ePublishing
ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL DOCUMENTATION CENTRE

**Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση**

Τόμ. 1 (2022)

7ο Πανελλήνιο Συνέδριο «Ένταξη και Χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία»

# Η αξιοποίηση κειμένου για την πρόβλεψη της επίδοσης με τη χρήση τεχνικών μηχανικής μάθησης: μια μελέτη περίπτωσης

*Βασιλική Ραγάζου, Χαράλαμπος Παπαδήμας, Ηλίας Καρασαββίδης*

**Βιβλιογραφική αναφορά:**

Ραγάζου Β., Παπαδήμας Χ., & Καρασαββίδης Η. (2023). 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*, *1*, 0809–0820. ανακτήθηκε από https://eproceedings.epublishing.ekt.gr/index.php/cetpe/article/view/5787

# Η αξιοποίηση κειμένου για την πρόβλεψη της επίδοσης με τη χρήση τεχνικών μηχανικής μάθησης: μια μελέτη περίπτωσης

**Ραγάζου Βασιλική, Παπαδήμας Χαράλαμπος, Καρασαββίδης Ηλίας**
ragazou@uth.gr, papadimas@uth.gr,ikaras@uth.gr
Παιδαγωγικό Τμήμα Προσχολικής Εκπαίδευσης, Πανεπιστήμιο Θεσσαλίας

## Περίληψη

Τα δεδομένα που καταγράφονται από Συστήματα Διαχείρισης Μάθησης (ΣΔΜ) για την πρόβλεψη της επίδοσης των φοιτητών έχουν προσελκύσει μεγάλη προσοχή την τελευταία δεκαετία. Συνήθως, ορισμένα από τα δεδομένα αυτά περιλαμβάνουν την παραγωγή διαφορετικών ειδών κειμένων. Ενώ οι φοιτητές παράγουν συστηματικά μεγάλο όγκο κειμένων σε ΣΔΜ, η προβλεπτική ικανότητα των μεταβλητών που παράγονται από κείμενο δεν έχει διερευνηθεί συστηματικά. Η παρούσα εργασία προτείνει τη χρήση κειμένων που δημιουργούνται από φοιτητές για την πρόβλεψη της επίδοσης τους μετά την παρακολούθηση σειράς βιντεοδιαλέξεων. Στην έρευνα συμμετείχαν 44 προπτυχιακές φοιτήτριες οι οποίες παρακολούθησαν έξι βιντεοδιαλέξεις και στη συνέχεια προχώρησαν στη συγγραφή σύντομων περιλήψεων για καθεμία. Από την επεξεργασία των περιλήψεων προέκυψαν δύο σύνολα μεταβλητών (ακατέργαστο και επεξεργασμένο) τα οποία τροφοδότησαν οκτώ αλγόριθμους μηχανικής μάθησης. Τα αποτελέσματα από την ανάλυση δείχνουν ότι τα κείμενα που παράγονται από φοιτητές μπορούν να αποτελέσουν ένα πολύ υποσχόμενο σύνολο μεταβλητών για την πρόβλεψη της μάθησης από βιντεοδιαλέξεις.

**Λέξεις κλειδιά:** μηχανική μάθηση, κείμενο, επεξεργασία φυσικής γλώσσας, ηλεκτρονική μάθηση, βιντεοδιαλέξεις

## Εισαγωγή

Τα τελευταία χρόνια, η Αναλυτική Δεδομένων (ΑΔ - Learning Analytics) υιοθετείται συστηματικά στην Τριτοβάθμια εκπαίδευση για τη βελτίωση της ποιότητας της ηλεκτρονικής μάθησης. Η ΑΔ περιλαμβάνει τη συλλογή και ανάλυση δεδομένων σχετικά με τις αλληλεπιδράσεις των μαθητών μέσα σε εκπαιδευτικά περιβάλλοντα με σκοπό τη βαθύτερη κατανόηση της τρέχουσας γνώσης των μαθητών και τη βελτίωση των μαθησιακών αποτελεσμάτων (Mangaroska & Giannakos, 2019). Στην Τριτοβάθμια Εκπαίδευση, η ΑΔ προσφέρει σημαντικά οφέλη στην εκπαιδευτική κοινότητα, π.χ. αξιοποίηση πληροφοριών σχετικά με το ιστορικό καταγραφής των ενεργειών κάθε μαθητή, που θα μπορούσε να αξιοποιηθεί για την κατασκευή μοντέλων μαθητή LA (Schumacher & Ifenthaler, 2018), με απώτερο σκοπό την παροχή προσωποποιημένης μάθησης (Pardo et al., 2022).

Στη βιβλιογραφία έχει προταθεί μια πληθώρα μεθόδων ΑΔ για την εκμαίευση πληροφοριών και την οπτικοποίηση δεδομένων όπως π.χ. η ανάλυση συστάδων, η ανάλυση δικτύου, η εξόρυξη κειμένου, κ.α. (Dawson et al., 2019; Matcha et al., 2020). Ωστόσο, κρίνεται επιτακτική η ανάγκη για επέκταση της έρευνας στο πεδίο της ηλεκτρονικής μάθησης (Gašević et al., 2015). Ουσιαστικά, ο στόχος κάθε συστήματος ηλεκτρονικής μάθησης είναι η υποστήριξη της μάθησης με βέλτιστο τρόπο. Στην ιδανική περίπτωση, αυτό θα απαιτούσε (α) τον προσδιορισμό της επίδοσης και (β) την αξιοποίηση αυτών των πληροφοριών για την υποστήριξη μαθητών που αντιμετωπίζουν δυσκολίες. Την τελευταία δεκαετία, οι προσεγγίσεις της ΑΔ έχουν διερευνήσει συστηματικά διαφορετικά σύνολα μεταβλητών για την πρόβλεψη της επίδοσης των μαθητών (Schumacher & Ifenthaler, 2018). Ο χρόνος που

Παναγιωτακόπουλος, Χ., Καρατράντου, Α., & Αρμακόλας, Σ. (Επιμ.), *Πρακτικά Εργασιών 7ου Πανελλήνιου Συνεδρίου «Ένταξη και Χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία»*, σ. 809-820, Τμήμα Επιστημών της Εκπαίδευσης και Κοινωνικής Εργασίας, Πανεπιστήμιο Πατρών, 16-18 Σεπτεμβρίου 2022. ISBN 978-618-83186-7-0.

αφιερώνεται στο διαδίκτυο, οι αλληλεπιδράσεις με τους συνομηλίκους, η κατάσταση προόδου της εκτέλεσης μιας εργασίας και ο συνολικός αριθμός αξιολογήσεων που ολοκληρώθηκαν είναι κοινές μεταβλητές ΑΔ που καταγράφονται από ΣΔΜ (Giannakos et al., 2014).

Εκτός των προαναφερθέντων χαρακτηριστικών, η συγγραφή κειμένων από τους μαθητές αποτελεί ένα νέο σύνολο μεταβλητών προς διερεύνηση στα πλαίσια της ηλεκτρονικής μάθησης. Για την επεξεργασία δεδομένων πρωτογενούς κειμένου έχουν υιοθετηθεί τεχνικές Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ- Natural Language Processing) δεδομένου ότι οι μαθητές συχνά παράγουν μεγάλο όγκο κειμένων κατά την αλληλεπίδραση τους με τα ΣΔΜ. Συνήθως, αυτές οι τεχνικές μετατρέπουν το κείμενο σε αριθμητικά διανύσματα τα οποία στη συνέχεια χρησιμοποιούνται ως είσοδοι σε αλγόριθμους Μηχανικής Μάθησης (ΜΜ) (π.χ. Bag-of-Words, Term Frequency–INBerse Document Frequency, Word Embeddings) (Li et al., 2020). Οι μέθοδοι ΕΦΓ έχουν εφαρμοστεί σε διάφορα πεδία όπως η εξόρυξη γνώσης, η ανάλυση συναισθήματος κ.α. Τα αποτελέσματα από τις τεχνικές ταξινόμησης κειμένων είναι πολύ υποσχόμενα (HaCohen-Kerner et al., 2020), ωστόσο επικεντρώνονται κυρίως στην εξέταση συγκεκριμένων μεταβλητών (π.χ. σύνολο λέξεων ανά μήνυμα, χρονόσημο μηνύματος κ.α.) και λιγότερο στη σημασιολογική τους ερμηνεία. Μια εξαίρεση στο πεδίο διερεύνησης κειμένου αποτελούν οι συζητήσεις μέσω των φόρουμ. Εμπειρικές έρευνες δείχνουν ότι το φόρουμ συζήτησης μπορεί να αποτελέσει ένα παράγοντα με σημαντική προβλεπτική ικανότητα στην επίδοση του μαθητή (Schumacher & Ifenthaler, 2018). Αξίζει να σημειωθεί, ότι η νέα τάση στο πεδίο της έρευνας στρέφεται στην εξέταση εννοιολογικών και συμπεριφορικών χαρακτηριστικών με σκοπό τη βαθύτερη κατανόηση της μαθησιακής συμπεριφοράς (Lan et al., 2017).

Η παρούσα εργασία εστιάζει στην παρουσίαση μιας νέας προσέγγισης διερευνώντας αφενός κείμενα (με τη μορφή της σύντομης περίληψης) που δημιουργούνται από φοιτητές ως μεταβλητές για την πρόβλεψη της επίδοσης τους κατά τη μάθηση διαμέσου βιντεοδιαλέξεων και αφετέρου την εξαγωγή μεταβλητών για την καταγραφή της κατανόησης του περιεχομένου από τους μαθητές. Ειδικότερα, η παρούσα μελέτη εστιάζει στις σύντομες περιλήψεις που συντάσσονται από τους φοιτητές μετά την παρακολούθηση βιντεοδιαλέξεων και εξετάζει συγκριτικά την ταξινόμηση χρησιμοποιώντας ως κριτήρια τα μέτρα ακρίβεια (accuracy) και F1 δύο συνόλων μεταβλητών: (α) ακατέργαστου 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 (ακατέργαστο κείμενο vs. επεξεργασμένο κείμενο) οδηγεί σε αποδοτικότερη ταξινόμηση της επίδοσης με κριτήρια αναφοράς τα μέτρα ακρίβειας και F1;

(β) ποιοι αλγόριθμοι ΜΜ επιτυγχάνουν την υψηλότερη απόδοση στην ταξινόμηση με κριτήρια αναφοράς τα μέτρα ακρίβεια και F1 ως συνάρτηση των χρησιμοποιούμενων μεταβλητών (ακατέργαστο κείμενο vs. επεξεργασμένο κείμενο);

## Μεθοδολογία έρευνας

### *Σχέδιο έρευνας και Συμμετέχοντες*

Η μελέτη υιοθετεί μια ερευνητική προσέγγιση που βασίζεται στον σχεδιασμό (Cobb et al., 2003; Collins et al., 2004) όντας προσαρμοσμένη σε επίπεδο ΑΔ (Rienties et al., 2017). Στη μελέτη συμμετείχαν 44 πρωτοετείς φοιτήτριες Παιδαγωγικού Τμήματος Προσχολικής Αγωγής σε περιφερειακό ΑΕΙ. Οι ηλικίες των συμμετεχόντων κυμαίνονταν μεταξύ 18 και 45 ετών (Μ

= 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*

Δημιουργήθηκαν 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). Στα πλαίσια της έρευνας χρησιμοποιήθηκε το ΣΔΜ Moodle σε ένα μάθημα του οποίου δημιουργήθηκε για τις ανάγκες της μελέτης μια γραμμή μάθησης. Πιο συγκεκριμένα, η γραμμή αυτή περιείχε μια ακολουθία μαθησιακών πόρων για τις έξι βιντεοδιαλέξεις με την εξής σειρά: βιντεοδιάλεξη, σύντομη περίληψη και κουίζ.

## *Μετρήσεις*

### Σύντομες περιλήψεις

Μετά την προβολή κάθε βιντεοδιάλεξης, οι φοιτήτριες κλήθηκαν να γράψουν μια σύντομη περίληψη της κάθε βιντεοδιάλεξης, έκτασης περίπου 100 λέξεων. Η σύνταξη και υποβολή των σύντομων περιλήψεων πραγματοποιούνταν απευθείας στο ΣΔΜ.

### Ερωτήσεις κατανόησης

Λόγω ανυπαρξίας ψυχομετρικών κλιμάκων για τη μέτρηση των εννοιών που κάλυπταν οι συγκεκριμένες βιντεοδιαλέξεις, δημιουργήθηκαν τρία είδη τεστ για τη μέτρηση της δηλωτικής γνώσης. Το προ-τεστ περιλάμβανε 14 ερωτήματα κλειστού τύπου (μέγιστη βαθμολογία 14) που αξιολογούσε την εξοικείωση με έννοιες των ψηφιακών μέσων [π.χ.: "Μια ψηφιακή εικόνα αποτελείται από εικονοστοιχεία" (Σωστό-Λάθος)]. Επίσης, δημιουργήθηκαν έξι τεστ δηλωτικής γνώσης για κάθε βιντεοδιάλεξη, κάθε ένα από τα οποία περιείχε 10 ερωτήματα κλειστού τύπου (μέγιστη βαθμολογία 10). Τέλος, δημιουργήθηκε ένα μετα-τεστ με 16 ερωτήματα κλειστού τύπου [π.χ., "Στην περίπτωση της προσομοίωσης φυσικών μέσων σε υπολογιστή, οι γλώσσες των μέσων αλληλεπιδρούν μεταξύ τους" (Σωστό-Λάθος)] το οποίο αξιολογούσε τη συνολική κατανόηση των φοιτητών από την παρακολούθηση όλων των βιντεοδιαλέξεων (μέγιστη βαθμολογία 16).

## *Διαδικασία*

Λόγω των περιορισμών που επιβλήθηκαν στα πλαίσια της πανδημίας COVID-19, η έρευνα πραγματοποιήθηκε διαδικτυακά. Η έρευνα εγκρίθηκε από την επιτροπή ακαδημαϊκής δεοντολογίας του τμήματος και οι φοιτήτριες συμπλήρωσαν τις αντίστοιχες δηλώσεις συναίνεσης για συμμετοχή. Στην αρχή του εξαμήνου οι φοιτήτριες ενημερώθηκαν για τον σκοπό της έρευνας και τις προϋποθέσεις συμμετοχής. Πριν από τη συμμετοχή στην έρευνα, δόθηκαν στις φοιτήτριες συγκεκριμένες οδηγίες για τον τρόπο πλοήγησης στο ΣΔΜ και την παρακολούθηση των βιντεοδιαλέξεων. Μετά την παρακολούθηση της πρώτης βιντεοδιάλεξης, ζητήθηκε από τις συμμετέχουσες να γράψουν μια συνοπτική περίληψη της διάλεξης και στη συνέχεια να συμπληρώσουν ένα τεστ δηλωτικής γνώσης. Η ίδια διαδικασία ακολουθήθηκε και για τις υπόλοιπες βιντεοδιαλέξεις. Με την ολοκλήρωση της έκτης βιντεοδιάλεξης, οι φοιτήτριες συμπλήρωσαν το μετα-τεστ. Η συνολική διάρκεια της έρευνας ήταν περίπου 3 ώρες και πραγματοποιήθηκε σε μία ενιαία συνεδρία. Ωστόσο, για να εξισορροπηθεί ο φόρτος εργασίας, δόθηκε η δυνατότητα ενός σύντομου διαλείμματος 15 λεπτών μετά την 3η βιντεοδιάλεξη (στα μισά της συνεδρίας) πριν προχωρήσουν στις υπόλοιπες.

## Ανάλυση

### *Τεχνικές Μηχανικής Μάθησης*

Αρχικά πραγματοποιήθηκε η εξαγωγή όλων των δεδομένων από το ΣΔΜ, η επεξεργασία και η εκκαθάριση τους πριν την εισαγωγή τους σε υπολογιστικό φύλλο. Για τις ανάγκες της έρευνας υιοθετήθηκε το οικοσύστημα ΜΜ που βασίζεται στην Python και ειδικότερα: α) η βιβλιοθήκη Pandas για την προετοιμασία και την αποθήκευση των δεδομένων και β) η βιβλιοθήκη Scikit-Learn (Pedregosa et al., 2011) χρησιμοποιήθηκε για την ΜΜ. Η βιβλιοθήκη spaCy (Explosion, 2022) χρησιμοποιήθηκε για την ΕΦΓ. Η διαδικασία απομαγνητοφώνησης των βιντεοδιαλέξεων σε κείμενο πραγματοποιήθηκε με τη βιβλιοθήκη αναγνώρισης λόγου Vosk (https://alphacephei.com/vosk/).

### *Επεξεργασία κειμενικών μεταβλητών*

Για τον σκοπό της ανάλυσης δεδομένων, σχεδιάστηκε και αναπτύχθηκε μια σειρά σεναρίων κώδικα σε γλώσσα προγραμματισμού Python. Παράλληλα, δημιουργήθηκαν συναρτήσεις προεπεξεργασίας δεδομένων για αφαίρεση στοιχείων κειμένου που δεν επρόκειτο να αξιολογηθούν κατά την ανάλυση των αλγορίθμων και κλήθηκαν όπου ήταν απαραίτητο. Το πρώτο σύνολο μεταβλητών περιλάμβανε αυτούσιες τις περιλήψεις που συνέταξαν οι φοιτήτριες, χωρίς δηλαδή κάποια προεπεξεργασία. Από την άλλη πλευρά, οι μεταβλητές που προέκυψαν από την επεξεργασία των περιλήψεων απαρτίζονται από 2 ομάδες μεταβλητών.

Η πρώτη ομάδα προέκυψε από την εξαγωγή χαρακτηριστικών του μέρους του λόγου (Part of Speech) και αποτελείται από συχνότητες α) λέξεων, β) προτάσεων, γ) ουσιαστικών, δ) ρημάτων και ε) αντικειμένων.

Η δεύτερη ομάδα βασίστηκε στην εξέταση της σημασιολογικής ομοιότητας κειμένων (semantic similarity) με τη χρήση της βιβλιοθήκης spaCy βασιζόμενο στο πρότυπο της ελληνικής γλώσσας. Η διαδικασία στηρίχτηκε στη σύγκριση 2 αντικειμένων spaCy. Το κείμενο της περίληψης που έγραψαν οι φοιτήτριες αποτέλεσε το ένα αντικείμενο ενώ το άλλο αντικείμενο ήταν το κείμενο της αντίστοιχης απομαγνητοφωνημένης διάλεξης. Εξετάστηκαν διάφοροι συνδυασμοί κειμένου για τη δημιουργία αντικειμένων όπως α) ακατέργαστο κείμενο, β) επεξεργασμένο κείμενο, γ) τμηματοποίηση κειμένου με βάση τα ουσιαστικά (noun – chunks) και δ) επεξεργασμένη τμηματοποίηση κειμένου με βάση τα ουσιαστικά. Μετά από διάφορες δοκιμές, διαπιστώθηκαν τρεις περιπτώσεις υψηλών τιμών ομοιότητας (cosine similarity). Συνεπώς, οι μεταβλητές που προέκυψαν ήταν:

α) μεταβλητή σύγκρισης επεξεργασμένου κειμένου περίληψης προς επεξεργασμένο κείμενο απομαγνητοφωνημένης βιντεοδιάλεξης,

β) μεταβλητή σύγκρισης τμηματοποιημένων ουσιαστικών (noun-chunks) κειμένου περίληψης προς τμηματοποιημένα ουσιαστικά απομαγνητοφωνημένου κειμένου απόδοσης βιντεοδιάλεξης και

γ) μεταβλητή σύγκρισης επεξεργασμένου κειμένου περίληψης προς επεξεργασμένου τμηματοποιημένων ουσιαστικών (noun–chunks) κειμένου απομαγνητοφωνημένης βιντεοδιάλεξης.

### *Αλγόριθμοι Μηχανικής Μάθησης για Ταξινόμηση επίδοσης*

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

**Πίνακας 1. Κατανομή συμμετεχόντων ανά βίντεο**

| Βιντεοδιάλεξη | Μέσος Όρος | Τυπική Απόκλιση | Διάμεσος | Επίδοση (αριθμός συμμετεχόντων) | |
|---|---|---|---|---|---|
| | | | | Χαμηλή | Υψηλή |
| 1 | 0.78 | 0.13 | 0.80 | 29 | 14 |
| 2 | 0.73 | 0.15 | 0.70 | 21 | 22 |
| 3 | 0.69 | 0.14 | 0.70 | 28 | 15 |
| 4 | 0.68 | 0.21 | 0.70 | 26 | 17 |
| 5 | 0.77 | 0.17 | 0.79 | 17 | 26 |
| 6 | 0.70 | 0.15 | 0.15 | 20 | 23 |

Το φίλτρο που χρησιμοποιήθηκε για τη δημιουργία των μεταβλητών επίδοσης ήταν η διάμεσος. Δημιουργήθηκαν με τον τρόπο αυτό δύο κλάσεις, μια χαμηλής επίδοσης και μια υψηλής επίδοσης αντίστοιχα.

Στον Πίνακα 1 παρουσιάζεται η επίδοση των φοιτητριών για κάθε βιντεοδιάλεξη καθώς και η κατανομή των συμμετεχόντων ανά κλάση. Όπως προκύπτει από τον πίνακα, οι κλάσεις είναι απολύτως ισορροπημένες μόνο στην περίπτωση δύο βιντεοδιαλέξεων (2 και 6), ενώ στις άλλες 4 περιπτώσεις υπάρχει ανισορροπία ως προς τον αριθμό των συμμετεχόντων (1, 3, 4 και 5). 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. Δεδομένης της ανισορροπίας των κλάσεων για τα 2/3 των βιντεοδιαλέξεων, επιλέξαμε δύο μέτρα για την αξιολόγηση της ταξινόμησης: ακρίβεια (accuracy) και F1.

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), 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.

## Αποτελέσματα

### *Αξιολόγηση ακρίβειας ταξινόμησης*

Ένα δεύτερο σενάριο Python αξιολόγησε την ακρίβεια ταξινόμησης ανά αλγόριθμο ΜΜ για κάθε βιντεοδιάλεξη χρησιμοποιώντας τα δύο σύνολα μεταβλητών. Η ακρίβεια ταξινόμησης (accuracy) αποτελεί ένα τυπικό μέτρο αξιολόγησης των αλγορίθμων 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.

Σχετικά με την προβλεπτική ικανότητα του ακατέργαστου κειμένου, ο αλγόριθμος RF έδωσε υψηλή τιμή ακρίβειας ταξινόμησης σε τρεις βιντεοδιαλέξεις ενώ ο αλγόριθμος SVC έχει την αμέσως καλύτερη απόδοση στην 1η και 2η βιντεοδιάλεξη.

Αξίζει να σημειωθεί ότι στη 2η βιντεοδιάλεξη παρατηρείται καλή απόδοση σχεδόν από όλους τους υπόλοιπους αλγόριθμους. Η 4η βιντεοδιάλεξη εμφάνισε την μικρότερη τιμή ακρίβειας στον αλγόριθμο LR ενώ οι 5η και 6η εμφάνισαν τις μικρότερες τιμές στους περισσότερους αλγόριθμους.

**Πίνακας 2. Συγκριτική αξιολόγηση της ακρίβειας ταξινόμησης μεταξύ αλγορίθμων ΜΜ για το σύνολο μεταβλητών του ακατέργαστου κειμένου**

| | Βιντεοδιδασκαλίες | | | | | |
|---|---|---|---|---|---|---|
| **Αλγόριθμοι** | **1** | **2** | **3** | **4** | **5** | **6** |
| LSVC | 0.6 | 0.8 | 0.6 | 0.6 | 0.4 | 0.6 |
| GB | 0.4 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 |
| AB | 0.6 | 0.6 | 0.6 | 0.6 | 0.4 | 0.6 |
| NB | 0.6 | 0.8 | 0.6 | 0.6 | 0.6 | 0.6 |
| SVC | 0.8 | 0.8 | 0.6 | 0.4 | 0.4 | 0.6 |
| RF | 0.8 | 1.0 | 0.8 | 0.6 | 0.6 | 0.4 |
| KNN | 0.6 | 0.6 | 0.8 | 0.6 | 0.4 | 0.6 |
| LR | 0.6 | 0.8 | 0.6 | 0.2 | 0.6 | 0.6 |

Σχετικά με την προβλεπτική ικανότητα των μεταβλητών που εξάχθηκαν από το επεξεργασμένο κείμενο, ο αλγόριθμος RFC έδωσε υψηλή τιμή ακρίβειας ταξινόμησης σε τρεις βιντεοδιαλέξεις.

Οι αλγόριθμοι LR, KNN, SVC παρουσίασαν υψηλή τιμή ακρίβειας ταξινόμησης σε περισσότερες από μια βιντεοδιαλέξεις. Η χαμηλότερη τιμή ακρίβειας ταξινόμησης παρουσιάστηκε στη 2η βιντεοδιάλεξη σε τρεις αλγόριθμους ενώ στις βιντεοδιαλέξεις 4, 5 και 6 η απόδοση της ακρίβειας των περισσότερων αλγορίθμων κυμάνθηκε στο 0.8.

**Πίνακας 3. Συγκριτική αξιολόγηση της ακρίβειας ταξινόμησης μεταξύ αλγορίθμων ΜΜ για το σύνολο μεταβλητών του επεξεργασμένου κειμένου**

| | Βιντεοδιδασκαλίες | | | | | |
|---|---|---|---|---|---|---|
| **Αλγόριθμοι** | **1** | **2** | **3** | **4** | **5** | **6** |
| LSVC | 0.6 | 0.4 | 0.6 | 0.6 | 0.6 | 1.0 |
| GB | 0.4 | 0.2 | 0.6 | 0.6 | 0.4 | 0.8 |
| AB | 0.6 | 0.2 | 0.8 | 0.6 | 0.6 | 0.4 |
| NB | 0.6 | 0.2 | 0.4 | 0.6 | 0.6 | 0.8 |
| SVC | 0.6 | 0.6 | 0.4 | 0.8 | 0.6 | 0.8 |
| RF | 0.6 | 0.4 | 0.8 | 0.8 | 0.6 | 0.8 |
| KNN | 0.6 | 0.6 | 0.6 | 0.8 | 0.4 | 0.8 |
| LR | 0.6 | 0.4 | 0.8 | 0.6 | 0.4 | 0.8 |

Η επόμενη φάση ανάλυσης εξέτασε την απόδοση των αλγορίθμων σε κάθε ένα σύνολο μεταβλητών ξεχωριστά.

Στον Πίνακα 4 παρουσιάζεται η υψηλότερη τιμή ακρίβειας ταξινόμησης σε κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών ακατέργαστου κειμένου.

Ο αλγόριθμος RF παρουσίασε την υψηλότερη τιμή ακρίβειας ταξινόμησης στη 2η βιντεοδιάλεξη ενώ σχεδόν σε όλες τις βιντεοδιαλέξεις εμφανίζει αντίστοιχα υψηλές τιμές ακρίβειας.

Οι αλγόριθμοι KNN και GB παρουσιάζουν υψηλές τιμές ακρίβειας ταξινόμησης σε τρεις βιντεοδιαλέξεις.

**Πίνακας 4. Υψηλότερη τιμή ακρίβειας ταξινόμησης για κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών ακατέργαστου κειμένου**

| Βιντεοδιάλεξη | Αλγόριθμος | Ακρίβεια (accuracy) | % βελτίωσης σε σχέση με την τυχαία ταξινόμηση |
|---|---|---|---|
| 1 | RF, SVC | 0.8 | 0.3 |
| 2 | RF | 1.0 | 0.5 |
| 3 | RF, KNN | 0.8 | 0.3 |
| 4 | NB, GB, KNN, RF, AB, LSVC | 0.6 | 0.1 |
| 5 | LR, RF, NB, GB | 0.6 | 0.1 |
| 6 | LR, KNN, SVC, GB, LSVC, AB | 0.6 | 0.1 |

Στον Πίνακα 5 παρουσιάζεται η υψηλότερη τιμή ακρίβειας ταξινόμησης σε κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών επεξεργασμένου κειμένου. Ο αλγόριθμος LSVC έδωσε την υψηλότερη τιμή ακρίβειας ταξινόμησης στην 6η βιντεοδιάλεξη και δίνει καλή απόδοση σε συνολικά τρεις βιντεοδιαλέξεις. Οι αλγόριθμοι RF και SVC έδωσαν την υψηλότερη τιμή ακρίβειας ταξινόμησης στις περισσότερες βιντεοδιαλέξεις.

**Πίνακας 5. Υψηλότερη τιμή ακρίβειας ταξινόμησης για κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών επεξεργασμένου κειμένου**

| Βιντεοδιάλεξη | Αλγόριθμος | Ακρίβεια (accuracy) | % βελτίωσης σε σχέση με την τυχαία ταξινόμηση |
|---|---|---|---|
| 1 | LR, KNN, RF, SVC, NB, AB, LSVC | 0.6 | 0.1 |
| 2 | KNN, SVC | 0.6 | 0.1 |
| 3 | LR, RF, AB | 0.8 | 0.3 |
| 4 | SVC, RF, KNN | 0.8 | 0.3 |
| 5 | RF, SVC, LSVC, AB, NB | 0.6 | 0.1 |
| 6 | LSVC | 1.0 | 0.5 |

Στον Πίνακα 6 παρουσιάζονται τα αποτελέσματα της σύγκρισης των τιμών ακρίβειας ταξινόμησης ακατέργαστου και επεξεργασμένου κειμένου ανεξαρτήτως αλγορίθμων. Οι συχνότητες ανά κατηγορία δείχνουν τις περιπτώσεις στις οποίες καταγράφηκε η μέγιστη ακρίβεια ταξινόμησης για κάθε σύνολο προβλεπτικής μεταβλητής ανά βιντεοδιάλεξη.

**Πίνακας 6.**

| 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 | 0 | 6 |
| 2 | 7 | 0 | 1 |
| 3 | 3 | 2 | 3 |
| 4 | 0 | 4 | 4 |
| 5 | 2 | 3 | 3 |
| 6 | 1 | 7 | 0 |
| **Σύνολο** | **15** | **16** | **17** |
| **(%)** | **31.25** | **33.33** | **35.42** |

### *Αξιολόγηση Ταξινόμησης με τη χρήση του μέτρου F1*

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. Το μέτρο της ορθότητας ισούται με τον λόγο των αληθώς θετικών κατηγοριοποιήσεων προς το άθροισμα των αληθώς θετικών και ψευδώς θετικών κατηγοριοποιήσεων (TP/TP+FP), δηλώνοντας το πόσες από τις κατηγοριοποιήσεις για τις οποίες έγινε πρόβλεψη ότι θα είναι θετικές είναι όντως θετικές. Το μέτρο της ανάκλησης ισούται με τον λόγο των αληθώς θετικών κατηγοριοποιήσεων προς το άθροισμα των αληθώς θετικών και των ψευδώς αρνητικών κατηγοριοποιήσεων (TP/TP+FN) ,δηλώνοντας την αναλογία των θετικών προβλέψεων που κατηγοριοποιήθηκαν ορθά. Το μέτρο F1 είναι ο αρμονικός μέσος όρος της ορθότητας και της ανάκλησης και υπολογίζεται με βάση τον τύπο: 2 (precision * recall)/ precision + recall. 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. Στον Πίνακα 7 παρουσιάζεται η υψηλότερη τιμή του μέτρου F1 σε κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών ακατέργαστου κειμένου. Ο αλγόριθμος LR έδωσε την υψηλότερη τιμή F1 σε τρεις βιντεοδιαλέξεις.

**Πίνακας 7. Υψηλότερη τιμή μέτρου F1 για κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών ακατέργαστου κειμένου**

| Βιντεοδιάλεξη | Αλγόριθμος | F1 |
|---|---|---|
| 1 | SVC | 0.67 |
| 2 | LR, RF, SVC, NB, LSCV | 0.80 |
| 3 | KNN | 0.80 |
| 4 | AB, GB | 0.50 |
| 5 | LR, RF, NB, GB | 0.75 |
| 6 | LR | 0.75 |

Στον Πίνακα 8 παρουσιάζεται η υψηλότερη τιμή του μέτρου F1 σε κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών επεξεργασμένου κειμένου. Οι αλγόριθμοι

KNN, SVC και NB έδωσαν τις υψηλότερες τιμές F1 σε τρεις βιντεοδιαλέξεις. Μάλιστα στη βιντεοδιάλεξη 6 η απόδοση του αλγορίθμου LSVC ήταν η μέγιστη δυνατή (1.00).

**Πίνακας 8. Υψηλότερη τιμή μέτρου F1 για κάθε βιντεοδιάλεξη χρησιμοποιώντας το σύνολο των χαρακτηριστικών επεξεργασμένου κειμένου**

| Βιντεοδιάλεξη | Αλγόριθμος | F1 |
|---|---|---|
| 1 | SVC | 0.67 |
| 2 | LR, RF, SVC, NB, LSCV | 0.80 |
| 3 | KNN | 0.80 |
| 4 | AB, GB | 0.50 |
| 5 | LR, RF, NB, GB | 0.75 |
| 6 | LR | 0.75 |

Στον Πίνακα 9 παρουσιάζονται οι υψηλότερες τιμές των μέτρων της ακρίβειας και F1 για κάθε βιντεοδιάλεξη για την περίπτωση του ακατέργαστου κειμένου. Όπως φαίνεται στον πίνακα, η διαφορά στις τιμές των δύο μέτρων κυμαίνεται από 0.00 έως 0.20. Στις βιντεοδιαλέξεις 1, 2 και 4 το μέτρο της ακρίβειας έχει υψηλότερη απόδοση σε σχέση με το μέτρο F1 ενώ το αντίστροφο ισχύσει για στις βιντεοδιαλέξεις 5 και 6.

**Πίνακας 9. Υψηλότερες τιμές μέτρων ακρίβειας F1 ανά βιντεοδιάλεξη για το σύνολο μεταβλητών ακατέργαστου κειμένου**

| Βιντεοδιάλεξη | Ακρίβεια | F1 | Διαφορά μέτρων Ακρίβεια – F1 |
|---|---|---|---|
| 1 | 0.8 | 0.67 | +0.13 |
| 2 | 1.0 | 0.80 | +0.20 |
| 3 | 0.8 | 0.80 | 0.00 |
| 4 | 0.6 | 0.50 | +0.10 |
| 5 | 0.6 | 0.75 | -0.15 |
| 6 | 0.6 | 0.75 | -0.15 |

Στον Πίνακα 10 παρουσιάζονται οι υψηλότερες τιμές για τα μέτρα ακρίβεια και F1 ανά βιντεοδιάλεξη για την περίπτωση του συνόλου των μεταβλητών επεξεργασμένου κειμένου. Η απόκλιση στις τιμές των μέτρων ακρίβεια και F1 κυμαίνεται από 0.00 έως 0.1 - με εξαίρεση την περίπτωση της βιντεοδιάλεξης 3 όπου η απόκλιση είναι πολύ μεγάλη (0.8) καθώς τη τιμή του F1 ήταν μηδενική. Στις βιντεοδιαλέξεις 1 και 3 το μέτρο ακρίβειας έχει καλύτερη απόδοση σε σχέση με το F1, ενώ το αντίθετο ισχύει στις βιντεοδιαλέξεις 2 και 5.. Τέλος, στις βιντεοδιαλέξεις 4 και 6 δεν υφίστανται διαφοροποιήσεις μεταξύ των δύο μέτρων.

**Πίνακας 10. Υψηλότερες τιμές μέτρων ακρίβειας F1 ανά βιντεοδιάλεξη για το σύνολο μεταβλητών επεξεργασμένου κειμένου**

| Βιντεοδιάλεξη | Ακρίβεια | F1 | Διαφορά μέτρων Ακρίβεια – F1 |
|---|---|---|---|
| 1 | 0.6 | 0.50 | +0.10 |
| 2 | 0.6 | 0.67 | -0.07 |
| 3 | 0.8 | 0.00 | +0.80 |
| 4 | 0.8 | 0.80 | 0.00 |
| 5 | 0.6 | 0.67 | -0.07 |
| 6 | 1.0 | 1.00 | 0.00 |

## Συζήτηση

Αναφορικά με το πρώτο ερευνητικό ερώτημα, τα αποτελέσματα δείχνουν ότι σε ποσοστό 35.42% επί του συνόλου των συγκρίσεων, δεν υφίστανται διαφορές στην ακρίβεια ταξινόμησης μεταξύ ακατέργαστου και επεξεργασμένου κειμένου – ανεξάρτητα από τον αλγόριθμο ΜΜ που χρησιμοποιήθηκε. Επίσης σε ποσοστό 33.33% επί του συνόλου, οι μεταβλητές που εξάχθηκαν από το επεξεργασμένο κείμενο οδήγησαν σε υψηλότερη ακρίβεια ταξινόμησης σε σχέση με το ακατέργαστο κείμενο. Τέλος, το ακατέργαστο κείμενο σημείωσε υψηλότερη ακρίβεια ταξινόμησης στο 31.25% των περιπτώσεων. Συνεπώς, σε ένα ποσοστό περίπου 70% επί του συνόλου των συγκρίσεων, οι προβλεπτικές μεταβλητές που εξάγονται από το επεξεργασμένο κείμενο επιτυγχάνουν την ίδια ή μεγαλύτερη ακρίβεια ταξινόμησης σε σχέση με το αντίστοιχο σύνολο των μεταβλητών που εξάγονται από το ακατέργαστο κείμενο. Αξίζει να σημειωθεί ότι στο σύνολο των έξι βιντεοδιαλέξεων, η μέση τιμή ακρίβειας ταξινόμησης των αλγορίθμων με την υψηλότερη τιμή είναι 0.73 και παραμένει ίδια και στις δύο κατηγορίες κειμένου. Αναφορικά με το μέτρο F1, τα αποτελέσματα στην περίπτωση του ακατέργαστου κειμένου κυμάνθηκαν σε υψηλά επίπεδα (0.75-0.80). Από την άλλη πλευρά, το επεξεργασμένο κείμενο σημείωσε τιμές F1 σε μέτρια επίπεδα, με εξαίρεση την τελευταία βιντεοδιάλεξη (1.00). Συνολικά, για την περίπτωση του ακατέργαστου κειμένου η σύγκριση των δυο μέτρων δείχνει την υπεροχή της ακρίβειας έναντι του F1 στις μισές βιντεοδιαλέξεις. Στην περίπτωση του επεξεργασμένου κειμένου, προκύπτει η υπεροχή της ακρίβειας έναντι του F1 σε μια μόνο βιντεοδιάλεξη χωρίς ωστόσο να παρατηρούνται μεγάλες διαφοροποιήσεις στις υπόλοιπες βιντεοδιαλέξεις.

Αναφορικά με το δεύτερο ερευνητικό ερώτημα, στην περίπτωση του ακατέργαστου κειμένου, ο αλγόριθμος RF έχοντας την υψηλότερη τιμή ακρίβειας ταξινόμησης δίνει υψηλές τιμές απόδοσης στο σύνολο των βιντεοδιαλέξεων, καταγράφοντας μια βελτίωση της τάξης του 30-50% σε σχέση με αυτό που θα αναμέναμε στην περίπτωση τυχαίας ταξινόμησης. Στην περίπτωση μεταβλητών που εξάγονται από το επεξεργασμένο κείμενο, ο αλγόριθμος LSVC με την υψηλότερη τιμή ακρίβειας ταξινόμησης δίνει υψηλές τιμές σε συνολικά τρεις βιντεοδιαλέξεις, με το αντίστοιχο ποσοστό βελτίωσης από την τυχαία ταξινόμηση να ανέρχεται από 30-50%. Ο αλγόριθμος RF δίνει συγκριτικά υψηλές τιμές ακρίβειας ταξινόμησης στο σύνολο των βιντεοδιαλέξεων. Επίσης, παρατηρήθηκε ότι στις βιντεοδιαλέξεις 3 και 5 η απόδοση των αλγορίθμων παραμένει η ίδια ανεξαρτήτως των μεταβλητών κειμένου που χρησιμοποιούνται. Αντίστοιχα, στις βιντεοδιαλέξεις 1 και 2 παρατηρούνται υψηλές τιμές ακρίβειας ταξινόμησης στην περίπτωση του ακατέργαστου κειμένου ενώ στις βιντεοδιαλέξεις 4 και 6 αντίστοιχη υπεροχή για τις μεταβλητές που προέρχονται από το επεξεργασμένο κείμενο. Στις περισσότερες βιντεοδιαλέξεις ο αλγόριθμος LR έχει συστηματικά υψηλή απόδοση με το μέτρο F1 σε σύγκριση με τους υπόλοιπους αλγόριθμους. Αντιθέτως, ο αλγόριθμος NB παρουσιάζει μια διαφορετική εικόνα, έχοντας μικρή απόδοση με το μέτρο F1 (0.00-0.50). Στην περίπτωση του επεξεργασμένου κειμένου, παρατηρείται μόνο σε μία βιντεοδιάλεξη υψηλή τιμή του F1 μέσω του αλγορίθμου LSCV. Στις υπόλοιπες βιντεοδιαλέξεις, οι τιμές του F1 διατηρούνται σε μεσαία επίπεδα (0.50-0.67). Αξιοπρόσεκτη είναι η μηδενική απόδοση των αλγορίθμων ΜΜ για το μέτρο F1 στην περίπτωση της βιντεοδιάλεξης 3 για τον λόγο που αναφέρθηκε παραπάνω.

Συμπερασματικά, τα μέτρα της ακρίβειας και του F1 αποτελούν τα πιο δημοφιλή στο πεδίο της ΜΜ για δυαδικές κατηγοριοποιήσεις. Δεδομένου ότι οι κλάσεις δεν ήταν ισορροπημένες σε όλες τις βιντεοδιαλέξεις, είναι αναμενόμενο ότι οι τιμές από το μέτρο F1 θα είναι πιο συντηρητικές σε σχέση με αυτές της ακρίβειας. Συγκεκριμένα, στην περίπτωση του ακατέργαστου κειμένου το μέτρο F1 είχε σταθερά υψηλή απόδοση σε σύγκριση με τις αντίστοιχες τιμές του F1 στην περίπτωση του επεξεργασμένου κειμένου, οι οποίες

κυμάνθηκαν σε μέτρια επίπεδα. Επίσης, οι μικρές διαφοροποιήσεις τιμών ανά βιντεοδιάλεξη των μέτρων ακρίβειας και F1 τόσο στην περίπτωση του ακατέργαστου κειμένου όσο και στο σύνολο των μεταβλητών από το επεξεργασμένο κείμενο, δείχνει ότι η ανισορροπία των κλάσεων δεν επηρέασε σημαντικά την απόδοση των αλγορίθμων. Οι προφανείς τρόποι επίλυσης της ανισορροπίας κλάσεων είναι είτε τα μεγαλύτερα δείγματα είτε η παρέμβαση σε επίπεδο διαμέσου κατά τη δημιουργία των δυαδικών μεταβλητών.

Εν κατακλείδι, ενώ μεταβλητές που εξάγονται από κείμενο έχουν χρησιμοποιηθεί για διάφορους σκοπούς (π.χ. Dang et al., 2020; Onan, 2021; Song et al., 2020), δεν έχει διερευνηθεί η συνεισφορά του κειμένου στην πρόβλεψη επίδοσης μέχρι τώρα. Ενώ τα αποτελέσματα της παρούσας μελέτης είναι πολύ υποσχόμενα, απαιτείται συστηματικότερη εξέταση της συνεισφοράς του κειμένου στην πρόβλεψη επίδοσης σε μελλοντικές έρευνες.

## Ευχαριστίες

Η παρούσα έρευνα συγχρηματοδοτήθηκε από την Ελλάδα και την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) μέσω του επιχειρησιακού προγράμματος «Ανάπτυξη Ανθρώπινου Δυναμικού, Εκπαίδευση και Διά Βίου Μάθηση 2014-2020» στα πλαίσια του έργου «Σχεδιασμός, Ανάπτυξη και Αξιολόγηση μιας Ευρετικής Μεθοδολογίας για τη Βελτίωση της Ποιότητας της Ηλεκτρονικής Μάθησης Διαμέσου Τεχνικών Μηχανικής Μάθησης» (MIS 5048955).

## Βιβλιογραφικές Αναφορές

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9–13. https://doi.org/10.3102/0013189X032001009.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2.

Dang, N. C., Moreno-García, M. N., & de la Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, 9(3), 483. https://doi.org/10.3390/electronics9030483.

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the Impact of Learning Analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 446–455. New York, NY, USA: ACM. https://doi.org/10.1145/3303772.3303784.

Explosion. (2022). *spaCy (v3.0) [Computer software]*. Retrieved May 23, 2022, from https://spacy.io/.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2014). Collecting and making sense of video learning analytics. *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 1–7. IEEE. https://doi.org/10.1109/FIE.2014.7044485

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLOS ONE*, 15(5), e0232525. https://doi.org/10.1371/journal.pone.0232525.

Lan, A. S., Brinton, C. G., Yang, T.-Y., & Chiang, M. (2017). Behavior-Based Latent Variable Model for Learner Engagement. *10th International Educational Data Mining Society*, 64–71.

Li, P., Mao, K., Xu, Y., Li, Q., & Zhang, J. (2020). Bag-of-Concepts representation for document classification based on automatic knowledge acquisition from probabilistic knowledge base. *Knowledge-Based Systems*, 193, 105436. https://doi.org/10.1016/j.knosys.2019.105436.

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. https://doi.org/10.1109/TLT.2018.2868673.

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. https://doi.org/10.1109/TLT.2019.2916802

Mayer, R. E. (2005). *Cognitive theory of multimedia learning* (T. C. handbook of multimedia Learning, Ed.).

Onan, A. (2021). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience*, 33(23). https://doi.org/10.1002/cpe.5909.

Pardo, A., Mirriahi, N., Gašević, D., & Dawson, S. (2022). A model for learning analytics to support personalization in higher education. *In Handbook of Digital Higher Education* (pp. 26–37). Edward Elgar Publishing. https://doi.org/10.4337/9781800888494.00012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

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). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_10.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397–407. https://doi.org/10.1016/j.chb.2017.06.030.

Song, C., Wang, X.-K., Cheng, P., Wang, J., & Li, L. (2020). SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis. *Knowledge-Based Systems*, 194, 105572. https://doi.org/10.1016/j.knosys.2020.105572.