



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

Τμήμα Πληροφορικής

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ακαδ. Έτος: 2022-2023

**Εξόρυξη δεδομένων και μηχανική μάθηση και εφαρμογές στην
εκπαίδευση**

Κοτρώτσιος Χρήστος Π2017109

Τριμελής Επιτροπή

Κ. Έξαρχος Θεμιστοκλής

Κ. Δουκάκης Σπυρίδων

Κ. Βραχάτης Αριστείδης

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέπον καθηγητή μου κ. Έξαρχο Θεμιστοκλή για την καθοδήγηση που μου προσέφερε και το χρόνο που διέθεσε δίνοντάς μου χρήσιμες συμβουλές και οδηγίες για την ολοκλήρωση της πτυχιακής μου εργασίας.

Στο ίδιο πλαίσιο ευγνωμοσύνης, θα ήθελα να ευχαριστήσω όλους τους καθηγητές του τμήματος Πληροφορικής για τη συμβολή τους στην επιστημονική και τεχνολογική μου συγκρότηση στα χρόνια της φοίτησής μου στο Τμήμα.

Οφείλω επίσης ένα μεγάλο ευχαριστώ σε όλους εκείνους που συνέβαλαν είτε πρακτικά είτε ψυχικά στην ολοκλήρωση της εργασίας μου.

Τέλος, ένα μεγάλο ευχαριστώ στους γονείς μου για την οικονομική τους υποστήριξη, καθώς και τους συγγενείς και τους φίλους για την ηθική υποστήριξη σε όλο το διάστημα των σπουδών μου.

Περίληψη

Η εξόρυξη δεδομένων και η μηχανική μάθηση είναι δύο στενά συνδεδεμένα πεδία που ασχολούνται με την εξαγωγή χρήσιμων πληροφοριών από μεγάλα σύνολα δεδομένων. Για τον λόγο αυτό τις τελευταίες δεκαετίες ερευνάται η χρησιμότητά των εφαρμογών τους πάνω στην εκπαίδευση, καθώς μπορεί να παρουσιαστούν πολύ χρήσιμες για την επίδοση των φοιτητών, την καλύτερη διεξαγωγή του μαθήματος, την αποτροπή της αποτυχίας των φοιτητών και για την διατήρηση τους μέσα στα εκπαιδευτικά ιδρύματα, για την καλύτερη αξιολόγησή τους, αλλά και για την καλύτερη αξιολόγηση των εκπαιδευτικών καθώς και για πολλούς άλλους λόγους. Έτσι στην συγκεκριμένη εργασία, γίνεται ανάλυση 16 σχετικών άρθρων με εφαρμογές στην εκπαίδευση, ώστε να γίνει κατανοητή η χρησιμότητά τους.

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες	3
Περίληψη	4
ΚΕΦΑΛΑΙΟ 1	7
Εισαγωγή	7
ΚΕΦΑΛΑΙΟ 2	9
Data Mining – Εξόρυξη Δεδομένων	9
Τί είναι το Data mining – Εξόρυξη Δεδομένων;	9
Machine Learning – Μηχανική Μάθηση	9
Τί είναι το Machine Learning – Μηχανική Μάθηση	9
Πώς λειτουργεί το Machine Learning;	9
Educational Data Mining	11
Τί είναι το Educational Data mining	12
Μηχανική Μάθηση στην εκπαίδευση	14
Learning Analytics	14
ΚΕΦΑΛΑΙΟ 3	16
Ανάλυση Βιβλιογραφίας	16
Data-mining techniques in detecting factors linked to academic achievement	18
Predicting drop-out from social behaviour of students	20
Students' LMS interaction patterns and their relationship with achievement: A case study in higher education	22
Mining Skills from Educational Data for Project Recommendations	24
Record of Undergraduates' Activities in Virtual Learning Enviroments	26
Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data	28
Three Interaction patterns on asynchronous online discussion behaviours: A methodological comparison	30
A Framework for Capturing Distinguishing User Interaction Behaviours in Novel Interfaces	32

Mining sentiments in SMS texts for teaching evaluation	34
Investigating student motivation in the context of a learning analytics intervention during a summer bridge program	36
The Q-matrix Method: Mining Student Response Data for Knowledge	38
Learning Log Database and Data Mining system for e-Learning	40
-On-Line Statistical Outlier Detection of irregular learning processes-	40
Patterns of User Activity in the Different Features of the Blackboard CMS across All Courses for an Academic Year at Blackboard CMS across All Courses for an Academic Year at Brigham Young University Brigham Young University	42
Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining	44
Preventing Student Dropout in Distance Learning Using Machine Learning Techniques	46
A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition	48
ΚΕΦΑΛΑΙΟ 4	50
Σύγκριση Βιβλιογραφίας	50
ΚΕΦΑΛΑΙΟ 5	56
Συμπεράσματα	56
Βιβλιογραφία	57

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Η εξόρυξη δεδομένων και η μηχανική μάθηση είναι δύο στενά συνδεδεμένα πεδία που ασχολούνται με την εξαγωγή χρήσιμων πληροφοριών από μεγάλα σύνολα δεδομένων. Η εξόρυξη δεδομένων περιλαμβάνει τη χρήση αλγορίθμων και στατιστικών μοντέλων για τον εντοπισμό προτύπων και τάσεων στα δεδομένα, ενώ η μηχανική μάθηση περιλαμβάνει την ανάπτυξη αλγορίθμων και μοντέλων που μπορούν να μάθουν από δεδομένα και να κάνουν προβλέψεις ή αποφάσεις με βάση αυτή τη μάθηση.

Στο πλαίσιο της σύγχρονης εκπαίδευσης, η εξόρυξη δεδομένων και η μηχανική μάθηση μπορούν να χρησιμοποιηθούν για να βοηθήσουν στη βελτίωση της αποτελεσματικότητας των εκπαιδευτικών συστημάτων και της απόδοσης μεμονωμένων μαθητών. Για παράδειγμα, η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί για την ανάλυση δεδομένων σχετικά με την απόδοση των μαθητών, τα δημογραφικά στοιχεία και τις μαθησιακές συνήθειες, προκειμένου να εντοπιστούν παράγοντες που σχετίζονται με την ακαδημαϊκή επιτυχία. Αυτές οι πληροφορίες μπορούν στη συνέχεια να χρησιμοποιηθούν για την ανάπτυξη εξατομικευμένων μαθησιακών σχεδίων για μεμονωμένους μαθητές ή για τον εντοπισμό περιοχών όπου το πρόγραμμα σπουδών ή οι μέθοδοι διδασκαλίας μπορεί να χρειαστεί να προσαρμοστούν προκειμένου να βελτιωθούν τα αποτελέσματα των μαθητών.

Η μηχανική μάθηση, από την άλλη πλευρά, μπορεί να χρησιμοποιηθεί για τη δημιουργία μοντέλων που μπορούν να προβλέψουν την απόδοση των μαθητών σε τεστ ή άλλες αξιολογήσεις με βάση τις προηγούμενες επιδόσεις τους και άλλους παράγοντες. Αυτά τα μοντέλα μπορούν στη συνέχεια να χρησιμοποιηθούν για την παροχή εξατομικευμένης ανατροφοδότησης σε μαθητές και δασκάλους, βοηθώντας στον εντοπισμό περιοχών όπου οι μαθητές μπορεί να χρειάζονται πρόσθετη υποστήριξη ή καθοδήγηση προκειμένου να επιτύχουν.

Μια άλλη σημαντική εφαρμογή της εξόρυξης δεδομένων και της μηχανικής μάθησης στην εκπαίδευση είναι η ανάπτυξη συστημάτων προσαρμοστικής μάθησης. Αυτά τα συστήματα χρησιμοποιούν αλγόριθμους μηχανικής μάθησης για να παρακολουθούν συνεχώς την πρόοδο του μαθητή και να προσαρμόζουν τη δυσκολία του υλικού που παρουσιάζεται με βάση την απόδοση του μαθητή. Αυτό μπορεί να βοηθήσει να κρατηθούν οι μαθητές αφοσιωμένοι και παρακινημένοι, και μπορεί επίσης να βοηθήσει να διασφαλιστεί ότι εργάζονται πάντα σε κατάλληλο επίπεδο πρόκλησης.

Συνολικά, η χρήση της εξόρυξης δεδομένων και της μηχανικής μάθησης στην εκπαίδευση μπορεί να βοηθήσει στη βελτίωση των αποτελεσμάτων των μαθητών παρέχοντας εξατομικευμένη ανατροφοδότηση και υποστήριξη και βοηθώντας στον εντοπισμό και την

αντιμετώπιση τυχόν αδυναμιών ή κενών στο εκπαιδευτικό σύστημα. Αυτές οι τεχνολογίες έχουν τη δυνατότητα να φέρουν επανάσταση στον τρόπο παροχής της εκπαίδευσης και χρησιμοποιούνται ήδη από πολλά σχολεία και πανεπιστήμια για να βοηθήσουν τους μαθητές να επιτύχουν.

Σκοπός της συγκεκριμένης εργασίας είναι ο σχολιασμός και η ανάλυση προηγούμενης βιβλιογραφίας με εφαρμογές εξόρυξης δεδομένων και μηχανικής μάθησης πάνω στην διαδικασία της εκπαίδευσης, ώστε να εξεταστεί ο αντίκτυπος των μεθόδων στις αποδόσεις των μαθητών-φοιτητών.

Αρχικά, αφού γίνει αναλυτική περιγραφή των όρων που χρησιμοποιούνται στην παρούσα εργασία, ώστε να είναι κατανοητή για τον αναγνώστη, γίνεται μια αναλυτική περιγραφή σε παρόμοια προηγούμενα άρθρα που και αυτά αναλύουν την σημασία της εξόρυξης δεδομένων στην εκπαίδευση και στη συνέχεια γίνεται ανάλυση 16 άρθρων που αφορούν την εξόρυξη δεδομένων αλλά και την μηχανική και εφαρμογές στην εκπαίδευση. Στη συνέχεια, γίνεται σύγκριση των τεχνικών εξόρυξης δεδομένων που χρησιμοποιήθηκαν στην εργασία αυτή, με την χρήση πίνακα. Τέλος, γίνεται εξαγωγή συμπερασμάτων.

ΚΕΦΑΛΑΙΟ 2

Data Mining – Εξόρυξη Δεδομένων

Τί είναι το Data mining – Εξόρυξη Δεδομένων;

Η εξόρυξη δεδομένων είναι η διαδικασία ταξινόμησης μέσω μεγάλων συνόλων δεδομένων για τον εντοπισμό προτύπων και σχέσεων που μπορούν να βοηθήσουν στην επίλυση προβλημάτων μέσω της ανάλυσης δεδομένων. Οι τεχνικές και τα εργαλεία εξόρυξης δεδομένων επιτρέπουν στους οργανισμούς να προβλέπουν τις μελλοντικές τάσεις και να λαμβάνουν πιο ενημερωμένες αποφάσεις.

Η εξόρυξη δεδομένων είναι ένα βασικό μέρος της ανάλυσης δεδομένων συνολικά και ένας από τους βασικούς κλάδους της επιστήμης δεδομένων, η οποία χρησιμοποιεί προηγμένες τεχνικές ανάλυσης για να βρει χρήσιμες πληροφορίες σε σύνολα δεδομένων. Σε πιο αναλυτικό επίπεδο, η εξόρυξη δεδομένων είναι ένα βήμα στη διαδικασία ανακάλυψης γνώσης σε βάσεις δεδομένων (KDD – Knowledge Discovery in Databases), μια μεθοδολογία επιστήμης δεδομένων για τη συλλογή, την επεξεργασία και την ανάλυση δεδομένων. Η εξόρυξη δεδομένων και το KDD αναφέρονται μερικές φορές εναλλακτικά, αλλά συνήθως θεωρούνται διαφορετικά πράγματα.

Machine Learning – Μηχανική Μάθηση

Τί είναι το Machine Learning – Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης (AI) και της επιστήμης των υπολογιστών που εστιάζει στη χρήση δεδομένων και αλγορίθμων για τη μίμηση του τρόπου με τον οποίο μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβειά της.

Πώς λειτουργεί το Machine Learning;

Παρόμοια με τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος αποκτά γνώση και κατανόηση, η μηχανική μάθηση βασίζεται σε δεδομένα, όπως δεδομένα εκπαίδευσης ή γραφήματα γνώσης, για να κατανοήσει οντότητες, τομείς και τις συνδέσεις μεταξύ τους. Με καθορισμένες οντότητες, μπορεί να ξεκινήσει η βαθιά μάθηση.

Η διαδικασία μηχανικής μάθησης ξεκινά με παρατηρήσεις ή δεδομένα, όπως παραδείγματα, άμεση εμπειρία ή οδηγίες. Αναζητά μοτίβα στα δεδομένα για να μπορεί αργότερα να βγάλει συμπεράσματα με βάση τα παραδείγματα που παρέχονται. Ο πρωταρχικός στόχος του

Machine Learning (ML) είναι να επιτρέψει στους υπολογιστές να μαθαίνουν αυτόνομα χωρίς ανθρώπινη παρέμβαση ή βοήθεια και να προσαρμόσουν τις ενέργειες ανάλογα.

Η ML έχει αποδειχθεί πολύτιμη επειδή μπορεί να λύσει προβλήματα σε ταχύτητα και κλίμακα που δεν μπορούν να αντιγραφούν μόνο από το ανθρώπινο μυαλό. Με τεράστια ποσά υπολογιστικής ικανότητας πίσω από μια μεμονωμένη εργασία ή πολλαπλές συγκεκριμένες εργασίες, οι μηχανές μπορούν να εκπαιδευτούν ώστε να αναγνωρίζουν μοτίβα και σχέσεις μεταξύ δεδομένων εισόδου και να αυτοματοποιούν τις διαδικασίες ρουτίνας.

Μέθοδοι Μηχανικής Μάθησης

Η μηχανική μάθηση προσφέρει σαφή οφέλη για τις τεχνολογίες τεχνητής νοημοσύνης. Ωστόσο υπάρχουν διάφορες προσεγγίσεις μηχανικής μάθησης για κάθε περίπτωση. Συγκεκριμένα υπάρχουν οι:

- Εποπτευόμενη μάθηση (Supervised learning)
- Μη Εποπτευόμενη μάθηση (Unsupervised learning)
- Ενισχυτική μάθηση (Reinforcement learning)

Πιο αναλυτικά, ο κύριος στόχος της Εποπτευόμενης Μάθησης είναι η χρήση επισημασμένων δεδομένων που παρέχονται στο σύστημα. Το μοντέλο εκπαιδεύεται να αναλύει τις λαμβανόμενες πληροφορίες μέχρι να μπορέσει να ταξινομήσει/προβλέψει το αποτέλεσμα με ακρίβεια.

Η Μη Εποπτευόμενη Μάθηση είναι ακριβώς το αντίθετο από την εποπτευόμενη μάθηση. Το μοντέλο λαμβάνει σημαντικό αριθμό μη καταχωρημένων δεδομένων και οργάνων που το βοηθούν να κατανοήσει τις ιδιότητες των πληροφοριών. Ως αποτέλεσμα, το σύστημα υιοθετεί ομαδοποίηση και συσχέτιση για να πλοηγηθεί στα δεδομένα και να τα οργανώσει με κατανοητό από τον άνθρωπο τρόπο.

Η προσέγγιση της Ενισχυτικής Μάθησης είναι τελείως διαφορετική από την εποπτευόμενη/μη εποπτευόμενη μάθηση. Η ενίσχυση της μηχανικής μάθησης χρησιμοποιεί την αλληλεπίδραση του συστήματος με το περιβάλλον συν ορισμένες ανταμοιβές για να τονώσει τις επιθυμητές ενέργειες.



Εικόνα 2.1 : Διάφοροι τύποι Μηχανικής Μάθησης

Educational Data Mining

Διάφορες έρευνες διεξάγονται αυτές τις μέρες στον τομέα της εξόρυξης δεδομένων. Το Educational Data Mining είναι ένα σημαντικό ερευνητικό πεδίο που ονομάζεται EDM, που χρησιμοποιεί πολλαπλούς αλγόριθμους για τη βελτίωση των εκπαιδευτικών αποτελεσμάτων και την επεξήγηση των εκπαιδευτικών διαδικασιών για περαιτέρω λήψη αποφάσεων.

Στην ενότητα αυτή, θα σχολιάσουμε μερικούς από τους αλγόριθμους εξόρυξης δεδομένων που εφαρμόζονται στον τομέα που σχετίζεται με την εκπαίδευση. Αυτοί οι αλγόριθμοι χρησιμοποιούνται για την εξόρυξη γνώσεων από εκπαιδευτικά δεδομένα και τη μελέτη των χαρακτηριστικών που μπορούν να προσθέσουν για να ενισχύσουν την απόδοση. Η μάθηση ξεκίνησε αρχικά στην τάξη και βασιζόταν σε συμπεριφορικά, ψυχολογικά και εποικοδομητικά μοντέλα. Τα μοντέλα συμπεριφοράς εξαρτώνται από παρατηρήσιμες αλλαγές στη συμπεριφορά του μαθητή για τον προσδιορισμό των μαθησιακών αποτελεσμάτων. Τα ψυχολογικά μοντέλα βασίζονται στην αποτελεσματική εμπλοκή του εκπαιδευτικού στη μαθησιακή διαδικασία. Στα εποικοδομητικά μοντέλα, οι μαθητές πρέπει να μάθουν από τις δικές τους από τις διάφορες διαθέσιμες πηγές.

Τί είναι το Educational Data mining

Το Educational Data Mining ή αλλιώς EDM, Μπορεί να οριστεί ως η τεχνική για την εύρεση των συγκεκριμένων τύπων δεδομένων που προέρχονται από το εκπαιδευτικό σύστημα και την εφαρμογή αυτών των τεχνικών για την καλύτερη κατανόηση των μαθητών και του συστήματος. Το EDM είναι η διαδικασία μετατροπής ακατέργαστων δεδομένων που λαμβάνονται από εκπαιδευτικά συστήματα, σε χρήσιμα δεδομένα που μπορούν να χρησιμοποιηθούν για τη λήψη αποφάσεων που βασίζονται σε δεδομένα. Η ανάπτυξη της εξόρυξης δεδομένων και της ανάλυσης στον τομέα της εκπαίδευσης ήταν σχετικά καθυστερημένη σε σύγκριση με άλλους τομείς. Ωστόσο, η εξόρυξη εκπαιδευτικών δεδομένων μέσω του Διαδικτύου είναι δύσκολη, λόγω των ειδικών χαρακτηριστικών του στα δεδομένα. Ενώ πολλοί τύποι δεδομένων έχουν διαδοχικές πτυχές, η διανομή εκπαιδευτικών πληροφοριών με την πάροδο του χρόνου έχει εξαιρετικά χαρακτηριστικά.

Οι τεχνικές EDM υπήρξαν καρποφόρες και δημιούργησαν μια σειρά φαινομένων σχετικά με τη μάθηση των μαθητών σε ηλεκτρονικές πλατφόρμες και την επίτευξη καλύτερης ακρίβειας με συνέπεια. Υπάρχουν σημαντικές πτυχές που πρέπει να εξεταστούν για να δικαιολογηθεί η εξαιρετική πρόοδος για τα εκπαιδευτικά δεδομένα, η οποία αναπτύσσει την αναγνώριση ότι δεν αποθηκεύονται όλα τα βασικά δεδομένα σε μία ροή δεδομένων.

Η έρευνα στην εκπαίδευση οδήγησε σε μερικές νέες εκπαιδευτικές βελτιώσεις. Οι καινοτομίες που βασίζονται σε υπολογιστές έχουν αλλάξει τον τρόπο που ζούμε και μαθαίνουμε. Σήμερα, η χρήση των πληροφοριών που συλλέγονται μέσω αυτών των εξελίξεων υποστηρίζει έναν δεύτερο γύρο μετασχηματισμού σε όλους τους τομείς και μάθησης με διάφορα επιτεύγματα. Η εξόρυξη δεδομένων είναι η πιο ισχυρή τεχνική με απίστευτες δυνατότητες για να βοηθήσει τα σχολεία και τα πανεπιστήμια να επικεντρωθούν στις πιο σημαντικές πληροφορίες στα σύνολα δεδομένων που έχουν συλλέξει σχετικά με τη συμπεριφορά των μαθητών και τους πιθανούς μαθητές.

Η εξόρυξη δεδομένων χρησιμοποιεί τα εργαλεία της για να βρει προηγουμένως άγνωστα μοτίβα και σχέσεις σε τεράστια σύνολα δεδομένων. Αυτά τα εργαλεία μπορούν να ενσωματώσουν στατιστικά μοντέλα, μαθηματικούς αλγόριθμους και τεχνικές μηχανικής μάθησης. Αυτές οι τεχνικές μπορούν να βρουν τις πληροφορίες μέσα στα δεδομένα που τα ερωτήματα και οι αναφορές δεν μπορούν να αποκαλύψουν αποτελεσματικά.

Συνοπτικά, όπως αναφέρουν και οι Romero & Ventura, 2007 [1], η εκπαιδευτική εξόρυξη δεδομένων είναι ένας ανερχόμενος τομέας που σχετίζεται με πολλούς καθιερωμένους τομείς έρευνας, όπως η ηλεκτρονική μάθηση, τα προσαρμοστικά υπερμέσα, τα ευφυή συστήματα διδασκαλίας, η εξόρυξη ιστού, η εξόρυξη δεδομένων κ.λπ. Η εφαρμογή της εξόρυξης δεδομένων στα εκπαιδευτικά συστήματα έχει συγκεκριμένες απαιτήσεις που δεν υπάρχουν στους άλλους τομείς, κυρίως την ανάγκη να λαμβάνονται υπόψη οι παιδαγωγικές πτυχές του μαθητή και του συστήματος.

Γενικά, υπάρχουν πολλές τεχνικές εφαρμογής εξόρυξης δεδομένων στην εκπαίδευση, ωστόσο, στη συγκεκριμένη εργασία εξηγώ και αναφέρομαι στις εξής:

Classification

Το Classification είναι μια συνάρτηση εξόρυξης δεδομένων που εκχωρεί στοιχεία σε μια συλλογή σε κατηγορίες ή κλάσεις-στόχους. Ο στόχος της ταξινόμησης είναι η ακριβής πρόβλεψη της κατηγορίας στόχου για κάθε περίπτωση στα δεδομένα.

Clustering

Το Clustering είναι η μέθοδος μετατροπής μιας ομάδας αφηρημένων αντικειμένων σε κλάσεις παρόμοιων αντικειμένων. Η ομαδοποίηση είναι μια μέθοδος κατάτμησης ενός συνόλου δεδομένων ή αντικειμένων σε ένα σύνολο σημαντικών υποκλάσεων που ονομάζονται συστάδες.

Visual Data Mining

Το Visual Data Mining (VDM) είναι η διαδικασία αλληλεπίδρασης και αναλυτικής συλλογιστικής με μία ή περισσότερες οπτικές αναπαραστάσεις αφηρημένων δεδομένων. Η διαδικασία μπορεί να οδηγήσει στην οπτική ανακάλυψη ισχυρών προτύπων σε αυτά τα δεδομένα ή να παρέχει κάποια καθοδήγηση για την εφαρμογή άλλων τεχνικών εξόρυξης δεδομένων και ανάλυσης. (Simoff, S.J. 2009)[2]

Statistics

Η στατιστική είναι ένα στοιχείο της εξόρυξης δεδομένων που παρέχει τα εργαλεία και τις τεχνικές ανάλυσης για την αντιμετώπιση μεγάλων ποσοτήτων δεδομένων. Είναι η επιστήμη της μάθησης από δεδομένα και περιλαμβάνει τα πάντα, από τη συλλογή και την οργάνωση έως την ανάλυση και την παρουσίαση δεδομένων.

Association Rule Mining

Το Association Rule Mining περιλαμβάνει τη χρήση μοντέλων μηχανικής μάθησης για την ανάλυση δεδομένων για μοτίβα ή συν-εμφανίσεις σε μια βάση δεδομένων. Προσδιορίζει συχνές συσχετίσεις εάν-τότε, οι οποίες είναι οι ίδιοι οι κανόνες συσχέτισης. Ένας κανόνας συσχέτισης έχει δύο μέρη: ένα προηγούμενο (αν) και ένα επακόλουθο (τότε).

Regression

Το Regression αναφέρεται σε έναν τύπο εποπτευόμενης τεχνικής μηχανικής εκμάθησης που χρησιμοποιείται για την πρόβλεψη οποιουδήποτε χαρακτηριστικού συνεχούς αξίας. Η παλινδρόμηση βοηθάει κάθε οργανισμό να αναλύσει τη μεταβλητή στόχο και τις σχέσεις μεταβλητής πρόβλεψης.

Sequential Pattern Mining

Το Sequential Pattern Mining είναι ένα θέμα της εξόρυξης δεδομένων που ασχολείται με την εύρεση στατιστικά σχετικών προτύπων μεταξύ παραδειγμάτων δεδομένων όπου οι τιμές παραδίδονται σε μια ακολουθία

Text Mining

Η τεχνική Text Mining, γνωστή και ως text data mining, είναι η διαδικασία μετατροπής μη δομημένου κειμένου σε δομημένη μορφή για τον εντοπισμό σημαντικών προτύπων και νέων γνώσεων. Εφαρμόζοντας προηγμένες αναλυτικές τεχνικές, όπως Naïve Bayes, Support Vector Machines (SVM) και άλλους αλγόριθμους βαθιάς μάθησης, οι οργανισμοί είναι σε θέση να εξερευνήσουν και να ανακαλύψουν κρυφές σχέσεις μέσα στα μη δομημένα δεδομένα τους.

Correlation

Το Correlation χρησιμοποιείται για τη μελέτη της εγγύτητας της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών, δηλαδή του βαθμού στον οποίο οι μεταβλητές συνδέονται μεταξύ τους.

Outlier Detection

Το Outlier Detection είναι η διαδικασία ανίχνευσης ακραίων τιμών ή ενός σημείου δεδομένων που απέχει πολύ από τον μέσο όρο και ανάλογα με το τι προσπαθείτε να επιτύχετε, ενδεχομένως αφαιρώντας ή επιλύοντάς τα από την ανάλυση για να αποφευχθεί οποιαδήποτε πιθανή παραμόρφωση.

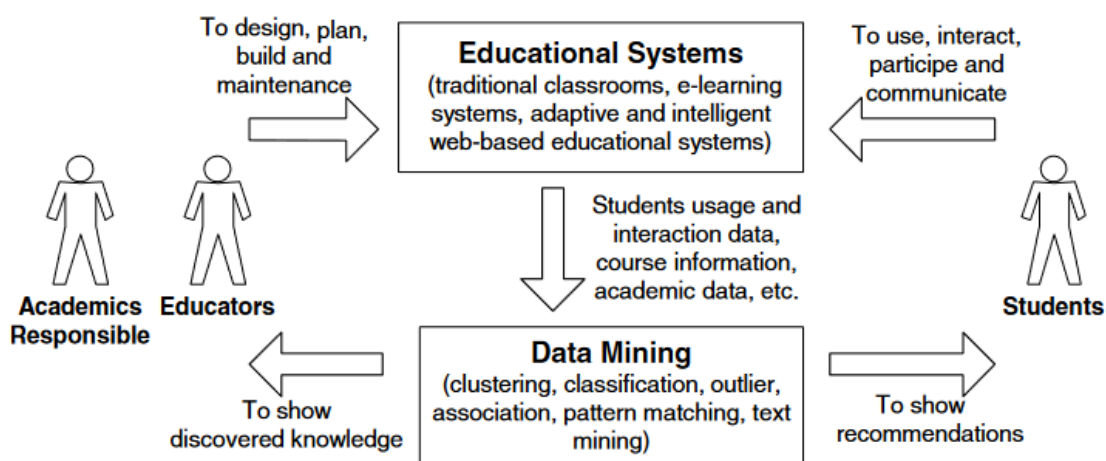
Μηχανική Μάθηση στην εκπαίδευση

Η μηχανική μάθηση είναι ένας ταχέως αναπτυσσόμενος τομέας που αλλάζει τον τρόπο που ζούμε και εργαζόμαστε. Στον τομέα της εκπαίδευσης, η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για την εξατομίκευση των μαθησιακών εμπειριών, τον εντοπισμό μαθησιακών κενών και τη βελτίωση των αποτελεσμάτων των μαθητών. Χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης, τα εκπαιδευτικά συστήματα μπορούν να αναλύσουν μεγάλο όγκο δεδομένων και να παρέχουν σε δασκάλους και μαθητές πληροφορίες και προτάσεις που μπορούν να βοηθήσουν στη βελτίωση της απόδοσης των μαθητών. Επιπλέον, η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για την αυτοματοποίηση της βαθμολόγησης και της ανατροφοδότησης, απελευθερώνοντας τους δασκάλους να επικεντρωθούν σε πιο σημαντικά καθήκοντα, όπως η εξατομικευμένη υποστήριξη στους μαθητές. Συνολικά, η χρήση της μηχανικής μάθησης στην εκπαίδευση έχει τη δυνατότητα να βελτιώσει σημαντικά την αποτελεσματικότητα του εκπαιδευτικού συστήματος και να βοηθήσει τους μαθητές να αξιοποιήσουν πλήρως τις δυνατότητές τους.

Learning Analytics

Τα αναλυτικά στοιχεία μάθησης είναι η χρήση δεδομένων, στατιστικών και ποσοτικών μεθόδων και εργαλείων οπτικοποίησης για την κατανόηση και τη βελτίωση της μάθησης και των περιβαλλόντων στα οποία εμφανίζεται. Αυτό το πεδίο περιλαμβάνει ένα ευρύ φάσμα

δραστηριοτήτων, συμπεριλαμβανομένης της συλλογής, ανάλυσης και ερμηνείας δεδομένων από διάφορες πηγές, όπως εκπαιδευτικά συστήματα, μέσα κοινωνικής δικτύωσης και αρχεία απόδοσης μαθητών. Τα αναλυτικά στοιχεία μάθησης μπορούν να χρησιμοποιηθούν για τον εντοπισμό τάσεων και προτύπων στη συμπεριφορά και τη μάθηση των μαθητών, παρέχοντας πληροφορίες που μπορούν να βοηθήσουν τους εκπαιδευτικούς να σχεδιάσουν πιο αποτελεσματικές στρατηγικές διδασκαλίας και μαθησιακά περιβάλλοντα. Επιπλέον, τα αναλυτικά στοιχεία μάθησης μπορούν να χρησιμοποιηθούν για την αξιολόγηση της αποτελεσματικότητας των εκπαιδευτικών παρεμβάσεων και προγραμμάτων, επιτρέποντας στους εκπαιδευτικούς να λαμβάνουν αποφάσεις βάσει δεδομένων σχετικά με τον τρόπο κατανομής πόρων και υποστήριξης των μαθητών. Συνολικά, τα αναλυτικά στοιχεία μάθησης μπορούν να βοηθήσουν στη βελτίωση της μάθησης και των επιδόσεων των μαθητών παρέχοντας στους εκπαιδευτικούς μια βαθιά κατανόηση των παραγόντων που επηρεάζουν την απόδοση των μαθητών.



Εικόνα 2.2 : Ο κύκλος εφαρμογής εξόρυξης δεδομένων σε εκπαιδευτικά συστήματα *Romero & Ventura, 2007 [1]*

ΚΕΦΑΛΑΙΟ 3

Ανάλυση Βιβλιογραφίας

Στο συγκεκριμένο κεφάλαιο θα γίνει αναλυτική περιγραφή 16 άρθρων που αφορούν τις εφαρμογές της εξόρυξης δεδομένων και της μηχανικής μάθησης στην εκπαίδευση, ώστε να κατανοήσουμε την χρησιμότητά τους στη σύγχρονη εκπαίδευση.

Προτού όμως γίνει ανάλυση των άρθρων, οφείλουμε να αναφερθούμε σε παρόμοιες εργασίες που έχουν γίνει στο παρελθόν και αφορούν την ανάλυση άρθρων με σκοπό την καλύτερη κατανόηση του θέματος και την παρουσίαση διαφόρων εφαρμογών πάνω στην εκπαίδευση.

Οι Romero και Ventura (2007)[1] παρουσιάζουν μια ανασκόπηση 81 έργων. Προσδιορίζουν την στατιστική-οπτικοποίηση και την εξόρυξη ιστού ως δύο τεχνικές DM για την ταξινόμηση της εφαρμογής του DM στο CBES (Computer Based Education Systems).

Όσον αφορά τις στατιστικές, προσδιορίζονται διάφορα εργαλεία και αναφέρονται επτά έργα EDM. Σχετικά με την οπτικοποίηση, τέσσερα έργα αναφέρονται και ένα εργαλείο αναγνωρίζεται. Όσον αφορά το web mining, χωρίζεται σε τρία είδη εργασιών:

1. Clustering, classification, και outlier detection
2. association rule και sequential pattern
3. text mining

Στα clustering, classification, and outlier detection αντιστοιχούν 15 εργασίες, στα association rule και sequential pattern αντιστοιχούν 14 εργασίες και στο text mining αντιστοιχούν 7 εργασίες.

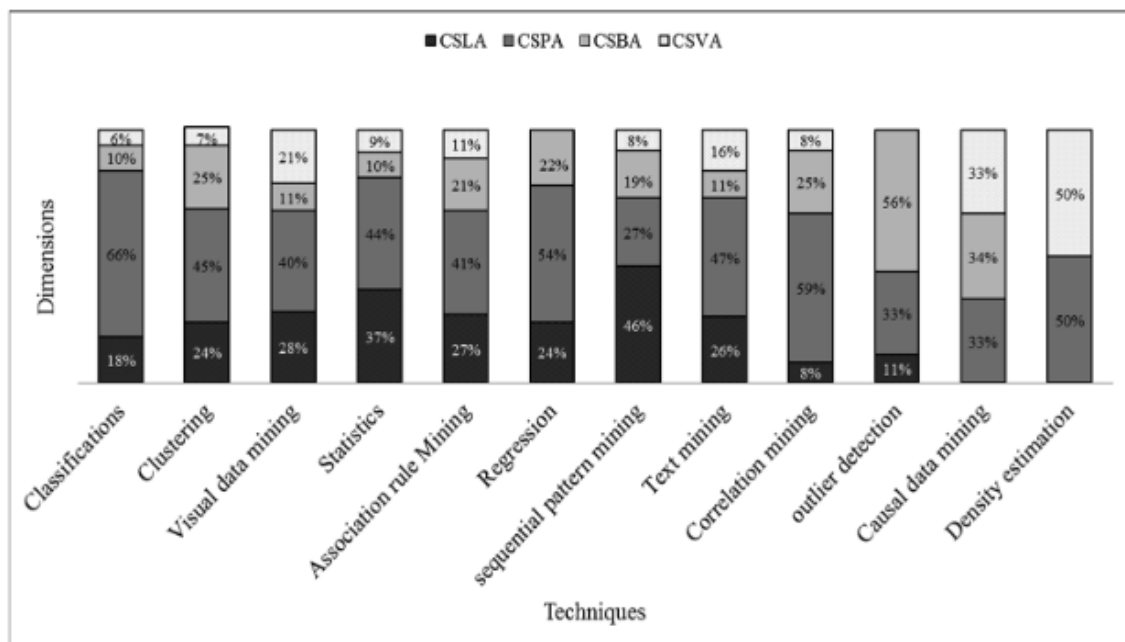
Στη συνέχεια οι Romero και Ventura(2010)[3] παρουσιάζουν μία ακόμη ανασκόπηση, βασισμένη στην προηγούμενή τους, προσθέτοντας επιπλέον 225 εργασίες.

Στη συνέχεια οι Baker και Yacef (2009)[4] παρουσιάζουν 45 εργασίες αναφοράς EDM, επαινούν την αναπτυσσόμενη ερευνητική κοινότητα EDM και ορίζουν το DM και το EDM. Η ανασκόπηση τους προσδιορίζει ορισμένους στόχους EDM, συμπεριλαμβανομένων των μοντέλων μαθητών, των μοντέλων γνώσης τομέα, της παιδαγωγικής βοήθειας και των επιπτώσεων στη μάθηση. Παρατίθενται 4, 3 και 4 σχετικές εργασίες, αντίστοιχα.

Οι Peña-Ayala, Domínguez, και Medel (2009)[5] παρέχουν 91 αναφορές για τα τρία θέματα CBES, DM και EDM. Για το πρώτο, λαμβάνονται υπόψη στρατηγικές όπως η εκπαίδευση με τη βοήθεια υπολογιστή, τα ευφυή συστήματα διδασκαλίας (ITS), τα συστήματα διαχείρισης μάθησης (LMS) και τα εκπαιδευτικά συστήματα που βασίζονται στο Διαδίκτυο (WBES). Αναγνωρίζονται πολυάριθμα μοντέλα, εργασίες και στρατηγικές για το DM, με το επίκεντρο της έρευνας να εστιάζεται σε μαθηματικές, βασισμένες σε κανόνες και τεχνικές μαλακών

υπολογισμών. Τα τέσσερα χαρακτηριστικά των έργων EDM είναι η μοντελοποίηση μαθητών, η διδασκαλία, το περιεχόμενο και η αξιολόγηση.

Σε μια πιο πρόσφατη ανάλυση, οι Aldowah, Al-Samarraie και Fauzy (2019)[6] , αναλύουν 402 άρθρα που αφορούν εφαρμογές EDM και Learning Analytics στην τριτοβάθμια εκπαίδευση τα οποία αρχικά ταξινομούνται σε 4 διαστάσεις, computer-supported learning analytics (CSLA), computer-supported predictive analytics (CSPA), computer-supported behavioral analytics (CSBA), και computer-supported visualization analytics (CSVA) και στη συνέχεια τα αντιστοιχισαν τις ακόλουθες τεχνικές data mining: classification (26.25%), clustering (21.25%), visual data mining (15%), statistics (14.25%), association rule mining (14%), regression (10.25%), sequential pattern mining (6.50%), text mining (4.75%), correlation mining (3%), outlier detection (2.25%), causal mining (1%), και density estimation (1%). Τα ποσοστά αντιστοιχούν στο πλήθος των άρθρων για την κάθε τεχνική.



Εικόνα 3.1 : Η χρήση των τεχνικών εξόρυξης δεδομένων ταξινομημένες στις 4 διαστάσεις Aldowah, Al-Samarraie και Fauzy (2019) [6]

Ακολουθεί η ανάλυση των 16 άρθρων της συγκεκριμένης εργασίας.

Άρθρο 1

Data-mining techniques in detecting factors linked to academic achievement

Fernando Martínez Abad & Alicia A. Chaparro Caso López

Μέσα απο το συγκεκριμένο άρθρο οι συγγραφείς παρουσιάζουν μια διαδικασία που χρησιμοποιείται για τον εντοπισμό παραγόντων που συνδέονται με την ακαδημαϊκή επίδοση σε αξιολογήσεις μεγάλης κλίμακας μέσω της εμφάνισης τεχνικών στατιστικής ανάλυσης που βασίζονται σε data mining στις επιστήμες της εκπαίδευσης και τη δυνατότητα που προσφέρουν για τον εντοπισμό μη ασήμαντων πληροφοριών σε μεγάλες βάσεις δεδομένων. Η μελέτη βασίζεται σε μη πειραματική, συγχρονική μελέτη με δείγμα 18.935 μαθητών Λυκείου από 99 εκπαιδευτικά ιδρύματα στην πολιτεία Baja California (Μεξικό). Οι πληροφορίες λήφθηκαν από το τεστ ENLACE που πρόκειται για απογραφικό τεστ που εφαρμόζεται σε όλα τα εκπαιδευτικά ιδρύματα στο Μεξικό, απο το δημοτικό μέχρι το λύκειο, καθώς και από context surveys που δόθηκαν σε μαθητές στη Baja California. Εφόσον γίνεται εφαρμογή μη πειραματικής μελέτης, ο προηγούμενος έλεγχος των μεταβλητών κατά τη συλλογή πληροφοριών δεν έχει κανένα ενδιαφέρον, καθώς πραγματοποιείται στο φυσικό περιβάλλον των υποκειμένων, ως προς τον καθορισμό της σχέσης μεταξύ των μεταβλητών μετά. Για την εφαρμογή τεχνικών ταξινόμησης χρησιμοποιήθηκαν τα δέντρα αποφάσεων, και τα αποτελέσματα δείχνουν ότι οι προσωπικοί παράγοντες είναι πιο ενδεικτικοί στην ακαδημαϊκή επίδοση, ακολουθούμενοι από σχολικούς και κοινωνικούς παράγοντες. Με βάση τον κύριο στόχο της μελέτης, τα αποτελέσματα που παρουσιάζονται καθιστούν δυνατό να δηλωθεί ότι οι τεχνικές που χρησιμοποιούνται είναι ένα «νέο» εργαλείο για την ανάλυση παραγόντων που σχετίζονται με την ακαδημαϊκή επίδοση.

Αν και έχει επιβεβαιωθεί ότι υπάρχει μεγάλος αριθμός τεχνικών ανάλυσης στην ανίχνευση παραγόντων που σχετίζονται με την ακαδημαϊκή επίδοση, οι τεχνικές Data Mining έχουν ελάχιστα διερευνηθεί. Σε αυτό το άρθρο, η χρήση αλγορίθμων classification που βασίζονται σε δέντρα αποφάσεων αποδείχθηκε χρήσιμη στον εντοπισμό παραγόντων που μπόρεσαν να εξηγήσουν την επίδοση των μαθητών. Αν και οι παράγοντες που προσδιορίζονται στον αλγόριθμο δεν είναι νέοι, ο τρόπος με τον οποίο είχαν διαμορφωθεί από το DM θεωρείται η κύρια συμβολή αυτής της μελέτης.

Δεδομένης της μεγάλης ποσότητας δεδομένων που δημιουργήθηκε, δεν ήταν δυνατό να εξαχθεί ένα διαχειρίσιμο δέντρο για τους παράγοντες για τους μαθητές. Ωστόσο, οι στρατηγικές μάθησης, αυτοεκτίμηση, χρήση ναρκωτικών, συνύπαρξη-βίας και πόροι στο σπίτι βρέθηκαν να είναι οι πιο σημαντικοί παράγοντες.

Όσον αφορά τα σχολεία, διαπιστώθηκε ότι ο σημαντικότερος παράγοντας ήταν η χρήση ναρκωτικών, που συμπεριλαμβάνει την μέση αντίληψη της βαρύτητας της χρήσης ναρκωτικών (συνολική κλίμακα με αρνητική έννοια) και την μέση αντίληψη της πραγματικής χρήσης ναρκωτικών στο σχολείο όπως αναφέρουν οι μαθητές. Φαίνεται, από τα αποτελέσματα ότι αυτός ο παράγοντας είναι κεντρικής σημασίας και θα πρέπει να

συμπεριλαμβάνεται μεταξύ των κυρίων προσωπικών, σχολικών και κοινωνικών παραγόντων που συνδέονται με την ακαδημαϊκή επίδοση.

Παράλληλα με τη χρήση ναρκωτικών, άλλοι παράγοντες που προωθούνται σε σχολεία με μέτρια, υψηλά ή άριστα επίπεδα ακαδημαϊκών επιδόσεων είναι η απόδοση υψηλής αξία στην ανάγνωση, στην αυτοεκτίμηση, στην υγεία και διατροφή και στις στρατηγικές μάθησης.

Συμπερασματικά, οι παράγοντες σε αυτή τη μελέτη είναι γενικά οι ίδιοι με αυτούς βρέθηκαν σε άλλες μελέτες. Έτσι, επιβεβαιώνεται ότι οι προσωπικοί παράγοντες έχουν μεγαλύτερη επιρροή από τους κοινωνικούς, οικογενειακούς ή σχολικούς παράγοντες όσον αφορά την ακαδημαϊκή επίδοση. Με δεδομένη τη στενή σχέση μεταξύ των αποτελεσμάτων που ελήφθησαν σε αυτή τη μελέτη και εκείνων που βρέθηκαν σε άλλες μελέτες που χρησιμοποιούν διαφορετικές στατιστικές τεχνικές, πρέπει να επισημανθεί ότι η κύρια συμβολή αυτής της μελέτης είναι ο εντοπισμός νέων τεχνικών για τη μελέτη παραγόντων που σχετίζονται με την ακαδημαϊκή επίδοση που βελτιστοποιεί τη διαδικασία με τη χρήση μεγάλων βάσεων δεδομένων, όπως συμβαίνει με μεγάλης κλίμακας αξιολογήσεις.

Άρθρο 2

Predicting drop-out from social behaviour of students

Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T., Popelinsky, L.

Στο συγκεκριμένο άρθρο οι συγγραφείς εστιάζουν στην πρόβλεψη της εγκατάλειψης των σπουδών και της αποτυχίας όταν τα δεδομένα των φοιτητών έχουν εμπλουτιστεί με δεδομένα που προέρχονται από την κοινωνική συμπεριφορά των φοιτητών. Έτσι, αυτά τα δεδομένα περιγράφουν κοινωνικές εξαρτήσεις που συγκεντρώθηκαν από e-mail και απο συνομιλίες σε πίνακες συζητήσεων, μεταξύ άλλων πηγών. Περιγράφουν ένα απόσπασμα από νέα χαρακτηριστικά, τόσο από τα δεδομένα φοιτητών όσο και από δεδομένα που αφορούν την συμπεριφορά, που αντιπροσωπεύονται από ένα κοινωνικό γράφημα το οποίο κατασκευάζουν οι συγγραφείς. Έπειτα εισάγουν μια νέα μέθοδο για την εκμάθηση ενός ταξινομητή για τη μελέτη πρόβλεψης αποτυχίας φοιτητών που χρησιμοποιεί μάθηση ευαίσθητη στο κόστος για να μειώσουν τον αριθμό των εσφαλμένα ταξινομημένων ανεπιτυχών φοιτητών. Δείχνουν ότι η χρήση δεδομένων κοινωνικής συμπεριφοράς οδηγεί σε σημαντική αύξηση της ακρίβειας πρόβλεψης.

Οι συγγραφείς χρησιμοποιούν δεδομένα μαθητών που έχουν αποθηκευτεί στο Πληροφοριακό Σύστημα του Πανεπιστημίου Masaryk (IS MU), που αποθηκεύει εκπαιδευτικά δεδομένα και περιλαμβάνει όλες τις πληροφορίες για τους φοιτητές και τις σπουδές τους, για τους δασκάλους και τα μαθήματα και παρέχει επίσης εργαλεία διαχείρισης εξετάσεων, αξιολόγηση διαδικτυακών δοκιμών και διάφορες μορφές επικοινωνίας, π.χ. πίνακες συζητήσεων. Χρησιμοποιούν μόνο ένα υποσύνολο πληροφοριών που είναι αποθηκευμένες στο IS MU που είναι σχετικό για την πρόβλεψη της επιτυχίας του φοιτητή, όπως βαθμολογίες σε τεστ ικανότητας για μελέτη, κερδισμένες μονάδες, μέσος όρος βαθμών, ή το φύλο. Το IS MU αποθηκεύει επίσης το πλήρες ιστορικό των αιτημάτων των χρηστών προς το σύστημα. Δεδομένα για την κοινωνική συμπεριφορά των μαθητών, όπως την συχνότητα διαπροσωπικής επικοινωνίας ή τον αριθμό των αμοιβαία κοινών αρχείων, μπορούν να παρατηρηθούν και να αποθηκευτούν είτε αμέσως, όταν χρησιμοποιείται η συγκεκριμένη λειτουργία συστήματος, ή αργότερα από το πλήρες ιστορικό των αιτημάτων των χρηστών που υπάρχει με τη μορφή του αρχείου καταγραφής πρόσβασης συστήματος. Η συγκεκριμένη έρευνα εξετάζει δεδομένα που αφορούν φοιτητές Εφαρμοσμένης Πληροφορικής που έγιναν δεκτοί στη σχολή Πληροφορικής του πανεπιστημίου Masaryk κατά τα έτη 2006, 2007 και 2008 που αντιστοιχεί συνολικά σε 775 φοιτητές

Το προαναφερθέν σύνολο 775 μαθητών είναι ο πυρήνας του εγωκεντρικού κοινωνικού δικτύου. Το δίκτυο αυτό δημιουργήθηκε από τους φοιτητές και τους άμεσους συμφοιτητές τους και τις μεταξύ τους σχέσεις. Οι σχέσεις αντικατοπτρίζουν τα πρότυπα των δεδομένων κοινωνικής συμπεριφοράς. Στη συνέχεια υπολογίζονται νέα χαρακτηριστικά μαθητή από τα δομικά χαρακτηριστικά του δικτύου και τα χαρακτηριστικά των άμεσων γειτόνων του μαθητή.

Όσον αφορά την πρόβλεψη εγκατάλειψης σπουδών οι συγγραφείς στοχεύσαν στην ανάπτυξη μιας ακριβούς μεθόδου για την πρόβλεψη της εγκατάλειψης που θα επέτρεπε επίσης την πρόβλεψη της εγκατάλειψης σε πρώιμο στάδιο των σπουδών. Η μέθοδος θα πρέπει να έχει όσο το λιγότερο δυνατά ψευδώς αρνητικά, δηλαδή μαθητές που δεν έχουν αναγνωριστεί ότι κινδυνεύουν να εγκαταλείψουν το σχολείο.

Σύμφωνα με τα αποτελέσματα βγαίνει το συμπέρασμα πως η επίδοση των φοιτητών φαίνεται να σχετίζεται κυρίως με τις κοινωνικές συνήθειες. Υποστηρίζει την υπόθεση ότι οι φοιτητές με μέτρια αποτελέσματα που όμως επικοινωνούν με φοιτητές με καλούς βαθμούς μπορούν να επιτύχουν-αποφοιτήσουν με μεγαλύτερη πιθανότητα από μαθητές με παρόμοια επίδοση αλλά δεν επικοινωνούν με επιτυχόντες μαθητές.

Ο κύριος στόχος αυτής της έρευνας ήταν η ανάπτυξη μιας μεθόδου για την εξόρυξη εκπαιδευτικών δεδομένων για την εκμάθηση ενός ταξινομητή ώστε να προβλέπει την επιτυχία μιας μαθητικής μελέτης και να επαληθεύει τη μέθοδο σε πραγματικά δεδομένα. Χρησιμοποιήθηκαν μέθοδοι DM και SNA(Social Network Analytics) για να λύσουν την εργασία. Οι συγγραφείς επαληθεύουν τη μέθοδο σε φοιτητές της Σχολής Πληροφορικής του Πανεπιστημίου Masaryk, αλλά τα δεδομένα που χρησιμοποιήθηκαν ήταν ανεξάρτητα από τη σχολή. Επομένως, η μέθοδος μπορεί να χρησιμοποιηθεί για οποιαδήποτε μονάδα πανεπιστημίου. Έδειξαν ότι τα δομημένα δεδομένα που αποκτήθηκαν μέσω link-based ανάλυση δεδομένων αύξησε σημαντικά την ακρίβεια της ταξινόμησης.

Χρησιμοποιήθηκαν μόνο τα δεδομένα που δεν είναι συγκεκριμένα για μια σχολή. Ωστόσο, για να αυξηθεί η ακρίβεια της ταξινόμησης θα ήταν χρήσιμο να εμπλουτιστούν τα δεδομένα με χαρακτηριστικά ειδικά για τη σχολή, π.χ. πληροφορίες σχετικά με συγκεκριμένες εξετάσεις που ο μαθητής πέρασε ή απέτυχε. Ένας άλλος πιθανός τρόπος μελλοντικής βελτίωσης μπορεί να είναι η εκμετάλλευση περισσότερων πληροφοριών από το κοινωνικό δίκτυο.

Άρθρο 3

Students' LMS interaction patterns and their relationship with achievement: A case study in higher education

Cerezo, R., Sánchez-Sántillán, M., Paule-Ruiz, M.P., Núñez, J.C.

Η χρήση των Συστημάτων Διαχείρισης Μάθησης (LMS) έχει αυξηθεί εκθετικά τα τελευταία αρκετά χρόνια και έχει καταλήξει να έχει ισχυρή επίδραση στη διαδικασία διδασκαλίας-μάθησης, ιδιαίτερα στην τριτοβάθμια εκπαίδευση. Η παρούσα μελέτη σκοπεύει να εξετάσει τις διαδικασίες ασύγχρονης μάθησης των μαθητών μέσω μιας προσέγγισης Εκπαιδευτικής Εξόρυξης Δεδομένων χρησιμοποιώντας δεδομένα που εξήχθησαν από τα αρχεία καταγραφής μαθητών που έκαναν χρήση του Moodle που ομαδοποιήθηκαν σύμφωνα με παρόμοιες συμπεριφορές σχετικά με την προσπάθεια, τον χρόνο εργασίας και την αναβλητικότητα. Στη συνέχεια, οι συμπεριφορές αντιστοιχίστηκαν με διαφορετικά επίπεδα επιτυχίας.

Αρχικά, ομαδοποιήθηκαν τα διαφορετικά πρότυπα ανάμειξης των μαθητών στη μαθησιακή διαδικασία σε ένα Σύστημα Διαχείρισης Μάθησης. Δεύτερον, οι διαφορετικές μεταβλητές που επιλέχθηκαν από τις εγγραφές του Moodle μελετήθηκαν για να διαπιστωθεί εάν ήταν εξίσου κατάλληλες για τη διαμόρφωση των ομάδων σπουδαστών. Τρίτον, εξετάστηκαν οι σχέσεις μεταξύ αυτών των προτύπων με τους τελικούς βαθμούς των μαθητών.

Μετά την ανάλυση των δεδομένων καταγραφής που συλλέχθηκαν από ένα μάθημα του Moodle 2.0 στο οποίο είχαν εγγραφεί 140 προπτυχιακοί φοιτητές, βρέθηκαν τέσσερα διαφορετικά πρότυπα μάθησης με διαφορετικούς τελικούς βαθμούς. Πρόσθετα αποτελέσματα έδειξαν ότι υπάρχουν μεταβλητές που σχετίζονται περισσότερο με την επίδοση και είναι πιο κατάλληλες για την ομαδοποίηση των μαθητών με βάση τις οποίες χαρακτηρίστηκαν οι διαφορετικές ομάδες, συγκεκριμένα, δύο Ομάδες Προσανατολισμού Εργασιών (κοινωνικά ή ατομικά εστιασμένες) και δύο Ομάδες Μη Προσανατολισμού Καθηκόντων (αναβλητικοί ή μη αναβλητικοί). Αυτά τα αποτελέσματα έχουν επιπτώσεις στο σχεδιασμό των παρεμβάσεων για τη βελτίωση των μαθησιακών διαδικασιών και των επιδόσεων των μαθητών στα LMS.

Πιο αναλυτικά σε αυτό το άρθρο οι συγγραφείς στοχεύουν να εξαγάγουν διαφορετικές ομάδες μαθητών με παρόμοια συμπεριφορά από τα αρχεία καταγραφής του Moodle και να τις αντιστοιχίσουν, αν είναι δυνατόν, με διαφορετικά επίπεδα επιτυχίας. Τα δεδομένα καταγραφής ανακτώνται από ένα μάθημα έντεκα εβδομάδων που βασίζεται στο Moodle. Το μάθημα έλαβε χώρα ως μέρος ενός υποχρεωτικού μαθήματος το τρίτο έτος πτυχιακού προγράμματος ψυχολογίας σε πανεπιστήμιο στη Βόρεια Ισπανία. Ο γενικός στόχος αυτής της εργασίας είναι να συμβάλει στη γνώση των αλληλεπιδράσεων των μαθητών με το LMS και την υποθετική τους αξία για την πρόβλεψη των επιτευγμάτων. Για το σκοπό αυτό τέθηκαν τρία ερευνητικά ερωτήματα:

1. Υπάρχουν διαφορετικά πρότυπα συμπεριφοράς προσπάθειας και αναβλητικότητας μεταξύ των μαθητών όταν μαθαίνουν μέσω LMS σε ένα αυθεντικό πλαίσιο;

2. Σε ποιο βαθμό οι μεταβλητές που επιλέγονται από τις εγγραφές του Moodle για τη διαμόρφωση των μοτίβων είναι εξίσου κατάλληλες για τη διαμόρφωση αυτών των συμπλεγμάτων;
3. Εάν τα μοτίβα προσπάθειας και αναβλητικότητας, που μετρώνται με βάση τα αρχεία καταγραφής του Moodle, σχετίζονται με τους τελικούς βαθμούς.

Για να απαντηθούν αυτές οι ερωτήσεις, χρησιμοποιήθηκαν διαφορετικές πηγές πληροφοριών με βάση τα δεδομένα που αποκτήθηκαν μετά την εργασία των μαθητών στο πρόγραμμα, όπως οι τελικοί βαθμοί τους και τα αρχεία καταγραφής των αλληλεπιδράσεων των μαθητών με το LMS που αναλύθηκαν. Ήταν δυνατό αλλά όχι απαραίτητο να χρησιμοποιήσουμε άλλη αρχιτεκτονική λογισμικού, όπως αρχιτεκτονική προσανατολισμένη στις υπηρεσίες, επειδή το Moodle περιλαμβάνει ήδη ένα εργαλείο—που ονομάζεται Log— που δημιουργεί αναφορές δραστηριότητας για κάθε χρήστη και αποθηκεύει τα δεδομένα στα αρχεία καταγραφής. +

Αφού τεθούν τα ερευνητικά ερωτήματα, περιγράφονται οι συμμετέχοντες και η διαδικασία, καθώς και η διαδικασία εξαγωγής μεταβλητών. Στη συνέχεια, η μεθοδολογία περιγράφεται σε δύο διαφορετικές φάσεις. Στο πρώτο στάδιο, η ανάλυση συστάδων ομαδοποιεί τους μαθητές σε διαφορετικές ομάδες και στο δεύτερο στάδιο, διεξάγεται ANOVA (Analysis of variance) για να παρατηρηθούν οι υποθετικές διαφορές μεταξύ των συστάδων. Στη συνέχεια, τα αποτελέσματα εμφανίζονται μετά τα ερευνητικά ερωτήματα.

Άρθρο 4

Mining Skills from Educational Data for Project Recommendations

Kordík, P., Kuznetsov, S.

Σε αυτό το άρθρο οι συγγραφείς παρουσιάζουν την μέθοδο που χρησιμοποιούν για να χαρτογραφήσουν μεμονωμένα θέματα στην οντολογία που δημιούργησαν. Παρουσιάζουν επίσης ένα δέντρο δεξιοτήτων. Αναφέρουν πως, υπάρχουν πολλές περισσότερες μέθοδοι για τον τρόπο περιγραφής της σχέσης μεταξύ θεμάτων και δεξιοτήτων. Έχουν αναπτύξει τη μέθοδό τους που βασίζεται στη συνεργασία με καθηγητές του πανεπιστημίου και βιομηχανικούς εταίρους, προκειμένου να παρουσιάσουν έναν βέλτιστο συμβιβασμό μεταξύ της πληρότητας και της απλής χρήσης. Το ίδιο το δέντρο θα ερευνάται και θα αναπτύσσεται συνεχώς. Το πανεπιστήμιο, παρόμοια με έναν ζωντανό οργανισμό, αναπτύσσεται συνεχώς, και γι' αυτό πρέπει να αναθεωρούνται οι λύσεις και να προσφέρονται νέες λύσεις προκειμένου να περιγράφεται καλύτερα η τρέχουσα εστίαση του πανεπιστημίου. Στην παρούσα εργασία παρουσιάζονται επίσης διαδικασίες που μεταφέρουν τους τελικούς βαθμούς των μαθητών σε βαθμολογίες και στη συνέχεια τη μετατροπή αυτών των βαθμολογιών σε επίπεδα δεξιοτήτων (αστέρια). Η όλη διαδικασία χωρίζεται σε 3 ξεχωριστές φάσεις. Αυτό το είδος διαχωρισμού είναι ευνοϊκό, καθώς επιτρέπει να αναπτύσσεται κάθε φάση ξεχωριστά χωρίς να επηρεάζεται η συμπεριφορά της όλης διαδικασίας.

Στην πρώτη φάση, παρουσιάζονται δύο αλγόριθμοι. Ο αλγόριθμος SimpleCompute αντιπροσωπεύει μια πολύ διαισθητική και άμεση προσέγγιση. Το μειονέκτημά του είναι το γεγονός ότι κάθε μάθημα έχει διαφορετικό επίπεδο δυσκολίας και επομένως είναι απαραίτητο να λαμβάνεται υπόψη η κατανομή των βαθμών εντός του μαθήματος. Ο δεύτερος αλγόριθμος είναι ο CumeDistCompute και αντιδρά καλύτερα στην εξάπλωση των βαθμών και έχει αποδειχθεί ότι είναι η καλύτερη επιλογή από τα δύο.

Στη δεύτερη φάση, παρουσιάζονται δύο αλγόριθμοι που συγκρίνουν τη βαθμολογία μεμονωμένων μαθητών και τη μετατρέπουν στο τελικό επίπεδο δεξιοτήτων. Το CompareMinMax είναι ένας απλός αλγόριθμος που συγκρίνει τη βαθμολογία κάθε μαθητή με τη μεγαλύτερη βαθμολογία στην ομάδα. Ο αλγόριθμος δεν αντιδρά στη συνολική εξάπλωση των μαθητών και αν τύχει να υπάρχει ένα εξαιρετικό άτομο, η αξία δεξιοτήτων του συνόλου δεν θα μειωθεί. Ένας άλλος αλγόριθμος είναι ο CompareSurvivalFunc που χρησιμοποιεί για μετασχηματισμούς τη λεγόμενη συνάρτηση Survival. Ο αλγόριθμος αντιδρά στη συνολική εξάπλωση των μαθητών σε ένα σύνολο και εξαλείφει εν μέρει την επίδραση που θα είχε ένα εξαιρετικό άτομο στο σετ. Οι διαφορές στους αλγόριθμους δεν είναι μεγάλες, αλλά το CompareSurvivalFunc είναι καλύτερο.

Αξιοσημείωτο είναι επίσης ότι οι περισσότεροι από τους μαθητές με χαμηλότερη βαθμολογία τείνουν να εισάγουν υψηλότερη αξία δεξιοτήτων, κάποιοι από αυτούς υπερεκτιμούν τις δεξιότητές τους και κάποιοι δεν έχουν ολοκληρώσει αρκετά μαθήματα για να πληρούν τις προϋποθέσεις για περισσότερα αστέρια.

Η τρίτη φάση χρησιμοποιεί έναν αλγόριθμο που βασίζεται στην «Ευκλείδεια απόσταση». Αυτός ο αλγόριθμος είναι πολύ απλός, αλλά δεν είναι σε θέση να προσδιορίσει με ακρίβεια την απόλυτη θέση ενός μαθητή. Οι συγγραφείς προσπαθούν να λύσουν αυτά και άλλα ζητήματα χρησιμοποιώντας τεχνολογίες της πληροφορίας καθώς και με διάφορες προσεγγίσεις δεδομένων.

Η κύρια συμβολή αυτής της εργασίας είναι ότι παρουσιάζει νέες προσεγγίσεις στον υπολογισμό και την παρουσίαση των δεξιοτήτων των μαθητών. Αυτή η προσέγγιση βασίζεται στη χαρτογράφηση των επιτευχθέντων αποτελεσμάτων των μαθητών και σε ένα τυποποιημένο δέντρο δεξιοτήτων. Χάρη σε αυτό, δημιουργούνται προφίλ μαθητών που μπορούν να παρουσιαστούν σε βιομηχανικούς συνεργάτες (πελάτες) όταν προτείνονται μαθητές για ένα συγκεκριμένο έργο. Αυτή η διαδικασία είναι πλέον αυτοματοποιημένη και χρησιμοποιώντας σχετικά απλούς αλγόριθμους οι συγγραφείς επιτυγχάνουν καλά αποτελέσματα συστάσεων σε σύγκριση με τις δεξιότητες αναφοράς των μαθητών. Το πλεονέκτημα της προσέγγισής τους, είναι ότι οι υπολογιστικές δεξιότητες ενημερώνονται αυτόματα όταν είναι διαθέσιμοι νέοι βαθμοί και οι μαθητές δεν χρειάζεται να προσαρμόσουν τις δεξιότητές τους με μη αυτόματο τρόπο. Σημειώνεται επίσης, ότι οι υπολογιστικές δεξιότητες είναι επίσης πιο αντικειμενικές από τις δεξιότητες που εισάγει ο μαθητής οι οποίες μπορεί να είναι έντονα προκατειλημμένες.

Όσον αφορά τα δεδομένα, οι συγγραφείς του άρθρου δημιούργησαν μια αποθήκη δεδομένων για την ενοποίηση δεδομένων από τα πανεπιστημιακά συστήματα. Το σύστημα είναι σχεδιασμένο σύμφωνα με την αρχιτεκτονική Inmona και είναι προσαρμοσμένο στις ανάγκες τους. Η αποθήκη δεδομένων καθιστά δυνατή την αυτόματη συλλογή δεδομένων, την ενοποίηση και την οπτικοποίηση τους με τη μορφή πινάκων εργαλείων και αναφορών ad-hoc. Επιπλέον, η αποθήκη δεδομένων δίνει τη δυνατότητα να αποθηκεύονται ιστορικά δεδομένα και έτσι είναι σε θέση να αναλαμβάνουν αναλύσεις που περιέχουν εκτεταμένες χρονικές περιόδους.

Άρθρο 5

Record of Undergraduates' Activities in Virtual Learning Enviroments

Alves, P., Miranda, L., Morais, C.

Σύμφωνα με τους συγγραφείς, τα εικονικά περιβάλλοντα μάθησης αποτέλεσαν αντικείμενο έρευνας πολλών ερευνητών και έχουν αντιπροσωπεύσει ένα στοιχείο τεχνολογικού και παιδαγωγικού εκσυγχρονισμού και ενημέρωσης σε πολλά εκπαιδευτικά ιδρύματα, συγκεκριμένα σε ιδρύματα τριτοβάθμιας εκπαίδευσης. Προκειμένου να αξιολογηθούν οι δυνατότητες των εργαλείων εικονικού περιβάλλοντος μάθησης καθώς και των προτύπων χρήσης τους, είναι σημαντικό να γνωρίζουμε τη σχέση που δημιουργούν οι προπτυχιακοί φοιτητές με αυτά τα περιβάλλοντα, συγκεκριμένα όσον αφορά την πρόσβασή τους στο περιβάλλον. Η αναγνώριση των προσβάσεων των μαθητών πραγματοποιήθηκε με βάση τον αριθμό των προσβάσεων σχετικά με τα αρχεία καταγραφής (logs) δραστηριοτήτων των εικονικών περιβαλλόντων μάθησης. Η χρήση ψηφιακών εργαλείων δεν αποτελεί από μόνη της μια προϋπόθεση που θα μπορούσε να είναι επαρκής για τη βελτίωση των παιδαγωγικών προσεγγίσεων στην οικοδόμηση της γνώσης. Επομένως, είναι σημαντικό να μελετηθούν οι προσβάσεις των προπτυχιακών σπουδαστών σχετικά με τα αρχεία καταγραφής δραστηριοτήτων του εικονικού περιβάλλοντος μάθησης καθώς και με τις δραστηριότητες που ανέπτυξαν με τα εργαλεία που παρέχονται σε αυτό το περιβάλλον.

Έτσι, το παρόν άρθρο εστιάζει στην ανάλυση της εξέλιξης των προσβάσεων σχετικά με τα αρχεία καταγραφής δραστηριοτήτων του εικονικού περιβάλλοντος λαμβάνοντας ως δείγμα προπτυχιακούς φοιτητές στα ακαδημαϊκά έτη μεταξύ 2009/2010 και 2013/2014. Οι κύριοι στόχοι αυτής της έρευνας είναι:

- Αξιολόγηση της εξέλιξης των προσβάσεων σχετικά με τα αρχεία καταγραφής δραστηριοτήτων του εικονικού περιβάλλοντος μάθησης.
- Ταξινόμηση της συχνότητας των προσβάσεων στο εικονικό περιβάλλον μάθησης.
- Αξιολόγηση του είδους των δραστηριοτήτων που αναπτύσσονται με τα εργαλεία στο εικονικό περιβάλλον μάθησης.

Η απάντηση στους αναφερόμενους στόχους θα ληφθεί από τα αρχεία καταγραφής δραστηριοτήτων του εικονικού μαθησιακού περιβάλλοντος από τους φοιτητές ενός πορτογαλικού ιδρύματος τριτοβάθμιας εκπαίδευσης, με δείγμα περίπου 7000 φοιτητές ετησίως, σε μια περίοδο πέντε ακαδημαϊκών ετών.

Η φύση αυτής της μελέτης είναι ουσιαστικά ποσοτική και η κύρια τεχνική συλλογής δεδομένων που χρησιμοποιήθηκε ήταν η ανασκόπηση στο γραφείο. Τα δεδομένα ελήφθησαν χρησιμοποιώντας υπολογιστικές διαδικασίες και συμβουλευτικές βάσεις δεδομένων που σχετίζονται με το εικονικό περιβάλλον μάθησης. Η διαδικασία συλλογής δεδομένων ακολούθησε την πολιτική απορρήτου του ιδρύματος, συγκεκριμένα όσον αφορά την εξουσιοδότηση και την πρόσβαση στα δεδομένα, καθώς και την εμπιστευτικότητα των χρηστών του εικονικού περιβάλλοντος μάθησης. Τα δεδομένα είναι συνεπή και έγκυρα,

καθώς είναι διαθέσιμα σε βάσεις δεδομένων που επικυρώνονται και παρακολουθούνται σωστά από φορείς που ανήκουν στο ίδρυμα που φοιτούν οι προπτυχιακοί φοιτητές. Επιλέχθηκε μια ποσοτική προσέγγιση επειδή οι μεταβλητές σε αυτή τη μελέτη υποθέτουν αριθμητικές τιμές και υπεροχή δόθηκε σε μια περιγραφική ανάλυση δεδομένων καθώς και στην εμφάνισή τους σε πίνακες και γραφήματα, καθιστώντας δυνατό τον έλεγχο της σχέσης μεταξύ των μεταβλητών μέσω στατιστικών διαδικασιών.

Τα αποτελέσματα έδειξαν ότι στο σύνολο των σχολών που αποτελούν το ίδρυμα, υπήρξε αύξηση στον αριθμό των προσβάσεων τα τρία πρώτα χρόνια και πτωτική τάση τα τελευταία δύο χρόνια. Ωστόσο, αυτή η τάση δεν παρατηρείται όταν αναλύουμε κάθε ένα από τις σχολές ξεχωριστά. Οι συγγραφείς χώρισαν την κατανομή των προσβάσεων σε πέντε ομάδες σύμφωνα με την ελάχιστη τιμή και πέντε εκατοστημόρια, λαμβάνοντας έτσι πέντε διαστήματα με περίπου τον ίδιο αριθμό δεδομένων, ταξινομώντας τους χρήστες ανάλογα με τη συχνότητα πρόσβασης στο εικονικό περιβάλλον μάθησης στις ακόλουθες κατηγορίες: πολύ χαμηλό, χαμηλό, μέτριο, υψηλό, πολύ υψηλό.

Οι δραστηριότητες που αναπτύσσονται στο εικονικό περιβάλλον μάθησης του ιδρύματος σχετίζονται με την πρόσβαση στα εργαλεία: Ειδοποιήσεις, Ιστολόγιο, Digital Locker, Ημερολόγιο, Συνομιλία, Στατιστικά, Φόρουμ, Μέλη Περιοχής, Μηνύματα, Πόροι, Δοκιμές, Εργασίες και Wiki. Η πιο συχνή δραστηριότητα σε κάθε ακαδημαϊκό έτος ήταν η «συμβουλευτική ενημέρωση».

Άρθρο 6

Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data

Aher, S.B., Lobo, L.

Στο συγκεκριμένο άρθρο οι συγγραφείς δείχνουν πώς οι τεχνικές εξόρυξης δεδομένων όπως ο αλγόριθμος ομαδοποίησης (clustering) και κανόνα συσχέτισης (association rule) είναι χρήσιμες στο Σύστημα Προτάσεων Μαθημάτων/Course Recommendation System που προτείνει το μάθημα στον μαθητή με βάση την επιλογή άλλων μαθητών για συγκεκριμένο σύνολο μαθημάτων που συλλέγονται από το Moodle. Ως αποτέλεσμα του Συστήματος Προτάσεων Μαθημάτων, μπορεί να προταθεί σε νέο φοιτητή που έχει εγγραφεί πρόσφατα για κάποιο μάθημα π.χ. Λειτουργικό Σύστημα, το νέο μάθημα που θα επιλεγεί π.χ. Κατανεμημένο σύστημα.

Πιο αναλυτικά, σε αυτή τη μελέτη έγινε σύγκριση του αποτελέσματος του association rule αλγορίθμου Apriori, που είναι ο υπάρχων αλγόριθμος στο εργαλείο εξόρυξης δεδομένων ανοιχτού κώδικα Weka, με το αποτέλεσμα του συνδυασμού του clustering αλγορίθμου και αλγορίθμου association rule χρησιμοποιώντας Simple K-means για clustering και τον association rule αλγόριθμο Apriori. Η προσέγγισή των συγγραφέων χρησιμοποιεί τον συνδυασμό τεχνικής ομαδοποίησης – Simple K-Means και τον association rule αλγόριθμο – Apriori και βρίσκει το αποτέλεσμα. Τα αποτελέσματα του συνδυασμένου αλγορίθμου είναι καλύτερα από το αποτέλεσμα που προέκυπτε χρησιμοποιώντας τον αλγόριθμο Apriori και άλλο συνδυασμό αλγορίθμου clustering. Το αποτέλεσμα χρησιμοποιώντας τον αλγόριθμο κανόνων συσχέτισης Apriori χρειάζεται το στάδιο προετοιμασίας δεδομένων καθώς και αν αυξήσουμε την υποστήριξη τότε ο αριθμός των κανόνων συσχέτισης είναι μικρότερος. Επίσης, τα αποτελέσματα που χρησιμοποιούν συνδυασμό της ομαδοποίησης μεγιστοποίησης προσδοκιών (Expectation Maximization Clustering) και του αλγορίθμου Apriori και του συνδυασμού του association rule αλγορίθμου Farthest First και Apriori δεν ταιριάζουν με τις πραγματικές επιλογές των μαθητών. Αλλά αν χρησιμοποιηθεί συνδυασμένη προσέγγιση χρησιμοποιώντας K-means και αλγόριθμο Apriori, τότε δεν χρειάζεται να χρησιμοποιηθεί το στάδιο προετοιμασίας δεδομένων και ο αριθμός των κανόνων συσχέτισης είναι μεγαλύτερος. Αυτή η συνδυασμένη προσέγγιση αυξάνει την ισχύ του κανόνα συσχέτισης. Έτσι, αυτό το Σύστημα Προτάσεων Μαθημάτων θα βοηθήσει τους μαθητές να επιλέξουν τον κατάλληλο συνδυασμό μαθημάτων σύμφωνα με τα ενδιαφέροντά τους. Αυτό το σύστημα μπορεί επίσης να χρησιμοποιηθεί για να προτείνει μαθήματα στον σπουδαστή στην εξ αποστάσεως εκπαίδευση.

Οι συγγραφείς αναφέρθηκαν σε βιβλιογραφία πολυάριθμων μελετών που έχουν γίνει πάνω στο σύστημα E-Learning που επικεντρώνεται στη συμπεριφορά του μαθητή. Ασχολούνται με διαφορετικές τεχνικές εξόρυξης δεδομένων για την ανάλυση της συμπεριφοράς των μαθητών, και στην συνέχεια προχώρησαν στην επεξήγηση των δεδομένων τους.

Σε αυτό το Σύστημα Προτάσεων Μαθημάτων, εξετάζονται 13 κατηγορίες μαθημάτων και 82 μαθήματα που σχετίζονται με δύο κλάδους, δηλαδή την Επιστήμη και Μηχανική Υπολογιστών (CSE) και την Τεχνολογία Πληροφορικής (IT) για τη συλλογή δεδομένων από φοιτητές. Μετά τη συλλογή των δεδομένων λαμβάνονται υπόψη μόνο εκείνοι οι φοιτητές που εγγράφονται για τουλάχιστον 5 μαθήματα και μόνο εκείνα τα μαθήματα που έχουν επιλεγεί από περισσότερα από 100 φοιτητές. Καθώς λαμβάνονται υπόψη 83 μαθήματα, ορισμένα μαθήματα ενδέχεται να μην επιλέγονται από τους φοιτητές. Έτσι, αφού σχηματιστεί ο τελικός πίνακας, λαμβάνονται υπόψη μόνο τα μαθήματα που έχουν επιλέξει τουλάχιστον 100 φοιτητές. Έτσι, τελικά, εξετάζονται 12 κατηγορίες μαθημάτων και 36 μαθήματα.

Μετά τη σύνδεση στο σύστημα Moodle, το όνομα χρήστη και ο κωδικός πρόσβασης του μαθητή επικυρώνονται. Εάν είναι έγκυρο, τότε ο φοιτητής περιηγείται στις κατηγορίες και τα μαθήματα των μαθημάτων. Μετά την περιήγηση σε αυτά τα μαθήματα, ο φοιτητής θα εγγραφεί στο σύνολο των μαθημάτων για τα οποία ενδιαφέρεται. Εδώ ο φοιτητής θα πρέπει να εγγραφεί σε τουλάχιστον πέντε μαθήματα, διαφορετικά τα δεδομένα του μαθητή δεν θα ληφθούν υπόψη κατά τη διαμόρφωση του τελικού πίνακα για να βρεθεί το αποτέλεσμα. Οι πληροφορίες εγγραφής για το σύνολο συγκεκριμένων μαθημάτων αποθηκεύονται στη βάση δεδομένων του Moodle. Επιλέγονται λοιπόν τα σχετικά δεδομένα από τη βάση δεδομένων του Moodle και σχηματίζεται ο τελικός πίνακας για να βρεθεί το αποτέλεσμα του Συστήματος Προτάσεων Μαθημάτων. Εδώ γίνεται χρήση του εργαλείου εξόρυξης δεδομένων ανοιχτού κώδικα Weka ώστε να βρεθεί το αποτέλεσμα

Άρθρο 7

Three Interaction patterns on asynchronous online discussion behaviours: A methodological comparison

Jo, I., Park, Y., Lee, H.

Ο σκοπός αυτής της μελέτης ήταν να αξιολογήσει τη διαδικτυακή συμπεριφορά των μαθητών, ειδικά σε ένα πλαίσιο AOD (Asynchronous Online Discussion), και να αναλύσει τις επιρροές στα ακαδημαϊκά αποτελέσματα. Για αυτό, οι συγγραφείς παρουσιάζουν την αποτελεσματικότητα της ενσωμάτωσης πολλαπλών μεθοδολογιών για την ανάλυση της ποιότητας ενός AOD. Τα αποτελέσματα της ανάλυσης P2S (Person to System), P2P (Person to Person) και P2C (Person to content) παρέχουν τις ακόλουθες μεθοδολογικές επιπτώσεις όσον αφορά την αξιολόγηση του AOD.

Στο πείραμα συμμετείχαν 43 μαθητές και ένας καθηγητής στο πλαίσιο του μαθήματος “Διοίκηση, Νομοθεσία και Πολιτική” του δεύτερου εξαμήνου, το 2013, σε ένα πανεπιστήμιο στην Νότια Κορέα.

Πρώτον, με βάση την ανάλυση P2S, μόνο η μεταβλητή επίσκεψης προσδιορίστηκε ως σημαντική μεταβλητή για την ακαδημαϊκή επίδοση των μαθητών στο AOD. Ωστόσο, ο χρόνος που δαπανήθηκε επί του συστήματος και η κανονικότητα στη συζήτηση ήταν άκυρα ως μεταβλητές πρόβλεψης. Αυτό συμβαίνει γιατί οι μαθητές προχωρούν στην κύρια εργασία στο περιβάλλον του τοπικού υπολογιστή και όχι απευθείας στον πίνακα συζητήσεων. Το MS Office ή το text pad είναι ένα αποτελεσματικό εργαλείο για αναθεώρηση. Ως εκ τούτου, η προεργασία για τη σύνταξη ενός μηνύματος συζήτησης μπορεί να προχωρήσει σε έναν τοπικό υπολογιστή και να φορτωθεί σε έναν πίνακα συζητήσεων μέσω αντιγραφής και επικόλλησης. Ως αποτέλεσμα, πόσος χρόνος αφιερώθηκε ή πόσο τακτικά συνδέονταν οι μαθητές ο πίνακας συζητήσεων φάνηκε να είναι ένα ασήμαντο θέμα προς ερμηνεία.

Δεύτερον, μέσω της ανάλυσης P2P, οι συγγραφείς ανακάλυψαν τα μοτίβα αλληλεπίδρασης των μαθητών, όπως η συχνότητα και η απεικόνιση των κεντρικών σημείων του δικτύου των μαθητών, ως σημαντικούς δείκτες για την πρόβλεψη της ακαδημαϊκής επιτυχίας. Επιπλέον, επαληθεύτηκε ότι η συγκέντρωση εκτός βαθμού μειώθηκε, αλλά το επίπεδο πυκνότητας αυξήθηκε όσο περνούσε ο καιρός. Αυτά υποδηλώνουν ότι όλο και περισσότεροι μαθητές συμμετείχαν στη συζήτηση και έγινε όλο και πιο ενεργοί με την πάροδο του χρόνου. Σε αντίθεση με την πρώτη περίοδο του εξαμήνου, περισσότεροι φοιτητές συμμετείχαν στη συζήτηση στην όψιμη περίοδο.

Τρίτον, τα αποτελέσματα P2C έδειξαν ότι τα περισσότερα από τα περιεχόμενα της συζήτησης ανήκουν στη φάση εξερεύνησης και ολοκλήρωσης. Ο δάσκαλος του μαθήματος προσπάθησε να κινήσει το ενδιαφέρον και το κίνητρο των μαθητών ως προκλητής, αλλά αυτό δεν έπαιξε ρόλο στην περίληψη των απόψεων που συζητήθηκαν. Σίγουρα δεν είναι εύκολη διαδικασία να δοθεί λύση με βάση τη σύνοψη της όλης συζήτησης. Αυτό υποδεικνύεται από το αποτέλεσμα ότι η φάση επίλυσης δεν εμφανίστηκε σε κανένα μήνυμα

συζήτησης. Ως εκ τούτου, συνιστάται κατάλληλη η παρέμβαση του εκπαιδευτή για την ολοκλήρωση ή τη σύνθεση της συζήτησης.

Κάθε μεθοδολογία είχε τα δυνατά της σημεία στην ανάλυση των προτύπων συζήτησης των μαθητών. Πρώτον, η ανάλυση P2S υποστήριξε την επισκόπηση της συνολικής δραστηριότητας συζήτησης μεταξύ των μαθητών, επειδή υπολογίζει πόσες φορές επισκέφτηκαν τον πίνακα συζητήσεων και πόσο συμμετείχαν στη συζήτηση. Τέτοια δεδομένα μπορούν να ερμηνευθούν ως ένας σημαντικός δείκτης για την ανάλυση των προτύπων συζήτησης των μαθητών. Δεύτερον, το P2P αναλύει βαθύτερα την αλληλεπίδραση των μαθητών μέσω του υπολογισμού της κεντρικότητας και της πυκνότητάς τους μέσα στο δίκτυο συζήτησης με βάση τα δεδομένα που ανιχνεύονται μέσω του υπολογιστή. Επειδή η δραστηριότητα συζήτησης είναι μια διαδικασία συνεργασίας που δεν περιλαμβάνει έναν μόνο παράγοντα αλλά παράγοντες σε ένα ολόκληρο δίκτυο, το P2S είναι ένα ισχυρό εργαλείο για την ανάλυση ενός μαθήματος AOD. Ωστόσο, τέτοιες ποσοτικές προσεγγίσεις ήταν περιορισμένες στην ικανότητά τους να αντικατοπτρίζουν συγκεκριμένο περιεχόμενο. Η γνωστική παρουσία στο P2C είναι μια πηγή για την αξιολόγηση της δραστηριότητας συζήτησης των μαθητών, ειδικά για το αν το περιεχόμενο της συζήτησής τους περιλάμβανε μια γνωστική διαδικασία υψηλού επιπέδου.

Συνοψίζοντας, δεν επαρκεί η χρήση μιας ενιαίας μεθοδολογίας για την ανάλυση για την αξιολόγηση AOD. Η λεπτομερής ανάλυση για το AOD θα οδηγήσει στην ανάπτυξη μιας εκπαιδευτικής στρατηγικής για αυτό. Αυτό τελικά συμβάλλει στη διευκόλυνση της πιο ενεργούς συμμετοχής των μαθητών και στη βελτίωση των μαθησιακών επιτευγμάτων τους στο μαθητοκεντρικό μαθησιακά περιβάλλοντα.

Άρθρο 8

A Framework for Capturing Distinguishing User Interaction Behaviours in Novel Interfaces

Kardan, S., Conati, C.

Η πρόοδος στην Επικοινωνία Ανθρώπου και Υπολογιστή βοηθά συνεχώς τη δημιουργία νέων διεπαφών για την υποστήριξη της εκπαίδευσης και της κατάρτισης. Λόγω της καινοτομίας αυτών των διεπαφών, μπορεί να είναι δύσκολο να κρίνουμε εκ των προτέρων ποιο σύνολο συμπεριφορών αλληλεπίδρασης με τους χρήστες ευνοεί τη μάθηση. Ο μακροπρόθεσμος στόχος των συγγραφέων είναι να επινοήσουν αυτόματες τεχνικές για την ανάλυση των αρχείων καταγραφής των αλληλεπιδράσεων με μια νέα εφαρμογή και τον εντοπισμό κατηγοριών τύπων χρηστών, τις συμπεριφορές ταυτοποίησής τους και πώς αυτές οι συμπεριφορές σχετίζονται με τη μάθηση. Επιπλέον, θέλουν να χρησιμοποιήσουν αυτές τις πληροφορίες για να δημιουργήσουν έναν χρήστη μοντέλο, δηλαδή να αναγνωρίζει αυτόματα τις συμπεριφορές των νέων χρηστών και να επιτρέπει στην εφαρμογή να παρέχει προσαρμοστική υποστήριξη κατά τη διάρκεια της αλληλεπίδρασης, εάν οι συμπεριφορές σχετίζονται με μη βέλτιστη απόδοση εργασιών.

Σε προηγούμενο άρθρο, οι συγγραφείς περιέγραψαν μια προσέγγιση μοντελοποίησης χρήστη απόδειξης της ιδέας που χρησιμοποιεί κανόνες ομαδοποίησης χωρίς επίβλεψη και συσχετισμού κλάσεων για τον εντοπισμό σχετικών τύπων/συμπεριφορών χρηστών από ένα υπάρχον σύνολο δεδομένων και βασίζεται σε αυτούς για την ταξινόμηση νέων χρηστών. Σε αυτό το άρθρο, τελειοποιούν αυτήν την απόδειξη της ιδέας σε ένα ολοκληρωμένο πλαίσιο μοντελοποίησης χρήστη που εξορθολογίζει τις φάσεις που είναι απαραίτητες για τη δημιουργία ενός ταξινομητή χρήστη από ένα αρχικό σύνολο δεδομένων ακατέργαστων αρχείων καταγραφής αλληλεπίδρασης. Η αρχική προσέγγιση αξιολογήθηκε σε ένα περιβάλλον για την υποστήριξη της εκμάθησης αλγορίθμων ΑΙ μέσω της εξερεύνησης διαδραστικών προσομοιώσεων. Εδώ, αξιολογούν το νέο πλαίσιο μοντελοποίησης χρηστών στο ίδιο περιβάλλον αλλά σε μεγαλύτερο σύνολο δεδομένων, παρέχοντας έτσι πιο πειστικά στοιχεία για την αποτελεσματικότητα της προσέγγισης.

Αφού γίνει αναφορά σε σχετικά άρθρα, παρουσιάζεται η γενική προσέγγιση μοντελοποίησης χρηστών, συμπεριλαμβανομένων βελτιώσεων από προηγούμενες εκδόσεις. Στη συνέχεια, συζητάμε μια εμπειρική αξιολόγηση του πλαισίου και ολοκληρώνουμε με μια συζήτηση για μελλοντική εργασία.

Σε αυτό το άρθρο, περιγράφεται ένα πλαίσιο μοντελοποίησης χρήστη που χρησιμοποιεί μη εποπτευόμενη ομαδοποίηση και Class Associating Mining για να ανακαλύψει και να αναγνωρίσει σχετικά μοτίβα αλληλεπίδρασης κατά την αλληλεπίδραση των μαθητών με το εκπαιδευτικό λογισμικό. Το πλαίσιο βελτιώνει μια προηγούμενη προσέγγιση απόδειξης της ιδέας προσθέτοντας λειτουργίες για πιο αποτελεσματική ομαδοποίηση και πιο βασική επιλογή ορισμένων από τις απαιτούμενες παραμέτρους. Μια εμπειρική αξιολόγηση του πλαισίου παρέχει στοιχεία ότι μπορεί να ομαδοποιήσει τους χρήστες σε σημαντικές ομάδες,

καθώς και να ταξινομήσει τους νέους χρήστες με ακρίβεια. Το πιο σημαντικό είναι ότι το πλαίσιο δημιουργεί κανόνες που παρέχουν μια λεπτομερή περιγραφή κοινών συμπεριφορών για χρήστες σε διαφορετικά συμπλέγματα. Αυτοί οι κανόνες φαίνεται να είναι κατάλληλοι για να καθοδηγήσουν προσαρμοστικές παρεμβάσεις στις οποίες στοχεύουν βελτίωση της αποτελεσματικότητας της αλληλεπίδρασης. Το επόμενο βήμα αυτής της εργασίας θα είναι η προσθήκη αυτών των προσαρμοστικών παρεμβάσεων στο εκπαιδευτικό λογισμικό που χρησιμοποιήθηκε ως δοκιμαστικό κρεβάτι για αυτήν την έρευνα, μια διαδραστική προσομοίωση που θα βοηθήσει τους μαθητές να κατανοήσουν έναν αλγόριθμο για την ικανοποίηση περιορισμών. Σχεδιάζουν οι συγγραφείς, επίσης να χρησιμοποιήσουν το πλαίσιο για τη δημιουργία classifier user models για άλλο εκπαιδευτικό λογισμικό, συμπεριλαμβανομένων διαδραστικών προσομοιώσεων για άλλους αλγόριθμους AI και ένα εκπαιδευτικό παιχνίδι για μαθηματικές δεξιότητες.

Τα δεδομένα για την αξιολόγησή συλλέχθηκαν μέσω μελετών χρηστών με βάση το πειραματικό πρωτόκολλο για τη μικροεφαρμογή CSP που περιγράφεται αναλυτικά στο άρθρο. Οι φοιτητές του πανεπιστημίου, οι οποίοι ήταν εξοικειωμένοι με τη βασική θεωρία γραφημάτων, αλλά δεν είχαν παρακολουθήσει μάθημα τεχνητής νοημοσύνης, εισήχθησαν στον αλγόριθμο AC3 και στη συνέχεια έκαναν μια προκαταρκτική δοκιμή για το θέμα. Στη συνέχεια, κάθε συμμετέχων χρησιμοποίησε τη μικροεφαρμογή σε δύο προβλήματα CSP και έγραψε ένα posttest. Το σύνολο δεδομένων που προκύπτει περιλαμβάνει αρχεία καταγραφής ενεργειών για 65 χρήστες, συνολικού αριθμού 13.078 ενεργειών για 62.752 δευτερόλεπτα αλληλεπίδρασης. Από αυτά τα αρχεία καταγραφής, υπολογίστηκαν:

- α) συχνότητα χρήσης κάθε ενέργειας διεπαφής
- β) μέσο όρο και τυπική απόκλιση της καθυστέρησης μεταξύ των ενεργειών.

Ο μέσος λανθάνων χρόνος είναι ένας δείκτης του χρόνου που δαπανάται για τον προβληματισμό μετά από μια ενέργεια και τον προγραμματισμό για την επόμενη, ενώ η τυπική απόκλιση του λανθάνοντος χρόνου δείχνει εάν ο χρήστης ήταν συνεπής ή επιλεκτικός ως προς το ποσό της παύσης μετά από κάθε ενέργεια.

Άρθρο 9

Mining sentiments in SMS texts for teaching evaluation

Leong, C.K., Lee, Y.H., Mak, W.K.

Στην εκπαίδευση, είναι χαρακτηριστικό να υπάρχει μια αξιολόγηση στο τέλος του μαθήματος στην οποία οι μαθητές μπορούν να εκφράσουν τις απόψεις τους για τους δασκάλους ή τους εκπαιδευτές καθώς και το περιεχόμενο και την παράδοση του μαθήματος. Αυτή η αξιολόγηση περιλαμβάνει γενικά μια κλίμακα Likert στην οποία οι μαθητές θα υποδείξουν τις απόψεις τους για το μάθημα ελέγχοντας πόσο έντονα συμφωνούν ή διαφωνούν με προσεκτικά κατασκευασμένες δηλώσεις σχετικά με τους δασκάλους, το περιεχόμενο του μαθήματος και την παράδοση του μαθήματος. Συνήθως, υπάρχουν πέντε επίπεδα: συμφωνώ απόλυτα, συμφωνώ, ουδέτερο, διαφωνώ και διαφωνώ απόλυτα.

Ο πολλαπλασιασμός των τηλεφώνων και η χρήση της υπηρεσίας σύντομων μηνυμάτων (SMS) μεταξύ των μαθητών προσφέρει μια ευκαιρία για μια τέτοια αξιολόγηση να διεξάγεται αμέσως μετά το μάθημα. Με αυτόν τον τρόπο, μπορεί να παρέχεται άμεση ανατροφοδότηση στους εκπαιδευτές και να γίνουν οι απαραίτητες ενέργειες κατά τη διάρκεια του μαθήματος.

Η υπηρεσία σύντομων μηνυμάτων (SMS) είναι ένα στοιχείο της σειράς προτύπων του Παγκόσμιου Συστήματος Κινητών Επικοινωνιών (GSM). Το SMS είναι ένα μέσο αποστολής σύντομων μηνυμάτων κειμένου που δεν υπερβαίνουν τους 160 χαρακτήρες (συμπεριλαμβανομένων των διαστημάτων) μεταξύ συσκευών κινητής τηλεφωνίας. Τα γραπτά μηνύματα SMS είναι η πιο διαδεδομένη εφαρμογή δεδομένων στον κόσμο, με ετήσιους όγκους SMS παγκοσμίως που εκτιμάται ότι αυξήθηκαν από 2,6 τρισεκατομμύρια το 2007 σε σχεδόν 5,5 τρισεκατομμύρια SMS στο τέλος του 2009. Ο όρος SMS χρησιμοποιείται συχνά ως συνώνυμο για όλους τους τύπους σύντομων μηνυμάτων κειμένου, καθώς και για τη δραστηριότητα δημιουργίας κειμένων SMS. Σε αυτό το άρθρο, χρησιμοποιείται ο όρος «κείμενα SMS» για να γίνει αναφορά στα μεμονωμένα μηνύματα κειμένου SMS που αποστέλλονται.

Αυτό το άρθρο εξετάζει τις ευκαιρίες και τις προκλήσεις της απόκτησης κριτικής της διδασκαλίας από τους μαθητές μέσω κειμένων SMS. Συγκεκριμένα, διερευνά τις πιθανές εφαρμογές της εξόρυξης κειμένου σε κείμενα SMS που χρησιμοποιούνται για την κριτική της διδασκαλίας. Σύμφωνα με τους συγγραφείς, η παρούσα εργασία είναι η πρώτη που διερευνά τέτοιες εφαρμογές στην εκπαιδευτική εξόρυξη δεδομένων.

Επιπλέον, η εργασία συμβάλλει επίσης στη βιβλιογραφία για την εξόρυξη κειμένου και την εξόρυξη συναισθημάτων. Σε αντίθεση με τα περισσότερα κείμενα, τα μηνύματα SMS είναι σχετικά σύντομα και προσφέρουν μια σειρά από προκλήσεις. Πρώτον, λόγω της φύσης του, οι απαντήσεις που δίνονται θα είναι διαφορετικές για διαφορετικούς χρήστες. Έτσι, εάν υπάρχουν 100 μαθητές σε μια τάξη, κανένας δύο μαθητές δεν θα έδινε την ίδια ανατροφοδότηση μέσω SMS ακόμα και όταν τηρούνται γλωσσικοί κανόνες και δεν γίνονται ορθογραφικά λάθη. Δεύτερον, τα μηνύματα SMS διαποτίζονται από συντομευμένες

συντομεύσεις (π.χ. οic), emoticons (π.χ. :)). Τρίτον, είναι σύνηθες τα κείμενα SMS να περιέχουν «θορύβους» όπως ορθογραφικά και γραμματικά λάθη, ημιτελείς προτάσεις, λανθασμένες εκκινήσεις, επαναλήψεις και λέξεις που γεμίζουν παύση, όπως um και uh. Η πρόκληση, σε αυτή την περίπτωση, είναι η εξαγωγή πληροφοριών από μια ανάλυση τέτοιων θορυβωδών και αδόμητων δεδομένων σε κείμενα SMS. Κατά τη διερεύνηση αυτών των προκλήσεων, η εργασία ανοίγει το δρόμο για περαιτέρω έρευνα προτείνοντας μια συστηματική προσέγγιση για την ανάλυση τέτοιων κειμένων SMS για εκπαιδευτικούς σκοπούς.

Σε αυτό το άρθρο, οι συγγραφείς διερευνούν την πιθανή εφαρμογή της εξόρυξης κειμένου και της εξόρυξης συναισθημάτων για την ανάλυση κειμένων υπηρεσίας σύντομων μηνυμάτων (SMS) στην αξιολόγηση διδασκαλίας. Σύμφωνα με τους συγγραφείς, αυτό το άρθρο είναι το πρώτο που εξερευνά τέτοιες εφαρμογές στην εκπαιδευτική εξόρυξη δεδομένων. Η προεπεξεργασία δεδομένων περιλαμβάνει την ανάγνωση, την ανάλυση και την κατηγοριοποίηση των κειμένων SMS. Αναπτύσσονται τρία μοντέλα: το βασικό μοντέλο, το «διορθωμένο» μοντέλο που προσαρμόζεται για ορθογραφικά λάθη και το μοντέλο «συναίσθημα» που επεκτείνει το «διορθωμένο» μοντέλο εκτελώντας εξόρυξη συναισθήματος. Στη συνέχεια, εφαρμόζεται ένα κριτήριο «ενδιαφέροντος» για να επιλεγεί το «συναισθηματικό» μοντέλο. Με βάση αυτό το μοντέλο, γίνεται συζήτηση των συναισθημάτων των μαθητών απέναντι στη διάλεξη. Τέλος, γίνεται διαπίστωση ότι τα αποτελέσματα της ανάλυσης είναι ανθεκτικά στην αφαίρεση ενός ή και των δύο τύπων ημιτελών κειμένων SMS. Κατά τη διερεύνηση αυτών των προκλήσεων, η εργασία ανοίγει το δρόμο για περαιτέρω έρευνα προτείνοντας μια συστηματική προσέγγιση για την ανάλυση τέτοιων κειμένων SMS για εκπαιδευτικούς σκοπούς. Ειδικότερα, η ανάλυση συναισθήματος του SMS παρέχει στον ομιλητή έγκαιρη πληροφόρηση για να μετρήσει τις θετικές και αρνητικές πτυχές της παράδοσης της διάλεξης και να προβληματιστεί σχετικά με την πιθανή βελτίωση των κειμένων SMS. Αυτό θα ήταν μια βελτίωση σε σχέση με την αξιολόγηση στο τέλος του μαθήματος, η οποία αποτυπώνει μόνο την απόδοση του διδάσκοντα στο τέλος του μαθήματος, οπότε δεν μπορούν να γίνουν διορθωτικές ενέργειες έως ότου το μάθημα προσφερθεί σε επόμενα εξάμηνα.

Άρθρο 10

Investigating student motivation in the context of a learning analytics intervention during a summer bridge program

Lonn, S., Aguilar, S.J., Teasley, S.D.

Η διατήρηση φοιτητών σε ένα πανεπιστήμιο θεωρείται κρίσιμο ζήτημα στην τριτοβάθμια εκπαίδευση εδώ και δεκαετίες. Η αποδοχή φοιτητών που αποτυγχάνουν να αποφοιτήσουν είναι καταστροφική για τους φοιτητές και έχει συνέπειες για τη θεσμική ευθύνη και τα σχετικά μοντέλα εσόδων. Κατά συνέπεια, τα μεταδευτεροβάθμια ιδρύματα ενδιαφέρονται ολοένα και περισσότερο να επενδύσουν σε βιώσιμα και επιτυχημένα μοντέλα για να αυξήσουν τη διατήρηση, ιδιαίτερα για ομάδες φοιτητών με ιστορικά χαμηλότερα ποσοστά αποφοίτησης, όπως φοιτητές πρώτης γενιάς και φοιτητές από οικογένειες χαμηλών κοινωνικοοικονομικών θέσεων.

Για την καταπολέμηση της μείωσης των ποσοστών διατήρησης για πληθυσμούς μαθητών που βρίσκονται σε κίνδυνο, πολλά ιδρύματα έχουν αναπτύξει προγράμματα καλοκαιρινών γεφυρών. Ιστορικά, αυτά τα προγράμματα έχουν σχεδιαστεί για να παρέχουν ακαδημαϊκή υποστήριξη και πληροφορίες σχετικά με τη ζωή στην πανεπιστημιούπολη, να προσανατολίζουν τους φοιτητές στη θεσμική κουλτούρα και να αναπτύσσουν την αυτοεκτίμηση και την αυτό-αποτελεσματικότητα των μαθητών που βρίσκονται σε κίνδυνο. Με απλά λόγια, «τα προγράμματα καλοκαιρινής γέφυρας προορίζονται να αντιμετωπίσουν σημαντικά κενά προετοιμασίας και επίτευξης που είναι εμφανή στην έρευνα σχετικά με τη διατήρηση και την επιμονή». Παρόλο που αυτά τα προγράμματα είναι εξαιρετικά ποικίλα ως προς το περιεχόμενο και την εφαρμογή τους, τα περισσότερα στοχεύουν στην ανάπτυξη των δεξιοτήτων μελέτης και διαχείρισης χρόνου των φοιτητών, στην ικανότητά τους να χρησιμοποιούν πανεπιστημιακές υπηρεσίες. Οι πρώιμες αξιολογήσεις και οι αξιολογήσεις των καλοκαιρινών προγραμμάτων γέφυρας ήταν σε μεγάλο βαθμό περιγραφικές, βασισμένες σε έρευνες ικανοποίησης μαθητών, ενώ πιο πρόσφατες έρευνες χρησιμοποίησαν δείγματα σύγκρισης μη συμμετεχόντων, διαχρονικές αναλύσεις και άλλες εμπειρικές τεχνικές.

Τα κύρια ερευνητικά ερωτήματα που καθοδηγούν αυτήν την παρούσα μελέτη είναι:

- Ερώτηση 1: Σε ποιο βαθμό, εάν υπάρχει, αλλάζουν οι προσανατολισμοί κινήτρων των μαθητών κατά τη διάρκεια ενός καλοκαιρινού προγράμματος γέφυρας;
- Ερώτηση 2: Ποιοι παράγοντες προβλέπουν τις αλλαγές στα κίνητρα, εάν υπάρχουν, που συμβαίνουν κατά τη διάρκεια ενός καλοκαιρινού προγράμματος γέφυρας;
- Ερώτηση 3: Ποια είναι η σχέση μεταξύ της χρήσης από τους συμβούλους ενός συστήματος έγκαιρης προειδοποίησης που υποστηρίζεται από αναλυτικά στοιχεία μάθησης και των ακαδημαϊκών επιδόσεων των μαθητών τους κατά τη διάρκεια ενός καλοκαιρινού προγράμματος γέφυρας;

Η συγκεκριμένη μελέτη είναι το αποτέλεσμα της συνεργασίας με το προσωπικό καλοκαιρινής γέφυρας. Οι συγγραφείς πιστεύουν ότι τέτοιες συνεργασίες είναι απαραίτητες

για την καλύτερη κατανόηση των διαφορετικών παραγόντων που επηρεάζουν τους προσανατολισμούς κινήτρων των μαθητών στο πλαίσιο του καλοκαιρινού μεταβατικού προγράμματος. Αυτοί οι παράγοντες μπορούν, με τη σειρά τους, να ενημερώσουν τους μελλοντικούς σχεδιασμούς των εργαλείων αναλυτικής μάθησης, έτσι ώστε τα νέα εργαλεία να περιλαμβάνουν μη γνωστικά καθώς και μέτρα ακαδημαϊκής απόδοσης για να βελτιώσουν τελικά τη μάθηση και τη διατήρηση των μαθητών.

Συμπερασματικά, αυτή η μελέτη διερεύνησε αλλαγές στους προσανατολισμούς ακαδημαϊκού κινήτρου των μαθητών κατά τη διάρκεια ενός προγράμματος γέφυρας και πώς χρησιμοποιήθηκε μια παρέμβαση βασισμένη σε μαθησιακά αναλυτικά στοιχεία από ακαδημαϊκούς συμβούλους για να ενημερώσουν τις πρόσωπο με πρόσωπο συναντήσεις τους με τους μαθητές. Τα αποτελέσματα της μελέτης δείχνουν ότι ο προσανατολισμός της γνώσης των μαθητών μειώθηκε κατά τη διάρκεια του προγράμματος και υποδεικνύουν ότι η έκθεση των μαθητών σε οθόνες της ακαδημαϊκής τους επίδοσης προβλέπει αρνητικά αυτήν την αλλαγή. Τα ευρήματα υποδεικνύουν ότι οι αντιλήψεις των μαθητών για τους στόχους τους και τις διαμορφωτικές επιδόσεις τους πρέπει να λαμβάνονται προσεκτικά υπόψη στο σχεδιασμό των παρεμβάσεων μαθησιακής ανάλυσης, καθώς τα εργαλεία που προκύπτουν μπορούν να επηρεάσουν τις ερμηνείες των μαθητών για τα δικά τους δεδομένα καθώς και την επακόλουθη ακαδημαϊκή τους επιτυχία.

Άρθρο 11

The Q-matrix Method: Mining Student Response Data for Knowledge

Barnes, T.

Η διδασκαλία με τη βοήθεια υπολογιστή έχει πολλά να προσφέρει στην επέκταση της προσβασιμότητας σε υψηλής ποιότητας εκπαίδευση. Ειδικότερα, είναι η ικανότητα προσαρμογής σε μεμονωμένους μαθητές που κάνει ελκυστικά τα ευφυή συστήματα διδασκαλίας. Ωστόσο, η πλειονότητα αυτών των συστημάτων απαιτεί την κατασκευή πολύπλοκων μοντέλων που ισχύουν μόνο για ένα συγκεκριμένο σεμινάριο σε ένα συγκεκριμένο πεδίο, απαιτώντας το χρόνο των ειδικών για τη δημιουργία και στη συνέχεια τη δοκιμή αυτών των μοντέλων σε μαθητές. Στην πραγματικότητα, αυτά είναι μόνο μερικά από τα μειονεκτήματα που αντιμετωπίζουν οι προγραμματιστές συστημάτων ITS(Intelligence Transportation Systems). Ένα συγκεκριμένο σύστημα, το REDEEM, κατασκευάστηκε για να βελτιώσει τον χρόνο που απαιτείται για τη δημιουργία ενός ITS και να επιτρέπει στους εκπαιδευτικούς να εφαρμόζουν τις δικές τους στρατηγικές διδασκαλίας σε ένα υπάρχον σύστημα εκπαίδευσης που βασίζεται σε υπολογιστή (CBT/Computer Based Training) και έχει αποδειχθεί πιο αποτελεσματικό από έναν ανθρώπινο δάσκαλο χωρίς εμπειρία στη βελτίωση των βαθμολογιών των μαθητών σε τεστ. Ένα άλλο σύστημα, το ASSERT, κατασκευάστηκε για να αντικαταστήσει την ανάγκη σχεδιασμού μοντέλων από εμπειρογνώμονες, χρησιμοποιώντας τελειοποίηση της θεωρίας για την εκμάθηση μοντέλων μαθητών από τη συμπεριφορά και τη δημιουργία ανατροφοδότησης για αποκατάσταση. Η μέθοδος q-matrix έχει τον ίδιο στόχο δηλαδή, να μπορεί να τροποποιήσει γρήγορα ένα υπάρχον εργαλείο CBT ώστε να είναι ένα απλό ITS. Η μέθοδος χρησιμοποιεί τεχνικές ανακάλυψης γνώσης και εξόρυξης δεδομένων για τη μοντελοποίηση της γνώσης των μαθητών και την άμεση αποκατάσταση της γνώσης. Σε αυτή την εργασία, οι συγγραφείς συζητάνε εν συντομία την ιστορία της μεθόδου q-matrix, περιγράφουν τον αλγόριθμο και τις αξιολογήσεις του q-matrix και παρουσιάζουν δύο πειράματα που καταδεικνύουν την εφαρμογή της μεθόδου στην κατανόηση της συμπεριφοράς των μαθητών και στην κατεύθυνση της μάθησης.

Σε αυτή την έρευνα, εφαρμόζεται η μέθοδος q-matrix σε μια μεγάλη ομάδα μαθητών και αναλύονται τα αποτελέσματα με διάφορους τρόπους. Οι υποθέσεις των συγγραφέων είναι:

- 1) οι ειδικοί και οι εξαγόμενοι πίνακες q θα διαφέρουν, αλλά οι εξαγόμενοι πίνακες q θα είναι χρήσιμοι για την ερμηνεία των δεδομένων μαθητή και
- 2) η μέθοδος q-matrix μπορεί να είναι αποτελεσματική στην καθοδήγηση της αποκατάστασης των μαθητών (τουλάχιστον τόσο αποτελεσματική όσο ένας μαθητής που κατευθύνει τη μάθησή του). Το σεμινάριο Binary Relations χορηγήθηκε σε περισσότερους από 100 μαθητές στο μάθημα των Διακριτών Μαθηματικών το φθινόπωρο του 2002, CSC 226, στο NC State University. Το σεμινάριο ήταν απαραίτητο για βαθμολόγηση στην τάξη, αλλά μόνο ένας βαθμός ολοκλήρωσης δόθηκε σε κάθε μαθητή. Οι μαθητές χωρίστηκαν τυχαία σε καθοδηγούμενες ή αυτοκαθοδηγούμενες ομάδες. Στην καθοδηγούμενη ομάδα, μετά τη λήψη ενός κουίζ, οι μαθητές κατευθύνονταν αυτόματα να εξετάσουν μια ερώτηση

που προσδιορίστηκε ότι ο μαθητής κατανοούσε το λιγότερο. Στην αυτοκαθοδηγούμενη ομάδα, ζητήθηκε από τους μαθητές να επιλέξουν ποια ερώτηση στο κουίζ κατανοούσαν λιγότερο, αλλά δεν ανακατευθύνθηκαν για να ελέγξουν κάποια συγκεκριμένα στοιχεία. Οι μαθητές και στις δύο ομάδες είχαν τη δυνατότητα να επανεξετάσουν όποια θέματα επιθυμούσαν είτε μετά από ανακατεύθυνση q-matrix είτε μετά από μια λίστα με την λιγότερο κατανοητή ερώτησή τους. Αφού οι μαθητές έλαβαν μέρος τα σεμινάρια, τα νέα δεδομένα τους χρησιμοποιήθηκαν για την εξαγωγή ενός νέου συνόλου πινάκων q, οι οποίοι συγκρίνονται με μοντέλα ειδικών, για τον έλεγχο της υπόθεσης ότι οι έννοιες των μαθητών δεν αντιστοιχούν πάντα στην ανάλυση των ειδικών. Αυτοί οι εξαγόμενοι πίνακες q εξετάζονται επίσης για να διερευνηθεί η υπόθεσή μας ότι μπορούν να χρησιμοποιηθούν για την κατανόηση της συμπεριφοράς των μαθητών. Τέλος, οι επιλογές που έκαναν οι μαθητές για αποκατάσταση συγκρίνονται με αυτές που έγιναν με τη μέθοδο q-matrix και παρουσιάζονται τα αποτελέσματα της έρευνας σχετικά με την αποκατάσταση.

Αυτή η έρευνα αντιπροσωπεύει μια αρχική μελέτη της αποτελεσματικότητας της χρήσης μιας τεχνικής εξόρυξης δεδομένων, της μεθόδου q-matrix, στην κατανόηση και την κατεύθυνση της μάθησης των μαθητών. Όπως είχε προβλεφθεί, οι ειδικοί και οι εξαγόμενοι πίνακες q δεν συμπίπτουν συχνά, αλλά οι απαντήσεις των μαθητών έγιναν κατανοητές με βάση τους εξαγόμενους πίνακες q. Οι συγγραφείς συγκρίναν επίσης τις ερωτήσεις που επέλεξαν να επανεξετάσουν οι αυτοκαθοδηγούμενοι μαθητές με τις ερωτήσεις που θα τους είχε επιλέξει η μέθοδος q-matrix. Διαπιστώθηκε ότι η μέθοδος του q-matrix συχνά επέλεγε τις ίδιες ερωτήσεις για επανεξέταση που επέλεξαν οι αυτοκαθοδηγούμενοι μαθητές για τον εαυτό τους. Βρέθηκε επίσης ότι οι μαθητές που επέλεξαν διαφορετική μέθοδο από τη μέθοδο του q-matrix θα μπορούσαν να έχουν ωφεληθεί από την ανασκόπηση μιας επιλεγμένης έννοιας του q-matrix. Ωστόσο, λόγω του μικρού μεγέθους δείγματος των αυτοκαθοδηγούμενων μαθητών, που επιλέγουν να επανεξετάσουν διαφορετικές ερωτήσεις από αυτές που θα είχαν προταθεί, οι οποίοι συμμετείχαν επίσης στην τελική εξέταση, αυτά τα ευρήματα είναι προκαταρκτικά και όχι οριστικά.

Άρθρο 12

Learning Log Database and Data Mining system for e-Learning -On-Line Statistical Outlier Detection of irregular learning processes-

Ueno, M., Nagaoka, K.

Πρόσφατα, η εξ αποστάσεως εκπαίδευση με τη χρήση ηλεκτρονικής μάθησης (E-Learning) είναι όλο και πιο δημοφιλής σε πραγματικές εκπαιδευτικές καταστάσεις. Ωστόσο, υπάρχει ένα πρόβλημα ότι η στρατηγική διδασκαλίας τείνει να είναι μονόδρομη, και έτσι μερικές φορές κάνει τους μαθητές να βαριούνται σε σύγκριση με τις συνήθεις μεθόδους διδασκαλίας. Για την επίλυση αυτών των προβλημάτων, είναι σημαντικό καθήκον να αναπτυχθεί μια νέα μεθοδολογία διδασκαλίας βασισμένη στην ηλεκτρονική μάθηση.

Από την άλλη πλευρά, είναι εύκολο να ληφθούν τεράστια δεδομένα ιστορικού μάθησης που έχουν αποθηκευτεί ως δεδομένα καταγραφής στο E-Learning. Σε αυτήν την περίπτωση, είναι σημαντικός ο τρόπος με τον οποίο αποθηκεύονται αυτά τα δεδομένα ή πώς χρησιμοποιούνται αυτά τα δεδομένα. Κάποιες έρευνες έχουν προταθεί από αυτή την άποψη. Ο Kawamura και ο Matsumoto πρότειναν μια μέθοδο ομαδοποίησης για μαθητές χρησιμοποιώντας δεδομένα καταγραφής e-Learning. Ο Matsui πρότεινε μια μέθοδο εξόρυξης δεδομένων που κατασκευάζει ένα δέντρο χρησιμοποιώντας τη μέθοδο ID3. Αυτό το άρθρο προτείνει μια μέθοδο διαδικτυακής ανίχνευσης ακραίων στοιχείων των ακανόνιστων μαθησιακών διαδικασιών των μαθητών χρησιμοποιώντας τα δεδομένα χρόνου απόκρισης των μαθητών για το περιεχόμενο της ηλεκτρονικής μάθησης. Τα μοναδικά χαρακτηριστικά αυτής της μεθόδου είναι τα εξής:

1. Προτείνει μια μέθοδο ανίχνευσης ακραίων τιμών χρησιμοποιώντας προγνωστική κατανομή Bayes.
2. Είναι διαθέσιμη για μικρό δείγμα,
3. Είναι βολική για να υπολογιστεί η προγνωστική κατανομή.
4. Η διαδικτυακή μάθηση πραγματοποιείται στον παγκόσμιο ιστό.
5. Βοηθά τη διδασκαλία με δύο τρόπους χρησιμοποιώντας αποτελέσματα εξόρυξης δεδομένων για τις διαδικασίες μάθησης των μαθητών.

Το σύστημα χρησιμοποιήθηκε για πραγματικές τάξεις. Τα αποτελέσματα δείχνουν την αποτελεσματικότητα του συστήματος. Οι συγγραφείς έχουν αναπτύξει ένα σύστημα πλατφόρμας ηλεκτρονικής μάθησης. Το σύστημα πλατφόρμας αποτελείται από :

1. Contents Presentation System (CPS),
2. Contents Database (CD),
3. Learning Histories Database (LHD), and
4. Data Mining System (DMS).

Το CPS ενσωματώνει διάφορα είδη περιεχομένων και παρουσιάζει τις ενσωματωμένες πληροφορίες στην ιστοσελίδα. Επιπλέον, το σύστημα παρουσιάζει ορισμένα τεστ που

επιβεβαιώνουν το βαθμό κατανόησης των μαθητών αμέσως μόλις ολοκληρωθεί το περιεχόμενο. Το CD είναι μια βάση δεδομένων που αποτελείται από διάφορα είδη μέσων, κείμενο, jpeg, mpeg και ούτω καθεξής. Η προτεινόμενη πλατφόρμα παρακολουθεί τις διαδικασίες μάθησης των μαθητών και τις αποθηκεύει ως δεδομένα καταγραφής στο LHD. Αρχικά, ο καθηγητής φτιάχνει τα περιεχόμενα που αφορούν τη διάλεξή του και τα αποθηκεύει στο CD. Στη συνέχεια, το CPS ενσωματώνει αυτόματα τα περιεχόμενα και τα παρουσιάζει στους μαθητές. Οι μαθητές μπορούν να τα μάθουν μέσω του Διαδικτύου. Τα δεδομένα καταγραφής ιστορικού μάθησης των μαθητών αποθηκεύονται στο LHD και αναλύονται στο DMS.

Το DMS παρουσιάζει τις ανατροφοδοτήσεις για τους εκπαιδευόμενους και τον καθηγητή αντίστοιχα. Ο καθηγητής μπορεί να γνωρίζει πληροφορίες σχετικά με τις διαδικασίες μάθησης των μαθητών και μπορεί να δώσει κάποια σχόλια ή οδηγίες στους μαθητές χρησιμοποιώντας e-mail. Η πλατφόρμα παρακολουθεί τις διαδικασίες μάθησης των μαθητών και τις αποθηκεύει ως δεδομένα καταγραφής στο LHD.

Άρθρο 13

Patterns of User Activity in the Different Features of the Blackboard CMS across All Courses for an Academic Year at Blackboard CMS across All Courses for an Academic Year at Brigham Young University Brigham Young University

Griffiths, M.E., Graham, C.R.

Στο παρελθόν, τα συστήματα διαχείρισης μαθημάτων (CMS/Course Management Systems) έχουν γίνει ένα αναπόσπαστο μέρος των περισσότερων ιδρυμάτων τριτοβάθμιας εκπαίδευσης. Ενώ έχει γίνει σημαντικός όγκος έρευνας που εξετάζει τη χρήση CMS σε επίπεδο μαθημάτων, έχει υπάρξει λιγότερη έρευνα που εξετάζει τη χρήση CMS σε επίπεδο ιδρύματος από την προοπτική της βάσης δεδομένων δραστηριότητας CMS. Αυτό το άρθρο μοιράζεται τα ευρήματα από μια μελέτη που αναλύει πάνω από 36 εκατομμύρια κλικ μαθητών και εκπαιδευτών εντός του Blackboard CMS σε ολόκληρη την πανεπιστημιούπολη του Πανεπιστημίου Brigham Young (BYU) κατά τη διάρκεια ενός πλήρους ακαδημαϊκού έτους. Αυτή η έρευνα αναφέρει συνολικά επίπεδα δραστηριότητας χρηστών σε όλη την πανεπιστημιούπολη, μοτίβα δραστηριότητας χρηστών στις διαφορετικές λειτουργίες του Blackboard και μοτίβα δραστηριότητας χρηστών στις διαφορετικές λειτουργίες του Blackboard σε διαφορετικά κολέγια. Η μελέτη συγκρίνει τα επίπεδα δραστηριότητας για χαρακτηριστικά CMS Blackboard που είναι κυρίως παιδαγωγικοί στίχοι, κυρίως διοικητικού χαρακτήρα.

Τα πρότυπα που συνοψίζονται σε αυτό το άρθρο παρέχουν ένα ορισμένο επίπεδο πληροφοριών από μια προοπτική επισκόπησης που μπορεί να χρησιμοποιηθεί για σκοπούς συζήτησης υψηλού επιπέδου, για μελλοντικές ερευνητικές συστάσεις και ως βασικό σύνολο στατιστικών που μπορούν να χρησιμοποιηθούν για σύγκριση με μελλοντικές μελέτες αυτού του είδους. Αυτή η ανάλυση δραστηριότητας CMS υψηλού επιπέδου μπορεί επίσης να ενημερώσει και να δημιουργήσει ερευνητικά ερωτήματα για τις πιο εμπειριστατωμένες μελέτες μεμονωμένων μαθημάτων. Αν και η προβολή υψηλού επιπέδου της δραστηριότητας CMS που αναφέρεται σε αυτή τη μελέτη δεν μπορεί να παράγει σταθερά συμπεράσματα σχετικά με τη μάθηση των μαθητών, τα μοτίβα που εμφανίζονται σε αυτό το επίπεδο είναι χρήσιμα για την καθοδήγηση και την ενημέρωση των λεπτομερών περιπτώσιολογικών μελετών της χρήσης ενός CMS από μαθητή και εκπαιδευτή. Είναι αδύνατο, για παράδειγμα, να γνωρίζουμε τις προθέσεις και τις παιδαγωγικές μεθόδους που χρησιμοποιούνται από τους εκπαιδευτές όλων των τμημάτων της τάξης που χρησιμοποιούν Blackboard, και μια λεπτομερής παρατήρηση και ανάλυση μερικών μόνο τμημάτων της τάξης θα ήταν ελάχιστα χρήσιμη στο πλαίσιο της ανάλυσης δεδομένων δραστηριότητας σε όλη την πανεπιστημιούπολη.

Αν και δεν μπορούν να γίνουν ακριβή συμπεράσματα χωρίς ένα πιο λεπτομερές επίπεδο ανάλυσης, υπάρχουν ορισμένες πιθανές επιπτώσεις που συζητούνται σε αυτή τη μελέτη. Η ανάλυση της χρήσης χαρακτηριστικών, η οποία δείχνει ότι τα έξι κορυφαία χαρακτηριστικά αντιπροσωπεύουν το 90% της συνολικής δραστηριότητας, μπορεί να υποδηλώνει ότι υπάρχει ανάγκη βελτίωσης της εκπαίδευσης του διδακτικού προσωπικού για τα λιγότερο

χρησιμοποιούμενα χαρακτηριστικά του CMS. Μπορεί επίσης να διαπιστωθεί ότι οι λιγότερο χρησιμοποιούμενες λειτουργίες είναι πιο δύσκολο να εφαρμοστούν και απαιτούν κάποια αναθεώρηση από τους παραγωγούς CMS.

Η υψηλή διαφοροποίηση στη χρήση των παιδαγωγικών χαρακτηριστικών σε διαφορετικά ιδρύματα μπορεί να προτείνει ότι το διδακτικό προσωπικό σε διαφορετικά ιδρύματα χρειάζεται περισσότερη κατάρτιση στο πώς να χρησιμοποιεί άλλα παιδαγωγικά χαρακτηριστικά που μπορεί επίσης να είναι χρήσιμα για την πειθαρχία τους.

Για να δημιουργηθούν ακριβέστερα συμπεράσματα, η έρευνα που βασίζεται σε πρότυπα που συζητούνται σε αυτή τη μελέτη θα πρέπει να διεξάγεται σε πιο λεπτομερές επίπεδο όπου υπάρχουν διαθέσιμα δεδομένα σε επίπεδο σπουδαστών και όχι σε δεδομένα επιπέδου μαθήματος. Συνιστάται η έρευνα δεδομένων σε επίπεδο μαθητή που ενημερώνεται από τα αποτελέσματα αυτής της μελέτης, ώστε να μπορούν να αναλυθούν οι αποκλίσεις και οι τυπικές αποκλίσεις εντός των τμημάτων των μαθημάτων και τα λεπτομερή πρότυπα δραστηριότητας των μαθητών να οδηγήσουν σε μεγαλύτερη κατανόηση των στρατηγικών και των συνηθειών μάθησης των μαθητών. Αυτό το επίπεδο έρευνας θα παράγει ένα λεπτομερές σύνολο αποτελεσμάτων που θα μπορούσαν ενδεχομένως να οδηγήσουν στη διαμόρφωση της χρήσης του Blackboard και άλλων CMS για τη μεγιστοποίηση των εκπαιδευτικών οφελών για τους μαθητές. Απαιτείται επίσης περαιτέρω έρευνα για να απαντηθεί το συνολικό ερώτημα της μέτρησης του οφέλους ενός CMS σε ένα ίδρυμα τριτοβάθμιας εκπαίδευσης.

Άρθρο 14

Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining

Nasiri, M., Minaei, B.

Σε αυτό το άρθρο, οι συγγραφείς περιγράφουν μια μελέτη περίπτωσης εξόρυξης εκπαιδευτικών δεδομένων με βάση τα δεδομένα που συλλέχθηκαν από το σύστημα διαχείρισης μάθησης (LMS) του κέντρου ηλεκτρονικής μάθησης και του ηλεκτρονικού εκπαιδευτικού συστήματος του Πανεπιστημίου Επιστήμης και Τεχνολογίας του Ιράν (IUST). Ο κύριος στόχος τους είναι να απεικονίσουν τις εφαρμογές του EDM στον τομέα της ηλεκτρονικής μάθησης και των διαδικτυακών μαθημάτων, εφαρμόζοντας ένα μοντέλο πρόβλεψης ακαδημαϊκών απολύσεων και επίσης Μέσο Όρο Βαθμών (GPA) των αποφοίτων φοιτητών. Η παρακολούθηση και η υποστήριξη νεοεισαχθέντων και πρωτοετών φοιτητών θεωρείται πολύ σημαντική σε πολλά εκπαιδευτικά ιδρύματα. Συνεπώς, εάν υπάρχουν μερικοί τρόποι εκτίμησης της πιθανότητας απόλυσης, εγκατάλειψης και άλλων προκλήσεων στη διαδικασία της αποφοίτησης, καθώς και ικανά εργαλεία για την πρόβλεψη του GPA ανά εξάμηνο, οι εργαζόμενοι του πανεπιστημίου θα μπορούν να σχεδιάσουν και να βελτιώσουν πιο αποτελεσματικές στρατηγικές για εκπαιδευτικά συστήματα ειδικά για ηλεκτρονικά εκπαιδευτικά συστήματα που περιλαμβάνουν λιγότερο γνωστά και πιο περίπλοκα προβλήματα. Για την επίτευξη του παραπάνω στόχου, έχει χρησιμοποιηθεί μια κοινή μεθοδολογία εξόρυξης δεδομένων που ονομάζεται CRISP. Τα αποτελέσματά των συγγραφέων δείχνουν ότι μπορούν να υπάρχουν μοντέλα με αυτοπεποίθηση για την πρόβλεψη των εκπαιδευτικών χαρακτηριστικών. Επί του παρόντος, υπάρχει αυξανόμενο ενδιαφέρον για την εξόρυξη δεδομένων και τα εκπαιδευτικά συστήματα, καθιστώντας την εξόρυξη εκπαιδευτικών δεδομένων ως μια νέα αναπτυσσόμενη ερευνητική κοινότητα.

Στην φάση της αρχικής συλλογής δεδομένων και την εξοικείωση και τον εντοπισμό προβλημάτων ποιότητας των δεδομένων οι συγγραφείς αναφέρουν πως, δεδομένου ότι το ηλεκτρονικό εκπαιδευτικό σύστημα του IUST έχει σχεδιαστεί μέσω μιας σχεσιακής βάσης δεδομένων και έχει υλοποιηθεί από Microsoft SQL Server, έπρεπε να αναλύσουν τις εγγραφές των πινάκων της βάσης δεδομένων και να κατανοήσουν τις μεταξύ τους σχέσεις. Στο έργο τους χρησιμοποιήθηκαν 3 κύριοι πίνακες, καθένας από τους οποίους έχει ορισμένα πεδία δεδομένων μαθητών, όπως ταυτότητα, εθνικότητα, ηλικία, εκπαιδευτικό ιστορικό, μαθήματα που πέρασαν, μαθήματα που απέτυχαν και διαγραμμένα μαθήματα. Επιπλέον, ορισμένα άλλα αναλυτικά γραφήματα και στατιστικά γραφήματα έχουν δημιουργηθεί από το SPSS Clementine 12.0 για καλύτερη αντίληψη των σχέσεων μεταξύ των πεδίων.

Στο κομμάτι της προετοιμασίας των δεδομένων αρχικά, χρειάστηκε να εξαγάγουν φοιτητές ηλεκτρονικής μάθησης από τους υπόλοιπους φοιτητές, καθώς όλα τα δεδομένα των φοιτητών αποθηκεύονται σε μια μοναδική βάση δεδομένων, συμπεριλαμβανομένων των φοιτητών που σπουδάζουν σε παραδοσιακά συστήματα και εκείνων που σπουδάζουν σε υβριδικά συστήματα. Χρησιμοποιώντας SQL queries, αυτό έγινε σε πρώτο στάδιο. Είναι επίσης σημαντικό να γίνει ανίχνευση του θορύβου στα δεδομένα που θα προκαλούσαν σφάλματα

αργότερα. Δεν ανιχνεύθηκε θόρυβος σε αυτό το στάδιο. Στη συνέχεια, στο στάδιο επιλογής χαρακτηριστικών, επιλέχθηκαν 13 πεδία που θα χρησιμοποιηθούν στη δημιουργία του μοντέλου. Πραγματοποιημένα μαθήματα, μη ολοκληρωμένα μαθήματα, αποτυχημένα και διαγραμμένα μαθήματα ήταν μεταξύ αυτών.

Επιπλέον, για να αποφευχθεί η υπερβολική προσαρμογή που προκαλεί αξιοσημείωτα σφάλματα στα αποτελέσματα, τρεις κατηγορίες απολυμένων μαθητών συγχωνεύθηκαν σε μία κατηγορία. Υπάρχουν πολλοί λόγοι για τους μαθητές που εγκαταλείπουν ημιτελή την εκπαίδευση. Στην ακαδημαϊκή απόλυση προστίθενται αυτολογισμοί αποχώρησης και κάποιες άλλες αιτίες γιατί θεωρούνται μειοψηφία σε σύγκριση με αυτήν.

Αυτή η εργασία έχει παρουσιάσει ανάλυση πραγματικών δεδομένων ενός συστήματος e-Learning. Η παλινδρόμηση και ο αλγόριθμος ταξινόμησης C5.0 αποκαλύπτουν ορισμένες αδυναμίες στη χρήση του EDM, καθώς εξαρτώνται σε μεγάλο βαθμό από την κατανομή των δεδομένων. Επομένως, μικρές διακυμάνσεις στα δεδομένα θα μπορούσαν να προκαλέσουν διαφορετικά συμπεράσματα. Σύμφωνα με αυτό το σημείο, η μελλοντική εργασία μπορεί να περιλαμβάνει τη βελτίωση αυτών των τεχνικών με άλλες μεθόδους εξόρυξης δεδομένων, όπως κανόνες συσχέτισης και εξόρυξη ακολουθιών. Ωστόσο, τα αποτελέσματα μιας τέτοιας εξόρυξης περιπτώσεων μπορούν να ανιχνεύσουν νέους τρόπους υποστήριξης και παρακολούθησης εκπαιδευτικών διαδικασιών σε πολύπλοκα συστήματα διαδικτυακών μαθημάτων. Τελικά, είναι σημαντικό να αξιολογηθεί και να επικυρωθεί η προτεινόμενη ανάλυση με άλλους τύπους δεδομένων μαζί με την κοινή χρήση των αποτελεσμάτων με τους ειδικούς για να ληφθούν τα σχόλιά τους. Υπό αυτή την έννοια, η μελλοντική εργασία μπορεί επίσης να περιλαμβάνει την εφαρμογή αυτής της ανάλυσης για δεδομένα διαφορετικών εκπαιδευτικών συστημάτων ώστε να βελτιωθούν οι μέθοδοι.

Άρθρο 15

Preventing Student Dropout in Distance Learning Using Machine Learning Techniques

Kotsiantis, S.B., Pierrakeas, C., Pintelas, P.E.

Ο περιορισμός της εγκατάλειψης είναι ζωτικής σημασίας στη διαδικτυακή μάθηση σε πανεπιστημιακό επίπεδο, επειδή τα ποσοστά εγκατάλειψης εκεί είναι αναμφίβολα μεγαλύτερα από εκείνα στα παραδοσιακά πανεπιστήμια. Επαγγελματικοί, ακαδημαϊκοί, υγειονομικοί, οικογενειακοί και προσωπικοί παράγοντες συμβάλλουν στην εγκατάλειψη, η οποία ποικίλλει ανάλογα με το εκπαιδευτικό μοντέλο που χρησιμοποιείται από το ίδρυμα που προσφέρει εξ αποστάσεως εκπαίδευση και τον επιλεγμένο τομέα σπουδών. Ανεξάρτητα από την απόσταση μεταξύ αυτών και των μαθητών τους σε ένα μάθημα εξ αποστάσεως εκπαίδευσης, οι δάσκαλοι πρέπει συνεχώς να τους ενθαρρύνουν. Ως εκ τούτου, ένα πρόγραμμα που θα μπορούσε να εντοπίσει αυτόματα μαθητές που έχουν μεγάλη πιθανότητα να εγκαταλείψουν το σχολείο (μαθητές που είναι επιρρεπείς στην εγκατάλειψη) είναι ζωτικής σημασίας, ώστε οι καθηγητές να μπορούν να λάβουν προληπτικά μέτρα για να μειώσουν τη διαρροή μαθητών.

Ο πρωταρχικός στόχος αυτής της έρευνας ήταν να προσδιορίσει εάν η χρήση προσεγγίσεων μηχανικής μάθησης σε αυτό το ζήτημα θα ήταν επωφελής. Ορισμένες μελέτες στερούνται της ικανότητας πρόβλεψης των μαθητών που είναι επιρρεπείς στην εγκατάλειψη και μπορούν να συσχετίσουν στατιστικά μόνο συγκεκριμένες δημογραφικές μεταβλητές μαθητών με τους μαθητές που εγκαταλείπουν το σχολείο. Για το σκοπό αυτό, συμπεριλάβαμε μια ποικιλία αλγορίθμων μάθησης στα δεδομένα που παρέχει το μάθημα «Πληροφορική» του Ελληνικού Ανοικτού Πανεπιστημίου. Στην πραγματικότητα, έχει αποδειχθεί ότι οι αλγόριθμοι μάθησης μπορούν να βοηθήσουν τους καθηγητές να εντοπίζουν με υψηλό βαθμό ακρίβειας μαθητές που είναι επιρρεπείς στην εγκατάλειψη ακόμη και στην αρχή του ακαδημαϊκού έτους χρησιμοποιώντας απλώς δημογραφικές πληροφορίες από τους μαθητές. Φυσικά, αυτή η ακρίβεια αυξάνεται όταν εισάγονται νέες πληροφορίες για το πρόγραμμα σπουδών κατά τη διάρκεια του ακαδημαϊκού έτους.

Ο πιο ακριβής αλγόριθμος για αυτό το πρόβλημα ήταν ο δεύτερος στόχος αυτής της μελέτης. Η Machine Learning κοινότητα φιλοδοξεί να είναι σε θέση να επιλέξει ή να δημιουργήσει τον καλύτερο αλγόριθμο που μπορεί να εφαρμοστεί σε όλες τις περιπτώσεις χρήσης και τομείς. Η θεωρητική έρευνα και τα πειραματικά ευρήματα, ωστόσο, δείχνουν ότι αυτό δεν είναι εφικτό. Επομένως, αντιπαραβάλαμε αυτούς τους αλγόριθμους μάθησης αιχμής στην πρόβλεψη της εγκατάλειψης των εισερχόμενων μαθητών. Ο αλγόριθμος Naive Bayes είχε την καλύτερη ακρίβεια μεταξύ των αλγορίθμων που εξετάστηκαν, κάτι που ήταν ένα ενδιαφέρον εύρημα. Τέλος, οι συγγραφείς δημιούργησαν ένα λειτουργικό πρωτότυπο ενός βοηθητικού προγράμματος που βασίζεται στο web για να βοηθήσουν αυτόν τον αλγόριθμο.

Τα δεδομένα που χρησιμοποιήθηκαν ως εκπαιδευτικό σετ για τη μελέτη μας προήλθαν από το μάθημα «Πληροφορική» του Ελληνικού Ανοικτού Πανεπιστημίου. Το Ελληνικό Ανοικτό Πανεπιστήμιο παρέχει εκπαίδευση πανεπιστημιακού επιπέδου μέσω της εξ αποστάσεως

εκπαίδευσης. Η ενότητα, η οποία ισοδυναμεί με τρεις ή τέσσερις διαλέξεις διάρκειας ακαδημαϊκού εξαμήνου, χρησιμεύει ως η βασική εκπαιδευτική μονάδα για το μάθημα «Πληροφορική» του ΕΑΠ καθώς και για οποιοδήποτε άλλο μεταπτυχιακό ή μεταπτυχιακό πρόγραμμα. Υπάρχουν 12 ενότητες στο μάθημα «Πληροφορική» (INF), το οποίο καταλήγει στην απόκτηση πτυχίου. Για το πρόγραμμα σπουδών «Εισαγωγή στην Πληροφορική» έχουν συγκεντρωθεί συνολικά 354 αρχεία μαθητών.

Σε αυτή την εργασία, έχει αποδειχθεί ότι οι αλγόριθμοι μάθησης προβλέπουν με ακρίβεια τη νέα εγκατάλειψη μαθητών, καθιστώντας τους ένα πολύτιμο εργαλείο στις προσπάθειες αποφυγής και στη συνέχεια μείωσης των ποσοστών εγκατάλειψης. Όταν γίνονται αρχικές προβλέψεις χρησιμοποιώντας μόνο τα δημογραφικά στοιχεία των μαθητών, η ακρίβεια φτάνει στο 63% και ανεβαίνει σε πάνω από 83% πριν από τα μισά του ακαδημαϊκού έτους. Η μέθοδος Naive Bayes βρέθηκε να είναι η καταλληλότερη αφού συγκρίθηκαν έξι αλγόριθμοι. Χρησιμοποιώντας αυτήν την τεχνική, δημιουργήθηκε μια δοκιμαστική έκδοση μιας διαδικτυακής υπηρεσίας βοήθειας που μπορεί να εντοπίσει μαθητές με υψηλό κίνδυνο εγκατάλειψης. Η συλλογή των δεδομένων τους, προέρχεται από το πρόγραμμα σπουδών «Εισαγωγή στην Πληροφορική», ωστόσο η πλειονότητα των συμπερασμάτων είναι κοινά και ισχύουν για την πλειονότητα των πανεπιστημιακών προγραμμάτων ηλεκτρονικής μάθησης. Μια διαδικασία επιλογής wrapper attributes και ο αλγόριθμος Naive Bayes θα χρησιμοποιηθούν σε μια μελλοντική μελέτη για την επίτευξη της μέγιστης ακρίβειας πρόβλεψης με τα λιγότερα εφικτά χαρακτηριστικά.

Άρθρο 16

A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition

Dech Thammasiri , Dursun Delen, Phayung Meesad , Nihat Kasap

Η αύξηση των ποσοστών της διατήρησης των φοιτητών είναι ένας μακροπρόθεσμος στόχος των πανεπιστημίων σε όλο τον κόσμο. Οι αρνητικές επιπτώσεις της φοιτητικής φθοράς είναι εμφανείς στους μαθητές, στους γονείς, στο πανεπιστήμιο και στην κοινωνία συνολικά. Ο θετικός αντίκτυπος της αυξημένης διατήρησης είναι επίσης προφανής: οι απόφοιτοι πανεπιστημίου είναι πιο πιθανό να έχουν καλύτερη καριέρα και υψηλότερο επίπεδο ζωής. Οι κατατάξεις των πανεπιστημίων, οι ομοσπονδιακοί φορείς χρηματοδότησης, οι κρατικές επιτροπές πιστώσεων και οι οργανισμοί διαπίστευσης προγραμμάτων ενδιαφέρονται για τα ποσοστά διατήρησης φοιτητών. Υψηλότερο ποσοστό διατήρησης, περισσότερες πιθανότητες για το ίδρυμα να καταταχθεί υψηλότερα, να εξασφαλίσει περισσότερα ομοσπονδιακά κεφάλαια, να διαπραγματευτεί ευνοϊκά για ιδιοποίηση και να έχει ευκολότερο δρόμο για τις διαπιστεύσεις προγραμμάτων. Εξαιτίας όλων αυτών των λόγων, οι διαχειριστές στην τριτοβάθμια εκπαίδευση αισθάνονται όλο και μεγαλύτερη πίεση για να σχεδιάσουν και να εφαρμόσουν στρατηγικές πρωτοβουλίες για την αύξηση των ποσοστών διατήρησης φοιτητών. Επιπλέον, τα πανεπιστήμια με υψηλά ποσοστά φθοράς αντιμετωπίζουν σημαντική απώλεια διδασκτρών και πιθανών συνεισφορών αποφοίτων. Ένα σημαντικό μέρος της φθοράς των φοιτητών συμβαίνει στο πρώτο έτος φοίτησης. Ως εκ τούτου, είναι σημαντικό να εντοπιστούν οι ευάλωτοι φοιτητές που είναι επιρρεπείς να εγκαταλείψουν τις σπουδές τους στο πρώτο έτος. Ο εντοπισμός των πρωτοετών φοιτητών σε κίνδυνο μπορεί να επιτρέψει στα ιδρύματα να προχωρήσουν καλύτερα και ταχύτερα προς την επίτευξη των στόχων διαχείρισης διατήρησης. Πολλές μέθοδοι μοντελοποίησης βρέθηκαν για να βοηθήσουν τα ιδρύματα να προβλέψουν τους μαθητές σε κίνδυνο, να σχεδιάσουν παρεμβάσεις, να κατανοήσουν καλύτερα και να αντιμετωπίσουν θεμελιώδη ζητήματα που προκαλούν εγκατάλειψη μαθητών και, τελικά, να αυξήσουν τα ποσοστά διατήρησης μαθητών.

Σε αυτό το άρθρο, οι συγγραφείς περιγράφουν την προσέγγισή τους στην κατασκευή ταξινομητών από μη ισορροπημένα σύνολα δεδομένων. Ένα σύνολο δεδομένων είναι ανισορροπημένο εάν οι κλάσεις (δηλαδή οι κατηγορίες ταξινόμησης) δεν αντιπροσωπεύονται σχεδόν εξίσου. Τα περισσότερα σύνολα δεδομένων του πραγματικού κόσμου είναι ανισορροπημένα, περιέχουν μεγάλο αριθμό κανονικών/αναμενόμενων παραδειγμάτων με μόνο ένα μικρό ποσοστό ακανόνιστων/απροσδόκητων παραδειγμάτων. Συχνά, αυτό που έχει ενδιαφέρον είναι η αναγνώριση/πρόβλεψη των ακανόνιστων/απροσδόκητων παραδειγμάτων. Οι τεχνικές μηχανικής μάθησης δεν είναι πολύ καλές στο να διακρίνουν/προβλέπουν λιγότερο αντιπροσωπευτική τάξη σε μη ισορροπημένα σύνολα δεδομένων. Ως εκ τούτου, απαιτείται μια εργασία εξισορρόπησης δεδομένων ως μέρος της φάσης προεπεξεργασίας δεδομένων. Σε αυτή τη μελέτη, συγκρίνονται τρεις τεχνικές εξισορρόπησης δεδομένων χρησιμοποιώντας τέσσερις δημοφιλείς μεθόδους ταξινόμησης μαζί με ένα μεγάλο σύνολο δεδομένων πραγματικού κόσμου, πλούσιο σε χαρακτηριστικά. Για να επιτύχουν, τα έργα διατήρησης φοιτητών πανεπιστημίου θα πρέπει να ακολουθούν μια διαδικασία πολλαπλών

φάσεων, η οποία μπορεί να ξεκινά με τον εντοπισμό, την αποθήκευση (σε βάσεις δεδομένων) και τη χρήση δεδομένων/χαρακτηριστικών σπουδαστών για την καλύτερη κατανόηση της υποκείμενης αιτίας και την πρόβλεψη των μαθητών που βρίσκονται σε κίνδυνο είναι πιθανό να εγκαταλείψουν το σχολείο και τελειώνει με την ανάπτυξη αποτελεσματικών και αποδοτικών μεθόδων παρέμβασης για τη διατήρησή τους. Σε μια τέτοια διαδικασία, τα αναλυτικά στοιχεία μπορούν να παίξουν τον πιο κρίσιμο ρόλο στον ακριβή προσδιορισμό των μαθητών με την υψηλότερη τάση να εγκαταλείψουν το πανεπιστήμιο καθώς και στην εξήγηση των παραγόντων που κρύβονται πίσω από το φαινόμενο. Επειδή οι μέθοδοι μηχανικής μάθησης (όπως αυτές που χρησιμοποιούνται σε αυτή τη μελέτη) είναι ικανές να μοντελοποιούν εξαιρετικά μη γραμμικές σχέσεις, πιστεύεται ότι είναι πιο κατάλληλες τεχνικές για την πρόβλεψη της περίπλοκης φύσης της φθοράς των μαθητών με υψηλό επίπεδο ακρίβειας. Τα αποτελέσματα αυτής της μελέτης δείχνουν ότι, εάν εφαρμοστούν κατάλληλες μέθοδοι προεπεξεργασίας σε αρκετά μεγάλα σύνολα δεδομένων με το πλούσιο σύνολο μεταβλητών, οι μέθοδοι ανάλυσης είναι ικανές να προβλέψουν τη φθορά των πρωτοετών φοιτητών με υψηλό επίπεδο ακρίβειας (έως και 90%).

ΚΕΦΑΛΑΙΟ 4

Σύγκριση Βιβλιογραφίας

Στο συγκεκριμένο κεφάλαιο γίνεται σύγκριση της βιβλιογραφίας που αναφέρεται στο προηγούμενο κεφάλαιο με σκοπό να δείξω την μέθοδο εξόρυξης δεδομένων που χρησιμοποιούν οι συγγραφείς των άρθρων. Συγκεκριμένα, τα άρθρα (1), (2), (14), (15), (16), χρησιμοποιούν την μέθοδο του Classification.

Τα άρθρα (3), (6), (8), (14), χρησιμοποιούν την μέθοδο του Clustering, τα άρθρα (4) και (10) χρησιμοποιούν Visual Data Mining, τα άρθρα (5), (10) και (13) χρησιμοποιούν την μέθοδο Statistics, τα (6) και (8) χρησιμοποιούν Association Rule Mining, το (7) χρησιμοποιεί Regression, το (8) χρησιμοποιεί Sequential Pattern Mining, το (9) χρησιμοποιεί Text Mining, το (11) χρησιμοποιεί Correlation Mining, και το (12) χρησιμοποιεί Outlier Detection.

Παρατηρείται λοιπόν, πως πολλές εφαρμογές χρησιμοποιούν και περισσότερες απο μια μεθόδους εξόρυξης δεδομένων όπως το άρθρο (8) που χρησιμοποιεί 3 μεθόδους, Clustering, Association Rule Mining, Sequential Pattern Mining.

Παρατηρείται επίσης ότι απο τα επιλεγμένα άρθρα η πιο η πιο χρησιμοποιούμενη μέθοδος εξόρυξης δεδομένων είναι αυτή του Classification, ακολουθούμενη απο το Clustering, Visual Data Mining και Statistics.

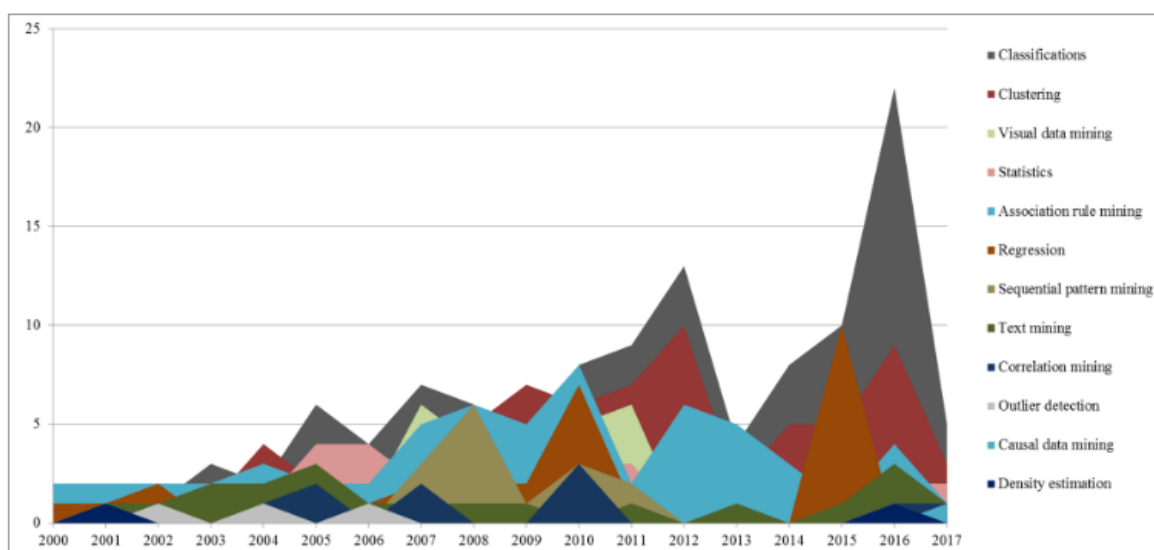
Ωστόσο το δείγμα των άρθρων που εξετάστηκαν στην συγκεκριμένη εργασία, είναι αρκετά μικρό ώστε να εξαχθεί κάποιο συμπέρασμα για την πιο διαδεδομένη μέθοδο εξόρυξης δεδομένων. Επιπλέον, υπάρχουν μέθοδοι οι οποίες δεν εξετάστηκαν σε αυτή την εργασία καθώς δεν υπάρχει όγκος άρθρων με εφαρμογές στην εκπαίδευση. Μερικές απ' τις μεθόδους αυτές είναι το Causal Mining και το Density Estimation.

Σύμφωνα με τους Aldowah, Al-Samarraie και Fauzy (2019) [6] οι οποίοι εργάστηκαν με παρόμοιο τρόπο και ανέλυσαν 402 άρθρα που αφορούν το Educational Data Mining και τα Learning Analytics, παρατήρησαν ότι οι μέθοδοι Classification, Clustering, Visual Data Mining και Statistics, παρουσιάζονται πιο συχνά απ' ότι οι άλλες μέθοδοι, με ποσοστό εμφάνισης 26.25%, 21.25%, 15% και 14.25% αντίστοιχα. Αναφέρουν επίσης ότι ορισμένες μέθοδοι, όπως η sequential pattern mining, η text mining, η correlation mining, η outlier detection, η causal mining, και η density estimation δεν χρησιμοποιούνται συνήθως λόγω της πολυπλοκότητας στην απόκτηση των απαραίτητων χαρακτηριστικών για τη ρύθμιση ή προσαρμοστεί στις ατομικές ανάγκες.

Το είδος του προβλήματος που πρέπει να λυθεί επηρεάζει σε μεγάλο βαθμό την τεχνική που χρησιμοποιείται ή εφαρμόζεται. Για παράδειγμα, πολλές μελέτες έχουν επικεντρωθεί στη δημιουργία εφαρμογών EDM που χρησιμοποιούν τεχνικές όπως classification, clustering,

regression, association rule mining, visual data mining και text mining για την πρόβλεψη, την παρακολούθηση και την αξιολόγηση της απόδοσης των μαθητών.

Η ανάλυση της μάθησης και της αλληλεπίδρασης των μαθητών σε διάφορες συνεργατικές δραστηριότητες, καθώς και η παροχή ευκαιριών μάθησης που ενσωματώνουν τις προηγούμενες γνώσεις των μαθητών για το αντικείμενο, έχουν γίνει όλα συχνά χρησιμοποιώντας τεχνικές όπως classification, clustering, association rule mining, text mining, visual data mining, και statistics.



Εικόνα 4.1: Η ανάπτυξη στην εφαρμογή τεχνικών εξόρυξης δεδομένων όλα αυτά τα χρόνια. Aldowah, Al-Samarraie και Fauzy (2019) [6]

Στην επόμενη σελίδα ακολουθεί σύγκριση των άρθρων που χρησιμοποιήθηκαν στην παρούσα εργασία.

Πίνακας 4.1: Σύγκριση Μεθόδων DM των άρθρων

ID	Συγγραφείς	Τίτλος Άρθρου	Μέθοδος
1	Fernando Martínez Abad & Alicia A. Chaparro Caso López	Data-mining techniques in detecting factors linked to academic achievement	Classification
2	Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T., Popelinsky, L	Predicting drop-out from social behaviour of students	Classification
3	Cerezo, R., Sánchez-Sántillán, M., Paule-Ruiz, M.P., Núñez, J.C	Students' LMS interaction patterns and their relationship with achievement: A case study in higher education	Clustering
4	Kordík, P., Kuznetsov, S.	Mining Skills from Educational Data for Project Recommendations	Visual Data Mining
5	Alves, P., Miranda, L., Morais, C.	Record of Undergraduates' Activities in Virtual Learning Enviroments	Statistics
6	Aher, S.B. Lobo, L.	Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data	Clustering, Association Rule Mining

Συνέχεια του Πίνακα 4.1: Σύγκριση Μεθόδων DM των άρθρων

ID	Συγγραφείς	Τίτλος Άρθρου	Μέθοδος
7	Jo, I., Park, Y., Lee, H.	Three interaction patterns on asynchronous online discussion behaviours: A methodological comparison	Regression
8	Kardan, S., Conati, C.	A Framework for Capturing Distinguishing User Interaction Behaviours in Novel Interfaces	Clustering, Association Rule Mining, Sequential Pattern Mining
9	Leong, C.K., Lee, Y.H., Mak, W.K.	Mining sentiments in SMS texts for teaching evaluation	Text Mining
10	Lonn, S., Aguilar, S.J., Teasley, S.D.	Investigating student motivation in the context of a learning analytics intervention during a summer bridge program	Visual Data Mining, Statistics
11	Barnes, T.	The Q-matrix Method: Mining Student Response Data for Knowledge	Correlation Mining
12	Ueno, M., Nagaoka, K.	Learning Log Database and Data Mining system for e-Learning -On-Line Statistical Outlier Detection of irregular learning processes-	Outlier Detection

Συνέχεια του Πίνακα 4.1: Σύγκριση Μεθόδων DM των άρθρων

ID	Συγγραφείς	Τίτλος Άρθρου	Μέθοδος
13	Griffiths, M.E., Graham, C.R.	Patterns of User Activity in the Different Features of the Blackboard CMS across All Courses for an Academic Year at Blackboard CMS across All Courses for an Academic Year at Brigham Young University Brigham Young University	Statistics
14	Nasiri, M., Minaei, B.	Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining	Clustering
15	Kotsiantis, S.B., Pierrakeas, C., Pintelas, P.E.	Preventing Student Dropout in Distance Learning Using Machine Learning Techniques	Classification
16	Dech Thammasiri , Dursun Delen, Phayung Meesad , Nihat Kasap	A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition	Classification

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα

Καθώς τις τελευταίες δεκαετίες, το ζήτημα της χρήσης της εξόρυξης δεδομένων και της μηχανικής μάθησης στην εκπαίδευση γίνεται ολοένα και πιο διαδεδομένο, οφείλουμε να αναλύσουμε τα ήδη υπάρχον έργα ώστε να βγάλουμε συμπεράσματα. Για τον λόγο αυτό στην εργασία αυτή αναλύω 16 σχετικά άρθρα και συμπεραίνω ότι η εξόρυξη δεδομένων και η μηχανική μάθηση είναι σημαντικές στην εκπαίδευση, επειδή μπορούν να βοηθήσουν στη βελτίωση της αποδοτικότητας και της αποτελεσματικότητας της διδασκαλίας και της μάθησης. Με τη βοήθεια αυτών των τεχνολογιών, τα εκπαιδευτικά ιδρύματα μπορούν να αναλύσουν μεγάλες ποσότητες δεδομένων για να εντοπίσουν πρότυπα και τάσεις, τα οποία στη συνέχεια μπορούν να χρησιμοποιηθούν για να προσαρμόσουν τις μεθόδους διδασκαλίας στις συγκεκριμένες ανάγκες των μεμονωμένων μαθητών. Αυτή η εξατομικευμένη προσέγγιση της εκπαίδευσης μπορεί να βοηθήσει στη βελτίωση των αποτελεσμάτων των μαθητών, καθώς και να κάνει τη μαθησιακή διαδικασία πιο ελκυστική και σχετική για τους μαθητές. Επιπλέον, η εξόρυξη δεδομένων και η μηχανική μάθηση μπορούν να χρησιμοποιηθούν για τον εντοπισμό περιοχών όπου οι μαθητές μπορεί να δυσκολεύονται, επιτρέποντας στους εκπαιδευτικούς να παρέχουν στοχευμένη υποστήριξη και παρέμβαση για να βοηθήσουν τους μαθητές να επιτύχουν.

Επιπλέον, η εξόρυξη δεδομένων και η μηχανική μάθηση μπορούν να χρησιμοποιηθούν για την αυτοματοποίηση πολλών κουραστικών και χρονοβόρων εργασιών που σχετίζονται με την εκπαίδευση, όπως η βαθμολόγηση εργασιών και η διαχείριση αρχείων μαθητών. Αυτό μπορεί να απελευθερώσει τους δασκάλους και να επικεντρωθούν σε πιο σημαντικά καθήκοντα, όπως ο σχεδιασμός μαθημάτων και η παροχή υποστήριξης στους μαθητές. Επιπλέον, η εξόρυξη δεδομένων και η μηχανική μάθηση μπορούν να βοηθήσουν στον εντοπισμό των πιο αποτελεσματικών μεθόδων διδασκαλίας και μαθησιακών υλικών, τα οποία μπορούν να μοιραστούν με άλλους εκπαιδευτικούς για να συμβάλλουν στη βελτίωση της συνολικής ποιότητας της εκπαίδευσης. Συνοπτικά, η εξόρυξη δεδομένων και η μηχανική μάθηση είναι σημαντικά εργαλεία στον τομέα της εκπαίδευσης, καθώς μπορούν να βοηθήσουν στη βελτίωση της αποδοτικότητας και της αποτελεσματικότητας της διδασκαλίας και της μάθησης και τελικά να οδηγήσουν σε καλύτερα αποτελέσματα των μαθητών.

Η συγκεκριμένη εργασία θα μπορεί να χρησιμοποιηθεί στο μέλλον για περεταίρω ανάλυση και περιγραφή άρθρων.

Βιβλιογραφία

- [1] Romero, C., & Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
- [2] Simoff, S.J. (2009). Visual Data Mining. In: LIU, L., ÖZSU, M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_1121
- [3] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on systems, man, and cybernetics, part C: applications* 2541 and reviews, 40(6), 601–618.
- [4] Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future vision. *Journal of Educational Data Mining*, 1(1), 1–15.
- [5] Peña-Ayala, A., Domínguez, R., & Medel, J. (2009). Educational data mining: a sample of review and study case. *World Journal of Educational Technology*, 2, 118–139.
- [6] Hanan Aldowah, Hosam Al-Samarraie, Wan Mohamad Fauzy, Educational data mining and learning analytics for 21st century higher education: A review and synthesis, *Telematics and Informatics*, Volume 37, 2019, Pages 13-49, ISSN 0736-5853, <https://doi.org/10.1016/j.tele.2019.01.007>.
- [7] Fernando Martínez Abad & Alicia A. Chaparro Caso López (2016): Data-mining techniques in detecting factors linked to academic achievement, *School Effectiveness and School Improvement*, DOI: 10.1080/09243453.2016.1235591
- [8] Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T., Popelinsky, L., 2012. Predicting Drop-Out from Social Behaviour of Students. *International Educational Data Mining Society*.
- [9] Cerezo, R., Sánchez-Sántillán, M., Paule-Ruiz, M.P., Núñez, J.C., 2016. Students' LMS interaction patterns and their relationship with achievement: a case study in higher education. *Comput. Educ.* 96, 42–54.
- [10] Kordík, P., Kuznetsov, S., 2015. Mining Skills from Educational Data for Project Recommendations. In: Paper presented at the International Joint Conference.
- [11] Alves, P., Miranda, L., Morais, C., 2015. Record of undergraduates' activities in virtual learning environments. In: Paper presented at the 14th European Conference on e-Learning.
- [12] Aher, S.B., Lobo, L., 2013. Combination of machine learning algorithms for recommendation of courses in E-Learning system based on historical data. *Knowledge-Based Syst.* 51, 1–14.

- [13] Jo, I., Park, Y., Lee, H., 2017. Three interaction patterns on asynchronous online discussion behaviours: a methodological comparison. *J. Comput. Assisted Learn.* 33 (2), 106–122.
- [14] Kardan, S., Conati, C., 2010. A framework for capturing distinguishing user interaction behaviors in novel interfaces. In: Paper presented at the Educational Data Mining 2011
- [15] Leong, C.K., Lee, Y.H., Mak, W.K., 2012. Mining sentiments in SMS texts for teaching evaluation. *Exp. Syst. Appl.* 39 (3), 2584–2589.
- [16] Lonn, S., Aguilar, S.J., Teasley, S.D., 2015. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Comput. Human Behav.* 47, 90–97.
- [17] Barnes, T., 2005. The Q-matrix method: Mining student response data for knowledge. In: Paper presented at the American Association for Artificial Intelligence 2005 Educational Data Mining Workshop.
- [18] Ueno, M., Nagaoka, K., 2002. Learning Log Database and Data Mining system for e-Learning-On-Line Statistical Outliers Detection of irregular learning processes. In: Paper presented at the IEEE Int'l. Conf. on Advanced Learning Technologies (ICALT).
- [19] Griffiths, M.E., Graham, C.R., 2009. Patterns of user activity in the different features of the Blackboard CMS across all courses for an academic year at Brigham Young University. *J. Online Learn. Teach.* 5 (2), 285.
- [20] Nasiri, M., Minaei, B., 2012. Predicting GPA and academic dismissal in LMS using educational data mining: a case mining. In: Paper presented at the E-Learning and E-Teaching (ICELET), 2012 Third International Conference on.
- [21] Kotsiantis, S.B., Pierrakeas, C., Pintelas, P.E., 2003. Preventing student dropout in distance learning using machine learning techniques. In: Paper presented at the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems.
- [22] Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2), 321–330.
doi:10.1016/j.eswa.2013.07.046