

Themes in Science and Technology Education

Vol 16, No 1-6 (2023)

Ηλεκτρονική Μάθηση και Ανοικτοί Εκπαιδευτικοί Πόροι: Θέματα και ερευνητικές τάσεις στην Ελλάδα



Μηχανική Μάθηση: Ο ρόλος του κειμένου στην πρόβλεψη της επίδοσης διαμέσου βιντεοδιαλέξεων

Βασιλική Ραγάζου, Χαράλαμπος Παπαδήμας, Ηλίας Καρασαββίδης

doi: [10.12681/thete.40002](https://doi.org/10.12681/thete.40002)

To cite this article:

Ραγάζου Β., Παπαδήμας Χ., & Καρασαββίδης Η. (2023). Μηχανική Μάθηση: Ο ρόλος του κειμένου στην πρόβλεψη της επίδοσης διαμέσου βιντεοδιαλέξεων. *Themes in Science and Technology Education*, 16(1-6), 19–38. <https://doi.org/10.12681/thete.40002>

Μηχανική Μάθηση: Ο ρόλος του κειμένου στην πρόβλεψη της επίδοσης διαμέσου βιντεοδιαλέξεων

Βασιλική Ραγάζου, Χαράλαμπος Παπαδήμας, Ηλίας Καρασαββίδης
ragazou@uth.gr, papadimas@uth.gr, ikaras@uth.gr

Παιδαγωγικό Τμήμα Προσχολικής Εκπαίδευσης, Σχολή Επιστημών του Ανθρώπου, Πανεπιστήμιο Θεσσαλίας

Περίληψη. Η παρούσα εργασία εξετάζει τη χρήση της Επεξεργασίας Φυσικής Γλώσσας με στόχο την πρόβλεψη της επίδοσης των φοιτητών σε περιβάλλοντα ηλεκτρονικής μάθησης. Παρόλο που η πρόοδος στον τομέα αυτόν είναι ενθαρρυντική, λείπουν μέχρι σήμερα συστηματικές ερευνητικές προσπάθειες να διερευνηθεί πλήρως το δυναμικό της Επεξεργασίας Φυσικής Γλώσσας για την ανάλυση και πρόβλεψη της επίδοσης. Στην έρευνα αυτή συμμετείχαν 85 φοιτητές, οι οποίοι παρακολούθησαν 6 βιντεοδιαλέξεις, για κάθε μία εκ των οποίων κλήθηκαν να συντάξουν μια μικρή περίληψη. Από την επεξεργασία των περιλήψεων προέκυψαν δύο σύνολα μεταβλητών (πρωτογενούς και επεξεργασμένου κειμένου) που τροφοδότησαν οκτώ αλγόριθμους μηχανικής μάθησης. Στις περισσότερες βιντεοδιαλέξεις παρατηρήθηκε μια μικρή διαφοροποίηση τιμών των μέτρων ακρίβειας ταξινόμησης και F1 ενώ η λογιστική παλινδρόμηση (LR) ήταν ο αλγόριθμος που επέφερε τα υψηλότερα επίπεδα ακρίβειας ταξινόμησης τόσο για το πρωτογενές όσο και για το επεξεργασμένο κείμενο. Τα αποτελέσματα της ανάλυσης αναδεικνύουν ότι οι περιλήψεις που δημιουργήθηκαν από τους φοιτητές μπορούν να αποτελέσουν ένα υποσχόμενο σύνολο χαρακτηριστικών για την πρόβλεψη της επίδοσης μέσω βιντεοδιαλέξεων.

Λέξεις κλειδιά: Επεξεργασία φυσικής γλώσσας, αναλυτική της μάθησης, μηχανική μάθηση, κειμενικά χαρακτηριστικά, ταξινόμηση, επίδοση, βιντεοδιαλέξεις

Εισαγωγή

Η Αναλυτική της Μάθησης (Learning Analytics) (AM) είναι ένα νέο διεπιστημονικό ερευνητικό πεδίο το οποίο συνδυάζει διάφορες ερευνητικές κοινότητες και παραδείγματα (Baker et al., 2021). Η AM περιλαμβάνει τη μέτρηση, συλλογή, ανάλυση και παρουσίαση δεδομένων για μαθητές και αποσκοπεί στην κατανόηση και τη βελτιστοποίηση της μάθησης (Siemens & Long, 2011). Ένας κύκλος AM περιλαμβάνει τέσσερα κύρια συστατικά στοιχεία: (1) μαθητές, οι οποίοι μελετούν σε διαδικτυακά μαθησιακά περιβάλλοντα, (2) δεδομένα, τα οποία συλλέγονται από συστήματα διαχείρισης μάθησης, (3) δείκτες, οι οποίοι προκύπτουν από την ανάλυση των δεδομένων του προηγούμενου βήματος και (4) παρεμβάσεις, που αποτελούν την ανατροφοδότηση που παρέχεται στους εμπλεκόμενους: μαθητές και εκπαιδευτές (Clow, 2012). Η μαθησιακή υποστήριξη των μαθητών είναι αδύνατη χωρίς τη συλλογή δεδομένων. Τα σύγχρονα Συστήματα Διαχείρισης Μάθησης (ΣΔΜ) παρέχουν προηγμένες δυνατότητες καταγραφής δεδομένων στα πλαίσια της αλληλεπίδρασης των μαθητών με τους πόρους που παρέχονται σε αυτά. Σε γενικές γραμμές μπορούμε να διακρίνουμε τρία επίπεδα συλλογής δεδομένων: (α) μικρο-επίπεδο (π.χ. αριθμός κλικ σε ένα δεδομένο πόρο), (β) μέσο-επίπεδο (π.χ. σώματα κειμένων) και (γ) μακρο-επίπεδο (π.χ. συλλογή δεδομένων σε ιδρυματικό επίπεδο) (Fischer et al., 2020). Τα ΣΔΜ αποτελούν την πιο σημαντική πηγή δεδομένων για την AM. Οι πιο συνηθισμένοι τύποι δεδομένων που συλλέγονται από τα ΣΔΜ περιλαμβάνουν τη συχνότητα χρήσης ενός πόρου, τη συμμετοχή σε φόρουμ συζήτησης, τον αριθμό των κοιζ που έχουν ολοκληρωθεί, τη σύνδεση και

αποσύνδεση στο σύστημα, την υποβολή εργασιών, τον αριθμό κλικ, τη συνολική διάρκεια προβολής περιεχομένου κ.λπ. (Fahd et al., 2022). Μετά τη συλλογή των δεδομένων ακολουθεί η ανάλυση τους και η λήψη αποφάσεων σχετικά με την υποστήριξη της μάθησης. Υπάρχουν πολλά παραδείγματα στη δημοσιευμένη βιβλιογραφία, όπου η ΑΜ έχει γνωρίσει επιτυχημένες εφαρμογές, όπως για παράδειγμα τον εντοπισμό φοιτητών που είναι στο όριο της αποτυχίας (Akçarpınar et al., 2019· Arnold & Pistilli, 2012· Chui et al., 2020), την πρόβλεψη φοιτητών που θα εγκαταλείψουν το μάθημα (Olive et al., 2019· Xing & Du, 2019), την πρόβλεψη τελικού βαθμού σε ένα μάθημα (Asif et al., 2017), την αξιολόγηση (Caspari-Sadeghi, 2023) και την παροχή ανατροφοδότησης (Pardo et al., 2018, 2019). Παρόλο που η δυναμική της ΑΜ είναι ευρέως αποδεκτή, η εμπειρική έρευνα δείχνει ότι τα εκπαιδευτικά συστήματα τείνουν να μην την αξιοποιούν συστηματικά για τη βελτίωση της διδακτικής και μαθησιακής διαδικασίας (Teasley, 2019). Πιο συγκεκριμένα, η θετική επίδραση της ΑΜ στη μάθηση φαίνεται να έχει μικρή εμπειρική τεκμηρίωση (Ferguson & Clow, 2017· Gašević et al., 2016· Guzmán-Valenzuela et al., 2021· Larrabee Sønderlund et al., 2019· Viberg et al., 2018). Για παράδειγμα, σε μια διεξοδική ερευνητική επισκόπηση πεδίου, οι Viberg et al. (2018) κατέληξαν στο συμπέρασμα ότι ένα πολύ μικρό ποσοστό των ερευνών (<10%) υποστηρίζει τη θετική επίδραση της ΑΜ στη βελτίωση της μάθησης. Μια άλλη ενδελεχής ερευνητική επισκόπηση διαπίστωσε ότι στην τρέχουσα μορφή της η ΑΜ αφορά περισσότερο την ανάλυση και λιγότερο τη μάθηση αυτή κάθε αυτή (Guzmán-Valenzuela et al., 2021).

Η ανατροφοδότηση στην ΑΜ

Η ανατροφοδότηση αποτελεί ουσιαστικά το πιο θεμελιώδες στοιχείο της διαδικασίας υποστήριξης της μάθησης. Στο πεδίο των Επιστημών της Μάθησης υπάρχει εκτεταμένη έρευνα, η οποία αποτυπώνει αναλυτικά τη σημασία της ανατροφοδότησης (Hattie & Timperley, 2007· Latifi et al., 2021· Shute, 2008· Wisniewski et al., 2020). Στο πεδίο της ΑΜ, η ανατροφοδότηση έχει πολλές διαφορετικές εκφάνσεις: αναστοχασμός, προσωποποίηση, πρόβλεψη, αξιολόγηση, προσαρμογή, πρόταση και έλεγχος (Banhashem et al., 2022· Bodily & Verbert, 2017). Γενικά, οι διάφορες ερευνητικές επισκοπήσεις δείχνουν ότι η υλοποίηση της ανατροφοδότησης στο πεδίο της ΑΜ κάθε άλλο παρά βέλτιστες είναι (Bodily & Verbert, 2017· Ramaswami et al., 2023). Για παράδειγμα οι Bodily και Verbert (2017) κατέληξαν στο συμπέρασμα πως μόνο το 18% των μελετών που αναλύθηκαν στην ερευνητική τους επισκόπηση παρείχαν ανατροφοδότηση, η οποία ήταν καθαρά περιγραφική: απλώς πληροφορούσε τους φοιτητές για το τρέχον επίπεδο της επίδοσής τους. Δεν υπάρχει αμφιβολία ότι η ανατροφοδότηση αυτού του τύπου μπορεί να αποβεί χρήσιμη για τους φοιτητές. Ας πάρουμε ενδεικτικά την περίπτωση ενός φοιτητή η επίδοσή του οποίου είναι κάτω από τον μέσο όρο της τάξης. Όταν ενημερωθεί για την επίδοσή αυτή, μπορεί να κινητοποιηθεί περαιτέρω για να μελετήσει πιο εστιασμένα το παρεχόμενο μαθησιακό υλικό και να υλοποιήσει τις εργασίες που έχουν ανατεθεί. Παρά τη δυνητική θετική συνεισφορά τους, θα πρέπει να επισημανθεί πως στην ουσία τους αυτές οι μορφές ανατροφοδότησης αποτυγχάνουν να αξιοποιήσουν το δυναμικό της ΑΜ.

Με βάση την εμπειρική βιβλιογραφία διαπιστώνονται πτυχές της ανατροφοδότησης, όπως αυτές υλοποιούνται στο πεδίο της ΑΜ, που είναι ιδιαίτερα προβληματικές: εστίαση σε επίδοση, διαφάνεια, γενικότητα.

Επίδοση σε κοιλίξ: Ο πιο συνηθισμένος τύπος ανατροφοδότησης που υιοθετείται στο πλαίσιο της ΑΜ περιλαμβάνει την παροχή πληροφοριών για την τρέχουσα επίδοση των φοιτητών. Η επίδοση αυτή μπορεί να χρησιμοποιηθεί τόσο για την πρόβλεψη του τελικού βαθμού όσο και για τον εντοπισμό των φοιτητών που είναι πολύ πιθανόν να αποτύχουν στο μάθημα. Η παραδοχή που υπάρχει πίσω από αυτόν τον τύπο ανατροφοδότησης είναι ότι η επίδοση

αντανακλά το επίπεδο κατανόησης των φοιτητών. Θα πρέπει να σημειωθεί ότι πληροφόρηση που αφορά αποκλειστικά και μόνο την επίδοση με τέτοιους συγκεντρωτικούς όρους είναι άκρως προβληματική. Για παράδειγμα, αν πληροφορηθεί ένας φοιτητής ότι η συνολική του βαθμολογία σε ένα κουίζ είναι 50%, τότε έχει μεν μια εικόνα της επίδοσης του, αλλά αυτό δεν σημαίνει ότι γνωρίζει τα συγκεκριμένα σημεία τα οποία απαιτούν βελτίωση. Δηλαδή η γνώση του συνολικού επιπέδου επίδοσης δεν μεταφράζεται αναγκαστικά σε απτές και συγκεκριμένες διαδρομές για τη βελτίωση του επιπέδου αυτού. Από την άποψη αυτή, μια συνολική βαθμολογία δεν αποτελεί επαρκή πληροφόρηση, καθώς οι φοιτητές πιθανόν να μην είναι σε θέση να εντοπίσουν τις συγκεκριμένες έννοιες που τους διέφυγαν. Παράλληλα, ο εντοπισμός των εννοιών με τις οποίες αντιμετώπισαν δυσκολίες απαιτεί προχωρημένα επίπεδα γνώσης, αυτορρύθμισης και μεταγνώσης, τα οποία θα επιτρέψουν στους φοιτητές να εντοπίσουν τα διάφορα εννοιολογικά κενά που έχουν, ώστε μακροπρόθεσμα να τα καλύψουν. Στη βιβλιογραφία εντοπίζονται μια σειρά από μελέτες που τείνουν να επισημαίνουν το συγκεκριμένο πρόβλημα που χαρακτηρίζει με την ανατροφοδότηση αυτού του είδους (Bodily & Verbert, 2017· Guillot et al., 2018· Jivet et al., 2017).

Διαφάνεια: Το είδος ανατροφοδότησης που παρέχεται από πολλά συστήματα ΑΜ αποτελεί επίσης ένα σημείο έντονο προβληματισμού στη σχετική βιβλιογραφία. Για παράδειγμα, στην ερευνητική επισκόπηση των Bodily και Verbert (2017), διαπιστώθηκε ότι παρόλο που οι μισές περίπου μελέτες περιλάμβαναν τη διατύπωση προτάσεων προς τους φοιτητές, πολλές από τις προτάσεις αυτές ήταν ασαφείς. Οι Matcha et al. (2020) υποστήριξαν πως θα πρέπει να υπάρχει ρητή αντιστοίχιση μεταξύ των έργων που εκτελούν οι φοιτητές και των εννοιών και των γνώσεων που αντιστοιχούν σε αυτά. Οι ερευνητές θεώρησαν ότι εφόσον δεν υπάρχει μια τέτοια αναλυτική αντιστοίχιση έργων με τις αντίστοιχες έννοιες, η όποια ανατροφοδότηση θα είναι δυνητικά περιορισμένη. Χαρακτηριστικό είναι επίσης το παράδειγμα της μελέτης των Silvola et al. (2021) οι οποίοι διαπίστωσαν ότι οι φοιτητές επιθυμούν τα συστήματα ΑΜ που παρέχουν λεπτομερή ανατροφοδότηση σχετικά με τη μαθησιακή πρόοδο τους. Μάλιστα, από την εν λόγω μελέτη προκύπτει ότι οι φοιτητές θεώρησαν ως πολύ σημαντική την ανατροφοδότηση που τους βοηθάει να προσδιορίσουν τα εννοιολογικά τους κενά, διευκολύνοντας τους με τον τρόπο αυτό να εργαστούν στη βελτίωση τους.

Γενικότητα: Σε κάποιες από τις μελέτες που έχουν πραγματοποιηθεί στο πεδίο της ΑΜ εφαρμόζονται σχήματα ανατροφοδότησης, τα οποία περιλαμβάνουν προσωποποιημένα μηνύματα προς τους φοιτητές. Ειδικότερα, τα μηνύματα αυτά βασίζονται είτε στην επίδοση των φοιτητών είτε στο επίπεδο εμπλοκής τους με τους μαθησιακούς πόρους. Παρά το γεγονός ότι τα προσωποποιημένα μηνύματα είναι πολύ σημαντικά, θα πρέπει να επισημανθεί πως ο παρεχόμενος τύπος ανατροφοδότησης είναι συνήθως γενερικού τύπου. Για παράδειγμα, στη μελέτη των Chen et al. (2021) σχεδιάστηκε ένας μηχανισμός προσωποποιημένης μαθησιακής ανατροφοδότησης για την υποστήριξη των φοιτητών. Ο μηχανισμός αυτός βασιζόταν αφενός στην ορθότητα των απαντήσεων στα κουίζ και αφετέρου στο χρόνο που απαιτήθηκε για να απαντηθούν αυτά. Οι ερευνητές σχεδίασαν τέσσερα επίπεδα ανατροφοδότησης τα οποία βασιζόταν σε όλους τους δυνατούς συνδυασμούς ορθότητας και ταχύτητας απάντησης: (α) σωστά- γρήγορα, (β) σωστά- αργά, (γ) λάθος-αργά και (δ) λάθος-γρήγορα. Στην περίπτωση φοιτητών που ολοκλήρωναν το κουίζ και το σύστημα διέγινωσε ότι ανήκαν στην κατηγορία της ελάχιστης κατανόησης (λάθος-αργά), η ανατροφοδότηση τους ενημέρωνε ότι είχαν περιορισμένη κατανόηση της ενότητας και πρότεινε την εκ νέου μελέτη του μαθησιακού υλικού για τη βελτίωση της κατανόησης. Παρόλο που το σύστημα τους ενημέρωνε σχετικά με τη μικρή κατανόηση του υλικού και πρότεινε την εκ νέου μελέτη του, η συγκεκριμένη ανατροφοδότηση δεν ανέφερε ρητά τις συγκεκριμένες έννοιες τις οποίες οι φοιτητές δεν μπόρεσαν να κατανοήσουν. Παράλληλα, η πρόταση για την μελέτη του υλικού δεν παρουσίαζε μια συγκεκριμένη πορεία δράσης για τους φοιτητές. Στη σχετική βιβλιογραφία μπορούν να εντοπιστούν αντίστοιχα παραδείγματα προσωποποιημένης

μάθησης κι ανατροφοδότησης, που είναι πολύ γενική και συγκεκριμένη (π.χ. Pardo et al., 2019).

Το βασικό πρόβλημα που προκύπτει με την ανατροφοδότηση αυτού του είδους έχει δύο ιδιαίτερες πτυχές. Πρώτον, οι φοιτητές ενημερώνονται για τη συνολική επίδοσή τους. Δεύτερον, γίνεται αναφορά σε έννοιες, τις οποίες πιθανόν να μην έχουν κατανοήσει επαρκώς. Ωστόσο, αυτό που απέτυχαν να κατανοήσουν ούτε επισημαίνεται αλλά ούτε και επεξηγείται. Συνεπώς, η ανατροφοδότηση βασίζεται αποκλειστικά και μόνο στην επίδοση. Παρόλο που η επίδοση προϋποθέτει εννοιολογική κατανόηση, η πληροφόρηση σχετικά με την επίδοση και μόνο δεν αντανάκλα αναγκαστικά την απαιτούμενη κατανόηση. Η επίδοση είναι μεν συνάρτηση της εννοιολογικής κατανόησης, αλλά μια απλή αριθμητική τιμή δεν αναπαριστά αναγκαστικά το επίπεδο εννοιολογικής κατανόησης. Οι προτροπές για εκ νέου μελέτη του υλικού είναι προφανώς χρήσιμες, ωστόσο δεν έχουν τον απαιτούμενο βαθμό αναλυτικότητας. Αυτό που απαιτείται είναι λεπτομερής πληροφόρηση σε σχέση με τις έννοιες τις οποίες δεν έχουν κατανοήσει οι φοιτητές, ώστε να μπορέσουν να εστιάσουν συγκεκριμένα σε αυτές. Σε πολλές περιπτώσεις μελετών, η ανατροφοδότηση που παρέχεται δεν είναι προσαρμοσμένη στις μαθησιακές ανάγκες των φοιτητών ούτε και αφορά ουσιαστικά τον βαθμό κατανόησης των εννοιών. Αν η ανατροφοδότηση που παρέχεται δεν εστιάζεται σε συγκεκριμένα στοιχεία της εννοιολογικής κατανόησης και δεν είναι αρκετά λεπτομερής, τότε οι φοιτητές ενημερώνονται μεν για την ύπαρξη ζητημάτων κατανόησης, αλλά δεν τους υποδεικνύεται ένας πολύ συγκεκριμένος δρόμος για να την αντιμετωπίσουν των ενδεχομένων εννοιολογικών κενών αυτών. Κατά συνέπεια, η ανατροφοδότηση που παρέχεται δεν είναι άμεσα και πρακτικά αξιοποιήσιμη.

Χαρακτηριστικά ανατροφοδότησης

Ιδεατά, στο πλαίσιο της ΑΜ η ανατροφοδότηση που παρέχεται θα πρέπει να πληροί δύο βασικά χαρακτηριστικά. Πρώτον, θα πρέπει να επισημαίνει στους φοιτητές εννοιολογικά κενά και ελλείψεις που τυχόν έχουν. Δεύτερον, θα πρέπει να παρέχει τα μέσα στους φοιτητές για να αντιμετωπίσουν αυτά τα εννοιολογικά κενά. Όπως επισημαίνεται χαρακτηριστικά από τους Matcha et al. (2020), η παρεχόμενη ανατροφοδότηση έχει νόημα για τους φοιτητές μόνο εφόσον πληρούνται οι δύο αυτές προϋποθέσεις.

Στην περίπτωση της ΑΜ ένα πλήρως αυτοματοποιημένο σύστημα ανατροφοδότησης, θα πρέπει αφενός να είναι σε θέση να αξιολογεί τη γνώση των φοιτητών και αφετέρου να παρέχει λεπτομερή και προσωποποιημένη ανατροφοδότηση. Τόσο η διάγνωση του επιπέδου κατανόησης όσο και η ανατροφοδότηση προϋποθέτουν δεδομένα, τα οποία μπορεί να προέρχονται είτε από την επίδοση σε διάφορα έργα (κουίζ) είτε από ενέργειες των φοιτητών (π.χ. προβολή πόρων). Εφόσον διαγνωστεί το επίπεδο κατανόησης των φοιτητών, το σύστημα ανατροφοδότησης θα μπορεί να προτείνει τρόπους αντιμετώπισης των προβλημάτων που εντοπίστηκαν στο προηγούμενο βήμα, προτείνοντας μια συγκεκριμένη μεθοδολογία αντιμετώπισης τους.

Στα παραδείγματα των δύο μελετών που προαναφέρθηκαν (Chen et al., 2021· Pardo et al., 2019), η γενική ανατροφοδότηση οφείλεται στον τύπο των δεδομένων πάνω στα οποία αυτή βασίζεται. Πιο συγκεκριμένα, η ανατροφοδότηση βασίστηκε στη βαθμολογία από τα κουίζ που συμπλήρωσαν οι φοιτητές. Παρότι όπως προαναφέραμε η πληροφορία αυτή είναι ιδιαίτερα χρήσιμη, για την αναλυτική διάγνωση του επιπέδου κατανόησης των φοιτητών απαιτούνται λεπτομερέστερα δεδομένα και όχι απλές αριθμητικές τιμές (π.χ. 64% επίδοση στο κουίζ). Η τυπική βαθμολογία ερωτημάτων που περιλαμβάνονται σε κουίζ είναι δυαδική, με αποτέλεσμα η συνολική επίδοση να προκύπτει ως άθροισμα των επιμέρους σωστών απαντήσεων, το οποίο συνήθως μεταφράζεται σε σχετικές τιμές. Αξίζει να επισημανθεί πως οι

αριθμητικές τιμές δεν είναι αρκετά αναλυτικές, ώστε να επιτρέπουν τον προσδιορισμό των πραγματικών επιπέδων κατανόησης των φοιτητών. Για παράδειγμα, ένας φοιτητής μπορεί να απαντήσει λανθασμένα σε κάποιο ερώτημα ενός κουίζ για πολύ διαφορετικούς λόγους. Από τη μία πλευρά, η εσφαλμένη απάντηση μπορεί να οφείλεται σε απλή έλλειψη γνώσεων σχετικά με την έννοια. Από την άλλη πλευρά, μια λανθασμένη απάντηση μπορεί να είναι δηλωτική μιας σοβαρής παρανόησης του υλικού. Συνεπώς, η τυπική δυαδική αναπαράσταση κατανόησης μιας έννοιας, δεν παρέχει την απαιτούμενη εννοιολογική ανάλυση, η οποία να ανοίγει το δρόμο για μια πιο εστιασμένη και λεπτομερή ανατροφοδότηση.

Στην εργασία αυτή εστιάζομαστε στο πρώτο από τα παραπάνω χαρακτηριστικά της ανατροφοδότησης, το οποίο αφορά την ακριβή, αναλυτική και αξιόπιστη αξιολόγηση του επιπέδου κατανόησης των φοιτητών. Ιδεατά, η αξιολόγηση του επιπέδου κατανόησης σε διαδικτυακά περιβάλλοντα αποτελεί ένα αντικείμενο το οποίο έχει διερευνηθεί διεξοδικά στην περίπτωση των ευφών διδακτικών συστημάτων (Corbett & Anderson, 1995· Khajah et al., 2016· Pavlik et al., 2009). Ωστόσο, θα πρέπει να επισημανθεί πως για λόγους της ΑΜ οι τεχνικές αξιολόγησης της γνώσης που έχουν υλοποιηθεί δεν είναι πρακτικά εφαρμόσιμες. Για τον λόγο αυτό στρεφόμαστε σε μια υποσχόμενη εναλλακτική πηγή δεδομένων, το κείμενο. Στο πλαίσιο της ΑΜ, το κείμενο συνιστά έναν τύπο δεδομένων που μπορεί να είναι εύκολα και άμεσα υλοποιήσιμος.

Το κείμενο ως πηγή δεδομένων στην αναλυτική της μάθησης

Στο πλαίσιο ΑΜ σε διαδικτυακά περιβάλλοντα ο πιο συνηθισμένος τύπος αυτοματοποιημένης μέτρησης του επιπέδου γνώσεων των φοιτητών είναι τα κουίζ. Ως μέτρηση, το κουίζ συνιστά ένα στατικό τρόπο καταγραφής της επίδοσης. Σε αντιδιαστολή με το κουίζ, το κείμενο αποτελεί ένα πιο δυναμικό τρόπο αποτύπωσης της κατανόησης των φοιτητών.

Στο πλαίσιο της αλληλεπίδρασής τους σε διαδικτυακά μαθησιακά περιβάλλοντα, οι φοιτητές δημιουργούν πολλά διαφορετικά είδη κειμένων: απαντήσεις σε ανοιχτού τύπου ερωτήσεις, περιλήψεις και γραπτές εργασίες. Τα τελευταία χρόνια έχει επισημανθεί η δυνητική σημασία του κειμένου για τους σκοπούς της ΑΜ. Οι Gašević et al. (2015) πρότειναν τη χρήση κειμένων που δημιουργούν οι φοιτητές, όπως είναι π.χ. οι διαδικτυακές συζητήσεις και τα ιστολόγια, ως τρόπο αντιμετώπισης των περιορισμών που προκύπτουν από την αποκλειστική εστίαση σε αξιολογήσεις, οι οποίες κατά κανόνα βασίζονται σε ανάκληση πληροφοριών από τη μνήμη. Όπως υποστηρίζουν οι ερευνητές, παράλληλα με τις διάφορες αξιολογήσεις που βασίζονται στην επίδοση, υπάρχει επιτακτική ανάγκη να εξεταστούν και δεδομένα τα οποία εστιάζονται στη διδακτική και μαθησιακή διαδικασία. Επιπρόσθετα, οι Mangaroska & Giannakos (2019) επισημάναν την έλλειψη ποιοτικών δεδομένων στο πλαίσιο των μεθόδων συλλογής δεδομένων που χρησιμοποιούνται στην αναλυτική της μάθησης. Γενικά, το γραπτό κείμενο θεωρείται ότι παρέχει μια μοναδική προοπτική στις γνωστικές διαδικασίες των φοιτητών στο πλαίσιο της μάθησης τους (Fischer et al., 2020).

Παρά την αναγνώριση του δυναμικού που παρέχει το κείμενο ως πηγή δεδομένων για τους σκοπούς της ΑΜ, η βιβλιογραφία δείχνει ότι η αξιοποίηση του κειμένου είναι περιορισμένη. Για παράδειγμα, η επισκόπηση των Banhashem et al. (2022) έδειξε πως μόνο των το 18% των μελετών που εξετάστηκαν περιλάμβαναν κάποιου τύπου ανάλυση κειμένου. Ακόμα και στις περιπτώσεις μελετών όπου διαπιστώθηκε η συστηματική αξιοποίηση του κειμένου (π.χ. Lee & Lim, 2016· Lim et al., 2020), η ανάλυση των κειμένων εστίαζε περισσότερο στην ανίχνευση των αντιλήψεων των φοιτητών παρά στην ανάλυση του πραγματικού επιπέδου επίδοσης και κατανόησης του μαθησιακού υλικού.

Η προοπτική αξιοποίησης του κειμένου ως μεθόδου μελέτη της εννοιολογικής κατανόησης έχει γνωρίσει διάδοση από τη δεκαετία του 1980. Από τη μία πλευρά, η έρευνα στη Γνωσιακή Επιστήμη δείχνει ότι ο γραπτός λόγος είναι ένα κατάλληλο εργαλείο για την αναπαράσταση της εννοιολογικής κατανόησης (Kintsch, 1988). Από την άλλη πλευρά, τα ερευνητικά ευρήματα από μετα-αναλύσεις δείχνουν ότι ο γραπτός λόγος μπορεί να αποτελέσει ένα πολύ αποτελεσματικό μαθησιακό εργαλείο (Bangert-Drowns et al., 2004· Graham et al., 2015, 2020). Επίσης, η δημιουργική αρχή της μάθησης με τα πολυμέσα υποστηρίζει την αποτελεσματικότητα της ενεργητικής προσέγγισης στη μάθηση (Fiorella et al., 2020· Mayer, 2021). Τέλος, στο μοντέλο ICAP (Chi & Wylie, 2014), η γραφή αντιστοιχίζεται τόσο στην ενεργή όσο και στη δημιουργική φάση του μοντέλου.

Όταν οι φοιτητές καλούνται να γράψουν μια μικρή περίληψη ενός κειμένου που διάβασαν ή μιας βιντεοδιάλεξης που παρακολούθησαν, θα πρέπει να επεξεργαστούν τις πληροφορίες που παρουσιάστηκαν και να σχηματίσουν ένα αντίστοιχο εννοιολογικό μοντέλο. Στο πλαίσιο της συγγραφής μιας περίληψης πρέπει να κάνουν μια επιλογή των εννοιών από αυτό το μοντέλο και να τις αναπαραστήσουν με γραπτό λόγο, ώστε να βγαίνει νόημα. Ως διαδικασία, ο γραπτός λόγος είναι μια παραγωγική δραστηριότητα που προωθεί την κατανόηση των εννοιών και, επομένως, μπορεί να χρησιμοποιηθεί ως μια μέθοδος παρακολούθησης και ελέγχου της μάθησης.

Δεδομένων των ραγδαίων εξελίξεων στον τομέα της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) την τελευταία δεκαετία (Hirschberg & Manning, 2015· Torfi et al., 2020), πολλές μελέτες έχουν εστιάσει στη χρήση τεχνολογιών ΕΦΓ την ανάλυση των κειμένων που δημιουργούν οι φοιτητές σε διαδικτυακά μαθησιακά περιβάλλοντα. Πιο συγκεκριμένα, η ΕΦΓ έχει χρησιμοποιηθεί σε πολλά διαφορετικά έργα όπως πρόβλεψη της επιτυχημένης ολοκλήρωσης μαθημάτων (Crossley et al., 2016· Robinson et al., 2016), ανίχνευση της γνωστικής παρουσίας (Andre et al., 2021· Hayati et al., 2019), προσδιορισμό της ανάγκης για άμεση παρέμβαση εκπαιδευτών από τα μηνύματα που δημοσιεύονται σε φόρουμ συζήτησης (Almatrafi et al., 2018), προσδιορισμός φοιτητών με ηγετικά χαρακτηριστικά σε φόρουμ συζήτησης (Moon et al., 2014), εντοπισμός σύγχυσης με βάση τα μηνύματα που δημοσιεύονται σε φόρουμ συζήτησης από φοιτητές (Agrawal et al., 2015· Atapattu & Falkner, 2018), διάκριση ενεργών από παθητικές συμπεριφορές (Atapattu et al., 2019) και πρόβλεψης επιπέδου κριτικής σκέψης (Yang et al., 2022). Στο βαθμό που μπορούμε να γνωρίζουμε από τη βιβλιογραφία, τεχνικές ΕΦΓ δεν έχουν χρησιμοποιηθεί άμεσα για την πρόβλεψη της επίδοσης των μαθητών σε διαδικτυακά μαθησιακά περιβάλλοντα.

Συλλογιστική

Με βάση τα παραπάνω, η προσέγγιση που υιοθετείται στην παρούσα μελέτη περιλαμβάνει τη χρήση κειμένων που δημιουργούν οι φοιτητές για την πρόβλεψη της επίδοσης τους. Ειδικότερα, χρησιμοποιήσαμε κείμενα που δημιουργούν οι φοιτητές με τη μορφή περιλήψεων, δεδομένου ότι η γραφή είναι ένας πιο δυναμικός και ενεργός τρόπος αποτύπωσης του επιπέδου της μάθησης των φοιτητών - σε σχέση πάντα με στατικές μετρήσεις όπως τα κουίζ. Η παραδοχή που υπάρχει πίσω από τη συγκεκριμένη βιβλιογραφία είναι ότι η σύνταξη μιας περίληψης μπορεί να εμπεριέχει χρήσιμο σήμα για την πρόβλεψη της εννοιολογικής κατανόησης, παρέχοντας με τον τρόπο αυτό έναν πολύ πιο διεξοδικό και εκλεπτυσμένο δείκτη μάθησης.

Για τον εντοπισμό του επιπέδου της εννοιολογικής κατανόησης χρησιμοποιήθηκε η κειμενική ομοιότητα. Πιο συγκεκριμένα, θεωρήσαμε ότι το κείμενο της βιντεοδιάλεξης συνιστά το αυθεντικό κείμενο αναφοράς και αναπαριστά το ιδεατό επιδιωκόμενο επίπεδο γνώσης για τους φοιτητές. Για κάθε περίληψη που συντάσσουν οι φοιτητές, υπολογίζεται η κειμενική

ομοιότητα με αυτό το κείμενο αναφοράς. Η παραδοχή που υπάρχει πίσω από αυτόν τον υπολογισμό είναι ότι, όσο πιο όμοια είναι η περιληψη που συντάσσουν οι φοιτητές με κάθε πρόταση του κειμένου της βιντεοδιάλεξης, τόσο πιο ακριβής είναι - δηλαδή τόσο πιο κοντά είναι στο επιθυμητό επίπεδο. Ουσιαστικά, χρησιμοποιήσαμε την ευθυγράμμιση των δύο ανυσμάτων, απομαγνητοφωνημένης βιντεοδιάλεξης και σύντομης περιληψης, ως ένα δείκτη προσεγγιστικής εννοιολογικής ορθότητας.

Αναπαραστάσεις κειμένου

Το κείμενο μπορεί να αναπαρασταθεί ανυσματικά με διάφορους τρόπους, με τη χρήση είτε στατικών είτε νευρωνικών προσεγγίσεων. Από την πλευρά της στατιστικής αναπαράστασης κειμένου, το κείμενο μοντελοποιείται ως άνυσμα συχνοτήτων με τη συχνότητα της κάθε λέξης να είναι ανάλογη της βαρύτητας που έχει η λέξη για το συγκεκριμένο έγγραφο εντός του οποίου συναντάται. Σε αυτό τον τύπο αναπαράστασης υιοθετείται η παραδοχή ότι οι λέξεις από τις οποίες απαρτίζεται ένα κείμενο είναι ανεξάρτητες μεταξύ τους, δηλαδή το κείμενο αποτελείται από ένα σύνολο λέξεων - προσέγγιση που είναι γνωστή ως σάκος λέξεων (Bag of Words - BOW) (Jurafsky & Martin, 2023· Manning et al., 2008).

Η αναπαράσταση κειμένου μέσω της προσέγγισης σάκου λέξεων είναι ιδιαίτερα απλή και πρακτική, έχοντας γνωρίσει μεγάλη διάδοση στο πεδίο της ΕΦΓ. Ωστόσο, τα ανύσματα που παράγονται από την προσέγγιση αυτή χαρακτηρίζονται από διάφορα σημαντικά μειονεκτήματα. Πρώτον, η διάσταση των ανυσμάτων που προκύπτουν εξαρτάται από το συνολικό πλήθος των λέξεων στο λεξιλόγιο: όσο περισσότερες λέξεις υπάρχουν στο σώμα κειμένων που εξετάζεται, τόσο μεγαλύτερη θα είναι η διάσταση των ανυσμάτων που δημιουργούνται. Κατά κανόνα, τα ανύσματα που δημιουργούνται μέσω της προσέγγισης αυτής έχουν πολύ μεγάλες διαστάσεις. Δεύτερον, οι περισσότερες λέξεις ενός σώματος κειμένων δεν εμφανίζονται σε ένα δεδομένο έγγραφο. Αυτό συνεπάγεται πως εκτός από πολυδιάστατα, τα ανύσματα που δημιουργούνται είναι αραιά (sparse), δεδομένου ότι η πλειοψηφία των τιμών είναι μηδέν. Τρίτον, οι στατιστικές αναπαραστάσεις κειμένου μέσω σάκου λέξεων δεν αναπαριστούν σημασιολογικά τις λέξεις. Η κάθε λέξη αναπαρίσταται με ένα μοναδικό άνυσμα, γεγονός που συνεπάγεται ότι λέξεις που είναι πολύ κοντά σημασιολογικά (π.χ. “ωραία” και “όμορφα”) θα έχουν εντελώς διαφορετικά ανύσματα. Το γεγονός αυτό καθιστά προβληματικό τον υπολογισμό της σημασιολογικής ομοιότητας κειμένων.

Από το 2010 και μετά συντελέστηκε μια επανάσταση στον τομέα της ΕΦΓ με τη μετάβαση από στατιστικές σε νευρωνικές προσεγγίσεις (Hirschberg & Manning, 2015· Torfi et al., 2020). Ειδικότερα, οι νευρωνικές προσεγγίσεις οδήγησαν σε μια διαφορετική κατηγορία διανυσματικών αναπαραστάσεων κειμένου, τις ενσωματώσεις λέξεων (word embeddings) (Jurafsky & Martin, 2023· Mikolov et al., 2013). Το κείμενο μοντελοποιείται ως άνυσμα σχέσεων μιας δεδομένης λέξης με τις γειτονικές της: το πλαίσιο συνεισφέρει καθοριστικά στη μοντελοποίηση μιας λέξης. Η βασική ιδέα πίσω από την νευρωνική αυτή προσέγγιση είναι ότι οι λέξεις που έχουν παρόμοια σημασία θα τείνουν να βρίσκονται πολύ κοντά στον πολυδιάστατο χώρο, όπου γίνεται η αναπαράστασή τους.

Σε σχέση με τις στατιστικές αναπαραστάσεις τύπου σάκος λέξεων, οι νευρωνικές αναπαραστάσεις χαρακτηρίζονται από σημαντικά πλεονεκτήματα και αποτελούν έναν πολύ πιο εκλεπτυσμένο τρόπο αναπαράστασης κειμένου. Πρώτον, τα ανύσματα που δημιουργούνται από τις νευρωνικές αναπαραστάσεις έχουν πολύ μικρότερες διαστάσεις (π.χ. 300). Παράλληλα, δεδομένου ότι ο υπολογισμός των διαστάσεων δε βασίζεται σε συχνότητες των λέξεων σε ένα δεδομένο κείμενο, τα ανύσματα αυτά είναι ιδιαίτερα πυκνά (dense) καθώς δεν περιλαμβάνουν μηδενικές τιμές. Δεύτερον, τα ανύσματα που παράγονται από τεχνητά

νευρωνικά δίκτυα αναπαριστούν σημασιολογικά τις λέξεις ενός κειμένου. Το γεγονός αυτό επιτρέπει την εξέταση της σημασιολογικής ομοιότητας δύο κειμένων με προηγμένο τρόπο. Πιο συγκεκριμένα, στην περίπτωση στατιστικής αναπαράστασης κειμένου με ανύσματα συχνότητας που προαναφέρθηκε, αν δεν υπάρχουν κοινοί όροι μεταξύ δύο κειμένων, τότε η κειμενική ομοιότητα θα είναι μηδέν - ακόμα και αν τα δύο κείμενα έχουν επακριβώς το ίδιο νόημα. Όπως είναι αντιληπτό, σε μια τέτοια περίπτωση η σημασιολογική κειμενική ομοιότητα δεν θα αντανakλά με ακρίβεια τον βαθμό στον οποίο τα δύο κείμενα έχουν την ίδια σημασία. Από την άλλη πλευρά, ο περιορισμός αυτός δεν ισχύει όταν η αναπαράσταση κειμένου υλοποιείται νευρωνικά, δηλαδή με ενσωματώσεις λέξεων. Αν δύο κείμενα έχουν το ίδιο νόημα, τότε οι ενσωματώσεις λέξεων μπορούν να ανιχνεύσουν αυτή την κειμενική ομοιότητα - ακόμα και στην ακραία περίπτωση που αυτά τα δύο κείμενα δεν έχουν κανένα κοινό όρο. Το γεγονός αυτό καθιστά τις ενσωματώσεις λέξεων πολύ πιο κατάλληλες για την αναπαράσταση κειμένου.

Θα πρέπει βέβαια να επισημανθεί πως, παρά τα πλεονεκτήματα αυτά, οι αναπαραστάσεις κειμένου που δημιουργούνται χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα όπως το Word2Vec (Mikolov et al., 2013), χαρακτηρίζονται από συγκεκριμένους περιορισμούς, που περιλαμβάνουν τη στατική αναπαράσταση λέξεων. Ωστόσο, για λόγους έκτασης, δε θα επεκταθούμε περαιτέρω στην παρούσα εργασία.

Εστίαση έρευνας και Ερευνητικά Ερωτήματα

Λαμβάνοντας υπόψη ότι η έρευνα δείχνει ότι ο συνδυασμός διαφορετικών τύπων κειμενικής αναπαράστασης οδηγεί σε υψηλότερα επίπεδα ταξινόμησης (Kim & Jeong, 2019), χρησιμοποιήσαμε διαφορετικούς τύπους κειμενικής αναπαράστασης. Επίσης, όπως αναφέρθηκε παραπάνω, το αποτέλεσμα του υπολογισμού της σημασιολογικής κειμενικής ομοιότητας εξαρτάται σε μεγάλο βαθμό από τον τύπο της αναπαράστασης του κειμένου που υιοθετείται. Στην παρούσα εργασία η παραδοχή που υπάρχει πίσω από τη χρήση διαφορετικών ανυσματικών παραστάσεων είναι ότι κάθε μία από αυτές δύναται να αποτυπώσει νοήματα με μοναδικούς τρόπους.

Η παρούσα εργασία παρουσιάζει μια νέα προσέγγιση που βασίζεται σε σύντομες περιλήψεις που δημιουργούν φοιτητές μετά την παρακολούθηση βιντεοδιαλέξεων. Η εργασία διερευνά την εξαγωγή μεταβλητών από τα κείμενα αυτά και τη χρήση τους για την πρόβλεψη της επίδοσης. Ειδικότερα, εξετάζονται δύο σύνολα μεταβλητών (α) πρωτογενούς κειμένου και (β) μεταβλητών που προκύπτουν από την επεξεργασία του κειμένου.

Τα ερευνητικά ερωτήματα που στοχεύει να απαντήσει η παρούσα μελέτη είναι τα ακόλουθα:

(α) Ποιο σύνολο μεταβλητών που εξάγονται από κείμενο (πρωτογενές έναντι επεξεργασμένο κείμενο) οδηγεί σε αποδοτικότερη ταξινόμηση της επίδοσης με κριτήρια αναφοράς τα μέτρα ακρίβειας και F1; Σύμφωνα με την υπάρχουσα βιβλιογραφία (Almatrafi et al., 2018· Dessì et al., 2019· Lin et al., 2009· Rossi & Gnawali, 2014· Wise et al., 2017) η εξαγωγή μεταβλητών από σώματα κειμένου (πρωτογενές ή επεξεργασμένο κείμενο) μπορεί να αποδώσει σημαντικά στοιχεία για την πρόβλεψη της επίδοσης. Η ακρίβεια και το μέτρο F1 αποτελούν σημαντικά μέτρα αξιολόγησης της απόδοσης ενός ΕΦΓ συστήματος (Huang et al., 2015).

(β) Ποιοι αλγόριθμοι Μηχανικής Μάθησης (MM) επιτυγχάνουν την υψηλότερη απόδοση στην ταξινόμηση με κριτήρια αναφοράς τα μέτρα ακρίβεια και F1 ως συνάρτηση των χρησιμοποιούμενων μεταβλητών (πρωτογενές έναντι επεξεργασμένο κείμενο); Η παρούσα μελέτη εστιάζει στην εξέταση οκτώ δημοφιλών αλγορίθμων MM (Logistic Regression, K-Nearest Neighbors, Random Forest, Support Vector Classifier, Naive Bayes, AdaBoost, GradientBoost και Linear Support Vector Classifier). Από μαθηματικής πλευράς, αυτοί οι ταξινομητές εφαρμόζουν διαφορετικές

τεχνικές που ποικίλλουν ως προς την πολυπλοκότητα, την ερμηνεία αποτελεσμάτων και την υπολογιστική ισχύ, γεγονός που καθιστά τη σύγκριση ενδιαφέρουσα.

Μέθοδος έρευνας

Συμμετέχοντες και πλαίσιο έρευνας

Η μελέτη υιοθετεί μια ερευνητική προσέγγιση που βασίζεται στο σχεδιασμό (Cobb et al., 2003· Collins et al., 2004), όντας προσαρμοσμένη σε επίπεδο AM (Rienties et al., 2017). Στη μελέτη συμμετείχαν 85 πρωτοετείς φοιτήτριες τμήματος εκπαίδευσης σε περιφερειακό ΑΕΙ. Οι ηλικίες των συμμετεχόντων κυμαίνονταν μεταξύ 18 και 45 ετών ($M=19,86$, $SD=4,8$). Το 35% των φοιτητριών ανέφεραν μέτριο επίπεδο εξοικείωσης με τις ΤΠΕ. Η συμμετοχή στη μελέτη ήταν εθελοντική, ενώ δόθηκε βαθμολογικό κίνητρο συμμετοχής.

Υλικά

Τα μαθησιακά υλικά που χρησιμοποιήθηκαν ήταν έξι βιντεοδιαλέξεις in vitro που κάλυπταν θεμελιώδεις πτυχές των ψηφιακών μέσων (Manovich, 2013). Για τον σχεδιασμό των βιντεοδιαλέξεων υιοθετήθηκαν οι αρχές της Γνωστικής Θεωρίας Πολυμεσικής Μάθησης (Mayer, 2005). Οι βιντεοδιαλέξεις ήταν σύντομες, διάρκειας μέχρι 10', ενώ κάθε μία πραγματεύονταν ένα διακριτό θέμα, όπως π.χ. η μετάβαση από τα αναλογικά στα ψηφιακά μέσα, η ψηφιακή σύνθεση κτλ. Για τις ανάγκες της έρευνας, αξιοποιήθηκε το ΣΔΜ Moodle για την καταγραφή και τη συλλογή των δεδομένων. Δημιουργήθηκε μια γραμμή μάθησης η οποία περιλάμβανε τη βιντεοδιάλεξη, ένα κουίζ δηλωτικής γνώσης και μια φόρμα για την υποβολή της σύντομης περίληψης.

Μετρήσεις

Τα δεδομένα που συλλέχθηκαν περιλάμβαναν (α) γραπτές περιλήψεις και (β) μέτρηση της δηλωτικής γνώσης. Αναφορικά με το πρώτο, μετά την παρακολούθηση της κάθε βιντεοδιάλεξης οι συμμετέχουσες κλήθηκαν να συντάξουν μια σύντομη περίληψη, η οποία κατά την κρίση τους αντικατόπτριζε τις κύριες έννοιες που περιλάμβανε η διάλεξη. Η ρητή οδηγία που δόθηκε στις συμμετέχουσες ήταν να γράψουν μια σύντομη περίληψη μέχρι 100 λέξεις, η οποία θεωρούσαν ότι αντικατόπτριζε τις έννοιες που είχαν παρουσιαστεί. Αναφορικά με το δεύτερο, για κάθε διάλεξη δημιουργήθηκε ένα τεστ δηλωτικής γνώσης το οποίο αποτελούνταν από 10 ερωτήσεις σε κλειστού τύπου (Σ-Λ). Η συνολική βαθμολογία υπολογίζονταν αθροίζοντας τις σωστές απαντήσεις στο τεστ αυτό, δεδομένου ότι κάθε σωστή απάντηση έπαιρνε ένα βαθμό.

Διαδικασία

Σε πρώτη φάση οι φοιτήτριες ενημερώθηκαν για τον σκοπό της έρευνας και τις προϋποθέσεις για τη συμμετοχή τους σε αυτή. Αφού εκδηλώθηκε ενδιαφέρον, αποστάλθηκε το σχετικό έντυπο συναίνεσης, το οποίο συμπληρώθηκε και επιστράφηκε. Με βάση τον κατάλογο που προέκυψε, δημιουργήθηκαν ατομικοί λογαριασμοί για κάθε μια από τις συμμετέχουσες στο ΣΔΜ.

Στη δεύτερη φάση ακολούθησε η διεξαγωγή της έρευνας, η οποία πραγματοποιήθηκε εξ ολοκλήρου διαδικτυακά και είχε συνολική διάρκεια 3 περίπου ωρών. Οι συμμετέχουσες συνδέονταν στο ΣΔΜ και ακολουθούσαν τη γραμμή μάθησης που είχε δημιουργηθεί, η οποία περιλάμβανε: (α) βιντεοδιάλεξη, (β) φόρμα σύνταξης της περίληψης και (γ) τεστ δηλωτικής γνώσης. Θα πρέπει να σημειωθεί ότι οι συμμετέχουσες δε μπορούσαν να προχωρήσουν στην επόμενη βιντεοδιάλεξη αν δεν είχαν ολοκληρώσει πλήρως την προηγούμενη, αν δηλαδή

προηγούμενως δεν είχαν συντάξει την περίληψη και είχαν απαντήσει σε όλα τα ερωτήματα του τεστ δηλωτικής γνώσης. Μετά την παρακολούθηση της πρώτης βιντεοδιάλεξης, τη σύνταξη της περίληψης για αυτή και την απάντηση του τεστ δηλωτικής γνώσης, η ίδια διαδικασία επαναλαμβάνονταν για τις υπόλοιπες 5 βιντεοδιαλέξεις.

Ανάλυση

Τα κείμενα που χρησιμοποιήθηκαν για τους σκοπούς της έρευνας περιλάμβαναν (α) τις απομαγνητοφωνημένες βιντεοδιαλέξεις και (β) τις περιλήψεις που είχαν συντάξει οι συμμετέχουσες. Χρησιμοποιήθηκε ο διαχωριστής της spaCy (Honnibal & Montani, 2017) για την τμηματοποίηση των κειμένων αυτών σε διακριτά στοιχεία (tokens) που αντιστοιχούσαν σε λέξεις. Η ανάλυση πραγματοποιήθηκε με τεχνικές MM και διακρίθηκε σε δύο στάδια: α) την επεξεργασία κειμενικών χαρακτηριστικών και β) την εφαρμογή αλγορίθμων MM.

Αναφορικά με την επεξεργασία κειμενικών χαρακτηριστικών, χρησιμοποιήθηκαν δύο τύποι αναπαραστάσεων κειμένου, στατιστικές και νευρωνικές. Αφενός δημιουργήθηκαν ανύσματα από συχνότητες λέξεων χρησιμοποιώντας την τυπική προσέγγιση σάκου λέξεων (BoW). Λαμβάνοντας υπόψη το σύνολο του λεξιλογίου σε βιντεοδιαλέξεις και περιλήψεις, τα αγωνίσματα που δημιουργήθηκαν είχαν περισσότερο από 1.5K διαστάσεις. Αφετέρου δημιουργήθηκαν ανύσματα κειμένου από ενσωματώσεις λέξης χρησιμοποιώντας τη βιβλιοθήκη spaCy (Honnibal & Montani, 2017), τα οποία είχαν 300 διαστάσεις. Η συγκεκριμένη βιβλιοθήκη παρέχει ενσωματώσεις λέξης για την ελληνική γλώσσα. Χρησιμοποιήθηκε το μεγάλο γλωσσικό μοντέλο της spaCy 3.5 (el_core_news_lg), το οποίο είχε εκπαιδευτεί σε ειδησεογραφικά κυρίως κείμενα (Explosion, 2022) και περιλάμβανε 500K μοναδικά ανύσματα.

Ως αποτέλεσμα της επεξεργασίας των κειμενικών αναπαραστάσεων, δημιουργήθηκαν δύο κατηγορίες κειμενικών χαρακτηριστικών (features), οι οποίες αποτέλεσαν τις ανεξάρτητες μεταβλητές της μελέτης (predictors).

Η πρώτη κατηγορία περιλάμβανε το ακατέργαστο κείμενο, το οποίο καταχώρησαν οι συμμετέχουσες ως απάντηση στην προτροπή να συντάξουν μια σύντομη περίληψη μετά την παρακολούθηση της βιντεοδιάλεξης. Το χαρακτηριστικό αυτό περιλαμβάνει στατιστικές αναπαραστάσεις κειμένου με τη μορφή του σάκου λέξεων (BoW). Θα πρέπει να σημειωθεί πως στην περίπτωση αυτή χρησιμοποιήθηκαν αυτούσιες οι περιλήψεις που συντάχθηκαν από τους φοιτητές, χωρίς δηλαδή να γίνει κάποια προηγούμενη κανονικοποίηση τους, π.χ. ορθογραφικές διορθώσεις, λημματοποίηση, αφαίρεση λέξεων κ.λπ. (Παναγιωτακόπουλος κ.α., 2022). Θεωρήσαμε το μη κανονικοποιημένο αυτό κείμενο ως την πιο ακατέργαστη μορφή χαρακτηριστικού, που μπορεί να χρησιμοποιηθεί ως προβλεπτική μεταβλητή. Στην προκειμένη περίπτωση, η κάθε λέξη αποτελεί και ένα χαρακτηριστικό το οποίο μπορεί να χρησιμοποιηθεί για σκοπούς πρόβλεψης. Η παραδοχή μας ήταν ότι αυτή η κατηγορία χαρακτηριστικών αποτελεί τη βάση αναφοράς, γύρω από την οποία μπορούν να γίνουν συγκρίσεις με πιο προηγμένες κατηγορίες χαρακτηριστικών, όπως αυτή που ακολουθεί.

Η δεύτερη κατηγορία κειμενικών χαρακτηριστικών περιλάμβανε το επεξεργασμένο κείμενο και αποτελούνταν από δύο γενικά είδη. Από τη μία πλευρά χρησιμοποιήθηκαν μεταβλητές που περιλάμβαναν συχνότητες (α) λέξεων, (β) προτάσεων, (γ) ουσιαστικών, (δ) ρημάτων και (ε) επιθέτων. Οι δύο πρώτες μεταβλητές αναπαριστούν γενικά γλωσσολογικά στοιχεία ενώ οι τρεις τελευταίες είναι πιο ειδικές και αναφέρονται σε μέρη του λόγου (Part of Speech - POS). Από την άλλη πλευρά, χρησιμοποιήθηκε η σημασιολογική κειμενική ομοιότητα (semantic textual similarity) μεταξύ της κάθε περίληψης που συντάχθηκε από τις συμμετέχουσες και της απομαγνητοφωνημένης βιντεοδιάλεξης. Συνεπώς, η δεύτερη κατηγορία κειμενικών χαρακτηριστικών αποτελούσε ένα συνδυασμό μεταξύ στατιστικών και νευρωνικών αναπαραστάσεων και περιλάμβανε την επεξεργασία των κειμένων.

Η μεταβλητή της επίδοσης διαχωρίστηκε σε δυο κατηγορίες χρησιμοποιώντας ως φίλτρο τη διάμεσο. Συνεπώς, δημιουργήθηκαν δυο ομάδες επίδοσης, μια χαμηλής επίδοσης και μια υψηλής επίδοσης αντίστοιχα. Η δυαδική αυτή μεταβλητή αποτέλεσε την εξαρτημένη μεταβλητή της έρευνας, σκοπός της οποίας ήταν η ταξινόμηση των φοιτητριών σε μια από αυτές τις ομάδες με την αξιοποίηση των κειμενικών χαρακτηριστικών των περιλήψεων τους.

Για τις ανάγκες της έρευνας χρησιμοποιήθηκαν οι παρακάτω αλγόριθμοι MM: Logistic Regression (LR), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Classifier (SVC), Naive Bayes (NB), AdaBoost (AB), GradientBoost (GB) και Linear Support Vector Classifier (LSVC). Κατά την εκτέλεση των αλγορίθμων ακολουθήθηκε η προσέγγιση 10-fold cross validation, χρησιμοποιώντας το 90% των δεδομένων για εκπαίδευση και το 10% των δεδομένων για έλεγχο.

Η απόδοση των αλγορίθμων αξιολογήθηκε με τα κριτήρια που υιοθετούνται ευρέως για την ταξινόμηση σε δύο ή περισσότερες κλάσεις, ακρίβεια (accuracy) και τιμή F1. Το μέτρο της ακρίβειας δηλώνει τον λόγο των ορθών ταξινομήσεων σε σχέση με το συνολικό πλήθος των ταξινομήσεων. Από την άλλη πλευρά, η μετρική F1 συνδυάζει δύο επιμέρους δείκτες, την ορθότητα (precision) και την ανάκληση (recall), αποτελώντας τον αρμονικό μέσο όρο τους (Manning, Raghavan & Schütze, 2008). Υπολογίζεται ως ο λόγος

$$2 * (\text{ορθότητα} * \text{ανάκληση}) / (\text{ορθότητα} + \text{ανάκληση})$$

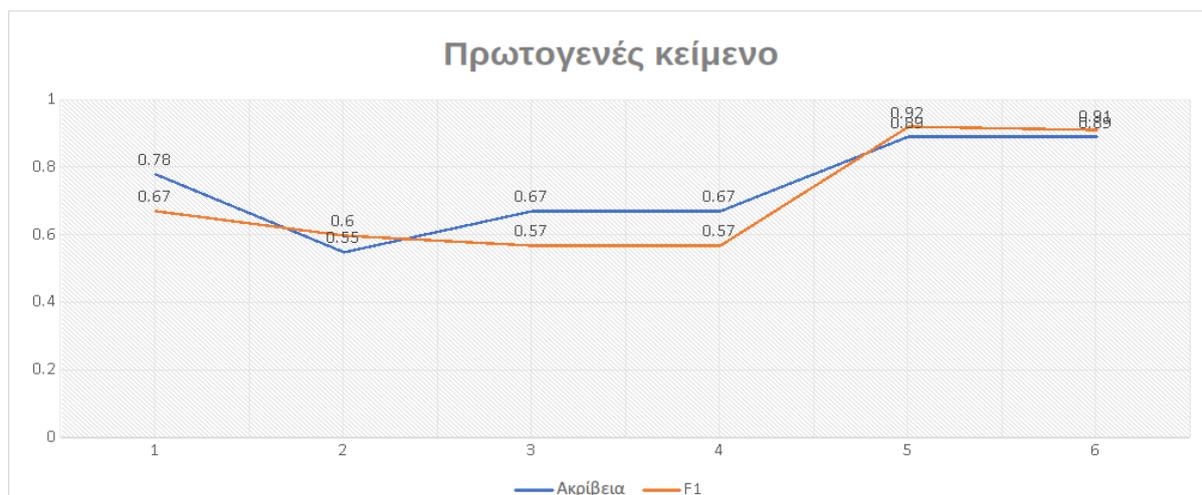
Το εύρος τιμών κυμαίνεται μεταξύ 0 και 1. Στις περιπτώσεις ανισορροπίας κλάσεων οι διάφορες παραλλαγές της μετρικής F1 είναι προτιμητέες έναντι της ακρίβειας, επειδή παρέχουν μια πιο σταθμισμένη εικόνα της ταξινόμησης.

Αποτελέσματα

Το πρώτο ερευνητικό ερώτημα αφορά την αποδοτικότερη ταξινόμηση της επίδοσης ως συνάρτηση των δύο χαρακτηριστικών κειμένου, πρωτογενές και επεξεργασμένο, χρησιμοποιώντας ως κριτήριο τα μέτρα ακρίβεια και F1 τιμή. Στον Πίνακα 1 αποτυπώνεται η συγκριτική εξέταση της απόδοσης των δύο χαρακτηριστικών κειμένου για το μέτρο της ακρίβειας ανεξαρτήτως αλγόριθμου ταξινόμησης. Σε ποσοστό περίπου 60% επί του συνόλου των συγκρίσεων, οι μεταβλητές που εξάχθηκαν από το πρωτογενές κείμενο οδήγησαν σε υψηλότερη ακρίβεια (accuracy) ταξινόμησης σε σχέση με το επεξεργασμένο κείμενο, το οποίο σημείωσε υψηλότερη ακρίβεια ταξινόμησης στο 17% των περιπτώσεων. Ενδιαφέρον έχει το γεγονός ότι στο 25% των περιπτώσεων δεν παρατηρήθηκαν διαφοροποιήσεις.

Πίνακας 1. Σύγκριση απόδοσης πρωτογενούς και επεξεργασμένου κειμένου

Βιντεοδιάλεξη	Αριθμός συγκρίσεων στις οποίες υπερτερεί το ακατέργαστο κείμενο	Αριθμός συγκρίσεων στις οποίες υπερτερεί το επεξεργασμένο κείμενο	Αριθμός συγκρίσεων στις οποίες δεν παρατηρείται διαφοροποίηση
1	3	0	5
2	4	4	0
3	6	1	1
4	6	0	2
5	3	1	4
6	6	2	0
Σύνολο	28	8	12
(%)	58,33	16,67	25,00



Σχήμα 1. Απόδοση ταξινόμησης χαρακτηριστικών πρωτογενούς κειμένου για τα μέτρα ακρίβεια και F1

Πίνακας 2. Σύγκριση απόδοσης πρωτογενούς και επεξεργασμένου κειμένου - μέτρο ταξινόμησης: F1

Βιντεοδιάλεξη	Αριθμός συγκρίσεων στις οποίες υπερτερεί το ακατέργαστο κείμενο	Αριθμός συγκρίσεων στις οποίες υπερτερεί το επεξεργασμένο κείμενο	Αριθμός συγκρίσεων στις οποίες δεν παρατηρείται διαφοροποίηση
1	2	6	0
2	0	8	0
3	4	3	1
4	4	3	1
5	4	4	0
6	6	2	0
Σύνολο	20	26	2

Αξίζει να σημειωθεί ότι η μέση τιμή ακρίβειας ταξινόμησης των αλγορίθμων με την υψηλότερη τιμή είναι 0,74 για το πρωτογενές κείμενο και 0,68 αντίστοιχα για το επεξεργασμένο. Οι αναλυτικές τιμές για τα μέτρα ακρίβειας και F1 για τα χαρακτηριστικά του πρωτογενούς κειμένου σε κάθε βιντεοδιάλεξη δίνονται στο Σχήμα 1. Για τη συγκεκριμένη κατηγορία κειμενικών χαρακτηριστικών, φαίνεται ότι οι τιμές των δύο μέτρων παρουσιάζουν μικρές αποκλίσεις που κυμαίνονται από 0,03 έως 0,12.

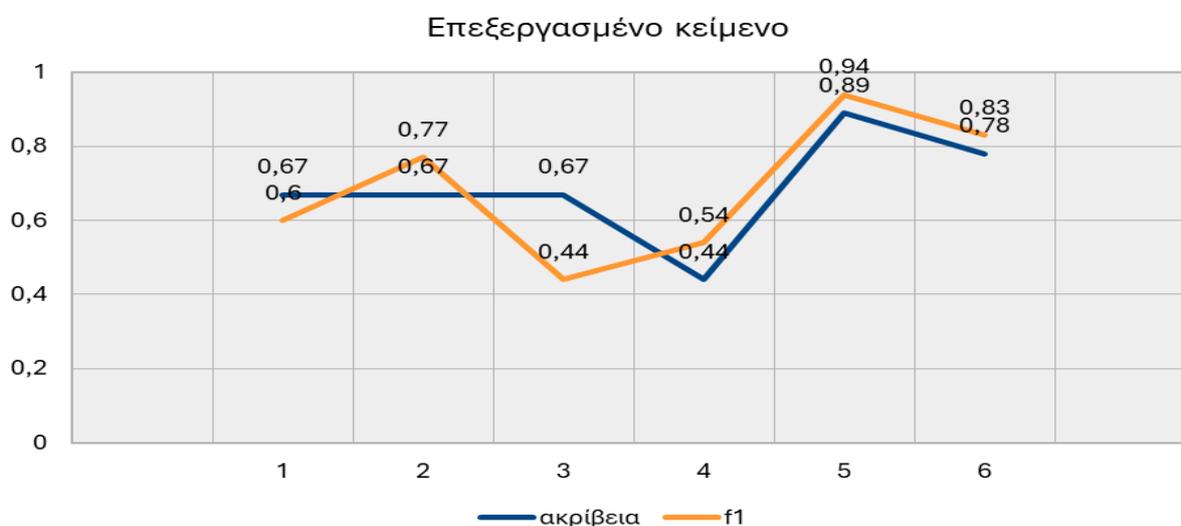
Αναφορικά για το μέτρο F1, οι μεταβλητές που εξάχθηκαν από το πρωτογενές σε σχέση με το επεξεργασμένο κείμενο έχουν σχετικά ισορροπημένη απόδοση, με πολύ μικρές διακυμάνσεις. Οι τιμές των συγκρίσεων ανεξαρτήτως αλγορίθμου MM παρουσιάζονται στον Πίνακα 2, όπου το επεξεργασμένο φαίνεται να υπερέχει ελαφρά σε σχέση με το ακατέργαστο κείμενο. Η μέση τιμή του μέτρου F1 της απόδοσης των αλγορίθμων με την υψηλότερη τιμή είναι 0,70 για το πρωτογενές κείμενο και 0,68 αντίστοιχα για το επεξεργασμένο.

Η σύγκριση μεταξύ της ακρίβειας ταξινόμησης και μέτρου F1 για το επεξεργασμένο κείμενο παρουσιάζεται στο Σχήμα 2. Στην περίπτωση αυτή οι αποκλίσεις μεταξύ των τιμών των δύο μέτρων είναι γενικά μικρές και κυμαίνονται από 0,05 μέχρι 0,10. Η μόνη εξαίρεση σε αυτό το μοτίβο είναι η 3η βιντεοδιάλεξη στην οποία καταγράφεται η μεγαλύτερη διαφορά μεταξύ των δύο μέτρων: 0,23. Επομένως, σε γενικές γραμμές διαπιστώνεται ένα παρόμοιο μοτίβο

αποκλίσεων μεταξύ των μέτρων ακρίβεια και F1 με το ακατέργαστο κείμενο, το οποίο παρουσιάστηκε παραπάνω.

Το δεύτερο ερευνητικό ερώτημα εστίαζε στους αλγόριθμους MM που επιτυγχάνουν την υψηλότερη απόδοση στην ταξινόμηση με κριτήρια αναφοράς τα μέτρα ακρίβεια και F1, χρησιμοποιώντας τα χαρακτηριστικά πρωτογενούς και επεξεργασμένου κειμένου. Τα αποτελέσματα από την ανάλυση ως προς το μέτρο ακρίβεια ταξινόμησης παρουσιάζονται στον Πίνακα 3 και Πίνακα 4. Ως βάση αναφοράς (baseline) θεωρούμε την τυχαία απόδοση (50%) και υπολογίζουμε το ποσοστό βελτίωσης με βάση αυτή. Όπως προκύπτει από τον πίνακα 3, στην περίπτωση του πρωτογενούς κειμένου οι αλγόριθμοι KNN, GB, SVC και LR παρουσιάζουν τις υψηλότερες τιμές ακρίβειας ταξινόμησης καταγράφοντας μια βελτίωση της τάξης από 17% έως 39% σε σχέση με την τυχαία ταξινόμηση.

Από την άλλη πλευρά, για την περίπτωση του επεξεργασμένου κειμένου οι αλγόριθμοι GB, NB και LR παρουσιάζουν τις υψηλότερες τιμές F1 και καταγράφουν μια βελτίωση από 17% έως 39% σε σχέση με την τυχαία ταξινόμηση όπως φαίνεται στον Πίνακα 4.



Σχήμα 2. Απόδοση ταξινόμησης χαρακτηριστικών επεξεργασμένου κειμένου για τα μέτρα ακρίβεια και F1

Πίνακας 3. Υψηλότερη απόδοση αλγορίθμων με χαρακτηριστικά πρωτογενούς κειμένου ανά βιντεοδιάλεξη - Μέτρο: Ακρίβεια

Βιντεοδιάλεξη	Αλγόριθμος	Ακρίβεια (accuracy)	% βελτίωση σε σχέση με την τυχαία ταξινόμηση
1	RFC, SVC, adaboost	0,78	0,28
2	RFC, NB, LSVC	0,55	0,05
3	LR, NB, LSVC	0,67	0,17
4	NB, Gradientboost	0,67	0,17
5	LR, Gradientboost	0,89	0,39
6	LR, KNN, SVC	0,89	0,39

Πίνακας 4. Υψηλότερη απόδοση αλγορίθμων με χαρακτηριστικά επεξεργασμένου κειμένου ανά βιντεοδιάλεξη - Μέτρο ακρίβεια

Βιντεοδιάλεξη	Αλγόριθμος	Ακρίβεια (accuracy)	% βελτίωσης σε σχέση με την τυχαία ταξινόμηση
1	LR, RF, LSVC	0,67	0,17
2	NB, LSVC	0,67	0,17
3	NB	0,67	0,17
4	LR, RF, SVC, Adaboost	0,44	
5	LR	0,89	0,39
6	NB, Gradientboost	0,78	0,28

Πίνακας 5. Υψηλότερη απόδοση αλγορίθμων με χαρακτηριστικά πρωτογενούς κειμένου ανά βιντεοδιάλεξη - Μέτρο F1

Βιντεοδιάλεξη	Αλγόριθμος	F1	Διαφορά μέτρων Ακρίβεια - F1
1	SVC	0,67	0,11
2	NB	0,60	-0,05
3	NB	0,57	0,10
4	NB	0,57	0,10
5	LR	0,92	-0,03
6	LR	0,91	-0,02

Πίνακας 6. Υψηλότερη απόδοση αλγορίθμων με χαρακτηριστικά επεξεργασμένου κειμένου ανά βιντεοδιάλεξη - Μέτρο F1

Βιντεοδιάλεξη	Αλγόριθμος	F1	Διαφορά μέτρων Ακρίβεια - F1
1	Adaboost	0,60	0,07
2	LSVC	0,77	-0,10
3	Adaboost, Gradientboost	0,44	0,23
4	SVC	0,54	-0,10
5	LR	0,94	-0,05
6	NB	0,83	-0,05

Τα αποτελέσματα αναφορικά με το μέτρο F1 παρουσιάζονται στον Πίνακα 5 και στον Πίνακα 6. Όπως προκύπτει από τον πίνακα, στην περίπτωση του πρωτογενούς κειμένου ο αλγόριθμος LR είχε την υψηλότερη απόδοση, καταγράφοντας μια βελτίωση της τάξης του 42% σε σχέση με την τυχαία ταξινόμηση. Για περίπτωση του επεξεργασμένου κειμένου, οι αλγόριθμοι LR και NB καταγράφουν μια βελτίωση από 33% έως 44% σε σχέση με την αναμενόμενη τυχαία ταξινόμηση.

Συζήτηση

Η ευρεία διαθεσιμότητα μεγάλων δεδομένων σε ποικίλα εκπαιδευτικά πλαίσια και ιδιαίτερα στην τριτοβάθμια εκπαίδευση, έχει οδηγήσει στην εμφάνιση του πεδίου της AM, η οποία αποσκοπεί στην υποστήριξη της μάθησης με τη χρήση συμπεριφορικών και μαθησιακών δεδομένων. Παρότι το μαθησιακό δυναμικό της AM είναι πολύ υποσχόμενο (Arnold & Pistilli, 2012· Asif et al., 2017· Caspari-Sadeghi, 2023· Chui et al., 2020· Olive et al., 2019), η αντίστοιχη εμπειρική τεκμηρίωση της δυναμικής αυτής συνεισφοράς είναι σχετικά περιορισμένη (Ferguson & Clow, 2017· Gašević et al., 2016· Guzmán-Valenzuela et al., 2021· Viberg et al., 2018). Η ανατροφοδότηση είναι ιδιαίτερα σημαντική για τη μάθηση (Hattie & Timperley, 2007), ενώ παράλληλα αποτελεί ένα θεμελιώδες συστατικό της AM (Banhashem et al., 2022). Ένα μεγάλο πρόβλημα που παρατηρείται με το είδος της ανατροφοδότησης που παρέχεται από τα συστήματα AM, είναι ότι η ανατροφοδότηση αυτή είναι πολύ συχνά γενερική, αγνοώντας ζητήματα κατανόησης ή αποτυγχάνοντας να παρέχει άμεσα αξιοποιήσιμη ανατροφοδότηση (Ramaswami et al., 2022· Bodily & Verbert, 2017).

Όπως υποστηρίξαμε στο πλαίσιο της βιβλιογραφικής επισκόπησης, ο κυριότερος λόγος για τον οποίο τα συστήματα AM παρέχουν γενερική ανατροφοδότηση είναι ότι η αξιολόγηση της γνώσης και κατανόησης των φοιτητών βασίζεται στην επίδοση τους στα διάφορα κουίζ που χορηγούνται μέσω των ΣΔΜ. Όπως αναλύσαμε, τα κουίζ αυτά παρέχουν μια πολύ ρηχή ανατροφοδότηση της γνώσης των μαθητών, καθώς συνιστούν μια αναπαράσταση της κατανόησής τους με δυαδικούς όρους. Από την βιβλιογραφία των Επιστημών της Μάθησης (Graham et al., 2015, 2020· Kintsch, 1988), προκύπτει ότι ο γραπτός λόγος αποτελεί μια πολύ σημαντική μέθοδο για την αποτύπωση της κατανόησης των μαθητών. Στην εργασία αυτή οι συμμετέχουσες κλήθηκαν να παρακολουθήσουν διαδικτυακά σύντομες βιντεοδιαλέξεις και στη συνέχεια να συντάξουν σύντομες περιλήψεις. Ακολούθως εξετάσαμε το κατά πόσο δύο κατηγορίες κειμενικών χαρακτηριστικών μπορούν να χρησιμοποιηθούν για την πρόβλεψη της επίδοσης φοιτητών.

Το πρώτο ερευνητικό ερώτημα εστίαζε στην αποδοτικότερη ταξινόμηση της επίδοσης χρησιμοποιώντας ως χαρακτηριστικά (features) πρωτογενές και επεξεργασμένο κείμενο και τα μέτρα ακρίβεια και F1 τιμή. Τα αποτελέσματα έδειξαν ότι για μέτρο της ακρίβειας το πρωτογενές κείμενο φαίνεται να οδηγεί σε καλύτερη απόδοση των αλγορίθμων ταξινόμησης σε σχέση με το επεξεργασμένο κείμενο. Από την άλλη πλευρά, σύμφωνα με το μέτρο F1, το επεξεργασμένο κείμενο φαίνεται ότι δίνει καλύτερες αποδόσεις ταξινόμησης σε σχέση με το πρωτογενές. Σε γενικές γραμμές στις περισσότερες βιντεοδιαλέξεις παρατηρήθηκε μια μικρή διαφοροποίηση τιμών των μέτρων ακρίβειας ταξινόμησης και F1.

Το δεύτερο ερευνητικό ερώτημα εξέτασε τους αλγόριθμους MM που επιτυγχάνουν την υψηλότερη απόδοση σε επίπεδο ταξινόμησης χρησιμοποιώντας ως κριτήρια αναφοράς τα μέτρα ακρίβεια και F1, τόσο για το πρωτογενές όσο και το επεξεργασμένο κείμενο. Τα αποτελέσματα έδειξαν ότι χρησιμοποιώντας το μέτρο ακρίβεια ταξινόμησης, τα χαρακτηριστικά αυτού του πρωτογενούς κειμένου οδήγησαν σε υψηλότερη ακρίβεια ταξινόμησης για τρεις από τις έξι συνολικά βιντεοδιαλέξεις. Η λογιστική παλινδρόμηση (LR) ήταν ο αλγόριθμος που επέφερε τα υψηλότερα επίπεδα ακρίβειας ταξινόμησης τόσο για το πρωτογενές όσο και για το επεξεργασμένο κείμενο. Παράλληλα οι αλγόριθμοι KNN και NB οδήγησαν σε βελτιώσεις σε σχέση με αυτό που θα θεωρούσαμε τυχαία απόδοση που κυμαινόταν από 17% έως και 39%. Χρησιμοποιώντας ως κριτήριο την τιμή F1, η εικόνα είναι κάπως διαφοροποιημένη καθώς το επεξεργασμένο κείμενο οδηγεί σε υψηλότερη ταξινόμηση σε δυο από τις έξι βιντεοδιαλέξεις. Οι αλγόριθμοι LR και NB φαίνονται να οδηγούν σε σημαντική βελτίωση από αυτό που θα αναμέναμε από την τυχαία ταξινόμηση.

Δεδομένου ότι δεν υπάρχει προηγούμενη έρευνα η οποία να εξετάζει συστηματικά τη δυνατότητα πρόβλεψης της μάθησης με τη χρήση αποκλειστικά και μόνο συνοπτικών περιλήψεων ως χαρακτηριστικά, είναι δύσκολη η πλαισίωση των ευρημάτων της παρούσας μελέτης. Σε γενικές γραμμές, τα αποτελέσματα ευθυγραμμίζονται με αντίστοιχα που αναφέρονται στη βιβλιογραφία για τα μέτρα της ακρίβειας ή του F1. 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· Hew et al., 2018). Ωστόσο, θα πρέπει να επισημανθεί ότι η παρούσα μελέτη διαφοροποιείται σημαντικά με όρους μεγέθους του σώματος κειμένου που χρησιμοποιήθηκε, αριθμού των χαρακτηριστικών και κανονικοποίησης κειμένου.

Η παρούσα μελέτη έχει ορισμένους σημαντικούς περιορισμούς, οι οποίοι θα πρέπει να ληφθούν υπόψη στην ερμηνεία των αποτελεσμάτων. Πρώτον, το δείγμα δεν ήταν ισορροπημένο ως προς το φύλο, επειδή λόγω της βολικής δειγματοληψίας το μεγαλύτερο ποσοστό συμμετεχόντων ήταν γυναίκες. Η μελλοντική έρευνα θα πρέπει να εξετάσει το κατά πόσο τα αποτελέσματα ισχύουν ανεξαρτήτως φύλου.

Δεύτερον, οι συμμετέχοντες προέρχονταν αποκλειστικά από φοιτητές τμημάτων κοινωνικών επιστημών. Η μελλοντική έρευνα θα πρέπει να διερευνήσει τον βαθμό στον οποίο τα ευρήματα της παρούσας μελέτης επιβεβαιώνονται με φοιτητές από άλλα επιστημονικά πεδία.

Τρίτον, παρότι ο συνολικός αριθμός των συμμετεχόντων είναι ικανοποιητικός, ο αριθμός των περιλήψεων που χρησιμοποιήθηκαν για την ανάλυση ήταν σχετικά μικρός (85 περιλήψεις). Κατά κανόνα, για τις ανάγκες της ΕΦΓ χρησιμοποιούνται μεγάλα σώματα κειμένων. Παρόλο που στη βιβλιογραφία υπάρχουν μελέτες που χρησιμοποιούν μικρά σώματα κειμένων, π.χ. 67 γραπτές εργασίες (Ulman, 2019), η εστίαση δε δίνεται στην επίδοση. Οι μελλοντικές έρευνες θα πρέπει να εξετάσουν σώματα κειμένων διαφορετικού εύρους, ώστε να διακριβωθεί εάν η ποιότητα του σήματος που εμπεριέχεται σε μικρά, μεσαία και μεγάλα σώματα κειμένων επιφέρει διαφοροποιήσεις στην πρόβλεψη της επίδοσης.

Τέταρτον, η αναπαράσταση των περιλήψεων πραγματοποιήθηκε με το μεγάλο γλωσσικό μοντέλο της spaCy για τα Ελληνικά, που αναπαριστά τις λέξεις με άνυσματα 300 διαστάσεων. Το κύριο μειονέκτημα των ενσωματώσεων αυτών είναι ότι μια λέξη αναπαριστάται πάντοτε με το ίδιο άνυσμα ανεξαρτήτως του πλαισίου στο οποίο βρίσκεται. Δεδομένου ότι η σημασία μιας λέξης εξαρτάται πάντοτε από το συγκεκριμένο της, οι στατικές αυτές αναπαραστάσεις λέξεων (static word embeddings) δεν είναι ιδιαίτερα εκλεπτυσμένες. Συνεπώς, οι μελλοντικές έρευνες θα πρέπει να διερευνήσουν μοντέλα ενσωματώσεων κειμένου όπως το SBERT (Reimers & Gurevych, 2019), το BGE (Chen et al., 2024· Xiao et al., 2023) και το E5 (Wang et al., 2023· 2024), τα οποία δημιουργούν πλαισιωμένες αναπαραστάσεις λέξεων (contextual word embeddings). Τα μοντέλα αυτά λαμβάνουν υπόψη το συγκεκριμένο συγκεκριμένο μιας λέξης για να δημιουργήσουν ένα σχετικό άνυσμα που θα αναπαριστά τη σημασία της.

Συνολικά, τα αποτελέσματα δείχνουν ότι η χρήση σύντομων περιλήψεων για την πρόβλεψη της επίδοσης από την παρακολούθηση εκπαιδευτικών βίντεο αποτελούν μια πολύ υποσχόμενη προσέγγιση που μπορεί να ανοίξει πολλαπλούς δρόμους στο πεδίο της ΑΜ. Η δυνατότητα ανατροφοδότησης σε πραγματικό χρόνο προϋποθέτει (α) διάγνωση του επιπέδου κατανόησης των φοιτητών και (β) προτάσεις για την κάλυψη τυχόν εννοιολογικών κενών. Η παρούσα μελέτη εστίασε στο πρώτο χρησιμοποιώντας τεχνικές από το πεδίο της ΕΦΓ για την εξαγωγή χαρακτηριστικών από τις περιλήψεις που συντάσσουν οι φοιτητές που παρακολούθησαν σύντομες βιντεοδιαλέξεις.

Από τα αποτελέσματα προκύπτει πως οι σύντομες περιλήψεις εμπεριέχουν σήμα που μπορεί να αξιοποιηθεί για την αξιολόγηση του επιπέδου κατανόησης των φοιτητών.

Η γενική εικόνα που διαμορφώνεται για το δυναμικό των δύο κατηγοριών χαρακτηριστικών κειμένου για την πρόβλεψη της επίδοσης είναι ιδιαίτερα θετική και μπορεί να συνεισφέρει καθοριστικά στη βελτίωση της μάθησης σε διαδικτυακά μαθησιακά περιβάλλοντα. Η τεράστια πρόοδος που συντελείται τόσο στη δημιουργία Μεγάλων Γλωσσικών Μοντέλων (Large Language Models) όσο και Μοντέλων Ενσωματώσεων Κειμένου (Text Embedding Models) διανοίγει μεγάλες προοπτικές για την αξιοποίηση των διαφόρων τύπων κειμένων που δημιουργούν οι φοιτητές σε περιβάλλοντα ηλεκτρονικής μάθησης. Απαιτείται εκτεταμένη και συστηματική περαιτέρω έρευνα για την αξιοποίηση των δυνατοτήτων που παρέχονται από την ΕΦΓ για την υποστήριξη της μάθησης σε διαδικτυακά μαθησιακά περιβάλλοντα.

Αναφορές

- 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*.
- Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1), 40.
- 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.
- Andre, M., Mello, R. F., Nascimento, A., Lins, R. D., & Gasevic, D. (2021). Toward Automatic Classification of Online Discussion Messages for Social Presence. *IEEE Transactions on Learning Technologies*, 14(6), 802-816.
- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270).
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194.
- Atapattu, T., & Falkner, K. (2018). Impact of Lecturer's Discourse for Students' Video Engagement: Video Learning Analytics Case Study of MOOCs. *Journal of Learning Analytics*, 5(3), 182-197.
- Atapattu, T., Falkner, K., Thilakarathne, M., Sivaneasharajah, L., & Jayashanka, R. (2019). *An Identification of Learners' Confusion through Language and Discourse Analysis*. ArXiv. <https://arxiv.org/pdf/1903.03286>.
- Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. *Computers & Education: Artificial Intelligence*, 2, 100021.
- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis. *Review of Educational Research*, 74(1), 29-58.
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489.
- Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 309-318).
- Caspari-Sadeghi, S. (2023). Learning assessment in the age of big data: Learning analytics in higher education. *Cogent Education*, 10(1).
- Chen, C.-M., Wang, J.-Y., & Hsu, L.-C. (2021). An interactive test dashboard with diagnosis and feedback mechanisms to facilitate learning performance. *Computers and Education: Artificial Intelligence*, 2, 100015.
- 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.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219-243.
- Chui, K. T., Fung, D. C. L., Lytras, M. D., & Lam, T. M. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior*, 107, 105584.
- Clow, D. (2012). The learning analytics cycle. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134-138.
- Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9-13.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(1), 15-42.
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, 4(4), 253-278.
- Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 6-14).

- Dessi, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2019). Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections. *Computers in Human Behavior*, 92, 468–477.
- 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.
- Explosion. (2022). *spaCy* (3.0). <https://spacy.io>
- Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, 27(3), 3743–3775.
- Ferguson, R., & Clow, D. (2017). Where is the evidence? *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 56–65).
- Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *WIREs Data Mining and Knowledge Discovery*, 9(6).
- Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology*, 112(5), 895–906.
- Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, 44(1), 130–160.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84.
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71.
- Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-Based Writing Practices and the Common Core. *The Elementary School Journal*, 115(4), 498–522.
- Graham, S., Kiuahara, S. A., & MacKay, M. (2020). The Effects of Writing on Learning in Science, Social Studies, and Mathematics: A Meta-Analysis. *Review of Educational Research*, 90(2), 179–226.
- Guillot, R., Seanosky, J., Guillot, I., Boulanger, D., Guillot, C., Kumar, V., Fraser, S. N., & Kinshuk, . (2018). Assessing Learning Analytics Systems Impact by Summative Measures. *Proceedings of the 18th International Conference on Advanced Learning Technologies, ICALT* (pp. 188–190). IEEE.
- Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: a preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education*, 18(1), 1–19.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Hayati, H., Chanaa, A., Idrissi, M. K., & Bennani, S. (2019). Doc2Vec & Naïve Bayes: Learners' Cognitive Presence Assessment through Asynchronous Online Discussion TQ Transcripts. *International Journal of Emerging Technologies in Learning*, 14(8), 70–81.
- Hew, K. F., Qiao, C., & Tang, Y. (2018). Understanding Student Engagement in Large-Scale Open Online Courses: A Machine Learning Facilitated Analysis of Student's Reflections in 18 Highly Rated MOOCs. *The International Review of Research in Open and Distributed Learning*, 19(3), 69–93.
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261–266.
- Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To Appear*, 7(1), 411–420.
- Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017* (pp. 82–96). Springer International Publishing.
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. NJ: Prentice Hall.
- Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). *How deep is knowledge tracing?*
- Kim, H., & Jeong, Y.-S. (2019). Sentiment Classification Using Convolutional Neural Networks. *Applied Sciences*, 9(11), 2347.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.
- Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618.
- Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology*, 52(2), 768–784.
- Lee, A., & Lim, T. M. (2016). Mining opinions from university students' feedback using text analytics. *Information Technology in Industry*, 4(1), 26–33.

- Lim, L.-A., Dawson, S., Gašević, D., Joksimović, S., Fudge, A., Pardo, A., & Gentili, S. (2020). Students' sense-making of personalised feedback based on learning analytics. *Australasian Journal of Educational Technology*, 36(6), 15-33.
- Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education*, 52(2), 481-495.
- Mangaroska, K., & Giannakos, M. (2019). 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). Xml retrieval. In *Introduction to Information Retrieval* (pp. 178-200). Cambridge University Press.
- Manovich. (2013). *Software Takes Command* (A & C Black, Ed.).
- Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226-245.
- Mayer, R. E. (2005). *Cognitive theory of multimedia learning* (T. C. handbook of multimedia Learning, Ed.).
- Mayer, R. E. (2021). Evidence-Based Principles for How to Design Effective Instructional Videos. *Journal of Applied Research in Memory and Cognition*, 10(2), 229-240.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781, 3781.
- Moon, S., Potdar, S., & Martin, L. (2014). Identifying student leaders from MOOC discussion forums through language influence. *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 15-20).
- Olive, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D. (2019). A Quest for a One-Size-Fits-All Neural Network: Early Prediction of Students at Risk in Online Courses. *IEEE Transactions on Learning Technologies*, 12(2), 171-183.
- Pardo, A., Bartimote, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., Moskal, A. C. M., Schulte, J., Siemens, G., & Vigentini, L. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, 5(3), 235-249.
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128-138.
- Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis—a new alternative to knowledge tracing. In V. Dimitrova & R. Mizoguchi (Eds.), *Proceedings of the 14th International Conference on Artificial intelligence in education* (pp. 531-538). IOS Press.
- Ramaswami, G., Susnjak, T., Mathrani, A., & Umer, R. (2023). Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. *Technology, Knowledge and Learning*, 28(3), 959-980.
- Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv preprint arXiv:1908.10084.
- Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? In *Big Data and Learning Analytics in Higher Education* (pp. 147-166). Springer International Publishing.
- 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* (pp. 383-387).
- Rossi, L. A., & Gnawali, O. (2014). Language independent analysis and classification of discussion threads in Coursera MOOC forums. In IEEE (Ed.), *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration* (pp. 654-661). IEEE.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30-40.
- Teasley, S. D. (2019). Learning analytics: where information science and the learning sciences meet. *Information and Learning Sciences*, 120(1/2), 59-73.
- Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2020). *Natural language processing advancements by deep learning: A survey*. arXiv preprint arXiv:2003.01200.
- Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, 29(2), 217-257.
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98-110.
- 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.
- Wise, A. F., Cui, Y., Jin, W., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education*, 32, 11-28.

- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.03087>
- Xiao, S., Liu, Z., Zhang, P., & Muennighof, N. (2023). *C-pack: Packaged resources to advance general chinese embedding*. arXiv preprint arXiv:2309.07597.
- Xing, W., & Du, D. (2019). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research, 57*(3), 547–570.
- Yang, J., Du, X., Hung, J.-L., & Tu, C. (2022). Analyzing online discussion data for understanding the student's critical thinking. *Data Technologies and Applications, 56*(2), 303–326.
- Παναγιωτακόπουλος, Χ., Τσαλίδης, Χ., Γάκης, Π., & Κόκκινος, Θ. (2022). *Υπολογιστική γλωσσολογία*. Αθήνα: Κάλυπος, Ανοικτές Ακαδημαϊκές Εκδόσεις.