

Εργασία 1^η

Διόνυσης Κοτζαΐτης

1. Διατυπώστε το στόχο της εργασίας/μελέτης (Section: METHODOLOGY)

Στόχος της εργασίας είναι η κατασκευή ενός προβλεπτικού μοντέλου, το οποίο χρησιμοποιεί δεδομένα (αποκλειστικά log files συνδέσεων ή αυτοματοποιημένων detector) μάθησης από συστήματα ηλεκτρονικής εκπαίδευσης μαθητών γυμνασίου/middle school (εν προκειμένω το ASSISTment, το οποίο αποτελεί σύστημα για εκμάθηση μαθηματικών). Το μοντέλο καλείται να προβλέψει, μέσα από δεδομένα όπως νωχελικότητα, σύγχυση, συγκέντρωση, gaming του συστήματος, απροσεξία και άλλες πληροφορίες/στατιστικά (πχ ποσοστό σωστών απαντήσεων), την πιθανότητα κάποιος μαθητής να συνεχίσει τις σπουδές του σε κάποιο πανεπιστήμιο.

2. Δώστε πληροφορίες για το λογισμικό/σύστημα που χρησιμοποιήθηκε, όπως: ονομασία, γνωστικό αντικείμενο, τύπος λογισμικού, ηλικία, δυνατότητες/χαρακτηριστικά λογισμικού κτλ.

Το λογισμικό που χρησιμοποιήθηκε για να λάβουν οι επιστήμονες τα δεδομένα, ονομάζεται ASSISTment. Είναι ένα δωρεάν web-based λογισμικό μαθηματικής υποβοηθούμενης εκμάθησης, το οποίο εξετάζει τις γνώσεις του μαθητή σε συγκεκριμένες μαθηματικές ερωτήσεις και δίνει στον καθηγητή την δυνατότητα πλήρους ελέγχου του μαθητή. Κάθε πρόβλημα συσχετίζεται με κάποια μαθηματική γνώση/skill. Σε περίπτωση σωστής απάντησης ο μαθητής περνάει σε επόμενη ερώτηση, ενώ σε περίπτωση λάθους το σύστημα προσφέρει βοηθητικές ερωτήσεις, που αποτελούν βήματα προς την λύση του αρχικού προβλήματος. Κατά την διάρκεια όλης της διαδικασίας, το σύστημα κρατάει στατιστικά στοιχεία για την διαδικασία μάθησης, όπως ποσοστό σωστών απαντήσεων ή αριθμό first actions.

3. Δεδομένα – συμμετέχοντες. Δώστε το πλήθος συμμετεχόντων, δημογραφικά στοιχεία συμμετεχόντων, χρονική διάρκεια εφαρμογής της παρέμβασης. Πλήθος και τύπος δεδομένων (πχ log files, απαντήσεις σε ερωτηματολόγια, παρατηρήσεις πεδίου (field observations) κτλ)

Εξετάστηκαν 3.747 μαθητές λυκείου από την περιοχή New England των ΗΠΑ, για τις περιόδους διδασκαλίας από 2004-2005 μέχρι 2006-2007 -με κάποιους απ' τους μαθητές να χρησιμοποιούν το λογισμικό μέχρι τις χρονιές 2007-2008 και 2008-2009. Οι μαθητές αυτοί προέρχονταν από 3 περιοχές, η μία περιοχή ήταν αστική με μαθητές «κατώτερων» κοινωνικών στρωμάτων και εισοδημάτων, ενώ οι άλλες δύο ήταν προάστεια με μαθητές από middle-class οικογένειες. Τα δεδομένα που χρησιμοποιήθηκαν ήταν από τα log files της εφαρμογής και αυτοματοποιημένα detectors, με πάνω από 2.107.108 «κινήσεις» μαθητών για την επίλυση 494.150 προβλημάτων (μ.ό. 132 προβλήματα/μαθητή) και από αυτά τα δεδομένα προήλθαν χαρακτηριστικά γνώσης, χαρακτήρα και συμπεριφοράς. Επίσης, χρησιμοποιήθηκαν στοιχεία γι' αυτούς τους 3.747 μαθητές, από το National Student Clearinghouse, για το αν εισήλθαν οι μαθητές στην τριτοβάθμια εκπαίδευση ή όχι (αποκλειστικά σαν labels για το μοντέλο).

4. Για τη μοντελοποίηση α) της γνώσης, β) της συμπεριφοράς (behavior state) και της γ) ψυχικής κατάστασης (affect state) ποια χαρακτηριστικά/δείκτες χρησιμοποιήθηκαν;

α. Για την μοντελοποίηση της γνώσης χρησιμοποιήθηκε το Bayesian Knowledge tracing (BKT), ένα μοντέλο το οποίο χρησιμοποιείται για την εκτίμηση της γνώσης ενός μαθητή. Προβλέπει την

δυσκολία του προβλήματος συγκρίνοντας το επίπεδο των ικανοτήτων που απαιτούνται για την επίλυση του με τις γνώσεις που έχει ο μαθητής, και υπολογίζει τις αντίστοιχες πιθανότητες επιτυχίας. Για να γίνει αυτή η διαδικασία απαιτείται για κάθε μαθητή και κάθε νέο πρόβλημα η γνώση 4 παραμέτρων:

- L0: Η αρχική πιθανότητα για να έχει το skill ο μαθητής.
 - T: η πιθανότητα εκμάθησης του skill την στιγμή που πρέπει να χρησιμοποιηθεί.
 - G: η πιθανότητα σωστής απάντησης, χωρίς γνώση του skill (τυχαιότητα)
 - S: η πιθανότητα λάθος απάντησης, ενώ γνωρίζει το skill.
- b. Για την μοντελοποίηση της συμπεριφοράς, χρησιμοποιούμε detector θετικής ή αρνητικής συμπεριφοράς ως προς το μάθημα και τον στόχο του. Δηλαδή, χρησιμοποιούμε detectors για 3 συναισθηματικές καταστάσεις:
- Συγκέντρωση.
 - Βαρεμάρα.
 - Σύγχυση.
- c. Αλλά και για 3 συμπεριφορές που υποδηλώνουν αποσύνδεση, τις:
- Εκτός εργασίες.
 - Gaming.
 - Απροσεξία.

5. Τι προσδιορίζει το μοντέλο Contextual Slip και πως χρησιμοποιήθηκε στη μελέτη αυτή;

Το μοντέλο αυτό προσδιορίζει την πιθανότητα ότι μια ενέργεια ενός συγκεκριμένου μαθητή υποδεικνύει απροσεξία ή λάθος βιασύνης, σε περίπτωση που ο μαθητής δώσει λάθος απάντηση. Η πιθανότητα αυτή προσδιορίζεται από τα συμφραζόμενα του προβλήματος, και εξαρτάται από κάθε πρόβλημα, το λάθος και από συγκεκριμένα χαρακτηριστικά του μαθητή, αλλά και από την κατάσταση από την οποία επήλθε το λάθος (πχ ταχύτητα απάντησης).

6. Ποιο μοντέλο χρησιμοποιήθηκε για την εκτίμηση της γνώσης των συμμετεχόντων;

Το BKT, έγινε αναφορά στο 4. Το σύστημα χρησιμοποιώντας τις BKT παραμέτρους αλλά και την πρώτη απάντηση του χρήστη, εξάγει επαναληπτικά την πιθανότητα ο χρήστης αυτός να κατέχει την ικανότητα -είτε εξ αρχής, είτε κατά την διάρκεια της διαδικασίας μάθησης- που απαιτείται για την επίλυση του κάθε προβλήματος.

7. Πως και από πού αντλήθηκαν οι μετρήσεις που αφορούσαν στους ανιχνευτές/ δείκτες (detectors) για τη συμπεριφορά και τη ψυχική κατάσταση των μαθητών;

Για να λάβουμε τις μετρήσεις για τους detectors, ακολουθήθηκε μια διαδικασία 2 βημάτων:

- Αρχικά, κάποιες on field observations (από άλλη πηγή), χρησιμοποιήθηκαν ως labels για τους μαθητές.
- Αυτές, συγχρονίστηκαν με βάση τα δεδομένα από τα log αρχεία του συστήματος.

Σημαντικό είναι να τονίσουμε ότι, για το μοντέλο χρησιμοποιήθηκαν αποκλειστικά δεδομένα από τα log files, για να μπορέσουν να υπάρξουν αποτελέσματα που μπορούν να κλιμακωθούν.

8. Με ποια μετρική αξιολογήθηκε η προβλεπτική ικανότητα των ανιχνευτών συμπεριφοράς;

Οι μετρικές Cohen's Kappa (δείχνει το βαθμό στον οποίο το μοντέλο να ανιχνεύσει την συναισθηματική κατάσταση σε σχέση με την τύχη) και την τιμή A' (είναι περίπου η μετρική AUC: Area under curve της ROC καμπύλης).

Πιο συγκεκριμένα για την συμπεριφορά χρησιμοποιήθηκε:

- Για τον detector off-task, ο αλγόριθμος REP-Tree αξιολογήθηκε με μετρικές $A'=0.819$ και $Kappa=0.506$.
- Για τον detector gaming, ο αλγόριθμος K^* αξιολογήθηκε με μετρικές $A'=0.802$ και $Kappa=0.370$.
- Για τον detector απροσεξίας, χρησιμοποιούμε ένα μοντέλο γραμμικής παλινδρόμησης από τις ετικέτες που προκύπτουν από το Contextual Slip μοντέλο. Το μοντέλο αυτό, γίνεται fitted με την χρήση του πακέτου RapidMiner και συγκεκριμένα με το M-5 prime. Τέλος, η αξιολόγηση γίνεται μέσω six-fold cross validation σε επίπεδο μαθητών. Δηλαδή συνεχόμενη εκπαίδευση του μοντέλου με νέους μαθητές για την και validation με το υπόλοιπο dataset.

9. Με ποια μετρική αξιολογήθηκε η προβλεπτική ικανότητα των ανιχνευτών ψυχικής κατάστασης;

Για την ψυχική κατάσταση χρησιμοποιήθηκε:

- Για τον detector της συγκέντρωσης, ο αλγόριθμος K^* αξιολογήθηκε με μετρικές την $A'=0.678$ και την $Kappa=0.358$.
- Για τον detector «βαρεμάρας», ο αλγόριθμος JRip αξιολογήθηκε με μετρικές $A'=0.632$ και $Kappa=0.229$.
- Για τον detector «βαρεμάρας», ο αλγόριθμος J48 αξιολογήθηκε με μετρικές $A'=0.736$ και $Kappa=0.274$.

10. Για κάθε μαθητή το ξεχωριστό στιγμιότυπο που δημιουργήθηκε ποιες είχε ως μεταβλητές πρόβλεψης και με τι τιμές;

Για κάθε μαθητή έχουμε ένα στιγμιότυπο το οποίο περιλαμβάνει μια μέση τιμή των χαρακτηριστικών συμπεριφοράς και ψυχικής κατάστασης, ενώ ταυτόχρονα περιέχει τα ποσοστά σωστών απαντήσεων και αριθμό πρώτων κινήσεων στο σύστημα ASSISTment.

11. Ποιο μοντέλο χρησιμοποιήθηκε για την πρόβλεψη αν ένας μαθητής θα συνεχίσει τις σπουδές του μετά το λύκειο; Οι συγγραφείς γιατί επέλεξαν αυτό το μοντέλο;

Για την κατασκευή του μοντέλου, χρησιμοποιήθηκε η πολυμεταβλητή λογιστική παλινδρόμηση. Αυτό χρησιμοποιείται γιατί υπάρχει μια σχέση διχοτόμησης (εισαγωγή ή όχι στο πανεπιστήμιο) , ενώ η σχέση ανάμεσα στα detectors και τα αποτελέσματα είναι μη γραμμική. Η χρήση της λογιστικής παλινδρόμησης βοηθάει στην καλή ερμηνεία των αποτελεσμάτων του μοντέλου. Ανεξάρτητες μεταβλητές αποτελούν τα αποτελέσματα των predictors, και κάποιες στατιστικές όπως οι σωστές απαντήσεις ή ο αριθμός πρώτων δράσεων, ενώ εξαρτημένη μεταβλητή ήταν η εισαγωγή στο πανεπιστήμιο.

12. Ποιες μεταβλητές συμμετείχαν στην προσαρμογή του μοντέλου πρόβλεψης αν ένας μαθητής θα συνεχίσει τις σπουδές του μετά το λύκειο;

Οι μεταβλητές που χρησιμοποιήθηκαν για την προσαρμογή του μοντέλου ήταν:

- Απροσεξία.
- Γνώση μαθητή.
- Σωστές απαντήσεις.
- Βαρεμάρα.
- Συγκέντρωση.
- Σύγχυση.
- Εκτός δράσης δραστηριότητα.
- Gaming the system δραστηριότητα.

- Αριθμός αρχικών πράξεων.

Εν συνεχεία, από αυτές τις μεταβλητές επιλέχθηκαν μόνο οι στατιστικά σημαντικές μέσω backwards elimination.

13. Γιατί όλες οι προβλεπτικές μεταβλητές τυποποιήθηκαν (standardized z-scores); (Section: RESULTS)

Όλες οι μεταβλητές τυποποιήθηκαν έτσι ώστε τα αποτελέσματα να είναι ενιαία και ευκόλως ερμηνεύσιμα, αλλά και για να χρησιμοποιηθεί το odds ratio.

14. Στον Table 1 τι παρουσιάζουν οι συγγραφείς; Ποιες παρατηρήσεις διατυπώνουν;

Στο table1 οι συγγραφείς παρουσιάζουν για κάθε μεταβλητή που εξετάζεται:

- Μέσες τιμές.
- Τυπικές διακυμάνσεις.
- Τυπικός μέσος λάθους.
- T-value.

Οι τιμές που εμφανίζονται, διαχωρίζονται ανάλογα με το αν ο μαθητής επέλεξε να συνεχίσει στο πανεπιστήμιο ή όχι. Οι συγγραφείς, μετά από την μελέτη των δεδομένων κατέληξαν στα:

- Υψηλότερες τιμές σε μετρικές μέσης γνωστικής ικανότητας, μέσης τιμής σωστών επιλογών, αριθμού πρώτων δράσεων, μέσης τιμής απροσεξίας και μέσης συγκέντρωσης, έχουν θετική συσχέτιση ως προς την επιλογή πανεπιστημιακής εκπαίδευσης.
- Από την άλλη υψηλές τιμές μέσης «βαρεμάρας», μέσης σύγχυσης, off-task και gaming, εμφανίζουν αρνητική συσχέτιση.

15. Τα αποτελέσματα που παρουσιάζονται στην ενότητα 3. Results και συγκεκριμένα στις 2 παραγράφους (σελ. 5, 2^η στήλη 2^η και 3^η παράγραφος) «These observations align with the individual effects of each feature on the prediction of college enrollment.... (έως) perhaps unsurprising given its relationship with poorer learning [17].» από πιο μοντέλο προέκυψαν;

Στις δύο παραγράφους αυτές τα αποτελέσματα προκύπτουν από πολλά μοντέλα λογιστικής παλινδρόμησης ανάμεσα στις μεταβλητές και στην εισαγωγή στο κολέγιο. Οι μεταβλητές αυτές δοκιμάζονται μια-μια στο μοντέλο, για να δούμε την μεμονωμένη επίδραση του χαρακτηριστικού όσον αφορά την πρόβλεψη.

16. Στις 2 παραγράφους, στις οποίες αναφέρεται το αμέσως προηγούμενο ερώτημα, παρουσιάζονται τα αποτελέσματα στη μορφή για πχ «CollegeEnrollment = 0.612 Correctness + 0.346, $\chi^2(df = 1, N = 3747) = 304.141, p < 0.00, 1 Odds Ratio (StudentKnowledge) = 1.722$)» από την προσαρμογή ποιου μοντέλου προκύπτουν αυτά. Τι εκφράζει ο συντελεστής κάθε μεταβλητής, τι εκφράζει το αποτέλεσμα $\chi^2(..)=....$ και με βάση πιο αποτέλεσμα προκύπτει το αντίστοιχο κάθε φορά συμπέρασμα, για πχ «there is a strong positive relationship between college enrollment and average correct answers» ή «... the same strong positive relationship ή positively

Τα αποτελέσματα αυτά προκύπτουν προσαρμόζοντας και μελετώντας την μονομεταβλητή λογιστική παλινδρόμηση, χρησιμοποιώντας ως εξαρτημένη μεταβλητή την εισαγωγή στο κολέγιο, και ως ανεξάρτητη κάθε φορά την κάθε μεταβλητή -predictor ή στατιστικό- που μελετάμε. Ο συντελεστής κάθε μεταβλητής υπάρχει, έτσι ώστε να προσδιορίσουμε την αύξηση της πιθανότητας εισαγωγής στο κολέγιο σε περίπτωση αύξησης της τιμής της μεταβλητής. Ο σταθερός όρος, δίνει την αρχική

πιθανότητα εισαγωγής στο κολέγιο, στην περίπτωση που η μεταβλητή είναι μηδενική. Το αποτέλεσμα, χ^2 (Wald and Sig) είναι το Wald χ^2 τεστ που ελέγχει την αρχική υπόθεση ότι η σταθερά (coefficient). Αν το sig του Wald test (δεν εμφανίζεται εδώ) είναι \leq του p-value, τότε απορρίπτεται η μηδενική υπόθεση, η οποία λέει ότι ο συντελεστής της μεταβλητής μας είναι 0, δηλαδή η μεταβλητή είναι στατιστικά μη σημαντική για το μοντέλο. Οι βαθμοί ελευθερίας df είναι οι βαθμοί ελευθερίας του Wald χ^2 τεστ (στο συγκεκριμένο τεστ ισούνται με τον αριθμό των κλάσεων -1). Με βάση λοιπόν την τιμή p-value, εάν η εκάστοτε ανεξάρτητη μεταβλητή είναι στατιστικά σημαντική, όταν προκύπτει p-value ≤ 0.001 , έχουμε ότι η μεταβλητή είναι στατιστικά σημαντική για την πρόβλεψη της συνέχισης των σπουδών στο πανεπιστήμιο, μάλιστα όταν συμβαίνει αυτό κοιτάμε το πρόσημο του coefficient. Θετικό πρόσημο, σημαίνει θετική συσχέτιση μεταξύ της εξαρτημένης και ανεξάρτητης μεταβλητής, ενώ αρνητικό, το αντίθετο. Όσο μεγαλύτερος ο συντελεστής της μεταβλητής, τόσο μεγαλύτερη η θετική ή αρνητική συσχέτιση.

17. Στον Table 2 τα αποτελέσματα που παρουσιάζουν οι συγγραφείς από ποιο μοντέλο προέκυψαν;

Τα αποτελέσματα του Table2 προέκυψαν από ένα μοντέλο λογιστικής παλινδρόμησης, το οποίο αξιολογήθηκε στο επίπεδο των μαθητών με 5-fold cross validation. Το μοντέλο αυτό, χρησιμοποιεί όλα τα predictors και τα στατιστικά που αναφέραμε.

18. Στον Table 3 τα αποτελέσματα που παρουσιάζουν οι συγγραφείς από ποιο μοντέλο προέκυψαν;

Τα αποτελέσματα του table3 προέκυψαν αφαιρώντας μη στατιστικά σημαντικές μεταβλητές, με backwards elimination, από τα αποτελέσματα του table2. Δηλαδή, ξεκινάει με όλες τις μεταβλητές, και κάνει αλληπάλληλα τεστ σημαντικότητας, αφαιρώντας τις μεταβλητές που δεν επηρεάζουν το μοντέλο μας. Χρησιμοποιεί τα: Student Knowledge, Correctness, Number of First Actions, Carelessness, Confusion, Boredom.

19. Στο τέλος της 5ης σελίδας αναφέρουν «This model was statistically significantly better than a null (intercept-only) model, $\chi^2(df = 9, N = 3747) = 390.146, p < 0.001$ » ενώ στην 6η σελίδα αναφέρει « This model is also statistically significantly better than the null model, $\chi^2(df = 6, N = 3747) = 386.502, p < 0.001$ ». Πως αντιλαμβάνεστε αυτό το σχόλιο; Από πού προκύπτει ότι το τελευταίο μοντέλο είναι καλύτερο από το προηγούμενο. Ο έλεγχος που διενεργούν ποιος είναι και γιατί χρειάζεται να γίνει;

Στην παράγραφο αυτή γίνονται 2 συγκρίσεις:

- Ανάμεσα στο μοντέλο του πίνακα2 (πλήρες μοντέλο λογιστικής παλινδρόμησης, με όλα τα predictors και τα στατιστικά να συμπεριλαμβάνονται) με ένα null μοντέλο (που περιέχει μόνο τον σταθερό όρο, άρα δεν έχει εξαρτημένες μεταβλητές).
- Ανάμεσα στο μοντέλο του πίνακα3 (τελικό μοντέλο λογιστικής παλινδρόμησης) με ένα null μοντέλο (που περιέχει μόνο τον σταθερό όρο, άρα δεν έχει εξαρτημένες μεταβλητές).

Τα αποτελέσματα που παίρνουμε είναι τα εξής:

- Για την πρώτη σύγκριση, διενεργώντας τεστ υποθέσεων προκύπτει στατιστικά σημαντική διαφορά ανάμεσα στο μοντέλο του table 2 και το null μοντέλο -Chi-square test σημαντικότητας με p-value <0.001 και 9 βαθμούς ελευθερίας (όσες οι μεταβλητές που δεν συμπεριλαμβάνονται στο null).
- Για την δεύτερη σύγκριση, διενεργώντας τεστ υποθέσεων προκύπτει στατιστικά σημαντική διαφορά ανάμεσα στο μοντέλο του table 3 και το null μοντέλο -Chi-square test σημαντικότητας με p-value <0.001 και 6 βαθμούς ελευθερίας (όσες οι μεταβλητές που δεν συμπεριλαμβάνονται στο null).

Η τελική σύγκριση προκύπτει από τις τιμές των cross validated A' και cross validated Kappa. Οι τιμές αυτές είναι σχετικά κοντά, παρ' όλα αυτά προτιμάμε το μοντέλο του πίνακα 3 καθώς είναι πιο απλό και γενικεύεται πιο εύκολα.

20. Ανάμεσα στα αποτελέσματα του Table 1 και Table 3 τι διαφορές παρατηρούνται για τις προβλεπτικές μεταβλητές; Πως ορμηνεύουν /εξηγούν τις διαφορές αυτές οι συγγραφείς;

Ανάμεσα στο 1^ο και το 3^ο table, εμφανίζονται διαφορές ανάμεσα στην απροσεξία, την σύγχυση και την «βαρεμάρα». Αυτές οι τρεις μεταβλητές, φαίνεται να αλλάζουν συσχέτιση σε σχέση με το αρχικό μοντέλο, όταν «προστίθενται» στο καινούργιο. Αυτό, σύμφωνα με τους συγγραφείς, μπορεί να συμβαίνει καθώς όταν αφαιρούμε μη επιτυγχόντες μαθητές έχουμε:

- Μαθητές που βαριούνται, γιατί το υλικό τους φαίνεται εύκολο. Άρα στο νέο μοντέλο φαίνεται θετική συσχέτιση.
- Μαθητές που χρησιμοποιούν την σύγχυση τους για να αποκτήσουν νέες γνώσεις, αντί να αποτύχουν. Άρα αλλάζει και εκεί η συσχέτιση.
- Ενώ, από τους επιτυγχόντες μαθητές, η συσχέτιση της απροσεξίας αλλάζει. Αρχικά, φαινόταν ότι η απροσεξία ήταν θετική ως προς την κολεγιακή επιτυχία, αυτό αλλάζει για τους επιτυγχόντες που παραμένουν απρόσεκτοι.

21. Ποια χαρακτηριστικά/μεταβλητές δεν συμπεριλήφθηκαν στο τελικό μοντέλο;

Δεν συμπεριλήφθηκαν στο τελικό μοντέλο τα:

- Συγκέντρωση ατόμου (engaged concentration).
- Off-Task.
- Gaming.

22. Δώστε την εξίσωση του τελικού μοντέλου πρόβλεψης

Από το μοντέλο 3 (τελικό μοντέλο), έχω:

Εισαγωγή_στο_πανεπιστήμιο=0.351+1.119*student_knowledge+0.698*correctness+0.261*Number_of_first_actions-1.145*carelessness+0.217*confussion+0.169*boredom

Αυτή η εξίσωση περνάει σαν όρος στην σιγμοειδή συνάρτηση, καθώς μιλάμε για λογιστική παλινδρόμηση.

23. Δώστε την ερμηνεία κάθε συντελεστή των προβλεπτικών μεταβλητών χρησιμοποιώντας το odds ratio

Στην λογιστική παλινδρόμηση μας, το odds ratio=P(να μπει στο κολέγιο ο μαθητής|predictor)/ P(να μπει στο κολέγιο ο μαθητής|να μην υπάρχει ο predictor). Σύμφωνα λοιπόν με το table3 έχουμε:

- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη student knowledge είναι 3,062 φορές μεγαλύτερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.
- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη Correctness (σωστές απαντήσεις) είναι 2,01 φορές μεγαλύτερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.
- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη αριθμού πρώτων δράσεων είναι 1,298 φορές μεγαλύτερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.
- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη απροσεξίας είναι 0,318 φορές μικρότερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.

- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη σύγχυσης είναι 1,242 φορές μεγαλύτερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.
- Η πιθανότητα κάποιος μαθητής να μπει στο κολέγιο αν έχει υψηλό δείκτη «βαρεμάρας» είναι 1,184 φορές μεγαλύτερη από το να μπει στο κολλέγιο με χαμηλό δείκτη.

24. Ποιοι ανιχνευτές συμπεριφοράς και ψυχικής κατάστασης συσχετίζονται με την εισαγωγή μαθητών στο κολλέγιο μετά το λύκειο;

Οι detectors που χρησιμοποιούνται μετά το middle school και σχετίζονται με την εισαγωγή ενός μαθητή στην τριτοβάθμια εκπαίδευση είναι:

- Η απροσεξία, όσον αφορά την συμπεριφορά
- Η νωχελικότητα και η σύγχυση, όσον αφορά την ψυχική κατάσταση.

25. Συνοψίζοντας ποια είναι η στατιστική μεθοδολογία που εφαρμόζει; (Section: DISCUSSION AND CONCLUSION)

Οι συγγραφείς χρησιμοποιούν την στατιστική μεθοδολογία της λογιστικής παλινδρόμησης.

26. Ποια είναι η ακρίβεια πρόβλεψης που πέτυχε το μοντέλο;

Το μοντέλο προβλέπει την είσοδο ενός μαθητή σε πανεπιστήμιο με ακρίβεια 68.6%

27. Ποιοι παράγοντες προβλέπουν την επιτυχή εισαγωγή ενός μαθητή στο κολέγιο; Τι ισχυρίζονται οι συγγραφείς;

Γενικά, οι συγγραφείς αναφέρουν ότι πολλοί κοινωνικοί λόγοι ευθύνονται για την εισαγωγή ενός μαθητή στο πανεπιστήμιο, όπως οικονομικοί λόγοι και η βοήθεια των γονέων και του σχολείου. Από τη άλλη, υπάρχουν τα ιδιαίτερα γνωστικά χαρακτηριστικά ενός ατόμου. Μέσα από την μελέτη τους οι συγγραφείς κατέληξαν ότι σημαντικότερο ρόλο παίζουν η «βαρεμάρα», η σύγχυση, η τυχαία χρήση του συστήματος (gaming) και η απροσεξία ενός μαθητή (κατά την χρήση του συστήματος ASSISTment), ενώ η επιτυχία στα μαθηματικά γυμνασίου φαίνεται να έχει θετική συσχέτιση ως προς την επιλογή να συνεχίσει την εκπαίδευση του.