανάγκης άμεσης ανταπόκρισης εκπαιδευτών σε φόρουμ συζήτησης (Almatrafi et al. 2018) και ανίχνευση σύγχυσης από μηνύματα σε φόρουμ συζήτησης (Agrawal et al., 2015). Ωστόσο, η ΕΦΓ για τη εξαγωγή χαρακτηριστικών από κείμενα φοιτητών/τριων με σκοπό την πρόβλεψη της επίδοσης και τη συνακόλουθη υποστήριξη της μάθησης δεν έχει απασχολήσει συστηματικά την ερευνητική κοινότητα. Υιοθετώντας νέες μεθόδους αναπαράστασης κειμένου, η παρούσα εργασία εξετάζει συγκριτικά την απόδοση τους για την πρόβλεψη της επίδοσης με σκοπό την υποστήριξη της μάθησης.

Αναπαράσταση κειμένου

Ιστορικά, η ανυσματική αναπαράσταση κειμένου περιλαμβάνει τρεις γενικές προσεγγίσεις: (α) δυαδικά ανύσματα (one-hot vectors), (β) ανύσματα συχνοτήτων (frequency vectors) και (γ) ανύσματα ενσωματώσεων (embedding vectors) (Manning et al,. 2008; Aggarwal & Zhai, 2012; Παναγιωτακόπουλος κ.α. 2023). Οι δύο πρώτες κατηγορίες ανυσμάτων χαρακτηρίζονται από μια σημαντική σειρά περιορισμών. Πρώτον, οι διαστάσεις των ανυσμάτων είναι συνάρτηση του πλήθους των λέξεων, με αποτέλεσμα τα ανύσματα που δημιουργούνται να έχουν πολύ μεγάλες διαστάσεις. Δεύτερον, τα ανύσματα είναι αραιά, καθώς οι συχνότητες (απόλυτες ή σταθμισμένες) είναι στην πλειοψηφία τους μηδενικές. Τρίτον, δε λαμβάνεται υπόψη η σειρά των λέξεων, στοιχείο άκρως προβληματικό, επειδή κατά κανόνα η σημασία μιας λέξης καθορίζεται από τη συγκεκριμένη ακολουθία των λέξεων εντός της οποίας απαντάται. Τέλος, οι λέξεις μοντελοποιούνται ως αυτόνομες και διακριτές οντότητες που δεν έχουν σχέση μεταξύ τους, χωρίς να αναπαριστάται η σημασιολογική τους συσχέτιση. Για παράδειγμα, δύο λέξεις που είναι κοντινές σε σημασία (π.χ. "δυνατός", "ισχυρός") θα βρίσκονται στην ίδια απόσταση στο σημασιολογικό χώρο σε σχέση με δύο λέξεις που έχουν εντελώς διαφορετική σημασία (π.χ. "δυνατός", "επιδημιολογικός").

Σχετικά με την τρίτη κατηγορία, οι περιορισμοί των παραπάνω αναπαραστάσεων αντιμετωπίστηκαν με τις νευρωνικές αναπαραστάσεις κειμένου μέσω του μοντέλου word2vec (Mikolov et al., 2013a; 2013b) που επέτρεψε την ταχύτατη δημιουργία ανυσμάτων λέξεων πολλών διαστάσεων από μεγάλα σώματα κειμένων που είναι γνωστά ως ενσωματώσεις λέξεων (word embeddings) ή ενσωματώσεις κειμένου γενικότερα (text embeddings). Πιο συγκεκριμένα, οι ενσωματώσεις είναι Μαθηματικά αντικείμενα που αναπαριστούν οντότητες διαφόρων τύπων (όπως π.χ. λέξεις) με μικρό αριθμό διαστάσεων. Συνιστούν ν-διάστατα ανύσματα, όπου κάθε διάσταση – θεωρητικά – αντιστοιχεί σε κάποιο χαρακτηριστικό της λέξης που αναπαριστά. Ενώ κάθε διάσταση θεωρητικά αποτυπώνει κάποιο χαρακτηριστικό ή ιδιότητα της λέξης, δεν μπορούμε να γνωρίζουμε ποιο χαρακτηριστικό αποτυπώνεται από μια συγκεκριμένη διάσταση. Κάθε λέξη αναπαρίσταται ως ένα σημείο σε ένα ν-διάστατο χώρο, ο οποίος χρησιμοποιείται για να αποτυπώσει τα χαρακτηριστικά της.

Οι ενσωματώσεις κειμένου (text embeddings) μπορούν να διακριθούν σε δύο γενικές κατηγορίες: (α) γενικές (ή στατικές) ενσωματώσεις κειμένου και (β) πλαισιωμένες ενσωματώσεις κειμένου. Οι γενικές ενσωματώσεις κειμένου δημιουργούνται από τους αλγορίθμους CBOW και Skip-gram του μοντέλου word2vec (Mikolov et al., 2013a; 2013b). Παρά την επανάσταση που έφερε στο πεδίο της ΕΦΓ το μοντέλο word2vec, οι γενικές ενσωματώσεις κειμένου χαρακτηρίζονται από έναν σημαντικό περιορισμό. Ειδικότερα, δημιουργείται μια καθολική αναπαράσταση του νοήματος μιας λέξης, που σημαίνει ότι μια λέξη θα έχει πάντα το ίδιο άνυσμα ανεξάρτητα από το πλαίσιο στο οποίο εμφανίζεται (static word embedding). Παρότι μια λέξη μπορεί φυσικά να έχει ένα γενικό νόημα, η ακριβής σημασία της εξαρτάται πάντοτε από το συγκείμενο της.

Η υπέρβαση του περιορισμού αυτού επιτυγχάνεται με τη χρήση πλαισιωμένων ενσωματώσεων λέξης (contextual word embeddings), οι οποίες δημιουργούν μια αναπαράσταση της λέξης που είναι συνάρτηση των υπολοίπων λέξεων, λαμβάνοντας με τον τρόπο αυτό υπόψη το υφιστάμενο πλαίσιο. Η έρευνα δείχνει ότι σε πολλά έργα ΕΦΓ οι πλαισιωμένες ενσωματώσεις λέξης έχουν καλύτερη απόδοση από τις αντίστοιχες στατικές ενσωματώσεις λέξης (Liu et al., 2019; Reimers et al., 2019), όπως επίσης ότι οι ενσωματώσεις προτάσεων αποδίδουν καλύτερα σε σχέση με τις ενσωματώσεις λέξεων (Cer et al., 2018).

Στη βιβλιογραφία καταγράφονται πολλές συστηματικές προσπάθειες δημιουργίας πλαισιωμένων ενσωματώσεων λέξεων και μεγαλύτερων κειμενικών ενοτήτων, όπως π.χ. οι προτάσεις. Αρχικά, επιχειρήθηκε η επέκταση του επιτυχημένου μοντέλου word2vec στη

δημιουργία ενσωματώσεων για ευρύτερες ενότητες κειμένου, όπως π.χ. προτάσεις (Mikolov et al., 2013b), παράγραφοι (Le & Mikolov, 2014) ή άλλων στατιστικών χαρακτηριστικών του κειμένου (Pennington et al., 2014). Ακολούθως, επιχειρήθηκε η εφαρμογή άλλων τύπων νευρωνικών δικτύων (Recurrent Neural Networks) στη δημιουργία αναπαραστάσεων κειμένου σε επίπεδο πέραν της λέξης (π.χ. Conneau et al. 2017; Logeswaran & Lee, 2018; Peters et al., 2018). Τέλος, καθοριστική ήταν η εμφάνιση της αρχιτεκτονικής των Μετασχηματιστών (Vaswani et al.,2017) που οδήγησε στη δημιουργία Μεγάλων Γλωσσικών Μοντέλων (ΜΓΜ) (Large Language Models), κωδικοποιητών (encoders) (π.χ. BERT: Devlin et al., 2019), αποκωδικοποιητών (π.χ. GPT: Radford et al., 2018) ή συνδυασμού τους (π.χ. T5: Raffel et al., 2020).

Σήμερα, η δημιουργία πλαισιωμένων ενσωματώσεων περιλαμβάνει δύο γενικές τάσεις. Πρώτον, επιχειρείται είτε η απευθείας εξαγωγή ενσωματώσεων κειμένου από Μεγάλα Γλωσσικά Μοντέλα (ΜΓΜ) (Large Language Models) είτε η χρήση τους για τη δημιουργία ενσωματώσεων κειμένου (π.χ. Jiang et al., 2023; Springer et al., 2024). Ωστόσο, προκύπτει πως οι ενσωματώσεις λέξεων που δημιουργούνται από ΜΓΜ είναι ακατάλληλες για τη δημιουργία αναπαραστάσεων που λαμβάνουν υπόψη τη σημασιολογική ομοιότητα κειμένων. Για παράδειγμα, οι Reimers και Gurevych (2019) έδειξαν ότι ο Μετασχηματιστής BERT αντιστοιχίζει προτάσεις σε έναν πολυδιάστατο χώρο, όπου οι τυπικές μετρήσεις σημασιολογικής ομοιότητας, όπως π.χ. η ομοιότητα συνημιτόνου, δεν έχουν πρακτική εφαρμογή.

Δεύτερον, αναπτύχθηκαν επί τούτου μοντέλα ενσωματώσεων κειμένου (text embedding models), τα οποία δεν εστιάζουν σε λέξεις αλλά σε ευρύτερες κειμενικές ενότητες όπως π.χ. SBERT (Reimers & Gurevych, 2019), SGPT (Muennighoff, 2022), ST5 (Ni et al., 2021) και GTE (Li et al., 2023). Ενώ ένα ΜΓΜ δέχεται ως είσοδο μια ακολουθία λέξεων και υπολογίζει την επόμενη πιθανή λέξη, ένα μοντέλο ενσωμάτωσης κειμένου δέχεται ως είσοδο μια ακολουθία λέξεων και υπολογίζει ένα άνυσμα σταθερής διάστασης, το οποίο συνιστά μια πλαισιωμένη αναπαράσταση της ακολουθίας αυτής. Οι επιδόσεις των μοντέλων ενσωμάτωσης που αναπτύσσονται είναι εντυπωσιακές τόσο σε επίπεδο σημασιολογικής κειμενικής ομοιότητας, όσο και σε επίπεδο άλλων τυπικών έργων του πεδίου της ΕΦΓ.

Ο κύριος περιορισμός αυτών των μοντέλων ενσωμάτωσης κειμένου, είναι ότι οι επιδόσεις τους αφορούν πρωτίστως την Αγγλική γλώσσα. Από τη μία πλευρά, οι Grave et al. (2018) επισημαίνουν ότι δημιουργία στατικών ενσωματώσεων λέξεων διαμέσου μοντέλων όπως το word2vec έχει ως αποτέλεσμα τη δημιουργία ανυσμάτων χαμηλής ποιότητας για γλώσσες οι οποίες δεν είναι διαδεδομένες (low resource languages). Από την άλλη πλευρά, οι Reimers και Gurevych (2020) επισημαίνουν ότι η πλειοψηφία των ΜΓΜ αλλά και των μοντέλων ενσωματώσεων είναι μονο-γλωσσικά και περιορίζονται στα Αγγλικά.

Γενικά, η δημιουργία ΜΓΜ και ενσωματώσεων κειμένου προϋποθέτει την ύπαρξη μεγάλων σωμάτων κειμένων, η συνηθέστερη πηγή των οποίων είναι το διαδίκτυο. Δεδομένου ότι η πιο κοινή γλώσσα στον παγκόσμιο ιστό είναι τα Αγγλικά, η διαθεσιμότητα Αγγλικών κειμένων είναι πολλαπλάσια της αντίστοιχης διαθεσιμότητας κειμένων σε άλλες γλώσσες. Για την αντιμετώπιση αυτής της ανάγκης έχουν αναπτυχθεί διάφορα πολυ-γλωσσικά μοντέλα ενσωμάτωσης κειμένου (multilingual text-embedding models), π.χ. mSBERT (Reimers & Gurevych, 2020), LaBSE (Feng et al., 2022), Nomic (Lee et al., 2024), BGE (Chen et al., 2024) και E5 (Wang et al., 2024).

Ενώ χρησιμοποιώντας διάφορα διαδεδομένα γλωσσολογικά κριτήρια (benchmarks, όπως π.χ. το MTEB) η έρευνα δείχνει πως η απόδοση των πολυ-γλωσσικών αυτών μοντέλων ενσωματώσεων κειμένου είναι πολύ υψηλή, οι δυνατότητες τους παραμένουν αχαρτογράφητες για το πεδίο της ΑΜ. Υπό αυτό το πρίσμα, οι Pijeira-Díaz et al. (2024a; 2024b)

επισήμαναν την επιτακτική ανάγκη διερεύνησης του δυναμικού των μοντέλων ενσωμάτωσης κειμένου εκτός της Αγγλικής γλώσσας, δηλαδή σε μη διαδεδομένες γλώσσες (low resource languages). Σε αυτό το πλαίσιο, προχώρησαν στην αξιολόγηση πολυ-γλωσσικών μοντέλων ενσωμάτωσης λέξεων που υποστηρίζουν την Ολλανδική γλώσσα. Ειδικότερα, η μελέτη των Pijeira-Díaz et al. (2024a) εξέτασε συγκριτικά τέσσερα μοντέλα ενσωμάτωσης κειμένου: spaCy (medium και large), FastText και ConceptNet NumberBatch. Από την άλλη πλευρά, οι Pijeira-Díaz et al. (2024b) εστίασαν στη συγκριτική απόδοση τριών μοντέλων ενσωμάτωσης κειμένου που βασίζονται στον Μετασχηματιστή BERT: SBERT, RobBERT και BERTje. Σε αμφότερες τις μελέτες εξετάστηκε η αποδοτικότητα των μοντέλων ενσωμάτωσης κειμένων με διάφορους αλγόριθμους ταξινόμησης Μηχανικής Μάθησης (ΜΜ) και Βαθιάς Μάθησης (ΒΜ) με υποσχόμενα αποτελέσματα για την Ολλανδική γλώσσα.

Σκοπός της μελέτης

Παρόλο που κάποια από τα πολυ-γλωσσικά μοντέλα ενσωμάτωσης που έχουν αναπτυχθεί πρόσφατα υποστηρίζουν Ελληνικά, απουσιάζει μια συστηματική διερεύνηση τους για τις ανάγκες της ΑΜ. Σε προηγούμενες εργασίες (Karasavvidis et al., 2022,) είχαμε εξετάσει τη δυνατότητα χρήσης κειμένων που δημιουργούν οι φοιτητές στα πλαίσια της ηλεκτρονικής μάθησης για την πρόβλεψη της επίδοσης τους από την παρακολούθηση βιντεοδιαλέξεων. Είχαμε διερευνήσει τόσο το δυναμικό των ανυσμάτων συχνοτήτων όσο και των ανυσμάτων ενσωματώσεων (Ραγάζου et al., 2022) αλλά και κειμενικής ομοιότητας για την πρόβλεψη της επίδοσης (Karasavvidis et al., 2022; Παπαδήμας et al., 2023). Στην παρούσα εργασία επεκτείνουμε την προγενέστερη έρευνα εστιάζοντας στη συστηματική αξιολόγηση πολυ-γλωσσικών μοντέλων ενσωμάτωσης λέξεων ή προτάσεων που υποστηρίζουν Ελληνικά. Ειδικότερα, επιχειρούμε να αξιολογήσουμε συγκριτικά τα διαθέσιμα μοντέλα ενσωμάτωσης κειμένου που είναι διαθέσιμα στα Ελληνικά για τις ανάγκες της ΑΜ σε περιβάλλοντα ηλεκτρονικής μάθησης.

Ο πρώτος στόχος της μελέτης είναι η συγκριτική αξιολόγηση των κειμενικών αναπαραστάσεων που δημιουργούνται από διαθέσιμα μοντέλα ενσωμάτωσης διαφορετικών διαστάσεων. Η υπόθεση εργασίας είναι ότι τα μοντέλα ενσωμάτωσης κειμένου που χρησιμοποιούν ανύσματα μεγαλύτερων διαστάσεων θα έχουν καλύτερη απόδοση στην ταξινόμηση της επίδοσης. Θεωρητικά, όσο μεγαλύτερη είναι η διάσταση ενός ανύσματος τόσο πιο εφικτή είναι η αναπαράσταση χαρακτηριστικών της λέξης με μεγαλύτερη λεπτομέρεια. Στη λογική ότι κάθε διάσταση συλλαμβάνει ένα ιδιαίτερο χαρακτηριστικό της λέξης, η ύπαρξη μεγάλων διαστάσεων συνεπάγεται δυνητικά την ικανότητα σύλληψης περισσότερων λεπτομερειών. Σύμφωνα με τους Mikolov et al. (2013a), η δημιουργία ανυσμάτων μεγάλων διαστάσεων επιτρέπει την ανίχνευση λεπτών σημασιολογικών διακρίσεων μεταξύ λέξεων. Σε αντίστοιχα συμπεράσματα καταλήγουν οι Turian et al. (2010) και οι Conneau et al. (2017).

Ο δεύτερος στόχος της μελέτης είναι η διερεύνηση της προβλεπτικής αξίας που έχουν τα μοντέλα ενσωμάτωσης κειμένου σε συνδυασμό με διάφορους αλγόριθμους Μηχανικής Μάθησης (ΜΜ). Για τις ανάγκες της μελέτης χρησιμοποιήθηκαν τα παρακάτω μοντέλα ενσωματώσεων κειμένου που υποστηρίζουν την Ελληνική γλώσσα: (α) spaCy (ref), (β) FastText (Grave et al., 2018), (γ) Nomic (Lee et al., 2024; Nussbaum et al., 2024), (δ) SBERT (Reimers & Gurevych, 2019; 2020), (ε) BGE (Xiao et al., 2023; Chen et al., 2024) και (στ) E5 και E5-Instruct (Wang et al., 2023; Wang et al., 2024). Τα μοντέλα της spaCy και FastText βασίζονται στην αρχιτεκτονική word2vec και δημιουργούν ανύσματα 300 διαστάσεων. Αμφότερα μοντέλα δημιουργούν ανύσματα σταθερού μήκους για προτάσεις μεταβλητού μήκους, υπολογίζοντας τον μέσο όρο των ανυσμάτων της κάθε λέξης που απαρτίζουν την πρόταση. Τα υπόλοιπα μοντέλα ενσωματώσεων δημιουργούν ανύσματα που βασίζονται στην

αρχιτεκτονική των Μετασχηματιστών, έχοντας είτε 768 (Nomic, SBERT) είτε 1024 διαστάσεις (BGE, E5, E5-Instruct).

Τα ερευνητικά ερωτήματα στα οποία επιχειρεί να απαντήσει η παρούσα μελέτη είναι τα εξής:

#1. *Πόσο αποτελεσματικά είναι τα διαθέσιμα ανοικτά πολυ-γλωσσικά μοντέλα ενσωμάτωσης κειμένου για τη νευρωνική αναπαράσταση ελληνικού κειμένου με όρους ταξινόμησης επίδοσης;*
#2. *Πόσο επηρεάζει η διάσταση του κάθε ανύσματος την αποτελεσματικότητα της ταξινόμησης;*
#3. *Ποιοι αλγόριθμοι μηχανικής μάθησης αποδίδουν περισσότερο με τα διαθέσιμα ανοικτά πολυ-γλωσσικά μοντέλα ενσωμάτωσης κειμένου;*

Μέθοδος

Συμμετέχοντες

Οι συμμετέχοντες στην έρευνα ήταν 254 φοιτήτριες και φοιτητές από τμήματα ανθρωπιστικών και θετικών επιστημών (Παιδαγωγικά, Πληροφορική) περιφερειακού ΑΕΙ της χώρας (87% γυναίκες, 13% άνδρες). Η ηλικιακή διακύμανση των συμμετεχόντων ήταν μεταξύ 18 και 45 ετών (M = 20.77, SD = 3.91). Παρότι η συμμετοχή στην έρευνα ήταν εθελοντική, καθώς οι φοιτήτριες και φοιτητές ανταποκρίθηκαν σε σχετική πρόσκληση, δόθηκε βαθμολογικό κίνητρο συμμετοχής.

Υλικά

Στο πλαίσιο της έρευνας, χρησιμοποιήθηκε το Σύστημα Διαχείρισης Μάθησης (ΣΔΜ) Moodle, το οποίο και παραμετροποιήθηκε σχετικά. Σχεδιάστηκε μια σειρά έξι βιντεοδιαλέξεων (διάρκειας 8 -12′) με θεματικές ενότητες των ψηφιακών μέσων. Ο σχεδιασμός των βιντεοδιαλέξεων βασίστηκε σε θεμελιώδεις αρχές της μάθησης με πολυμέσα.

Μετρήσεις

Τα δεδομένα που συλλέχθηκαν από το ΣΔΜ αφορούσαν τις βαθμολογικές επιδόσεις των φοιτητών σε τεστ δηλωτικής γνώσης και τις γραπτές περιλήψεις μετά από κάθε παρακολούθηση βιντεοδιάλεξης. Το τεστ δηλωτικής γνώσης περιλάμβανε δέκα ερωτήματα κλειστού τύπου τα οποία κάλυπταν το περιεχόμενο της εκάστοτε βιντεοδιάλεξης. Η βαθμολογία ήταν δίτιμη (για κάθε σωστή απάντηση η βαθμολογία ήταν 1 βαθμός, ενώ για κάθε λανθασμένη η βαθμολογία ήταν 0). Συνολικά, η μεταβλητή «επίδοση» προέκυψε από το άθροισμα των τιμών κάθε τεστ. Στη γραπτή περίληψη οι συμμετέχοντες κλήθηκαν να γράψουν ένα κείμενο συνολικής έκτασης 100 λέξεων στο οποίο αποτύπωναν την κατανόησή τους για τα θέματα της βιντεοδιάλεξης που παρακολούθησαν.

Διαδικασία

Η έρευνα διεξάχθηκε εξ αποστάσεως μεταξύ 2020-2023, λόγω των περιορισμών που είχαν επιβληθεί κατά τη διάρκεια της πανδημίας Covid-19 και περιλάμβανε τρεις γενικές φάσεις. Στην πρώτη φάση, οι φοιτητές ενημερώθηκαν για τον σκοπό και τις προϋποθέσεις συμμετοχής. Στη συνέχεια εκδήλωσαν ενδιαφέρον συμμετοχής μέσω μιας ηλεκτρονικής φόρμας και παρέλαβαν μέσω της υπηρεσίας ηλεκτρονικού ταχυδρομείου τα διαπιστευτήρια εισόδου και τις οδηγίες πρόσβασης στο Σύστημα Διαχείρισης Μάθησης (ΣΔΜ). Στη δεύτερη φάση οι συμμετέχοντες συμπλήρωσαν ερωτηματολόγια δημογραφικού τύπου, εξοικείωσης με τις ΤΠΕ και απόψεων/στάσεων αναφορικά με τη γνωστική δυσκολία των βίντεο διαλέξεων και την παρώθηση. Το περιεχόμενο των βιντεοδιαλέξεων ήταν οργανωμένο σε έξι θεματικές

ενότητες. Κάθε ενότητα περιλάμβανε διαδοχικά τρεις πόρους: (α) τη βιντεοδιάλεξη, (β) τη φόρμα συγγραφής της περίληψης και (γ) το τεστ δηλωτικής γνώσης. Οι συμμετέχοντες μπορούσαν να προχωρήσουν στο επόμενο βήμα μόνο αφού ολοκλήρωναν το προηγούμενο (π.χ. δεν μπορούσαν να απαντήσουν το τεστ δηλωτικής γνώσης πριν γράψουν την περίληψη). Η ίδια διαδικασία ακολούθησε και για τις υπόλοιπες 5 βιντεοδιαλέξεις. Στην τρίτη φάση οι φοιτητές που ολοκλήρωσαν επιτυχώς τη διαδικασία, έλαβαν ευχαριστήριο μήνυμα για τη συμμετοχή τους. Η συνολική διάρκεια της έρευνας ήταν περίπου τρεις ώρες.

Ανάλυση

Ενσωματώσεις Κειμένου

Αρχικά, πραγματοποιήθηκε εξαγωγή όλων των δεδομένων από το ΣΔΜ και η αποθήκευση τους για την περαιτέρω επεξεργασία . Το πρωτογενές υλικό της μελέτης αποτέλεσαν οι περιλήψεις των βιντεοδιαλέξεων που έγραψαν οι συμμετέχοντες. Για τις ανάγκες της μελέτης υιοθετήσαμε αποκλειστικά νευρωνικές αναπαραστάσεις κειμένου. Στον Πίνακα 1 παρουσιάζονται τα ΜΓΜ (spaCy) και μοντέλα ενσωματώσεων κειμένου που χρησιμοποιήθηκαν για την αναπαράσταση των περιλήψεων.

Πίνακας 1. Μοντέλα Ενσωμάτωσης

Μοντέλο Ενσωμάτωσης	Περιγραφή	Τύπος Ενσωμάτωσης	Διαστάσεις
Spacy (Explosion, 2024)	el_core_news_lg	Στατική	300
FastText (Grave et al., 2018))		Στατική	300
SBERT (Reimers & Gurevych, 2019; 2020)	sn-xlm-roberta-base-snli-mnli-anli-xnli	Πλαισιωμένη	768
Nomic (Lee et al., 2024; Nussbaum et al., 2024)	nomic-embed-text-v1.5	Πλαισιωμένη	768
E5 (Wang et al., 2023; Wang et al., 2024)	multilingual-e5-large-instruct multilingual-e5-large	Πλαισιωμένη	1024
BGE (Xiao et al., 2023; Chen et al., 2024)	BGE-m3	Πλαισιωμένη	1024

Θα πρέπει να σημειώσουμε πως όλα τα μοντέλα ενσωμάτωσης κειμένου που χρησιμοποιήσαμε είναι ανοικτά και ελεύθερα προσβάσιμα στην ερευνητική κοινότητα, δηλαδή διατίθενται τουλάχιστον τα βάρη τους. Κατά περίπτωση, μπορεί παράλληλα να παρέχονται επιπλέον στοιχεία, όπως π.χ. ο κώδικας που χρησιμοποιήθηκε για την εκπαίδευση τους ή η βάση κειμένων πάνω στην οποία εκπαιδεύτηκαν . Τα μοντέλα διατίθενται είτε από ιστοχώρους διαμοίρασης της κοινότητας (όπως είναι π.χ. το Huggingface ή το GitHub) είτε απευθείας από τους αντίστοιχους φορείς, εργαστήρια ή εταιρείες δημιουργίας τους (π.χ. spaCy, FastText). Γενικά, όλα τα μοντέλα ενσωμάτωσης κειμένου που χρησιμοποιήσαμε είναι διαθέσιμα χωρίς συνδρομή ή άλλο περιορισμό. Δε χρησιμοποιήσαμε ΜΓΜ ή μοντέλα ενσωμάτωσης που είναι κλειστά και διατίθενται συνδρομητικά μέσω Προγραμματιστικής Διεπαφής Χρήστη (API) (όπως είναι π.χ. τα μοντέλα της Google ή της OpenAI).

Δείκτες κεντρικής τάσης και διασποράς για την επίδοση

Η επίδοση των συμμετεχόντων σε κάθε βιντεοδιάλεξη χρησιμοποιήθηκε για τη δημιουργία δυαδικών μεταβλητών. Το φίλτρο που χρησιμοποιήθηκε για τη δημιουργία των μεταβλητών

αυτών ήταν η διάμεσος. Δημιουργήθηκαν δύο κλάσεις συμμετεχόντων, μια χαμηλής επίδοσης και μια υψηλής επίδοσης αντίστοιχα. Στον Πίνακα 2 παρουσιάζεται η επίδοση των φοιτητών για κάθε βιντεοδιάλεξη καθώς και η κατανομή των συμμετεχόντων ανά κλάση. Όπως προκύπτει από τον πίνακα, οι κλάσεις είναι απολύτως ισορροπημένες μόνο σε μια βιντεοδιάλεξη (2η), ενώ στις υπόλοιπες 5 περιπτώσεις υπάρχει ανισορροπία ως προς τον αριθμό των συμμετεχόντων. Ενδιαφέρον δε έχει το γεγονός ότι σε κάποιες περιπτώσεις υπάρχει μεγαλύτερος αριθμός συμμετεχόντων στην κλάση της χαμηλής επίδοσης ενώ σε άλλες στην κλάση της υψηλής επίδοσης.

Πίνακας 2. Επίδοση και κατανομή συμμετεχόντων ανά βιντεοδιάλεξη

Βιντεοδιάλεξη	Μέσος Όρος	Τυπική απόκλιση	Διάμεσος	Αριθμός συμμετεχόντων	
1	0.79	0.13	0.70	95[a]	159[b]
2	0.75	0.15	0.75	127[a]	127[b]
3	0.71	0.12	0.70	156[a]	98[b]
4	0.68	0.19	0.69	102[a]	152[b]
5	0.80	0.17	0.80	137[a]	117[b]
6	0.71	0.14	0.70	108[a]	146[b]

[a]: Χαμηλή επίδοση, b: Υψηλή επίδοση

Δεδομένης της ανισορροπίας των κλάσεων για τις περισσότερες βιντεοδιαλέξεις, επιλέξαμε δύο μέτρα για την αξιολόγηση της ταξινόμησης: ακρίβεια (accuracy) και F1 τα οποία παρουσιάζονται στην επόμενη ενότητα. Οι αλγόριθμοι ταξινόμησης ΜΜ που χρησιμοποιήθηκαν για την κατηγοριοποίηση είναι οι εξής: Logistic Regression (LR), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Classifier (SVC), Naive Bayes (NB) και Linear Support Vector Classifier (LSVC) Κατά την εκτέλεση των αλγορίθμων, χρησιμοποιήθηκε το 80% των δεδομένων για εκπαίδευση και το 20% των δεδομένων για έλεγχο.

Αποτελέσματα

Η ακρίβεια ταξινόμησης (accuracy) αποτελεί ένα τυπικό μέτρο αξιολόγησης και υπολογίζει τις σωστές ταξινομήσεις που πραγματοποιούνται στο σύνολο των παρατηρήσεων. Ο υπολογισμός της γίνεται με βάση τον λόγο TP+TN / P + N, ενώ το εύρος των τιμών ορίζεται σε ένα διάστημα από 0 έως 1.

Όπως προαναφέρθηκε, ένα από τα μέτρα αξιολόγησης που χρησιμοποιούνται στην περίπτωση ανισορροπίας κλάσεων είναι το F1. Το συγκριτικό πλεονέκτημα που παρουσιάζει έναντι του μέτρου της ακρίβειας είναι ότι λαμβάνει υπόψη ταυτόχρονα δύο άλλα μέτρα, αυτά της ορθότητας (precision) και ανάκλησης (recall) αντίστοιχα. Σημειωτέον, πως το μέτρο F1 είναι πολύ πιο ευαίσθητο - και επομένως συντηρητικό - σε σχέση με το μέτρο της ακρίβειας καθότι εάν είτε η ορθότητα είτε η ανάκληση είναι μηδέν, τότε ο παραπάνω αριθμητής θα είναι μηδέν με συνέπεια η συνολική τιμή του μέτρου F1 να είναι μηδέν.

Οι μέσες τιμές ακρίβειας ταξινόμησης για κάθε μοντέλο ενσωμάτωσης πρότασης ανά βιντεοδιάλεξη παρουσιάζονται στον Πίνακα 3. Οι υψηλότερες μέσες τιμές ακρίβειας ταξινόμησης για κάθε μοντέλο ενσωμάτωσης καταγράφηκαν για τα μοντέλα ενσωμάτωσης SBERT (4η και 5η) και spaCy (1η και 3η). Τα μοντέλα ενσωμάτωσης FastText, E και E5 Instruct

σημείωσαν τις υψηλότερες μέσες τιμές ταξινόμησης σε μία βιντεοδιάλεξη αντίστοιχα (2η, 5η και 6η).

Πίνακας 3. Μέση τιμή ακρίβειας ταξινόμησης για κάθε μοντέλο ενσωμάτωσης ανά βιντεοδιάλεξη

Βιντεοδιάλεξη	Spacy 300[d]	FastText 300[d]	Nomic 768[d]	SBERT 768[d]	BGE 1024[d]	E5 Instruct 1024[d]	E5 1024[d]
1	**0,64**	0,61	0,56	0,63	0,6	0,58	0,62
2	0,52	0,53	0,48	0,52	0,53	0,55	**0,56**
3	**0,59**	0,57	0,55	0,56	0,58	0,58	0,57
4	0,62	0,63	0,64	**0,65**	0,62	0,63	0,62
5	0,56	0,56	0,56	**0,61**	0,6	**0,61**	0,58
6	0,6	**0,63**	0,59	0,62	0,6	0,61	0,6

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Ο Πίνακας 4 παρουσιάζει τη μέση τιμή ταξινόμησης της μετρικής F1 ανά βιντεοδιάλεξη. Όπως προκύπτει, το μοντέλο της spaCy οδήγησε στην υψηλότερη μέση τιμή F1 που καταγράφηκε σε μία από τις βινεοδιαλέξεις (1η). Ωστόσο, όλα τα μοντέλα ενσωμάτωσης χαρακτηρίστηκαν από σχετικά υψηλές τιμές για την 1η βιντεοδιάλεξη. Επίσης, το μοντέλο ενσωμάτωσης FastText οδήγησε στις υψηλότερες μέσες τιμές ταξινόμησης για δύο από τις βιντεοδιαλέξεις (4η και 6η).

Πίνακας 4. Μέση τιμή ταξινόμησης μετρικής F1 για κάθε μοντέλο ενσωμάτωσης ανά βιντεοδιάλεξη

Βιντεοδιάλεξη	Spacy 300[d]	FastText 300[d]	Nomic 768[d]	SBERT 768[d]	BGE 1024[d]	E5 Instruct 1024[d]	E5 1024[d]
1	**0,75**	0,72	0,69	0,73	0,71	0,70	0,72
2	0,57	0,59	0,53	0,52	0,58	0,57	**0,61**
3	0,32	0,27	0,24	0,24	0,26	0,28	0,27
4	0,71	**0,73**	0,72	0,72	0,70	0,70	0,70
5	0,46	0,48	0,42	**0,49**	0,47	0,48	0,44
6	0,70	**0,73**	0,68	0,69	0,69	0,69	0,69

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Στον Πίνακα 5 παρουσιάζονται οι τιμές της ακρίβειας ταξινόμησης ανά αλγόριθμο για το σύνολο των βιντεοδιαλέξεων. Από τον πίνακα προκύπτει ότι οι υψηλότερες μέσες τιμές ταξινόμησης επιτεύχθηκαν από τους συνδυασμούς (α) του αλγορίθμου RF με μοντέλο ενσωμάτωσης πρότασης E5 Instruct και (β) του αλγορίθμου SVC με το μοντέλο ενσωμάτωσης FastText. Αξίζει επίσης να σημειωθεί ότι ο συνδυασμός του αλγόριθμου RF με σχεδόν όλα τα μοντέλα ενσωμάτωσης οδήγησε σε σταθερά υψηλές τιμές ακρίβειας ταξινόμησης.

Πίνακας 5. Μέση τιμή ακρίβειας ταξινόμησης για κάθε αλγόριθμο και μοντέλο ενσωμάτωσης στο σύνολο των βιντεοδιαλέξεων

Αλγόριθμος	Spacy 300[d]	FastText 300[d]	Nomic 768[d]	SBERT 768[d]	BGE 1024[d]	E5 Instruct 1024[d]	E5 1024[d]
LR	0,58	0,58	0,50	0,60	0,60	0,58	**0,62**
RF	0,62	0,62	0,60	0,64	0,61	**0,65**	0,64
DT	0,54	0,51	0,56	0,53	**0,57**	0,55	0,56
SVC	0,62	**0,65**	0,60	0,62	0,61	0,64	0,61
KNN	0,57	**0,58**	0,54	**0,58**	**0,58**	0,57	0,54
NB	**0,62**	0,61	0,57	**0,62**	0,57	0,59	0,60

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Στον Πίνακα 6 παρατίθεται η μέση τιμή μετρικής F1 στο σύνολο των βιντεοδιαλέξεων ανά μοντέλο ενσωμάτωσης. Όπως φαίνεται στον πίνακα, ο αλγόριθμος NB σημείωσε την υψηλότερη μέση τιμή F1 (0.67) με το μοντέλο της spaCy, ενώ γενικά είχε υψηλές μέσες τιμές F1 με όλα τα μοντέλα. Επίσης, ο συνδυασμός του αλγορίθμου RF με δύο μοντέλα ενσωμάτωσης (spaCy και FastText) οδήγησε στην υψηλότερη μέση τιμή F1.

Πίνακας 6. Μέση τιμή μετρικής F1 για κάθε μοντέλο ενσωμάτωσης στο σύνολο των βιντεοδιαλέξεων

Αλγόριθμος	Spacy 300[d]	FastText 300[d]	Nomic 768[d]	SBERT 768[d]	BGE 1024[d]	E5 Instruct 1024[d]	E5 1024[d]
LR	0,59	0,60	0,49	0,58	0,59	0,57	**0,61**
RF	**0,60**	**0,60**	0,54	0,58	0,56	0,58	0,58
DT	0,54	0,54	**0,57**	0,50	0,55	0,54	0,55
SVC	0,59	**0,61**	0,55	0,57	0,57	0,60	0,57
KNN	0,61	**0,63**	0,59	0,60	0,61	0,58	0,56
NB	**0,67**	0,66	0,61	0,66	0,62	0,62	0,64

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Δεδομένου ότι η μέση τιμή επηρεάζεται δυσανάλογα από ακραίες τιμές, εξετάζουμε παρακάτω τη μέγιστη τιμή που σημειώθηκε είτε στο μέτρο της ακρίβειας είτε σε αυτό της F1. Παρόλο που οι μέγιστες τιμές δεν είναι εξίσου αντιπροσωπευτικές με τη μέση τιμή, προχωρούμε στην παράθεση αυτή επειδή είναι δηλωτικές του μέγιστου δυνατού δυναμικού που παρέχουν οι ενσωματώσεις κειμένου για την πρόβλεψη της επίδοσης . Ο Πίνακας 7 παρουσιάζει την υψηλότερη τιμή ακρίβειας ταξινόμησης που καταγράφηκε ανά μοντέλο ενσωμάτωσης για κάθε βιντεοδιάλεξη. Στην 1η βιντεοδιάλεξη η υψηλότερη τιμή ακρίβειας ταξινόμησης (0.67) σημειώθηκε με διαφορετικούς αλγορίθμους, αλλά κυρίως με τους RF και SVC. Η υψηλότερη τιμή ακρίβειας που παρατηρήθηκε στη 2η βιντεοδιάλεξη (0.63) σημειώθηκε από δύο αλγόριθμους ταξινόμησης, ο ένας εκ των οποίων πάλι ήταν ο RF. Στην 3η βιντεοδιάλεξη η υψηλότερη τιμή ακρίβειας που καταγράφηκε (0.65) ήταν από τον αλγόριθμο LR. Η υψηλότερη τιμή ακρίβειας ταξινόμησης (0.71) για την 4η βιντεοδιάλεξη σημειώθηκες με τρεις διαφορετικούς ταξινομητές (SVC, NB και RF). Στη βιντεοδιάλεξη 5 η υψηλότερη τιμή ακρίβειας που παρατηρήθηκε (0.71) ήταν σε συνδυασμό με τον ταξινομητή NB. Τέλος, η υψηλότερη τιμή ακρίβειας που καταγράφηκε στη βιντεοδιάλεξη 6 (0.69) περιλάμβανε διάφορους ταξινομητές, με συνηθέστερο τον RF.

Πίνακας 7. Υψηλότερη τιμή ακρίβειας ταξινόμησης για κάθε μοντέλο ενσωμάτωσης ανά βιντεοδιάλεξη

Βιντεοδιάλεξη	Spacy 300d	FastText 300d	Nomic 768d	SBERT 768d	BGE 1024d	E5 Instruct 1024d	E5 1024d
1	KNN (0,67)	SVC (0,67)	RF, NB, DC (0,63)	RF (0,67)	RF (0,67)	RF, SVC (0,67)	SVC (0,67)
2	NB (0,59)	SVC (0,61)	SVC (0,53)	RF, LR (0,57)	DT, SVC (0,59)	SVC (0,61)	RF, DT (0,63)
3	KNN (0,63)	LR (0,65)	RF, SVC, DC (0,61)	RF, SVC (0,61)	KNN (0,65)	LR (0,65)	RF (0,61)
4	RF, NB (0,65)	SVC (0,71)	SVC (0,69)	NB (0,71)	KNN (0,67)	RF (0,71)	RF (0,67)
5	NB (0,67)	SVC (0,65)	NB (0,63)	NB (0,71)	LR (0,67)	RF, NB (0,67)	LR (0,67)
6	RF, SVC (0,69)	RF, SVC, KNN (0,67)	RF, SVC, NB (0,63)	RF, KNN (0,69)	DT (0,65)	RF (0,67)	LR (0,69)

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Τέλος, ο Πίνακας 8 δίνει τις υψηλότερες τιμές F1 που καταγράφηκαν με κάθε αλγόριθμο ταξινόμησης για κάθε μοντέλο ενσωμάτωσης ανά βιντεοδιάλεξη. Η υψηλότερη τιμή F1 για την 1η βιντεοδιάλεξη (0.78) σημειώθηκε από τους ταξινομητές SVC και RF. Ιδιαίτερο ενδιαφέρον έχει το γεγονός ότι οι ταξινομητές αυτοί είχαν πολύ υψηλή τιμή F1 με όλα ανεξαιρέτως τα μοντέλα ενσωματώσεων, είτε λέξης είτε πρότασης. Ο συνδυασμός του μοντέλου FastText με τον αλγόριθμο SVC οδήγησε στην υψηλότερη τιμή F1 για την 4η βιντεοδιάλεξη. Οι βιντεοδιαλέξεις 3 και 5 χαρακτηρίστηκαν από τις υψηλότερες τιμές F1 από τους αλγορίθμους KNN και NB αντίστοιχα σε συνδυασμό με διάφορα μοντέλα ενσωμάτωσης. Τέλος, η βιντεοδιάλεξη 6 χαρακτηρίστηκε από την υψηλότερη τιμή F1 (0.78) από διάφορους ταξινομητές, με τον RF να έχει τη μεγαλύτερη συχνότητα σε συνδυασμό με spaCy και SBERT.

Πίνακας 8. Υψηλότερη τιμή μετρικής F1 για κάθε μοντέλο ενσωμάτωσης ανά βιντεοδιάλεξη

Βιντεοδιάλεξη	Spacy	FastText	Nomic	SBERT	BGE	E5 Instruct	E5
1	SVC (0,78)	SVC (0,78)	RF, DC (0,77)	RF (0,78)	RF (0,78)	RF, SVC (0,78)	RF, SVC (0,78)
2	DC (0,66)	DC (0,66)	DC (0,66)	DC (0,66)	DC (0,66)	DC (0,66)	DC (0,66)
3	KNN (0,49)	LR (0,55)	KNN (0,48)	KNN (0,41)	KNN, NB (0,44)	DT (0,47)	NB (0,45)
4	DC (0,76)	SVC (0,78)	DC (0,76)	DC (0,76)	DC (0,76)	RF (0,77)	DC (0,76)
5	NB (0,73)	NB (0,68)	NB (0,68)	NB (0,72)	NB (0,62)	NB (0,7)	NB (0,68)
6	RF, SVC (0,78)	RF (0,77)	RF, SVC (0,73)	RF (0,78)	DC (0,72)	RF (0,75)	RF (0,75)

[d]: Διαστάσεις διανύσματος ενσωμάτωσης πρότασης. Η υψηλότερη τιμή για κάθε μοντέλο ενσωμάτωσης δίνεται με έντονη μορφοποίηση

Συμπεράσματα

Παρά τις καταιγιστικές εξελίξεις στο πεδίο της ΕΦΓ, αυτές έχουν παραμείνει αναξιοποίητες στο πλαίσιο της ΑΜ. Η αξιοποίηση των κειμένων διαφόρων τύπων που δημιουργούν φοιτητές για τις ανάγκες μαθημάτων είναι πολύ περιορισμένη – ειδικά αναφορικά με την πρόβλεψη της μάθησης. Σε αυτό το πλαίσιο, η παρούσα εργασία υιοθέτησε νέες μεθόδους αναπαράστασης κειμένου εξετάζοντας συγκριτικά τη χρήση τους για την πρόβλεψη της επίδοσης.

Το πρώτο ερευνητικό ερώτημα εστίασε στην αποτελεσματικότητα των διαθέσιμων πολυγλωσσικών μοντέλων για την πρόβλεψη της επίδοσης. Χρησιμοποιώντας ως κριτήριο τη μέση ακρίβεια ταξινόμησης, η συνολική εικόνα που προκύπτει δεν είναι ικανοποιητική, καθώς λίγα μοντέλα ενσωμάτωσης είχαν μέση τιμή ταξινόμησης μεγαλύτερη από 0.60. Ωστόσο, η εικόνα που προκύπτει από τη μέση τιμή F1 ήταν πολύ πιο υποσχόμενη, δεδομένου ότι κάποια μοντέλα ενσωμάτωσης είχαν τιμή μεγαλύτερη του 0.70 σε κάποιες βιντεοδιαλέξεις. Όπως προαναφέρθηκε, ο δείκτης F1 είναι καταλληλότερος λόγω της ανισορροπίας των κλάσεων για κάθε βιντεοδιάλεξη. Θα πρέπει να σημειωθεί ότι η γενική εικόνα που προκύπτει είναι κοντινή στους αντίστοιχους δείκτες που αναφέρουν οι Pijeira-Díaz et al. (2024a) για Ολλανδικά μοντέλα ενσωμάτωσης κειμένου. Ωστόσο, τα έργα που αφορούν την παραπάνω μελέτη είναι εντελώς διαφορετικά σε σχέση με τα αντίστοιχα που χρησιμοποιήθηκαν στην παρούσα εργασία.

Το δεύτερο ερευνητικό ερώτημα εξέτασε την αποδοτικότητα της ταξινόμησης της επίδοσης ως συνάρτηση της διάστασης των ενσωματώσεων κειμένου. Στη βάση της βιβλιογραφίας, η παραδοχή που κάναμε ήταν ότι τα ανύσματα μεγαλύτερων διαστάσεων θα οδηγούσαν σε μεγαλύτερη ακρίβεια ταξινόμησης, δεδομένου ότι οι μεγάλες διαστάσεις μπορούν δυνητικά να αποτυπώσουν περισσότερες λεπτομέρειες (Mikolov et al., 2013a; Conneau et al., 2017). Τα αποτελέσματα από την ανάλυση δεν επιβεβαιώνουν αυτή την υπόθεση εργασίας. Αντίθετα, σε αρκετές περιπτώσεις μοντέλα όπως η spaCy και το FastText που έχουν ανύσματα με τη μικρότερη διάσταση οδήγησαν σε υψηλότερες μέσες τιμές ακρίβειας και F1. Παρόλο που σε κάποιες περιπτώσεις βιντεοδιαλέξεων και με κάποιους συνδυασμούς αλγορίθμων τα μοντέλα

με μεσαίες (768) και μεγάλες (1024) διαστάσεις ενσωματώσεων σημείωσαν υψηλές τιμές ακρίβειας ταξινόμησης και F1, δεν προκύπτει ένα ξεκάθαρο μοτίβο.

Τέλος, το τρίτο ερευνητικό ερώτημα εστίασε στην αποδοτικότητα των αλγορίθμων ΜΜ σε συνδυασμό με τα διαθέσιμα μοντέλα ενσωμάτωσης κειμένου. Η μέση ακρίβεια ταξινόμησης για τους διάφορους αλγορίθμους, ήταν σχετικά χαμηλή, στοιχείο που δείχνει ότι χρησιμοποιώντας την ακρίβεια ως κριτήριο, η εικόνα που προκύπτει δεν είναι ικανοποιητική. Σε πολλές περιπτώσεις καταγράφηκε μέση τιμή ακρίβειας πάνω από 0.60, μέση τιμή που δεν είναι ιδιαίτερα ενθαρρυντική. Αντίστοιχη είναι η εικόνα όταν εξετάζονται οι μέσες τιμές του δείκτη F1. Ωστόσο, οι υψηλότερες τιμές ακρίβειας και F1 που καταγράφηκαν δείχνουν ότι οι συνδυασμοί κάποιων αλγορίθμων με κάποια μοντέλα ενσωμάτωσης μπορούν δυνητικά να δώσουν τιμές αντίστοιχες με αυτές των Pijeira-Díaz et al. (2024a). Στη συγκεκριμένη μελέτη η απόδοση των ταξινομητών που χρησιμοποιήθηκαν κυμάνθηκε σε μέσα και υψηλά επίπεδα, συγκρίσιμα σε πολλές περιπτώσεις με τους δείκτες της παρούσας εργασίας. Ωστόσο, θα πρέπει να επισημάνουμε μερικές σημαντικές διαφοροποιήσεις σε σχέση με την παρούσα εργασία. Πρώτον, δεν είναι άμεσα εφικτή η σύγκριση Ολλανδικών με Ελληνικά μοντέλα ενσωμάτωσης, παρόλο που και στις δύο περιπτώσεις πρόκειται για μη διαδεδομένες γλώσσες. Δεύτερον, η έκταση των σωμάτων κειμένων που χρησιμοποίησαν οι Pijeira-Díaz et al. (2024a) ήταν πιο εκτεταμένη συγκριτικά με τις 254 περιλήψεις που αναλύθηκαν στην παρούσα εργασία. Τρίτον, τα κείμενα που χρησιμοποιήθηκαν στη μελέτη των Pijeira-Díaz et al. (2024a) είχαν κατηγοριοποιηθεί και αξιολογηθεί από ειδικούς, διαδικασία που είναι χρονοβόρα, επίπονη και συνεπάγεται μεγάλο κόστος. Τέταρτον, το έργο που οι υιοθέτησαν στη μελέτη τους Pijeira-Díaz et al. (2024a) ήταν πολύ καλά προσδιορισμένο, σε αντιδιαστολή με την παρούσα εργασία, όπου οι συμμετέχοντες προχώρησαν στη συγγραφή περιλήψεων.

Συμπερασματικά, η συγκριτική αξιολόγηση ανοικτών πολυ-γλωσσικών μοντέλων ενσωμάτωσης κειμένου που υποστηρίζουν την Ελληνική γλώσσα ήταν υποσχόμενη και σε κάποιο βαθμό ευθυγραμμίζεται με αποτελέσματα πρόσφατων μελετών σε άλλες γλώσσες. Για παράδειγμα, κάποιες μελέτες αναφέρουν επίπεδα ταξινόμησης που κυμαίνονται μεταξύ 60% και 67% (El Aouifi et al., 2021), ενώ άλλες μελέτες αναφέρουν σχεδόν τέλεια επίπεδα ταξινόμησης(Ferreira-Mello et al., 2019). Παρόλα αυτά, η εικόνα που προκύπτει από τα αποτελέσματα της παρούσας μελέτης συνιστά ότι η απόδοση των μοντέλων ενσωμάτωσης όπως παρέχονται πιθανότατα δεν επαρκεί για τις ανάγκες της ΑΜ. Επομένως, απαιτείται συστηματική προσαρμογή (fine-tuning) των μοντέλων ενσωμάτωσης κειμένων στην Ελληνική γλώσσα, ώστε να καταστεί δυνατή η αξιοποίηση των εξελίξεων στο πεδίο της ΕΦΓ για την υποστήριξη της μάθησης στο πλαίσιο της ΑΜ.

Βιβλιογραφικές Αναφορές

Aggarwal, C. & Zhai, C.C. (2012). (Eds). Mining text data. Springer.

Agrawal, A., Venkatraman, J., Leonard, S., & Paepcke, A. (2015). YouEDU: Addressing Confusion in MOOC Discussion Forums by Recommending Instructional Video Clips. International Educational Data Mining Society.

Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education*, 118, 1-9.

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 100489.

Baroni, M., Dinu, G., & Kruszewski, G. (2014, June). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers*) (pp. 238-247).

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5, 135-146.

Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., ... & Kurzweil, R. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). BGE m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

El Aouifi, H., El Hajji, M., Es-Saady, Y., & Douzi, H. (2021). Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. *Education and Information Technologies*, 26(5), 5799–5814. https://doi.org/10.1007/s10639-021-10512-4.

Explosion (2024). Available: https://spacy.ioFeng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2022). Language-agnostic BERT sentence embedding. *arXiv preprint arXiv:2007.01852*.

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. WIREs Data Mining and Knowledge Discovery, 9(6). https://doi.org/10.1002/widm.1332.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*.

Karasavvidis, I., Papadimas, C., & Ragazou, V. (2022). Student-generated texts as features for predicting learning from video lectures: An initial evaluation. *Themes in eLearning*, 15, 21-45.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188-1196). PMLR.

Lee, C., Roy, R., Xu, M., Raiman, J., Shoeybi, M., Catanzaro, B., & Ping, W. (2024). NV-Embed: Improved Techniques for Training LLMs as Generalist Embedding Models. *arXiv preprint arXiv:2405.17428*.

Li, Z., Zhang, X., Zhang, Y., Long, D., Xie, P., & Zhang, M. (2023). Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*.

Liu, N. F., Gardner, M., Belinkov, Y., Peters, M. E., & Smith, N. A. (2019). Linguistic knowledge and transferability of contextual representations. *arXiv preprint arXiv:1903.08855*.

Logeswaran, L., & Lee, H. (2018). An efficient framework for learning sentence representations. *arXiv preprint arXiv:1803.02893*.

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in learning and education. *EDUCAUSE Review*, 31–40.

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 12(4), 516-534.

Manning, C. D., Raghavan, P. & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Muennighoff, N. (2022). SGPT: GPT sentence embeddings for semantic search. *arXiv preprint arXiv:2202.08904*.

Ni, J., Abrego, G. H., Constant, N., Ma, J., Hall, K. B., Cer, D., & Yang, Y. (2021). Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv preprint arXiv:2108.08877*.

Nussbaum, Z., Morris, J. X., Duderstadt, B., & Mulyar, A. (2024). Nomic embed: Training a reproducible long context text embedder. *arXiv preprint arXiv:2402.01613*.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. *ArXiv, abs/1802.05365*.

Pijeira-Díaz, H. J., Braumann, S., van de Pol, J., van Gog, T., & de Bruin, A. B. (2024a). Towards adaptive support for self-regulated learning of causal relations: Evaluating four Dutch word vector models. *British Journal of Educational Technology*.

Pijeira-Díaz, H. J., Subramanya, S., van de Pol, J., & de Bruin, A. (2024b). Evaluating Sentence-BERT-powered learning analytics for automated assessment of students' causal diagrams. *Journal of Computer Assisted Learning*.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140), 1-67.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv preprint arXiv:1908.10084*.

Reimers, N., & Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:2004.09813*.

Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I. (2019). Classification and clustering of arguments with contextualized word embeddings. *arXiv preprint arXiv:1906.09821*.

Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016). Forecasting student achievement in MOOCs with natural language processing. Proceedings *of the Sixth International Conference on Learning Analytics & Knowledge* – LAK '16, 383–387. New York, New York, USA: ACM Press. https://doi.org/10.1145/2883851.2883932.

Schütze, H. (1992). Word space. *Advances in Neural Information Processing Systems*, 5.

Turian, J., Ratinov, L., & Bengio, Y. (2010). Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 384-394).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2023). Improving text embeddings with large language models. *arXiv preprint arXiv:2401.00368*.

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). Multilingual e5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*.

Xiao, S., Liu, Z., Zhang, P., & Muennighof, N. (2023). C-pack: Packaged resources to advance general chinese embedding. *arXiv preprint arXiv:2309.07597*.

Παναγιωτακόπουλος, Χ., Τσαλίδης, Χ., Γάκης, Π., & Κόκκινος, Θ. (2023). Υπολογιστική γλωσσολογία: Από τον προγραμματισμό μέχρι τη διδακτική πράξη [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. http://hdl.handle.net/11419/8638.

Παπαδήμας, Χ., Ραγάζου, Β., & Καρασαββίδης, Η. (2023). Η Επεξεργασία Φυσικής Γλώσσας για την πρόβλεψη της επίδοσης: Μια διερευνητική αξιολόγηση δύο συνόλων χαρακτηριστικών. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 314-321.

Ραγάζου, Β., Παπαδήμας, Χ., & Καρασαββίδης, Η. (2022). Η αξιοποίηση κειμένου για την πρόβλεψη της επίδοσης με τη χρήση τεχνικών μηχανικής μάθησης: μια μελέτη περίπτωσης. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, 809-820.