

ΑΛΕΞΑΝΔΡΕΙΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ
ΙΔΡΥΜΑ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αναλυτική Επεξεργασία
Δεδομένων Της Υπηρεσίας
ΠΥΘΙΑ

Φοιτητής
ΨΥΤΤΕΛΗΣ ΘΕΟΔΩΡΟΣ

Επιβλέπων Καθηγητής
ΔΗΜΗΤΡΙΟΣ ΔΕΡΒΟΣ

Αριθμός Μητρώου
113813

Συνεπιβλέπων
Σ. ΟΥΓΙΑΡΟΓΛΟΥ *PhD*

31 Δεκεμβρίου 2018



Abstract

At the Information Technology Department of the Thessaloniki ATEI the effort to improve the quality of the educational content and that of the teaching/training practice is intense and continuous. This is dictated by the rapid pace at which technology evolves and its impact on the job market requirements for I.T. skills and qualifications. The present thesis project focuses on the analytical processing of undergraduate and graduate exams data. The latter first undergo extensive preparation (cleansing and homogenization) for the results of the later data processing stages to be reliable and trustworthy. Next comes the statistical data processing stage. The first set of results obtained turn out to be particularly useful in sketching the students assessment 'profile' of each one course from one academic semester to the next and from one academic year to its next one. An objective set out as such from the beginning of the project has been to apply data mining algorithms in the processing of students examinations data. As it has turned out, the data preparation and the subsequent statistical processing stages have required more time and effort than what was originally expected. In this respect, the data mining stage has been restricted to association rules mining. Still, the rules discovered are informative and useful in the direction of assessing the quality of the department's undergraduate and graduate course curricula. In addition, the results obtained can be utilized in devising and information service dor the students to plan their study in a way that maximizes their benefit from enrolling in the degree program. In this respect, the present thesis project may be taken to comprise a first, pilot attempt to apply data mining techniques to students examinations data at a typical Higher Education Institution.

Περίληψη

Στο ΑΤΕΙ/Θ, στο Τμήμα Μηχανικών Πληροφορικής Τ.Ε. ειδικότερα, καταβάλλεται συνεχής προσπάθεια για τη βελτίωση του παρεχόμενου εκπαιδευτικού έργου και του επιπέδου σπουδών και κατάρτισης των αποφοίτων του. Αυτό επιτάσσουν η διαρκής εξέλιξη της τεχνολογίας και ο αντίκτυπός της στις απαιτήσεις της σύγχρονης αγοράς εργασίας. Αντικείμενο της παρούσας πτυχιακής εργασίας συνιστά η επεξεργασία δεδομένων βαθμολογιών εξετάσεων προπτυχιακών μαθημάτων και μεταπτυχιακών μαθημάτων του Τμήματος. Τα παραπάνω βαθμολογικά δεδομένα υπέστησαν πρώτα μία εκτεταμένη προεπεξεργασία καθαρισμού, διόρθωσης και ομογενοποίησης. Παράλληλα, το στάδιο της προεπεξεργασίας των δεδομένων περιελάμβανε και περιγραφικού τύπου στατιστική επεξεργασία από την οποία προέκυψαν εξαιρετικά χρήσιμα στοιχεία για τη βαθμολογική 'συμπεριφορά' του κάθε ενός μαθήματος από την μία εξεταστική περίοδο στην επόμενη της και από το ένα ακαδημαϊκό έτος στο επόμενο του. Απώτερο στόχο αποτελεί η επεξεργασία των βαθμολογικών δεδομένων με αλγόριθμους Εξόρυξης Πληροφορίας (Data Mining). Προς την κατεύθυνση αυτή, χρησιμοποιήθηκε η τεχνική της παραγωγής συνδυαστικών κανόνων (association rules) με τη χρήση του αλγόριθμου Apriori. Στην εργασία γίνεται αναφορά για την πληροφορία η οποία εξήχθη και για τη χρησιμότητά της για τη βελτίωση του παρεχόμενου εκπαιδευτικού έργου και των υπηρεσιών πληροφόρησης των φοιτητών του Τμήματος. Με αυτήν την έννοια, η παρούσα εργασία συμβάλλει στην προτυποποίηση της χρήση τεχνικών εξόρυξης πληροφορίας από τα βαθμολογικά δεδομένα ενός τριτοβάθμιου εκπαιδευτικού ιδρύματος.

Ευχαριστίες

Θα ήθελα να εκφράσω την βαθύτατη ευγνωμοσύνη μου στους επιβλέποντες καθηγητές μου, κύριο Δημήτριο Δέρβο και κύριο Στέφανο Ουγιάρογλου για τον προσωπικό χρόνο που αφιέρωσαν, τη συνεχή καθοδήγηση, την υπομονή και την επιμονή τους. Η συμβολή τους ήταν καθοριστική για την ολοκλήρωση αυτής της πτυχιακής εργασίας.

Περιεχόμενα

1	Εισαγωγή	6
2	Εργαλεία και Τεχνικές	8
2.1	Η γλώσσα Προγραμματισμού R και το RStudio	8
2.2	Περιγραφική Στατιστική	11
2.3	Εξόρυξη Δεδομένων	11
2.4	Εξόρυξη Κανόνων Συσχέτισης	12
2.5	Πληροφορίες για τα Δεδομένων	12
3	Περιγραφική Στατιστική	13
3.1	Σύνολο Δεδομένων	13
3.2	Αθροιστικά Στατιστικά	13
3.3	Αποτελέσματα	15
3.3.1	Προπτυχιακό Πρόγραμμα Σπουδών	15
3.3.2	Μεταπτυχιακό Πρόγραμμα Σπουδών	22
4	Εξόρυξη Κανόνων Συσχέτισης	28
4.1	Σύνολο Δεδομένων	28
4.2	Καθαρισμός και Προετοιμασία Δεδομένων	28
4.3	Εξαγωγή Κανόνων Συσχέτισης με τον Αλγόριθμο Apriori	30
4.3.1	Ο αλγόριθμος Apriori	30
4.3.2	Εφαρμογή του Apriori	32
4.4	Αποτελέσματα	34
5	Επίλογος	45
	Παράρτημα Α	46
	Αναφορές	54

Κατάλογος Σχημάτων

1	Παράδειγμα Γραφήματος	9
2	Προπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά εξεταστική περίοδο	16
3	Προπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά ακαδημαϊκό έτος	18
4	Προπτυχιακό Μάθημα: Προσπάθειες (άξονας Try) και ποσοστά επιτυχίας/αποτυχίας εξετασθέντων (άξονας Percentage)	21
5	Μεταπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά εξεταστική Περίοδο	23
6	Μεταπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά ακαδημαϊκό έτος	25
7	Μεταπτυχιακό Μάθημα: Προσπάθειες (άξονας Try) και ποσοστά επιτυχίας/αποτυχίας εξετασθέντων (άξονας Percentage) . .	27
8	Διάγραμμα Ροής Αλγορίθμου	31
9	Γράφημα διασποράς για διάφορες τιμές των <i>support</i> , <i>confidence</i> και <i>lift</i>	36
10	Γράφημα διασποράς για διάφορες τιμές <i>support</i> , <i>confidence</i> ανά μήκος κανόνα	38

Κατάλογος Πινάκων

1	Αντιστοίχιση διαβαθμισμένης κατηγορίας με διαστήματα βαθμολογίας	29
2	Ποσοστό γραπτών ανά διαβαθμισμένη κατηγορία	29
3	Αριθμός κανόνων για διάφορες τιμές των <i>maxlen</i> , <i>support</i> και <i>confidence</i>	33
4	Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση <i>καλά</i>	43
5	Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση <i>πολύ καλά</i>	44
6	Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση <i>άριστα</i>	44
7	Τίτλοι Μαθημάτων των Π3 και Π4 προγραμμάτων σπουδών με τα αντίστοιχά τους στο Π5.	46

1 Εισαγωγή

Τα εκπαιδευτικά ιδρύματα διαχειρίζονται δεδομένα τα οποία μπορούν να χρησιμοποιηθούν για την αξιολόγηση και τη βελτίωση της εκπαιδευτικής διαδικασίας. Πιο συγκεκριμένα, είναι οι πληροφορίες των φοιτητών όπως η ηλικία, το φύλλο ή ακόμα εάν φοίτησαν σε γενικό ή επαγγελματικό λύκειο. Επίσης στοιχεία για την επίδοση των φοιτητών στα μαθήματα, τον βαθμό πτυχίου ακόμα και το πόσα χρόνια διήρκεσαν οι σπουδές τους. Όλα τα παραπάνω δεδομένα μπορούν να αποκτηθούν από το πληροφοριακό σύστημα (ΠΣ) του τμήματος.

Η γνώση που κρύβεται μέσα στα ακαδημαϊκά δεδομένα μπορεί να παίζει σημαντικό ρόλο στο σύνολο των ενεργειών που απαιτούνται για την αναβάθμιση της εκπαιδευτικής διαδικασίας των ιδρυμάτων[EIH09]. Οι εκπαιδευτικοί, μέσω της ανάλυσης δεδομένων[DBP05], έχουν την δυνατότητα να προσεγγίσουν τα γραπτά των φοιτητών από διαφορετικές πλευρές και κατά επέκταση να κατανοήσουν καλύτερα την βαθμολογική συμπεριφοράβαθμολογική συμπεριφορά των φοιτητών ή ακόμα και να εντοπίσουν πιθανές αιτίες που συνδράμουν στην διακύμανση της επίδοσης τους. Οι εκπαιδευτικοί λαμβάνοντάς υπόψιν την καινούργια πληροφορία θα είναι σε θέση να προσαρμόζουν την εκπαιδευτική διαδικασία πάνω στις, όλο και περισσότερο μεταβαλλόμενες, ανάγκες των σύγχρονων φοιτητών.

Τα εκπαιδευτικά ιδρύματα για να βελτιώσουν την αποτελεσματικότητα και την συνολική εκπαιδευτική τους διαδικασία πρέπει πρώτα να αξιολογήσουν τις ήδη υπάρχουσες μεθόδους διδασκαλίας που χρησιμοποιούν [RV07]. Με αφορμή την εσωτερική αξιολόγηση του τμήματος και την ανάγκη για εύρεση νέων τρόπων και τεχνικών για την εκμετάλλευση των δεδομένων προς όφελος των φοιτητών, μας δημιουργήθηκε η ανάγκη και το κίνητρο για την πραγματοποίηση αυτής της εργασίας.

Σε αυτή την εργασία εφαρμόζουμε ένα συνδυασμό τεχνικών ανάλυσης με σκοπό να δούμε τα βαθμολογικά δεδομένα των φοιτητών από διαφορετικές οπτικές γωνίες. Αρχικά με περιγραφική στατιστική [Sta18] για να εξάγουμε χρήσιμη πληροφορία σχετικά με την επίδοση των φοιτητών στα μαθήματα. Επίσης, σε αυτό το κομμάτι δημιουργήσαμε γραφικές παραστάσεις για την πληρέστερη κατανόηση των αποτελεσμάτων. Έπειτα συνεχίζουμε την ανάλυση με την εξαγωγή κανόνων συσχέτισης [KK05] με σκοπό να βρούμε συσχετίσεις και πρότυπα σχετικά με μαθήματα που οι φοιτητές τείνουν να πετυχαίνουν υψηλή επίδοση. Αξιοποιώντας τις δυνατότητες της χρήσης τεχνικών εξόρυξης πληροφορίας στα βαθμολογικά δεδομένα του εκπαιδευτικού ιδρύματος, θεωρούμε ότι συμμετέχουμε ενεργά στη βελτίωση των εκπαιδευτικών διαδικασιών του

τμήματος.

Η εργασία οργανώθηκε ως εξής: Στην ενότητα 2 παρουσιάζονται τα εργαλεία και οι τεχνικές που χρησιμοποιήθηκαν. Στο ενότητα 3 αναλύεται η περιγραφική στατιστική και παρουσιάζονται τα αποτελέσματά της. Στην ενότητα 4 παρουσιάζεται ο αλγόριθμος εξόρυξης κανόνων συσχέτισης Apriori και η εφαρμογή του στα ακαδημαϊκά δεδομένα. Τέλος στην ενότητα 5 βρίσκονται τα συμπεράσματα αυτής της εργασίας και προτάσεις για μελλοντικές εργασίες.

2 Εργαλεία και Τεχνικές

2.1 Η γλώσσα Προγραμματισμού R και το RStudio

Η R [R C17] είναι ένα λογισμικό ανοιχτού κώδικα, για την στατιστική και αναλυτική επεξεργασία δεδομένων. Περιλαμβάνει μια γλώσσα προγραμματισμού, συνδεσιμότητα με άλλες γλώσσες και δυνατότητα εντοπισμού σφαλμάτων. Η R έχει πολλές ομοιότητες με την γλώσσα S [Cha91] και θεωρείται ως μια διαφορετική υλοποίησή της. Βασικοί τύποι που παρέχει η R για τον χειρισμό των δεδομένων είναι:

- vectors
- matrices
- data frames
- arrays
- lists
- factors

Ένα από τα βασικά πλεονεκτήματα της R είναι η επεκτασιμότητά της. Συγκεκριμένα, μπορεί ο καθένας να προσθέσει λειτουργικότητα με την δημιουργία πακέτου ή να χρησιμοποιήσει στον κώδικά του κάποιο από τα ήδη υπάρχοντα (πχ. το ggplot2 για την παραγωγή γραφικών).

Στον παρακάτω κώδικα φαίνεται η ευκολία με την οποία μπορεί κάποιος να χρησιμοποιήσει ένα υπάρχον πακέτο:

```
1 # Download package
2 install.packages(ggplot2)
3 # Use package
4 library(ggplot2)
```

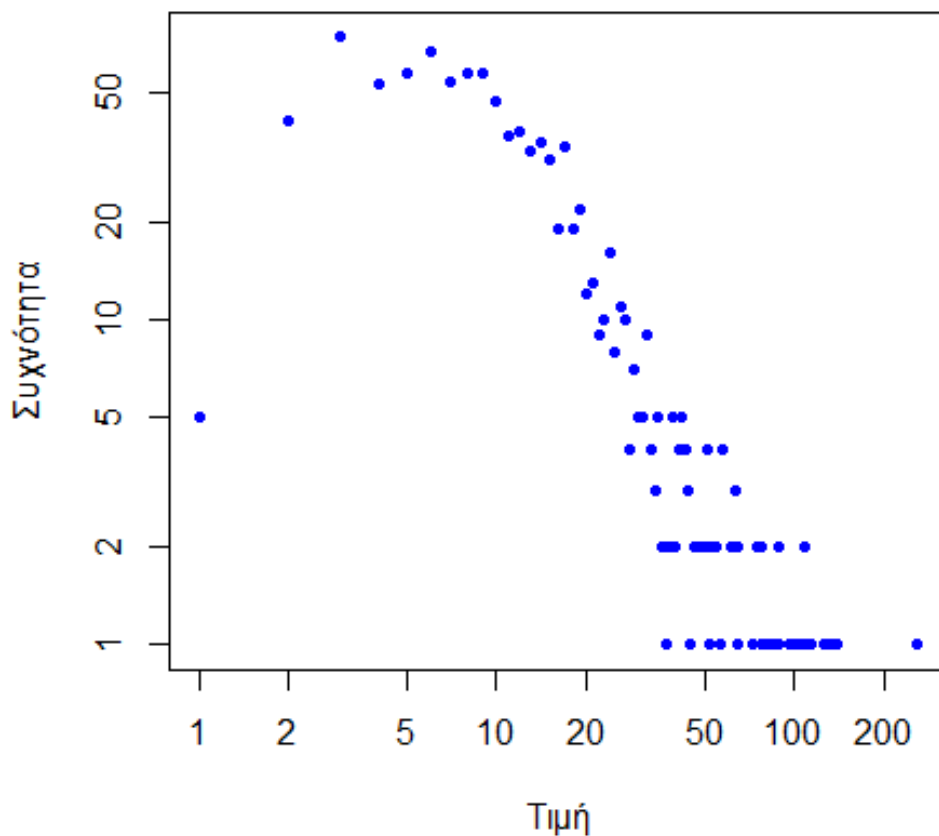
Επίσης, ένα άλλο θετικό είναι ότι η R παρέχει ενσωματωμένη συλλογή από εργαλεία για την αναλυτική επεξεργασία δεδομένων και την εξόρυξη χρήσιμης πληροφορίας.

Ένα από τα μειονεκτήματα της R είναι ότι, καθώς είναι ανοιχτού κώδικα λογισμικό, δεν γίνεται κάποιος έλεγχος ποιότητας στα πακέτα, με συνέπεια να υπάρχουν πακέτα χαμηλής ποιότητας. Επίσης οι εντολές στην R δεν δίνουν αρκετή σημασία στην διαχείριση της μνήμης, με αποτέλεσμα μερικές φορές να καταναλώνεται όλη η διαθέσιμη μνήμη του υπολογιστή.

Παράδειγμα κώδικα R για την παραγωγή γραφικών (Σχήμα 1):

```
1 # Get a random log-normal distribution
2 r <- rlnorm(1000)
3
4 # Get the distribution without plotting it using tighter breaks
5 h <- hist(r, plot=F, breaks=c(seq(0,max(r)+1, .1)))
6
7 # Plot the distribution using log scale on both axes, and use
8 # blue points
9 plot(h$counts, log="xy", pch=20, col="blue",
10      main="Log-normal distribution",
11      xlab="Value", ylab="Frequency")
```

Λογαριθμική Κανονική Κατανομή



Σχήμα 1: Παράδειγμα Γραφήματος

Για την διεκπεραίωση της παρούσας εργασίας χρησιμοποιήθηκαν τα παρακάτω πακέτα:

- *arules* [Hah+11] Παρέχει λειτουργικότητα για την εξόρυξη κανόνων συσχέτισης.
- *arulesViz* [Hah18] Παρέχει λειτουργικότητα για την παραγωγή γραφικών από τους κανόνες συσχέτισης.
- *ggplot2* [Wic09] Πακέτο για την παραγωγή γραφικών.
- *tidyverse* [Wic17b] Είναι συλλογή πακέτων, ειδικά σχεδιασμένα για ανάλυση δεδομένων.
- *stringr* [Wic17a] Πακέτο για τον χειρισμό συμβολοσειρών (*Strings*)
- *plyr* [Wic11] Παρέχει εργαλεία για τον διαχωρισμό, εφαρμογή και συνδυασμό δεδομένων.
- *extrafont* [Cha14] Πακέτο για τον χειρισμό γραμματοσειρών.
- *xlsx* [Dra14] Πακέτο για τον χειρισμό (λειτουργίες *read/write*) αρχείων *Excel*.

Η χρησιμοποίηση της *R* έγινε μέσω του *RStudio*[RSt15]. Το *RStudio* είναι ένα ανοιχτού κώδικα προγραμματιστικό περιβάλλον για την χρησιμοποίηση της *R* (Integrated Development Environment-IDE). Μέσω του γραφικού του περιβάλλοντος καθιστά εύκολο τον χειρισμό των αντικειμένων και των δομών αποθήκευσης δεδομένων της *R*. Επίσης παρέχει ενσωματωμένες δυνατότητες για τον χειρισμό μεγάλων έργων λογισμικού που αποτελούνται από πολλά αρχεία καθώς επιτρέπει την εύκολη πρόσβαση σε αυτά μέσα από ένα ενοποιημένο γραφικό περιβάλλον. Ένα ακόμα πλεονέκτημα του *RStudio* μπορεί να θεωρηθεί η αυτό-συμπλήρωση κώδικα καθώς βοηθάει σημαντικά στην γρηγορότερη ανάπτυξη λογισμικού. Επιπλέον, σημαντικό εργαλείο που παρέχει το *RStudio* είναι ο γρήγορος εντοπισμός σφαλμάτων (*debugging*) στον κώδικα.

2.2 Περιγραφική Στατιστική

Η περιγραφική στατιστική [Sta18] είναι ένα θεμελιώδες στάδιο στην ανάλυση δεδομένων. Συγκεκριμένα περιλαμβάνει συνοπτικά στατιστικά όπως μέσο όρο, ελάχιστη, μέγιστη τιμή κ.α. Επίσης, περιλαμβάνει μεθόδους οπτικοποίησης των δεδομένων όπως οι γραφικές παραστάσεις. Είναι σημαντική καθώς μέσω αυτής εξάγονται χρήσιμες πληροφορίες οι οποίες βοηθούν στην βαθύτερη κατανόηση των δεδομένων. Στην εργασία αυτή μέσω της περιγραφικής στατιστικής προσπαθήσαμε να δούμε καλύτερα την επίδοση των φοιτητών στα μαθήματα. Έτσι θα μπορέσουμε να οργανώσουμε καλύτερα τα δεδομένα για το δεύτερο στάδιο της εργασίας: την Εξόρυξη Κανόνων Συσχέτισης.

2.3 Εξόρυξη Δεδομένων

Εξόρυξη δεδομένων (Data Mining [HKP12]) ονομάζεται η διαδικασία εξαγωγής χρήσιμης πληροφορίας ανάμεσα από μεγάλους όγκους δεδομένων. Με άλλα λόγια μπορούμε να πούμε ότι το Data Mining είναι η εξόρυξη γνώσης από ακατέργαστα δεδομένα. Κάποιοι υποστηρίζουν ότι εξόρυξη γνώσης από βάσεις δεδομένων (KDD) [OP12] και εξόρυξη γνώσης (Data Mining) είναι ισοδύναμες έννοιες. Ωστόσο υπάρχει και η άποψη ότι η εξόρυξη γνώσης από δεδομένα (Data Mining) αποτελεί μέρος της εξόρυξη γνώσης από βάσεις δεδομένων (KDD) [FPS96].

Τα βασικά στάδια της Εξόρυξη δεδομένων είναι:

- Πρώτο στάδιο είναι η συλλογή των δεδομένων από διάφορες πηγές. Αυτές μπορεί να είναι βάσεις δεδομένων, αρχεία δεδομένων ή ακόμα και ο παγκόσμιος ιστός.
- Επόμενο βήμα είναι ο καθαρισμός και μετασχηματισμός των δεδομένων (αν χρειάζεται). Για παράδειγμα, κανονικοποίηση τιμών, διαγραφή λανθασμένων τιμών, αναπλήρωση τιμών που λείπουν κ.α. Αξίζει να σημειωθεί ότι ο καθαρισμός των δεδομένων είναι ένα από τα σημαντικότερα στάδια της ανάλυσης καθώς παίζει καθοριστικό ρόλο στην ποιότητα των αποτελεσμάτων. Επίσης περιλαμβάνει τη δημιουργία νέων μεταβλητών, εάν αυτό κρίνεται απαραίτητο. Αυτή η διαδικασία μπορεί να καταλαμβάνει μέχρι και το 80% του συνολικού χρόνου της ανάλυσης.
- Στη συνέχεια ακολουθεί η μοντελοποίηση, η επιλογή και η εφαρμογή των κατάλληλων τεχνικών εξόρυξης. Βασικές τεχνικές εξόρυξης πληροφορίας από δεδομένα είναι οι εξής:

– Κανόνες Συσχέτισης [HD15],[KK05]

- Συσταδοποίηση [Ber02]
 - Κατηγοριοποίηση [FZR18]
 - Παλινδρόμηση [SOC16]
- Τελευταίο στάδιο είναι η ερμηνεία και η αξιολόγηση των αποτελεσμάτων. Η εξόρυξη πληροφορίας βρίσκει εφαρμογή σε διάφορους τομείς όπως η εκπαίδευση [RV07], [DBP05], [OP12] η οικονομία [CC08], οι επιχειρήσεις λιανικής πώλησης [NXC09] και ο αθλητισμός [SKC10].

2.4 Εξόρυξη Κανόνων Συσχέτισης

Η εξόρυξη κανόνων συσχέτισης είναι μια από τις πιο γνωστές και καλά μελετημένες τεχνικές εξόρυξης γνώσης από δεδομένα [AIS93]. Σκοπό έχει να ανακαλύψει κανόνες ανάμεσα στα δεδομένα που ικανοποιούν κάποια μεγέθη σημαντικότητας (support, confidence, lift). Οι κανόνες συσχέτισης είναι κατάλληλοι για την εξόρυξη γνώσης από μη αριθμητικά δεδομένα.

Οι παραγόμενοι κανόνες έχουν την μορφή Εάν/Τότε (If/then) και αποτελούνται από το σώμα (Εάν) και την κεφαλή (Τότε) μέρος.

Για παράδειγμα:

- **Εάν** ένας πελάτης αγοράζει ψωμί **Τότε** είναι πιθανότερο να αγοράζει και γάλα.

Οι εξόρυξη κανόνων συσχέτισης είναι επίσης γνωστή και ως *Market Basket Analysis* καθώς αρχικά εφαρμόστηκε στο χώρο των επιχειρήσεων λιανικής [Heg03] έχοντας ως στόχο την εύρεση προτύπων σχετικά με τον εντοπισμό προϊόντων όπου η αγορά του ενός προκαλεί την αγορά του άλλου. Υπάρχουν διάφοροι αλγόριθμοι για την εξόρυξη κανόνων συσχέτισης, μεταξύ των πιο γνωστών είναι ο Apriori τον οποίο και χρησιμοποιούμε στην παρούσα εργασία. Σκοπός μας είναι να ανακαλύψουμε κανόνες της μορφής:

- **Εάν** ένας φοιτητής έχει πολύ καλή επίδοση στο μάθημα X , **Τότε** τείνει να έχει άριστη επίδοση στο μάθημα Y .

2.5 Πληροφορίες για τα Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν για την παρούσα εργασία αποκτήθηκαν από το πληροφοριακό σύστημα του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Θεσσαλονίκης το οποίο ονομάζεται 'Πυθία'. Συγκεκριμένα περιλαμβάνει τις βαθμολογίες των προπτυχιακών φοιτητών του τμήματος Μηχανικών Πληροφορικής σε όλες τις εξεταστικές περιόδους από το χειμερινό εξάμηνο του 2003 έως και το χειμερινό του 2017. Επιπλέον σε αυτά τα δεδομένα υπάρχουν

πληροφορίες για τους φοιτητές καθώς επίσης και πληροφορίες για τα μαθήματα. Σε αυτό το χρονικό διάστημα που καλύπτουν τα δεδομένα έχουν εξεταστεί 3210 διαφορετικοί φοιτητές.

Τα δεδομένα για το μεταπτυχιακό πρόγραμμα σπουδών περιλαμβάνουν τα γραπτά των φοιτητών από το χειμερινό εξάμηνο του ακαδημαϊκού έτους 2013-2014 έως και το εαρινό εξάμηνο του 2016-2017 ακαδημαϊκού έτους. Το μεταπτυχιακό πρόγραμμα σπουδών περιλαμβάνει δέκα μαθήματα. Κατά τη χρονική περίοδο των τεσσάρων αυτών ακαδημαϊκών ετών έλαβαν μέρος σε εξετάσεις 83 φοιτητές.

3 Περιγραφική Στατιστική

3.1 Σύνολο Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν αφορούν τα ακαδημαϊκά έτη 2015-2016 και 2016-2017 στα οποία εφαρμόζεται το Π5 πρόγραμμα σπουδών. Το Π5 αποτελείται από 45 θεωρίες και 25 εργαστήρια. Κατά τη χρονική περίοδο αυτή πραγματοποιήθηκαν 8 εξεταστικές (4 ανά ακαδημαϊκό έτος). Συγκεκριμένα, μία εξεταστική για το εαρινό εξάμηνο, μία για το χειμερινό, μία εμβόλιμη στο χειμερινό εξάμηνο και τέλος η εξεταστική του Σεπτεμβρίου. Σε αυτό το διάστημα έλαβαν μέρος 1090 φοιτητές και βαθμολογήθηκαν 24615 γραπτά. Από το σύνολο της πληροφορίας που είχαμε διαθέσιμο κρατήσαμε τα εξής:

- ID Φοιτητή
- Τίτλος Μαθήματος
- Εξεταστική Περίοδος
- Ακαδημαϊκό Έτος
- Βαθμός (που βαθμολογήθηκε το γραπτό)

3.2 Αθροιστικά Στατιστικά

Αρχικά βγάλαμε αθροιστικά στατιστικά για κάθε μάθημα ξεχωριστά ανά ακαδημαϊκό έτος και ανά εξεταστική περίοδο. Πιο αναλυτικά, βρήκαμε το συνολικό αριθμό γραπτών, επίσης υπολογίσαμε το ποσοστό από τα γραπτά τα οποία βαθμολογήθηκαν στο διάστημα $[0,1]$. Θεωρήσαμε ότι τα γραπτά αυτά αφορούν φοιτητές οι οποίοι δεν αφιέρωσαν καθόλου χρόνο για να προετοιμαστούν για

την εξέταση του συγκεκριμένου μαθήματος. Ως εκ τούτου θα επηρέαζαν αρνητικά την ποιότητα των αποτελεσμάτων μας. Έπειτα υπολογίσαμε το ποσοστό των γραπτών που βαθμολογήθηκαν με βαθμό μεγαλύτερο ή ίσο του 5, και τέλος τον μέσο βαθμό των γραπτών. Στις δύο τελευταίες μετρήσεις δεν λήφθηκαν υπόψη τα γραπτά που βαθμολογήθηκαν στο διάστημα $[0,1]$. Όλα τα παραπάνω αποτελέσματα τα συμπεριλάβαμε σε γραφήματα πυκνότητας (π.χ. Σχήμα 2β') τα οποία δείχνουν τους βαθμούς με τους οποίους έχει περισσότερες πιθανότητες να βαθμολογηθεί ένας φοιτητής στο κάθε μάθημα με βάση τα δεδομένα.

Ένας άλλος τρόπος για να προσεγγίσουμε ποσοτικά τον βαθμό δυσκολίας των μαθημάτων, ήταν για κάθε μάθημα ξεχωριστά να υπολογίσουμε το ποσοστό των φοιτητών που πέρασαν και το ποσοστό αυτών που δεν πέρασαν το μάθημα σε σχέση με τον αριθμό των εξετάσεων που έλαβαν μέρος. Αναλυτικότερα, να βρούμε τι ποσοστό από τους φοιτητές που έχουν εξεταστεί στο συγκεκριμένο μάθημα, έχει εξεταστεί μόνο μία φορά και το πέρασε, τι ποσοστό έχει εξεταστεί μια φορά αλλά δεν το έχει περάσει ακόμα, τι ποσοστό το πέρασε με την δεύτερη φορά, τι ποσοστό έχει εξεταστεί δύο φορές αλλά δεν το έχει περάσει ακόμα κ.ο.κ.

Για να υπολογίσουμε τα παραπάνω ποσοστά ορίσαμε ως αφετηρία το ακαδημαϊκό έτος 2015-2016.

Αρχικά ομαδοποιήσαμε τα δεδομένα ανά μάθημα, στη συνέχεια για κάθε μάθημα ξεχωριστά κρατήσαμε τους φοιτητές που έχουν εξεταστεί τουλάχιστον μία φορά στο συγκεκριμένο μάθημα. Έπειτα για κάθε ένα από τους παραπάνω φοιτητές πήραμε τους βαθμούς από όλα τα γραπτά του φοιτητή στο συγκεκριμένο μάθημα. Επόμενο βήμα ήταν να ελέγξουμε αν ένας φοιτητής έχει εξεταστεί πρώτη φορά στο μάθημα το 2015-2016. Όταν η παραπάνω υπόθεση ήταν αληθής ελέγξαμε αν υπάρχει βαθμολογία μεγαλύτερη ή ίση του πέντε, καθώς αυτό σημαίνει ότι έχει περάσει το μάθημα, αν υπήρχε το προσθέσαμε στο ποσοστό των φοιτητών που έχουν περάσει το μάθημα με αριθμό προσπαθειών όσες είναι και ο συνολικός αριθμός των γραπτών του φοιτητή στο συγκεκριμένο μάθημα. Την ίδια διαδικασία ακολουθήσαμε και για φοιτητές που έδωσαν το μάθημα πρώτη φορά το 2016-2017. Τέλος τα παραπάνω αποτελέσματα τα οπτικοποιήσαμε με τη βοήθεια Γραφημάτων Ράβδων (barplots)(π.χ το Σχήμα 4).

3.3 Αποτελέσματα

3.3.1 Προπτυχιακό Πρόγραμμα Σπουδών

Στο Σχήμα 2α', το όνομα του μαθήματος και το ακαδημαϊκό έτος αναγράφονται στην πρώτη γραμμή στο επάνω αριστερό άκρο της εικόνας: *Συστήματα Διαχείρισης Βάσεων Δεδομένων, 2015-2016*. Στη δεύτερη γραμμή της ίδιας λεζάντας αναφέρεται ότι κατά την πρώτη εξεταστική περίοδο του εαρινού εξαμήνου προσήλθαν για να εξεταστούν δεκατρείς (13) από τους εγγεγραμμένους στο μάθημα φοιτητές. Κανείς από τους δεκατρείς φοιτητές δεν βαθμολογήθηκε στο διάστημα[0,1]. Το ποσοστό των φοιτητών που βαθμολογήθηκαν στο διάστημα[0,1] παραλείπεται κατά τον υπολογισμό των δύο επόμενων δύο στατιστικών τιμών: του ποσοστού των φοιτητών οι οποίοι βαθμολογήθηκαν με βαθμό μεγαλύτερο ή ίσο της προβιβάσιμης τιμής πέντε (5.0, 30.%) και τον μέσο βαθμό στην εν λόγω εξέταση[τρία κόμμα δύο, 3.2].

Αντίστοιχα για την πρώτη (Α) εξεταστική του χειμερινού εξαμήνου του μαθήματος συμμετείχαν στις εξετάσεις εκατό εβδομήντα επτά (177) φοιτητές/τριες, ένα ποσοστό (24.3%) βαθμολογήθηκε στο διάστημα[0,1]. Το ποσοστό επιτυχίας στην εξέταση ήταν (44.5%) και η μέση τιμή του βαθμού ήταν 3.6.

Για την την δεύτερη (Β) εξεταστική του εαρινού εξαμήνου του μαθήματος συμμετείχαν στις εξετάσεις δύο φοιτητές/τριες, και οι δύο βαθμολογήθηκαν με βαθμό μεγαλύτερο ή ίσο του πέντε (5.0) και ο μέσος βαθμός των γραπτών τους ήταν 5.5.

Για την την δεύτερη (Β εξεταστική) του χειμερινού εξαμήνου (εμβόλιμη) του μαθήματος συμμετείχαν στις εξετάσεις σαράντα πέντε (45) φοιτητές/τριες. Από αυτούς ένα ποσοστό (31.2 %) βαθμολογήθηκαν στο διάστημα[0, 1]. Το ποσοστό επιτυχίας στην εξέταση ήταν 15.2% και η μέση τιμή του βαθμού ήταν 2.2.

Ανάλογα, στο Σχήμα 2β' για το μάθημα *Συστήματα Διαχείρισης Βάσεων Δεδομένων* κατά το ακαδημαϊκό έτος 2016-2017.

Για την πρώτη εξεταστική του χειμερινού εξαμήνου (Α ΧΕΙΜ) προσήλθαν για να εξεταστούν 153 φοιτητές/τριες. Από αυτούς ένα ποσοστό 2% βαθμολογήθηκε στο διάστημα[0,1]. Το ποσοστό επιτυχίας ήταν 46% και η μέση τιμή του βαθμού ήταν 4.7.

Για την δεύτερη εξεταστική του χειμερινού εξαμήνου (Β ΧΕΙΜ) προσήλθαν για να εξεταστούν 76 φοιτητές/τριες. Από αυτούς ένα ποσοστό 5.3% βαθμολογήθηκε στο διάστημα[0,1]. Το ποσοστό επιτυχίας ήταν 52.8% και η μέση τιμή του βαθμού ήταν 4.2.

Για την πρώτη εξεταστική του εαρινού εξαμήνου (Α ΕΑΡ) προσήλθε για να εξεταστεί 1 φοιτητής/τρια ο οποίος βαθμολογήθηκε με 6.2.

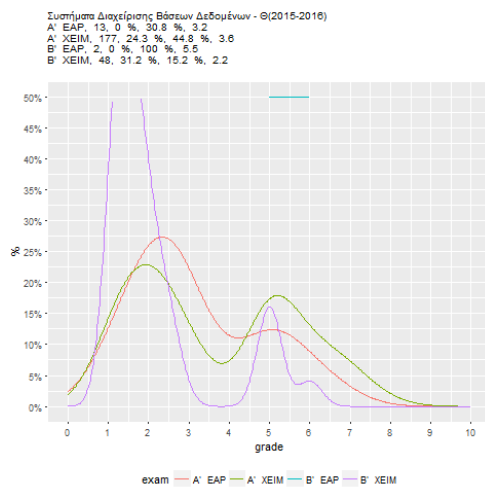
Παρακάτω ο κώδικας R που παράγει τα γραφήματα κατανομής συχνότητας

του Σχήματος 2.

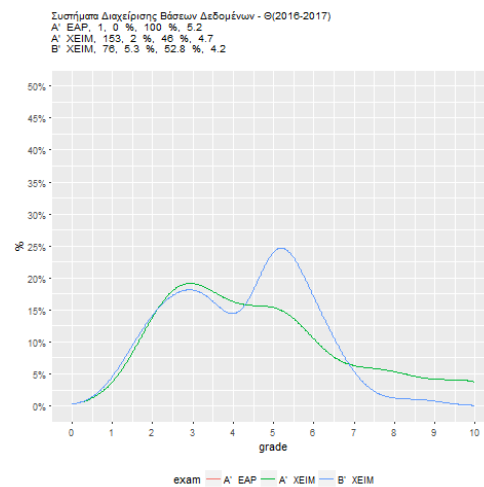
```

1 ggplot(course_df, aes(x=grade, colour=exam))+labs(y="%")+
2   ggtitle(title)+
3   stat_density(geom="line", position="identity")+
4   scale_y_continuous(labels=scales::percent, limits = c(0,.5),
5   breaks = seq(0,0.5,0.05))+
6   scale_x_continuous("grade", limits = c(0,10), breaks = seq
7   (0,10,1))+
8   theme(plot.title = element_text(size=10),
9         legend.position = "bottom")

```



(α') Ακαδημαϊκό έτος 2015 - 2016



(β') Ακαδημαϊκό έτος 2016 - 2017

Σχήμα 2: Προπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά εξεταστική περίοδο

Στο Σχήμα 3α' φαίνεται η συνάθροιση των ποσοστών επιτυχίας και αποτυχίας στο μάθημα *Συστήματα Διαχείρισης Βάσεων Δεδομένων* για το ακαδημαϊκό έτος 2015-2016. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν διακόσια (240) γραπτά. Το 24.2% των τελευταίων βαθμολογήθηκαν με βαθμό στο διάστημα[0, 1] και δεν λογίζονται στη συνέχεια. Πρόκειται για φοιτητές/τριες "επισκέπτες" (ίσως σε μεγάλο τυπικό εξάμηνο σπουδών) η αξιολόγηση των οποίων θεωρείται ότι συνιστά ένα είδος "θορύβου" στην αξιολόγηση που αποσκοπεί να αναδείξει το βαθμό δυσκολίας του μαθήματος. Το 39% των υπόλοιπων γραπτών βαθμολογήθηκαν με προβιβάσιμο βαθμό και ο μέσος βαθμός με τον οποίο βαθμολογήθηκε γραπτό του συγκεκριμένου μαθήματος το συγκεκριμένο ακαδημαϊκό έτος 3.4

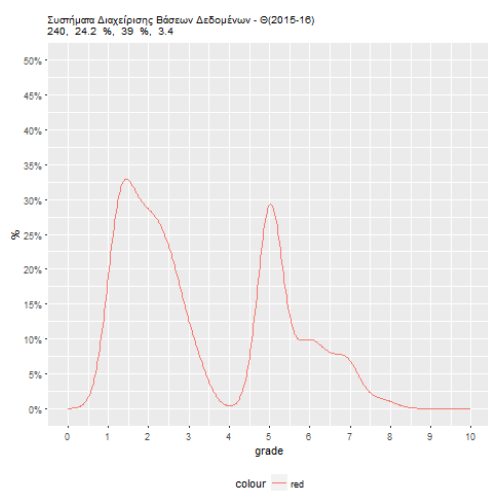
Αντίστοιχα στο Σχήμα 3β' φαίνεται η συνάθροιση των ποσοστών επιτυχίας και αποτυχίας στο μάθημα *Συστήματα Διαχείρισης Βάσεων Δεδομένων* για το ακαδημαϊκό έτος 2016-2017. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν διακόσια (230) γραπτά. Το 3% των τελευταίων βαθμολογήθηκαν με βαθμό στο διάστημα[0, 1] και δεν λογίζονται στη συνέχεια. Το 48.4% των υπόλοιπων γραπτών βαθμολογήθηκαν με προβιβάσιμο βαθμό και ο μέσος βαθμός με τον οποίο βαθμολογήθηκε γραπτό του συγκεκριμένου μαθήματος το συγκεκριμένο ακαδημαϊκό έτος 4.5.

Παρακάτω ο κώδικας *R* που παράγει τα γραφήματα κατανομής συχνότητας ανά ακαδημαϊκό του Σχήματος 3.

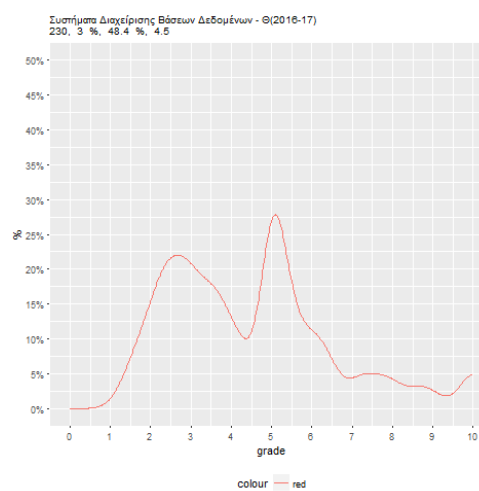
```

1 ggplot(course_df, aes(x=grade, color= "red"))+labs(y="%")+
2   ggtitle(title)+
3   stat_density(geom="line", position="identity", adjust = 0.5)+
4   scale_y_continuous(labels=scales::percent, limits = c(0,.5),
5     breaks = seq(0,0.5,0.05))+
6   scale_x_continuous("grade", limits = c(0,10), breaks = seq
7     (0,10,1))+
8   theme(plot.title = element_text(size=10),
9     legend.position = "bottom")

```



(α') Ακαδημαϊκό έτος 2015 - 2016



(β') Ακαδημαϊκό έτος 2016 - 2017

Σχήμα 3: Προπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά ακαδημαϊκό έτος

Το Σχήμα 4α' αναφέρεται στο (θεωρητικό) μάθημα *Συστήματα Διαχείρισης Βάσεων Δεδομένων*. Αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2015-16. Το σύνολο των εν λόγω φοιτητών συμπεριλαμβάνει εξίσου νέους (πρωτοεξεταζόμενους) και παλιότερους φοιτητές. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2015-2016 έως και την Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα, ένα ποσοστό 32% το πέρασαν (με την πρώτη) και ένα ποσοστό 23.3% από αυτούς που εξετάστηκαν μία μόνον φορά στο μάθημα και δεν το έχουν περάσει ακόμη (δηλ. έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17).

Η δεύτερη στήλη του ιστογράμματος πληροφορεί ότι: από τους φοιτητές οι οποίοι εξετάστηκαν το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα, ένα ποσοστό 13.5% το πέρασαν (με τη δεύτερη) και ένα ποσοστό 12.4% από αυτούς εξετάστηκαν δύο φορές στο μάθημα και δεν το έχουν περάσει ακόμη (έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-2017).

Η τρίτη στήλη του ιστογράμματος πληροφορεί ότι: από τους φοιτητές οι οποίοι εξετάστηκαν το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα, ένα ποσοστό 9.3% το πέρασαν (με τη τρίτη) φορά και ένα ποσοστό 4.1% από αυτούς εξετάστηκαν τρεις φορές στο μάθημα και δεν το έχουν περάσει ακόμη (έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-2017).

Η τέταρτη στήλη του ιστογράμματος πληροφορεί ότι: από τους φοιτητές οι οποίοι εξετάστηκαν το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα, ένα ποσοστό 2.6% το πέρασαν (με τη τέταρτη) φορά και ένα ποσοστό 1.6% από αυτούς εξετάστηκαν τέσσερις φορές στο μάθημα και δεν το έχουν περάσει ακόμη (έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-2017).

Η πέμπτη στήλη του ιστογράμματος μας πληροφορεί ότι: από τους φοιτητές οι οποίοι εξετάστηκαν το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα, το πέρασαν όλοι όσοι εξετάστηκαν για πέμπτη φορά.

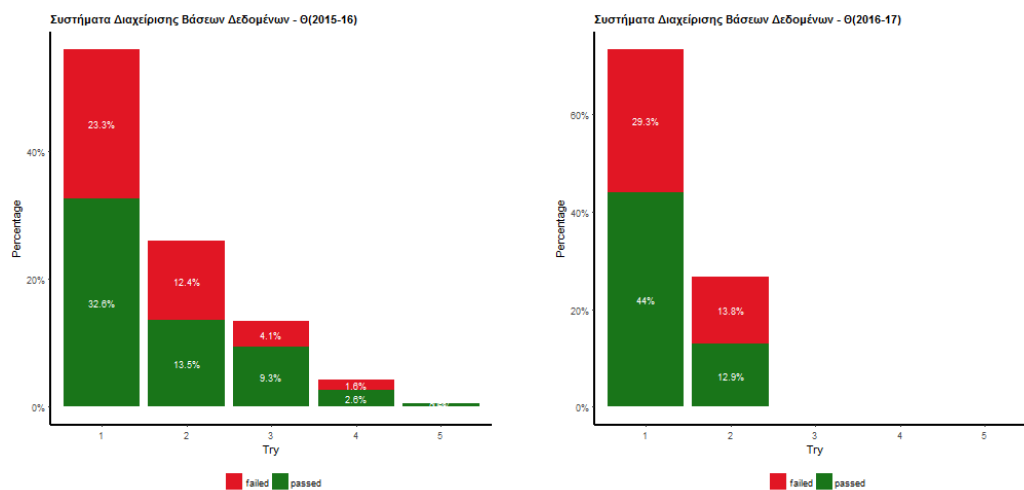
Αντίστοιχα, το Σχήμα 4β' αναφέρεται στο (θεωρητικό) μάθημα *Συστήματα Διαχείρισης Βάσεων Δεδομένων*. Αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2016-17. Το σύνολο των εν λόγω φοιτητών συμπεριλαμβάνει εξίσου νέους (πρωτοεξεταζόμενους) και παλιότερους φοιτητές. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2016-17 έως και την Β' εξεταστική περίοδο του ίδιου ακαδημαϊκού έτους 2016-17. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν το ακαδημαϊκό έτος 2016-17 στο εν λόγω μάθημα, ένα ποσοστό 44% το πέρασαν (με την πρώτη) και ένα ποσοστό 29.3% από αυτούς που εξετάστηκαν μία μόνον φορά στο μάθημα

και δεν το έχουν περάσει ακόμη(δηλ. έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17).

Η δεύτερη στήλη του ιστογράμματος πληροφορεί ότι: από τους φοιτητές οι οποίοι εξετάστηκαν το ακαδημαϊκό έτος 2016-17 στο εν λόγω μάθημα, ένα ποσοστό 12.9% το πέρασαν (με τη δεύτερη) και ένα ποσοστό 13.8% από αυτούς, εξετάστηκαν δύο φορές στο μάθημα και δεν το έχουν περάσει ακόμη(έως και τη Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-2017).

Παρακάτω ο κώδικας *R* που παράγει τα γραφήματα του Σχήματος 4.

```
1 barpl <-ggplot()+geom_bar(data=Editdf ,aes(x=try , fill=status ,
2   y=percent) , stat = "identity")
3 barpl <- barpl + geom_text(data=Editdf ,aes(x = try , y = pos ,
4   label = paste0(percent ,"%")) , size=3,color="white")
5 barpl <- barpl +theme(legend.position="bottom" ,
6   legend.direction="horizontal" , legend.title = element_blank())
7
8 barpl <- barpl + labs(x = "Try" , y = "Percentage" ) +
9   scale_y_continuous(labels = dollar_format(suffix = "%" ,
10  prefix = "")) +
11  ggtitle(paste0(stud_passed2015[i ,1] , "(2015-16)"))+
12  scale_fill_manual(values=fill) +
13  theme(plot.title = element_text(size=11,face="bold"))+
14  theme(axis.line = element_line(size=1, colour = "black") ,
15  panel.grid.major = element_blank() , panel.grid.minor =
16  element_blank() ,
17  panel.border = element_blank() , panel.background = element_
  blank())
```



(α') Ποσοστά φοιτητών που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2015 - 2016

(β') Ποσοστά φοιτητών που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2016 - 2017

Σχήμα 4: Προπτυχιακό Μάθημα: Προσπάθειες (άξονας Try) και ποσοστά επιτυχίας/αποτυχίας εξετασθέντων (άξονας Percentage)

3.3.2 Μεταπτυχιακό Πρόγραμμα Σπουδών

Στην υποενότητα αυτή παρουσιάζονται τα αποτελέσματα για το Μεταπτυχιακό Πρόγραμμα Σπουδών από το ακαδημαϊκό έτος 2013 - 2014 έως και το 2016 - 2017 για το μάθημα *Ανάκτηση Πληροφοριών στο Διαδίκτυο*

Στο Σχήμα 5α' απεικονίζονται τα στατιστικά για το ακαδημαϊκό έτος 2013-2014. Στη λεζάντα αναφέρεται ότι: Για την πρώτη εξεταστική του εαρινού εξαμήνου(A EAP) προσήλθαν για να εξεταστούν 29 φοιτητές/τριες. Από αυτούς ένα ποσοστό 6.9% βαθμολογήθηκε στο διάστημα[0, 1]. Το ποσοστό επιτυχίας ήταν 74.1% και η μέση τιμή του βαθμού ήταν 5.3.

Για την δεύτερη εξεταστική του εαρινού εξαμήνου(B EAP) προσήλθαν για να εξεταστούν 10 φοιτητές/τριες. Από αυτούς τους φοιτητές κανείς δεν βαθμολογήθηκε στο διάστημα[0, 1].Το ποσοστό επιτυχίας ήταν 80% και η μέση τιμή του βαθμού ήταν 5.9.

Στο Σχήμα 5β' απεικονίζονται τα στατιστικά για το ακαδημαϊκό έτος 2014-2015. Στη λεζάντα αναφέρεται ότι: Για την πρώτη εξεταστική του εαρινού εξαμήνου(A EAP) προσήλθαν για να εξεταστούν 21 φοιτητές/τριες. Από αυτούς τους φοιτητές κανείς δεν βαθμολογήθηκε στο διάστημα[0, 1]. Το ποσοστό επιτυχίας ήταν 85.7% και η μέση τιμή του βαθμού ήταν 6.5.

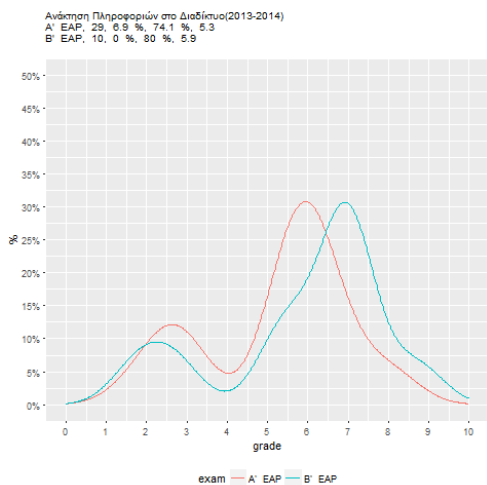
Για την δεύτερη εξεταστική του εαρινού εξαμήνου(B EAP) προσήλθαν για να εξεταστούν 3 φοιτητές/τριες. Από αυτούς τους φοιτητές κανείς δεν βαθμολογήθηκε στο διάστημα[0, 1].Το ποσοστό επιτυχίας ήταν 100% και η μέση τιμή του βαθμού ήταν 5.5.

Στο Σχήμα 5γ' απεικονίζονται τα στατιστικά για το ακαδημαϊκό έτος 2015-2016. Στη λεζάντα αναφέρεται ότι: Για την πρώτη εξεταστική του εαρινού εξαμήνου(A EAP) προσήλθαν για να εξεταστούν 21 φοιτητές/τριες. Από αυτούς τους φοιτητές όλοι πέρασαν το μάθημα και η μέση τιμή του βαθμού ήταν 6.7.

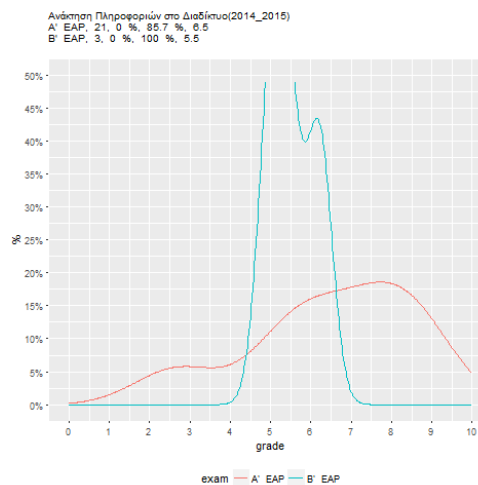
Στο Σχήμα 5δ' απεικονίζονται τα στατιστικά για το ακαδημαϊκό έτος 2016-2017. Στη λεζάντα αναφέρεται ότι: Για την πρώτη εξεταστική του εαρινού εξαμήνου(A EAP) προσήλθαν για να εξεταστούν 20 φοιτητές/τριες. Από αυτούς ένα ποσοστό 5% βαθμολογήθηκε στο διάστημα[0,1]. Το ποσοστό επιτυχίας ήταν 78.9% και η μέση τιμή του βαθμού ήταν 5.9.

Για την δεύτερη εξεταστική του εαρινού εξαμήνου(B EAP) προσήλθαν για να εξεταστούν 5 φοιτητές/τριες. Από αυτούς τους φοιτητές κανείς δεν βαθμολογήθηκε στο διάστημα[0, 1].Το ποσοστό επιτυχίας ήταν 60% και η μέση τιμή του βαθμού ήταν 4.2.

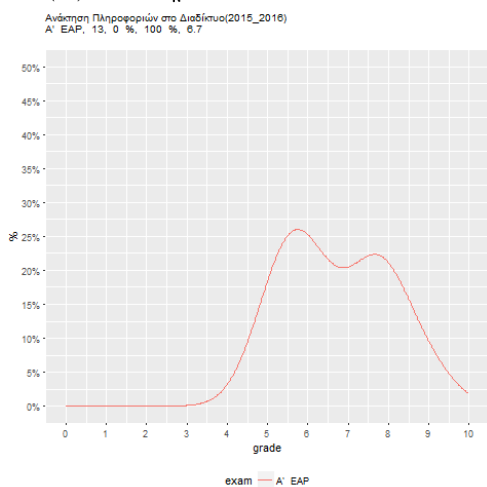
Αναλυτική Επεξεργασία Δεδομένων Της Υπηρεσίας ΠΥΘΙΑ



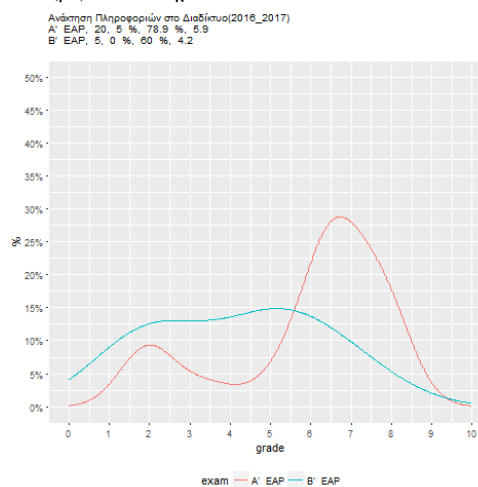
(α') Ακαδημαϊκό Έτος 2013-2014



(β') Ακαδημαϊκό Έτος 2014-2015



(γ') Ακαδημαϊκό Έτος 2015-2016



(δ') Ακαδημαϊκό Έτος 2016-2017

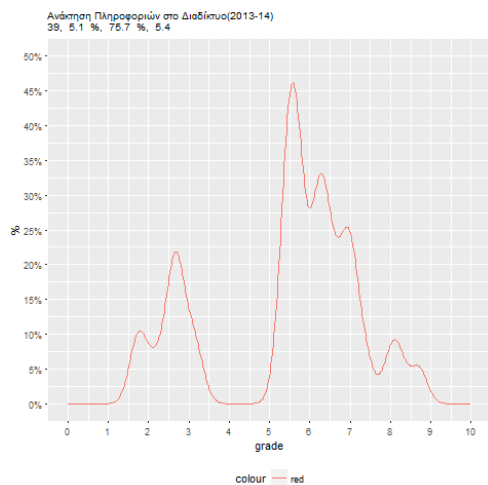
Σχήμα 5: Μεταπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά εξεταστική Περίοδο

Στο Σχήμα 6α' φαίνεται η συνάθροιση των ποσοστών επιτυχίας και αποτυχίας στο μεταπτυχιακό μάθημα *Ανάκτηση Πληροφοριών στο Διαδίκτυο* για το ακαδημαϊκό έτος 2013-2014. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν 39 γραπτά. Το 5.1% των τελευταίων βαθμολογήθηκαν με βαθμό στο διάστημα $[0, 1]$ και δεν λογίζονται στη συνέχεια. Το 75.7% των υπόλοιπων γραπτών βαθμολογήθηκαν με προβιβάσιμο βαθμό και η μέση τιμή του βαθμού ήταν 5.4.

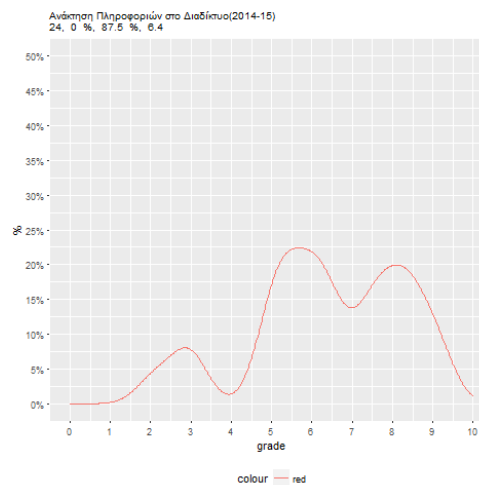
Στο Σχήμα 6β' φαίνεται η συνάθροιση των ποσοστών επιτυχίας και αποτυχίας στο για το ακαδημαϊκό έτος 2014-2015. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν 24 γραπτά. Κανένα από τα γραπτά δεν βαθμολογήθηκε στο διάστημα $[0, 1]$. Το 87.5% των γραπτών βαθμολογήθηκαν με προβιβάσιμο βαθμό και η μέση τιμή του βαθμού ήταν 6.4.

Στο Σχήμα 6γ' φαίνεται η συνάθροιση των ποσοστών επιτυχίας και αποτυχίας στο για το ακαδημαϊκό έτος 2015-2016. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν 13 γραπτά. Όλα τα γραπτά βαθμολογήθηκαν με προβιβάσιμο βαθμό και η μέση τιμή του βαθμού ήταν 6.7 .

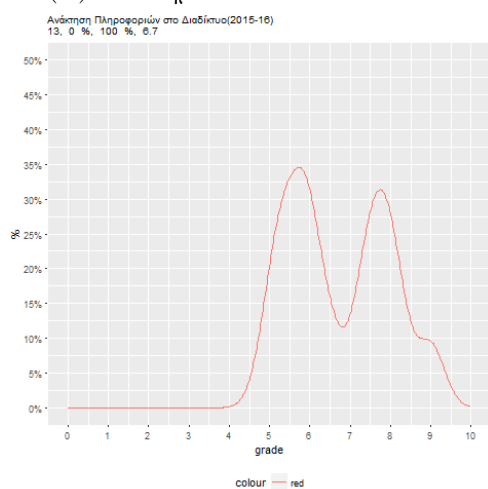
Στο Σχήμα 6δ' φαίνεται η συνάθροιση των ποσοστών για το ακαδημαϊκό έτος 2016-2017. Στο σύνολο των εξεταστικών περιόδων του ακαδημαϊκού έτους, βαθμολογήθηκαν 25 γραπτά. Το 4% των τελευταίων βαθμολογήθηκαν με βαθμό στο διάστημα $[0, 1]$ και δεν λογίζονται στη συνέχεια. Το 75% των υπόλοιπων γραπτών βαθμολογήθηκαν με προβιβάσιμο βαθμό και η μέση τιμή του βαθμού ήταν 5.5.



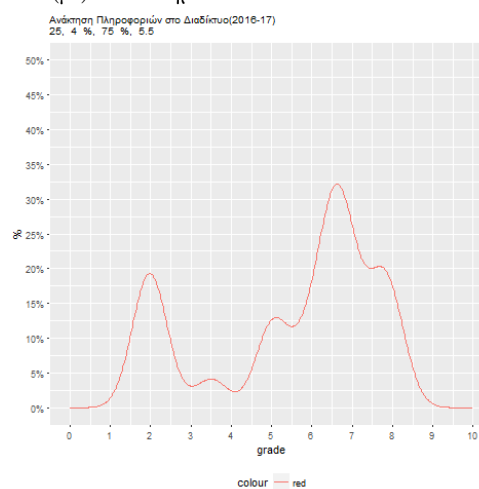
(α') Ακαδημαϊκό Έτος 2013-2014



(β') Ακαδημαϊκό Έτος 2014-2015



(γ') Ακαδημαϊκό Έτος 2015-2016



(δ') Ακαδημαϊκό Έτος 2016-2017

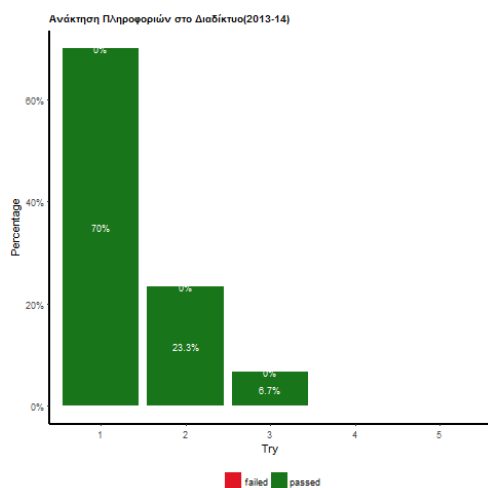
Σχήμα 6: Μεταπτυχιακό Μάθημα: Κατανομή συχνότητας βαθμών ανά ακαδημαϊκό έτος

Το Σχήμα 7α' αναφέρεται στο μεταπτυχιακό μάθημα *Ανάκτηση Πληροφοριών στο Διαδίκτυο*. Αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2013-14. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2013-2014 έως και την Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2013-14 στο εν λόγω μάθημα, το 70% το πέρασε (με την πρώτη) το 23.3% το πέρασε με την δεύτερη φορά και το 6.7% με την τρίτη φορά.

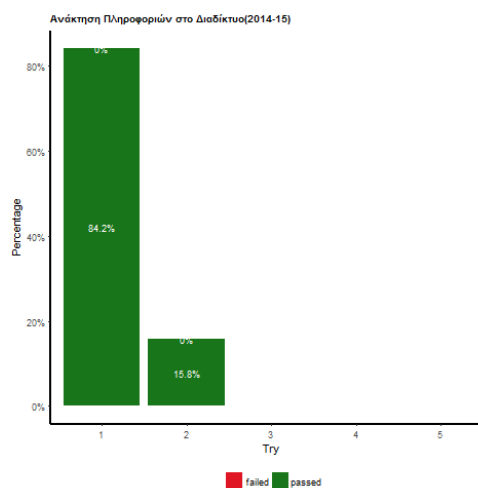
Το Σχήμα 7β' αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2014-15. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2014-2015 έως και την Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2014-15 στο εν λόγω μάθημα, το 84.2% των φοιτητών το πέρασε (με την πρώτη) φορά, ενώ το 15.8% το πέρασε με την δεύτερη φορά.

Το Σχήμα 7γ' αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2015-16. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2015-2016 έως και την Β' εξεταστική περίοδο του ακαδημαϊκού έτους 2016-17. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2015-16 στο εν λόγω μάθημα το πέρασαν όλοι οι φοιτητές με την πρώτη φορά.

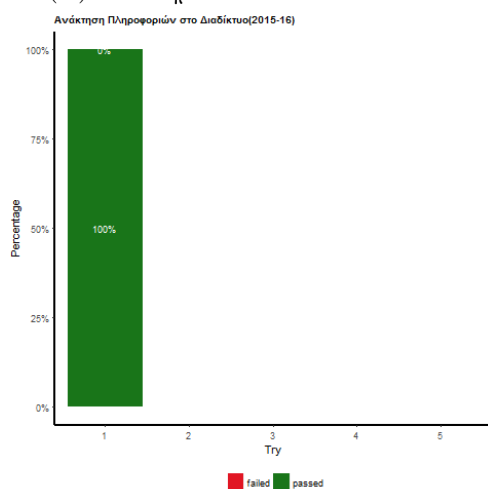
Το Σχήμα 7δ' αφορά φοιτητές/τριες που ξεκίνησαν να εξετάζονται στο εν λόγω μάθημα το ακαδημαϊκό έτος 2016-17. Ξεκινά να τους παρακολουθεί από την αρχή του ακαδημαϊκού έτους 2016-2017 έως και την Β' εξεταστική περίοδο του ίδιου ακαδημαϊκού έτους. Τα ποσοστά που αναγράφονται στην πρώτη στήλη του ιστογράμματος διαβάζονται ως εξής: από τους φοιτητές που εξετάστηκαν πρώτη φορά το ακαδημαϊκό έτος 2016-17 στο εν λόγω μάθημα, το 75% των φοιτητών το πέρασε (με την πρώτη) φορά, το 15% το πέρασε με την δεύτερη φορά, ενώ ένα ποσοστό 10% έχει εξεταστεί στο συγκεκριμένο μάθημα δύο φορές αλλά δεν το έχει περάσει.



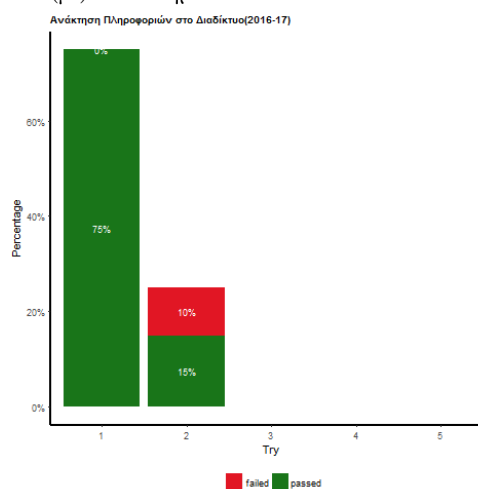
(α') Ακαδημαϊκό Έτος 2013-2014



(β') Ακαδημαϊκό Έτος 2014-2015



(γ') Ακαδημαϊκό Έτος 2015-2016



(δ') Ακαδημαϊκό Έτος 2016-2017

Σχήμα 7: Μεταπτυχιακό Μάθημα: Προσπάθειες (άξονας Try) και ποσοστά επιτυχίας/αποτυχίας εξετασθέντων (άξονας Percentage)

4 Εξόρυξη Κανόνων Συσχέτισης

4.1 Σύνολο Δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήσαμε για αυτό το μέρος της ανάλυσης αποτελείται από 166461 εγγραφές(γραπτά) και 8 μεταβλητές(πληροφορίες για τα γραπτά):

- ID Φοιτητή
- ID Μαθήματος
- Τίτλος Μαθήματος
- Εξεταστική
- Ακαδημαϊκό Έτος
- Βαθμός
- Κατεύθυνση Φοιτητή
- Τύπος Μαθήματος(Θεωρία ή Εργαστήριο)

Περιλαμβάνει δεδομένα για τις εξεταστικές περιόδους από το ακαδημαϊκό έτος 2003-2004 έως και το χειμερινό εξάμηνο του 2017-2018. Σε αυτό το χρονικό διάστημα υπάρχουν 3 διαφορετικά προπτυχιακά προγράμματα σπουδών το Π3,Π4 και Π5. Οι φοιτητές που εξετάστηκαν στα 15 ακαδημαϊκά έτη είναι 3210. Κατά τη διάρκεια των τριών διαφορετικών προγραμμάτων σπουδών υπάρχουν μαθήματα που αφαιρέθηκαν ή άλλαξαν όνομα. Επίσης υπάρχουν μικτά μαθήματα που τους αφαιρέθηκε το εργαστηριακό μέρος. Περισσότερες λεπτομέρειες για τον χειρισμό των παραπάνω περιπτώσεων υπάρχουν στην επόμενη ενότητα.

4.2 Καθαρισμός και Προετοιμασία Δεδομένων

Προεπεξεργαστήκαμε τα δεδομένα ώστε να μπορέσουμε να τα εισάγουμε στον αλγόριθμο εξόρυξης κανόνων. Συγκεκριμένα, προσθέσαμε μία καινούργια μεταβλητή η οποία έχει για κάθε μάθημα από το Π3, Π4 το αντίστοιχό τους στο Π5 (Παράρτημα Α). Με τον τρόπο αυτό μαθήματα που έχουν αλλάξει όνομα θα λογίζονται σαν ένα από τον αλγόριθμο. Στη συνέχεια ομαδοποιήσαμε όλα τα γραπτά ανά φοιτητή και για κάθε ένα από αυτούς πραγματοποιήσαμε την εξής διαδικασία:

- Κρατήσαμε μόνο τα γραπτά που βαθμολογήθηκαν με βαθμό μεγαλύτερο ή ίσο του πέντε(δηλαδή ο φοιτητής πέρασε το μάθημα).

- Επίσης κρατήσαμε τα μαθήματα που είναι μόνο θεωρία και από τα μικτά εκείνα που ο φοιτητής έχει περάσει και τα δυο μέρη: Θεωρία και Εργαστήριο.
- Από μαθήματα που ήταν μικτά σε προηγούμενα προγράμματα σπουδών και στο Π5 έχει αφαιρεθεί το εργαστήριο, λάβαμε υπόψιν μόνο τη θεωρία.
- Στη συνέχεια υπολογίσαμε τον τελικό βαθμό για τα μικτά μαθήματα, ο οποίος είναι 60% θεωρία και 40% εργαστήριο.
- Αν ο φοιτητής έδωσε το μάθημα περισσότερες από δύο φορές, αφαιρέσαμε 0,3 από το βαθμό του για κάθε επιπλέον προσπάθεια.
- Ο βαθμός που προέκυψε από τα προηγούμενα βήματα τον κατηγοριοποιήσαμε ώστε να είναι δυνατή η μετατροπή των δεδομένων σε συναλλαγές (transactions) όπως απαιτεί ο αλγόριθμος. Για την κατηγοριοποίηση έγινε ο μετασχηματισμός των συνεχών τιμών σε διαβαθμισμένες τιμές σύμφωνα με το Πίνακα 1.

Διάστημα Βαθμολογίας	Διαβαθμισμένη Κατηγορία
[0, 6.5]	μέτρια
[6.6, 7.5]	καλά
[7.7, 8.5]	πολύ καλά
[8.6, 10]	άριστα

Πίνακας 1: Αντιστοίχιση διαβαθμισμένης κατηγορίας με διαστήματα βαθμολογίας

Στον Πίνακα 2 παρουσιάζεται το ποσοστό των γραπτών ανά την διαβαθμισμένη κατηγορία στην οποία εντάσσονται.

Διαβαθμισμένη Κατηγορία	Ποσοστό Γραπτών
μέτρια	63%
καλά	18%
πολύ καλά	10%
άριστα	7,5%

Πίνακας 2: Ποσοστό γραπτών ανά διαβαθμισμένη κατηγορία

4.3 Εξαγωγή Κανόνων Συσχέτισης με τον Αλγόριθμο Apriori

4.3.1 Ο αλγόριθμος Apriori

Ο αλγόριθμος Apriori [ΑΙΣ93] είναι ένας από τους πιο γνωστούς αλγορίθμους για την εξόρυξη κανόνων συσχέτισης. Βασική δομή στην οποία στηρίζεται ο αλγόριθμος είναι τα στοιχειοσύνολα. Ο Apriori βασίζεται στην αρχή ότι όλα τα υποσύνολα ενός στοιχειοσυνόλου που εμφανίζεται συχνά, θα είναι και αυτά με την σειρά τους συχνά. Συγκεκριμένα, έστω συχνό στοιχειοσύνολο $\{A,B,\Gamma,\Delta\}$ με βάση την αρχή που στηρίζεται ο Apriori, τότε και τα υποσύνολα $\{A\}$, $\{B\}$, $\{\Gamma\}$, $\{\Delta\}$ θα είναι επίσης συχνά. Επίσης για κάθε μη συχνό υποσύνολο, πρέπει όλα τα υπερσύνολα στα οποία περιέχεται να είναι και αυτά μη συχνά.

Βασικά μεγέθη του Apriori είναι η υποστήριξη (support), εμπιστοσύνη (confidence) και ανύψωση (lift). Συγκεκριμένα, έστω X , Y τυχαία στοιχειοσύνολα, η υποστήριξη(X) ισούται με τον αριθμό των συναλλαγών στις οποίες υπάρχει το X δια τον συνολικό αριθμό των συναλλαγών.

Η εμπιστοσύνη ορίζεται ως: $confidence(X \rightarrow Y) = \frac{support(XUY)}{confidence(X)}$ και δηλώνει πόσο πιθανό είναι να επιλεγθεί το στοιχειοσύνολο Y δεδομένου ότι έχει επιλεγθεί πρώτα το στοιχειοσύνολο X .

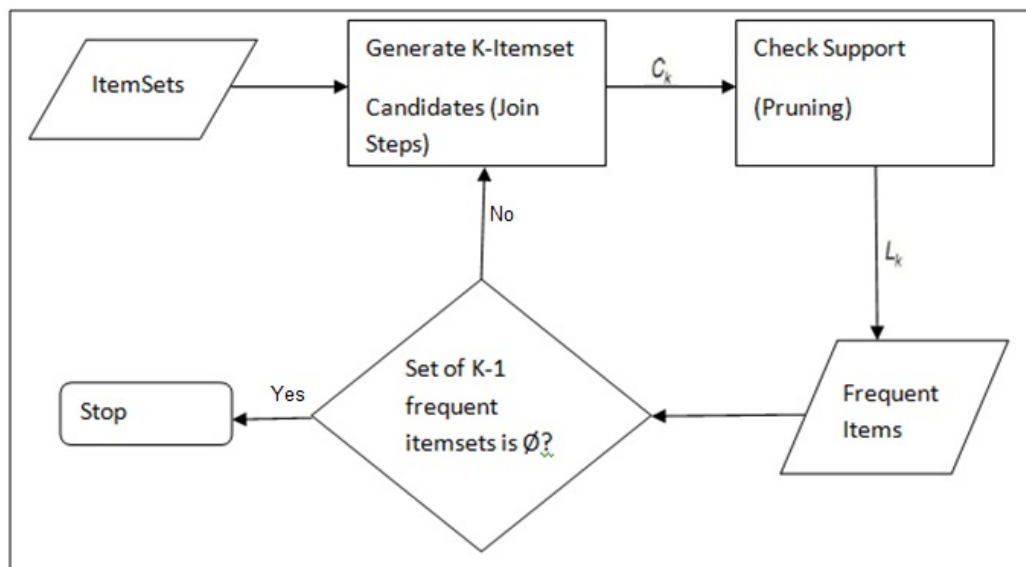
Η ανύψωση ορίζεται ως: $lift(X \rightarrow Y) = \frac{support(XUY)}{support(X)*support(Y)}$ και δηλώνει πόσο πιθανό είναι να επιλεγθεί το στοιχειοσύνολο Y δεδομένου ότι έχει επιλεγθεί το στοιχειοσύνολο X λαμβάνοντας υπόψη και το πόσο συχνά εμφανίζεται στις συναλλαγές το Y .

Η διαδικασία εξόρυξης κανόνων με τον Apriori χωρίζεται σε δύο κυρίως βήματα:

1. Εύρεση των συχνών στοιχειοσυνόλων (δηλαδή εκείνα που ικανοποιούν την ελάχιστη υποστήριξη).
2. Εξόρυξη συχνών και αξιόπιστων κανόνων συσχέτισης.
 - (α') Εξόρυξη κανόνων από συχνά στοιχειοσύνολα.
 - (β') Επιλογή αυτών που η εμπιστοσύνη είναι μεγαλύτερη ή ίση από εκείνη τις ελάχιστης προκαθορισμένης.

Για παράδειγμα, έστω ότι έχουμε μια βάση δεδομένων με X συναλλαγές. Ο αλγόριθμός διαβάσει τον πίνακα με τις συναλλαγές. Ο πίνακας θα διαβαστεί το πολύ όσες φορές είναι και ο αριθμός των στοιχείων που περιέχει. Στο πρώτο

πέρασμα του πίνακα ο αλγόριθμος υπολογίζει την υποστήριξη των στοιχειοσυνόλων, έπειτα κρατάει αυτά που ικανοποιούν την προκαθορισμένη ελάχιστη υποστήριξη. Σε κάθε επόμενο βήμα χρησιμοποιούνται τα στοιχειοσύνολα του προηγούμενου περάσματος για να δημιουργηθούν καινούργια υποψήφια συχνά στοιχειοσύνολα. Στο τέλος κάθε βήματος υπολογίζεται ποια στοιχειοσύνολα πληρούν την προκαθορισμένη υποστήριξη έτσι ώστε να χρησιμοποιηθούν στο επόμενο βήμα. Λαμβάνοντας υπόψη την βασική αρχή του Apriori απορρίπτεται ένας μεγάλος αριθμός στοιχειοσυνόλων τα οποία είναι υπερσύνολα μη συχνών συνόλων. Η παραπάνω διαδικασία τερματίζεται όταν δεν υπάρχει πλέον κάποιιο στοιχειοσύνολο που να ικανοποιεί τον περιορισμό της ελάχιστης υποστήριξης. Στο Σχήμα 8 φαίνεται το διάγραμμα ροής τους αλγορίθμου.



Σχήμα 8: Διάγραμμα Ροής Αλγορίθμου

Παρακάτω παρουσιάζεται η διαδικασία του Apriori σε μορφή ψευδοκώδικα:

Τέλος ο Apriori παρουσιάζει δύο σημαντικά μειονεκτήματα: πρώτων η παραγωγή μεγάλου αριθμού κανόνων χωρίς ενδιαφέρον και δεύτερον η παραγωγή μεγάλου μήκους στοιχειοσυνόλων όπως αναλυτικά αναφέρει ο Hegland[2].

Algorithm 1 Apriori

C_k : Candidate itemset of size k
 L_k : frequent itemset of size k
 $L_1 = \{\text{frequent items}\}$
for $k = 1; L_k \neq \emptyset; k++$; **do**
 C_{k+1} = candidates generated from L_k
 for each transaction t in database **do**
 Increment the count of all candidates in C_{k+1}
 Those are contained in t
 end for
 L_{k+1} = candidates in C_{k+1} with min_support
end for
return $\bigcup_k L_k$

4.3.2 Εφαρμογή του Apriori

Η διαδικασία που ακολουθήσαμε για την εξόρυξη κανόνων συνοψίζεται στα παρακάτω βήματα:

1. Ρύθμιση παραμέτρων (support, conf, maxlen).
2. Εισαγωγή συναλλαγών στον Apriori με τις προκαθορισμένες παραμέτρους.
3. Αφαίρεση πλεοναζόντων κανόνων (pruning).
4. Φιλτράρισμα και συλλογή κανόνων που μας ενδιαφέρουν.

Πρώτο βήμα είναι η ρύθμιση των παραμέτρων του Apriori. Ορίσαμε το support και το confidence σε διάφορες τιμές. Όσο μικρότερες είναι αυτές οι τιμές τόσο περισσότεροι είναι οι κανόνες που παράγει ο Apriori. Ο αριθμός των κανόνων ανάλογα με αυτές τις παραμέτρους μπορεί να κυμαίνεται από μερικές χιλιάδες μέχρι κάποια εκατομμύρια οι περισσότεροι εκ των οποίων είναι θόρυβος (χαμηλού ενδιαφέροντος). Η επιλογή τιμών πολύ κοντά στο μηδέν (στο support) οδηγούν σε τεράστια σύνολα παραγόμενων κανόνων τα οποία είναι δύσκολο να χειριστεί ένας συμβατικός ηλεκτρονικός υπολογιστής.

Ενδεικτικά στον πίνακα 3 οι αριθμοί των κανόνων που παρήχθησαν για διάφορες τιμές των support, confidence και maxlen:

maxlen	support	confidence	rules
10	0.01	0.55	out of memory
5	0.01	0.55	787814
7	0.01	0.6	4118656
7	0.01	0.5	4596925

Πίνακας 3: Αριθμός κανόνων για διάφορες τιμές των maxlen, support και confidence.

Επόμενο βήμα ήταν η εισαγωγή των δεδομένων στον Apriori με τις κατάλληλες παραμέτρους. Σκοπός ήταν να βρούμε σε ποια μαθήματα οι φοιτητές τείνουν να έχουν υψηλή βαθμολογία δεδομένου ότι τα έχουν πάει καλά σε κάποιο συγκεκριμένο μάθημα. Το 63% των γραπτών είχαν ως ανηγμένο βαθμό *μέτρια*. Λαμβάνοντας υπόψιν αυτό το γεγονός επιλέξαμε πολύ μικρό support(0,005) ώστε να μπορέσει ο αλγόριθμος να εντοπίσει κανόνες για επίδοση *καλά*, *πολύ καλά* και *άριστα*. confidence ορίσαμε 0,5 ώστε να επιλέξουμε εμείς τους κανόνες με βάση το lift.

Στη συνέχεια αφαιρέσαμε τους πλεονάζοντες κανόνες. Όπως αναφέρει ο Bayardo (200) Πλεονάζον είναι ένας κανόνας εάν υπάρχει ένας πιο γενικός κανόνας με ίδιο ή μεγαλύτερο confidence . Συνεπώς, ένας πιο συγκεκριμένος κανόνας είναι πλεονάζον εάν είναι μόνο ίσος ή λιγότερο προγνωστικός από έναν πιο γενικό κανόνα. Ένας κανόνας είναι πιο γενικός αν έχει το ίδιο δεξιό μέρος και ένα ή περισσότερα στοιχεία λιγότερα στο αριστερό μέρος'. Για την αφαίρεση των πλεοναζόντων κανόνων στο RStudio χρησιμοποιήσαμε την μέθοδο *is.redundant()* από το πακέτο *arules*

Τελευταίο βήμα ήταν το φιλτράρισμα των κανόνων ώστε να αφαιρέσουμε τον θόρυβο και κανόνες που δεν μας ενδιαφέρουν. Σκοπός της εργασίας ήταν να βρούμε κανόνες για μαθήματα στα οποία οι φοιτητές τείνουν να έχουν υψηλές επιδόσεις. Η πρώτη ενέργεια ήταν να βρούμε μόνο τους κανόνες από τους οποίους το δεξιό μέρος περιλάμβανε κάποιο μάθημα το οποίο ο φοιτητής είχε *καλή*, *πολύ καλή* ή *άριστη* επίδοση. Στη συνέχεια από αυτούς τους κανόνες κρατήσαμε εκείνους που είχαν περισσότερες από δέκα εμφανίσεις(count) στις συναλλαγές και lift μεγαλύτερο ή ίσο του 4 καθώς θεωρήσαμε ότι κανόνες με μικρότερες τιμές θα είναι αμφιβόλου αξίας.

4.4 Αποτελέσματα

Το Σχήμα 9ά μας πληροφορεί ότι: ο αλγόριθμος(Apriori) εξόρυξε 3549 κανόνες συσχέτισης για στοιχειοσύνολα που στο δεξί τους μέρος έχουν μαθήματα με επίδοση *καλά* (δηλαδή κάθε κουκίδα είναι ένας κανόνας της μορφής:

$$\{ \dots \} \implies \{ \text{μάθημα} \rightarrow \text{καλά} \}$$

Ο άξονας *x* μας πληροφορεί για το support των κανόνων και ο άξονας *y* για το confidence. Στο δεξιό μέρος του σχήματος υπάρχει η στήλη που έχει την πληροφορία για την τιμή του lift (όσο πιο κόκκινο το χρώμα της κουκίδας

τόσο μεγαλύτερη η τιμή του lift). Στο σχήμα παρατηρούμε ότι όσο μικρότερο είναι το support του κανόνα τόσο μεγαλύτερο είναι το lift.

Το Σχήμα 9β' μας πληροφορεί ότι: ο αλγόριθμος(Apriori) εξόρυξε 230 κανόνες συσχέτισης για στοιχειοσύνολα που στο δεξί τους μέλος έχουν μαθήματα με επίδοση πολύ καλά (δηλαδή κάθε κουκίδα είναι ένας κανόνας της μορφής:

$$\{ \dots \} \implies \{ \text{μάθημα} \rightarrow \text{πολύ καλά} \}$$

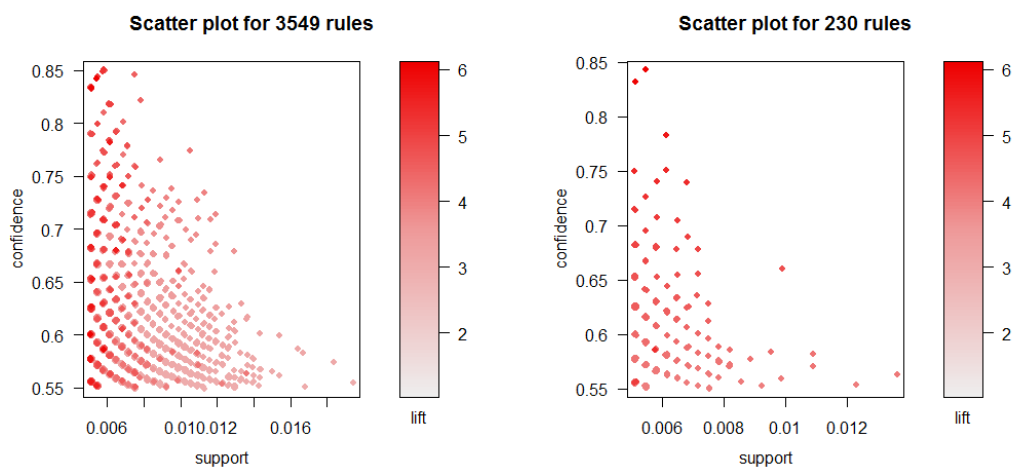
Ο άξονας x μας πληροφορεί για το support των κανόνων και ο άξονας y για το confidence. Στο δεξιό μέρος του σχήματος υπάρχει η στήλη που έχει την πληροφορία για την τιμή του lift (όσο πιο κόκκινο το χρώμα της κουκίδας τόσο μεγαλύτερη η τιμή του lift).

Αντίστοιχα και για το Σχήμα 9γ' ο αλγόριθμος(Apriori) εξόρυξε 684 κανόνες συσχέτισης για στοιχειοσύνολα που στο δεξί τους μέλος έχουν μαθήματα με επίδοση άριστα(δηλαδή κάθε κουκίδα είναι ένας κανόνας της μορφής:

$$\{ \dots \} \implies \{ \text{μάθημα} \rightarrow \text{άριστα} \}$$

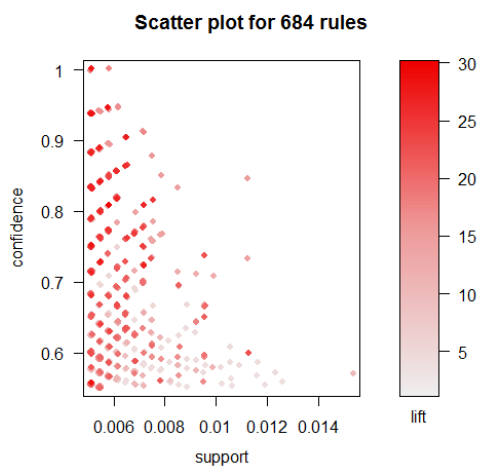
Ο άξονας x μας πληροφορεί για το support των κανόνων και ο άξονας y για το confidence. Στο δεξιό μέρος του σχήματος υπάρχει η στήλη που έχει την πληροφορία για την τιμή του lift(όσο πιο κόκκινο το χρώμα της κουκίδας τόσο μεγαλύτερη η τιμή του lift). Στο σχήμα παρατηρούμε ότι όσο μικρότερο είναι το confidence του κανόνα τόσο μεγαλύτερο είναι το lift.

```
1 # Scatter Plot Rules
2 library('arulesViz')
3 plot(rules_good, measure = c("support", "confidence"),
4      shading = "lift")
5 plot(rules_great, measure = c("support", "confidence"),
6      shading = "lift")
7 plot(rules_best, measure = c("support", "confidence"),
8      shading = "lift")
```



(α') Κανόνες: Δεξί Μέλος καλά

(β') Κανόνες: Δεξί Μέλος πολύ καλά



(γ') Κανόνες: Δεξί Μέλος άριστα

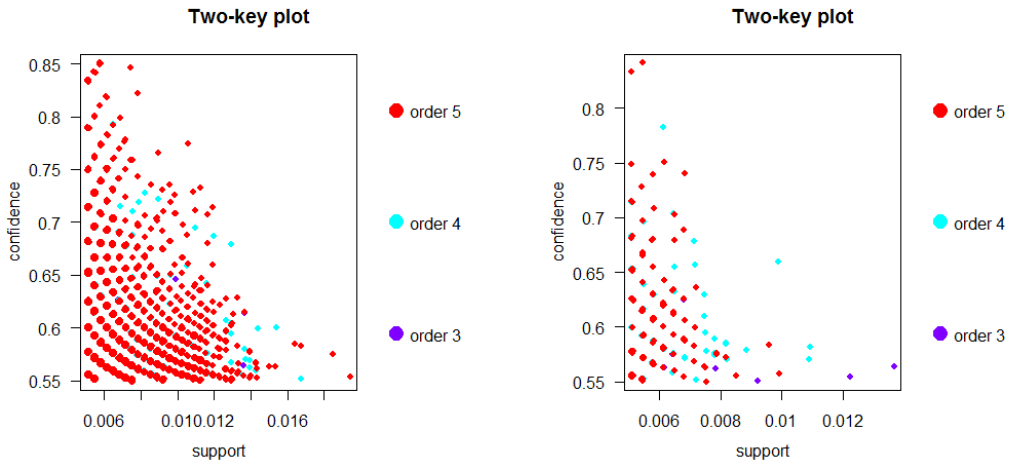
Σχήμα 9: Γράφημα διασποράς για διάφορες τιμές των *support*, *confidence* και *lift*

Το Σχήμα 10α΄ μας δίνει πληροφορίες για τις τιμές των support και confidence των παραγόμενων κανόνων όπου στο δεξί τους μέλος υπάρχει μάθημα με επίδοση *καλά* ανάλογα με το μήκος τους (αριθμός στοιχείων που αποτελούνται). Με κόκκινο χρώμα είναι οι κανόνες με μήκος πέντε (5), με γαλάζιο οι κανόνες με μήκος τέσσερα (4) και με μοβ οι κανόνες με μήκος τρία (3).

Το Σχήμα 10β΄ μας δίνει πληροφορίες για τις τιμές των support και confidence των παραγόμενων κανόνων όπου στο δεξί τους μέλος υπάρχει μάθημα με επίδοση *πολύ καλά* ανάλογα με το μήκος τους (αριθμός στοιχείων που αποτελούνται). Με κόκκινο χρώμα είναι οι κανόνες με μήκος πέντε (5), με γαλάζιο οι κανόνες με μήκος τέσσερα (4) και με μοβ οι κανόνες με μήκος τρία (3).

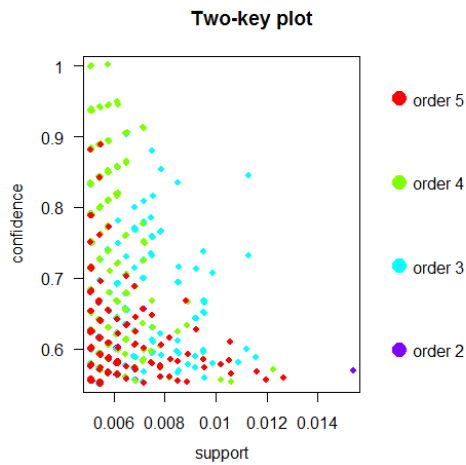
Το Σχήμα 10γ΄ μας δίνει πληροφορίες για τις τιμές των support και confidence των παραγόμενων κανόνων όπου στο δεξί τους μέλος υπάρχει μάθημα με επίδοση *άριστα* ανάλογα με το μήκος τους (αριθμός στοιχείων που αποτελούνται). Με κόκκινο χρώμα είναι οι κανόνες με μήκος πέντε (5), με ανοιχτό πράσινο κανόνες με μήκος τέσσερα (4), με γαλάζιο οι κανόνες με μήκος τρία (3) και με μοβ οι κανόνες με μήκος δύο (2).

```
1 # Plots number of items contained in the rule
2 library('arulesViz')
3 plot(rules_good, method = "two-key plot")
4 plot(rules_great, method = "two-key plot")
5 plot(rules_best, method = "two-key plot")
```



(α') Κανόνες: Δεξί Μέλος καλά

(β') Κανόνες: Δεξί Μέλος πολύ καλά



(γ') Κανόνες: Δεξί Μέλος άριστα

Σχήμα 10: Γράφημα διασποράς για διάφορες τιμές *support*, *confidence* ανά μήκος κανόνα

Παρακάτω παρουσιάζονται οι 10 (από τους συνολικούς 133 κανόνες) ισχυρότεροι κανόνες με βάση το lift οι οποίοι στο δεξί τους μέλος έχουν μάθημα με επίδοση καλά. Το support, confidence και lift παρουσιάζονται αναλυτικά στον Πίνακα 4.

1. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = πολύ καλά, Διακριτά Μαθηματικά = μέτρια, Πληροφοριακά Συστήματα II = άριστα} \Rightarrow {Μηχανική Μάθηση = **καλά**}
2. {Διακριτά Μαθηματικά = μέτρια, Δομές Δεδομένων και Ανάλυση Αλγορίθμων = καλά, Μεθοδολογίες Προγραμματισμού = μέτρια, Συστήματα Διαχείρισης Βάσεων Δεδομένων = καλά} \Rightarrow {Ανάπτυξη Διαδικτυακών Συστ. και Εφαρμογών = καλά}
3. {Γλώσσες και Τεχνολογίες Ιστού = μέτρια, Διακριτά Μαθηματικά = μέτρια, Δίκτυα Η/Υ = μέτρια, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = πολύ καλά} \Rightarrow {Αλγοριθμική και Προγραμματισμός = πολύ καλά}
4. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια, Πληροφοριακά Συστήματα I = μέτρια, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = μέτρια, Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
5. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια, Γλώσσες και Τεχνολογίες Ιστού = μέτρια, Πληροφοριακά Συστήματα I = μέτρια, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = πολύ καλά}
6. {Μαθηματική Ανάλυση και Γραμμική Άλγεβρα = μέτρια, Μεθοδολογίες Προγραμματισμού = μέτρια, Συστήματα Διαχείρισης Βάσεων Δεδομένων = καλά, Τεχνολογία Βάσεων Δεδομένων = καλά} \Rightarrow {Εισαγωγή στην Πληροφορική = πολύ καλά}
7. {Διακριτά Μαθηματικά = μέτρια, Πληροφοριακά Συστήματα I = μέτρια, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα, Τηλεπικοινωνίες και Δίκτυα Υπολογιστών = μέτρια} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = πολύ καλά}

8. {Πληροφοριακά Συστήματα II = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = μέτρια, Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
9. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = μέτρια, Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
10. {Αλγοριθμική και Προγραμματισμός = πολύ καλά,Επιχειρησιακή Έρευνα = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}

Παρακάτω παρουσιάζονται οι 10 (συνολικοί κανόνες 13) ισχυρότεροι κανόνες με βάση το lift οι οποίοι στο δεξί τους μέλος έχουν μάρτυρα με επίδοση πολύ καλά. Το support, confidence και lift παρουσιάζονται αναλυτικά στον Πίνακα 5.

1. {Γλώσσες και Τεχνολογίες Ιστού = μέτρια,Διακριτά Μαθηματικά = μέτρια,Δίκτυα Η/Υ = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = πολύ καλά} \Rightarrow {Αλγοριθμική και Προγραμματισμός = πολύ καλά}
2. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια,Πληροφοριακά Συστήματα I = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = μέτρια,Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
3. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια,Γλώσσες και Τεχνολογίες Ιστού = μέτρια,Πληροφοριακά Συστήματα I = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = πολύ καλά}
4. {Μαθηματική Ανάλυση και Γραμμική Άλγεβρα = μέτρια,Μεθοδολογίες Προγραμματισμού = μέτρια,Συστήματα Διαχείρισης Βάσεων Δεδομένων = καλά,Τεχνολογία Βάσεων Δεδομένων = καλά} \Rightarrow {Εισαγωγή στην Πληροφορική = πολύ καλά}
5. {Διακριτά Μαθηματικά = μέτρια,Πληροφοριακά Συστήματα I = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα,Τηλεπικοινωνίες και Δίκτυα Υπολογιστών = μέτρια} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = πολύ καλά}

6. {Πληροφοριακά Συστήματα II = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες κ Τεχνικές = μέτρια, Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
7. {Αρχές Σχεδίασης Λειτουργικών Συστημάτων = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = μέτρια,Τεχνολογία Βάσεων Δεδομένων = πολύ καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
8. {Αλγοριθμική και Προγραμματισμός = πολύ καλά,Επιχειρησιακή Έρευνα = μέτρια,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = καλά} \Rightarrow {Μεθοδολογίες Προγραμματισμού = πολύ καλά}
9. {Δίκτυα Η/Υ = μέτρια,Εισαγωγή στα Λειτουργικά Συστήματα = άριστα,Πληροφοριακά Συστήματα II = μέτρια} \Rightarrow {Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανιών Χρήστη = πολύ καλά}
10. {Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανιών Χρήστη = καλά,Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών = καλά,Δομές Δεδομένων και Ανάλυση Αλγορίθμων = μέτρια,Εισαγωγή στα Λειτουργικά Συστήματα = μέτρια} \Rightarrow {Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα = πολύ καλά }

Παρακάτω παρουσιάζονται οι 10 (συνολικοί κανόνες 133) ισχυρότεροι κανόνες με βάση το lift οι οποίοι στο δεξί τους μέλος έχουν μάθημα με επίδοση άριστα. Το support, confidence και lift παρουσιάζονται αναλυτικά στον Πίνακα 6.

1. {Μηχανική Μάθηση = άριστα,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Αντικειμενοστραφής Προγραμματισμός = άριστα}
2. {Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών = άριστα,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Τεχνολογία Βάσεων Δεδομένων = άριστα}
3. {Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών = άριστα,Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Αντικειμενοστραφής Προγραμματισμός = άριστα}
4. {Αλγοριθμική και Προγραμματισμός = άριστα,Ευφυή Συστήματα = άριστα} \Rightarrow {Αντικειμενοστραφής Προγραμματισμός = άριστα}

5. {Μεθοδολογίες Προγραμματισμού = άριστα, Πληροφοριακά Συστήματα II = άριστα} \Rightarrow {Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών = άριστα}
6. {Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών = άριστα, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = άριστα}
7. {Μεθοδολογίες Προγ. = άριστα, Τεχνολογία Βάσεων Δεδομένων = άριστα} \Rightarrow {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = άριστα}
8. {Εισαγωγή στα Λειτουργικά Συστήματα = άριστα, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow {Τεχνολογία Βάσεων Δεδομένων = άριστα}
9. {Μεθοδολογίες Προγραμματισμού = άριστα, Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές = άριστα} \Rightarrow { Αντικειμενοστραφής Προγραμματισμός = άριστα}
10. {Δομές Δεδομένων και Ανάλυση Αλγορίθμων = άριστα, Εισαγωγή στα Λειτουργικά Συστήματα = άριστα} \Rightarrow {Μεθοδολογίες Προγραμματισμού = άριστα}

Πίνακας 4: Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση καλά

<i>support</i>	<i>confidence</i>	<i>lift</i>	<i>count</i>
0.00862069	0.714285714	9.206349206	10
0.00862069	0.625	8.055555556	10
0.00862069	0.555555556	8.055555556	10
0.00862069	0.555555556	7.581699346	10
0.00862069	0.769230769	7.561929596	10
0.00862069	0.555555556	7.323232323	10
0.00862069	0.555555556	7.323232323	10
0.009482759	0.55	7.25	11
0.009482759	0.733333333	7.209039548	11
0.012068966	0.7	6.881355932	14

Ο Πίνακας 4, 5 και 6 απεικονίζουν τις τιμές των *support*, *confidence*, *lift* για τους δέκα ισχυρότερους κανόνες που εξόρυξε ο *Apriori*. Οι κανόνες είναι ταξινομημένοι κατά φθίνουσα σειρά με βάση το *lift*. Η στήλη *count* περιέχει τον αριθμό των συνολικών εμφανίσεων του κανόνα. Συγκεκριμένα ο πίνακας 4 περιέχει τους κανόνες όπου το δεξί τους μέρος περιέχει κάποιο μάθημα με επίδοση 'καλά'. Ο πίνακας 5 περιέχει τους κανόνες όπου το δεξί τους μέρος περιέχει κάποιο μάθημα με επίδοση 'πολύ καλά'. Ο πίνακας 6 περιέχει τους κανόνες όπου το δεξί τους μέρος περιέχει κάποιο μάθημα με επίδοση 'άριστα'.

Πίνακας 5: Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση πολύ καλά

<i>support</i>	<i>confidence</i>	<i>lift</i>	<i>count</i>
0.00862069	0.555555556	7.581699346	10
0.00862069	0.769230769	7.561929596	10
0.00862069	0.555555556	7.323232323	10
0.00862069	0.555555556	7.323232323	10
0.009482759	0.55	7.25	11
0.009482759	0.733333333	7.209039548	11
0.012068966	0.7	6.881355932	14
0.00862069	0.666666667	6.553672316	10
0.00862069	0.555555556	5.461393597	10

Πίνακας 6: Οι δέκα (10) ισχυρότεροι κανόνες για επίδοση άριστα

<i>support</i>	<i>confidence</i>	<i>lift</i>	<i>count</i>
0.00862069	0.714285714	18.83116883	10
0.00862069	0.666666667	17.57575758	10
0.00862069	0.666666667	17.57575758	10
0.00862069	0.666666667	17.57575758	10
0.00862069	0.588235294	17.05882353	10
0.00862069	0.666666667	16.45390071	10
0.00862069	0.666666667	16.45390071	10
0.011206897	0.619047619	16.32034632	13
0.011206897	0.619047619	16.32034632	13

5 Επίλογος

Στην εργασία αυτή, στο κεφάλαιο 3 προσπαθήσαμε να κατανοήσουμε καλύτερα την βαθμολογική συμπεριφορά των φοιτητών αλλά και να προσεγγίσουμε ποσοτικά το βαθμό δυσκολίας των μαθημάτων. Στη συνέχεια στο κεφάλαιο 4 με την εφαρμογή του αλγόριθμου *Apriori* και την εξόρυξη κανόνων συσχέτισης προσπαθήσαμε να κατανοήσουμε σε πια μαθήματα οι φοιτητές τείνουν να έχουν υψηλές επιδόσεις. Χρησιμοποιήθηκε μόνο η μέθοδος της εξόρυξης κανόνων συσχέτισης για την εξόρυξη χρήσιμης πληροφορίας, ωστόσο πρέπει να ληφθεί υπόψιν ότι θα μπορούσαν να χρησιμοποιηθούν και άλλες μέθοδοι εξόρυξης δεδομένων για μία πιο ολοκληρωμένη αξιοποίηση των δεδομένων.

Πέρα από την εξόρυξη γνώσης από τα ακαδημαϊκά δεδομένα, είναι πολύ σημαντικό αυτή η γνώση να εφαρμόζεται άμεσα στην εκπαιδευτική διαδικασία. Συνεπώς μελλοντικές εργασίες πρέπει να έχουν ως αντικείμενό τους τη δημιουργία νέων τρόπων για την αποτελεσματική εφαρμογή της, ήδη εξορυγμένης γνώσης, στα εκπαιδευτικά ιδρύματα.

Επόμενο βήμα αυτής της εργασίας θα ήταν αυτή η καινούργια πληροφορία να χρησιμοποιηθεί ως βάση για τη δημιουργία μιας πλατφόρμας, της οποίας στόχος θα είναι να προτείνει πιθανές κατευθύνσεις στους φοιτητές με βάση τα μαθήματα που είχαν υψηλές επιδόσεις. Η ανάλυση δεδομένων σε εκπαιδευτικά δεδομένα είναι ένας τομέας γεμάτος προκλήσεις αλλά ταυτόχρονα και ζωτικής σημασίας, για την βελτίωση της εκπαιδευτικής διαδικασίας των πανεπιστημιακών ιδρυμάτων τα επόμενα χρόνια.

Παράρτημα Α

Πίνακας 7: Τίτλοι Μαθημάτων των Π3 και Π4 προγραμμάτων σπουδών με τα αντίστοιχά τους στο Π5.

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
ΑΓΓΛΙΚΗ ΟΡΟΛΟΓΙΑ	
Αλγοριθμική & Προγραμματισμός - Ε	Αλγοριθμική και Προγραμματισμός - Ε
Αλγοριθμική & Προγραμματισμός - Θ	Αλγοριθμική και Προγραμματισμός - Θ
Αλληλεπίδραση Ανθρώπου-Μηχανής & Ανάπτυξη Διεπιφανειών Χρήστη - Ε	Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανειών Χρήστη - Ε
Αλληλεπίδραση Ανθρώπου-Μηχανής & Ανάπτυξη Διεπιφανειών Χρήστη - Θ	Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανειών Χρήστη - Θ
Ανάπτυξη Διαδικτυακών Συστημάτων & Εφαρμογών - Ε	Ανάπτυξη Διαδικτυακών Συστ. και Εφαρμογών - Ε
Ανάπτυξη Διαδικτυακών Συστημάτων & Εφαρμογών - Θ	Ανάπτυξη Διαδικτυακών Συστ. και Εφαρμογών - Θ
ΑΝΑΠΤΥΞΗ ΔΙΕΠΙΦΑΝΕΙΩΝ ΧΡΗΣΤΗ-Ε	Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανειών Χρήστη - Ε
ΑΝΑΠΤΥΞΗ ΔΙΕΠΙΦΑΝΕΙΩΝ ΧΡΗΣΤΗ-Θ	Αλληλεπίδραση Ανθρώπου-Μηχανής και Ανάπτυξη Διεπιφανειών Χρήστη - Θ
ΑΝΑΠΤΥΞΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΕΦΑΡΜΟΓΩΝ-Ε	Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πληροφ. Συστημάτων και Εφαρμογών - Ε
ΑΝΑΠΤΥΞΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΕΦΑΡΜΟΓΩΝ-Θ	Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πληροφ. Συστημάτων και Εφαρμογών - Θ
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πλ. Συστημάτων & Εφαρμογών - Ε	Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πληροφ. Συστημάτων και Εφαρμογών - Ε
Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πλ. Συστημάτων & Εφαρμογών - Θ	Ανάπτυξη και Διαχείριση Ολοκληρωμένων Πληροφ. Συστημάτων και Εφαρμογών - Θ
Αντικειμενοστρεφής Προγραμματισμός - Ε	Αντικειμενοστρεφής Προγραμματισμός - Ε
Αντικειμενοστρεφής Προγραμματισμός - Θ	Αντικειμενοστρεφής Προγραμματισμός - Θ
Αριθ. Ανάλυση και Προγ/μος Επιστημονικών Εφαρμογών - Ε	Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών - Ε
Αριθ. Ανάλυση και Προγ/μος Επιστημονικών Εφαρμογών - Θ	Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών - Θ
ΑΡΙΘΜ ΑΝΑΛΥΣΗ & ΠΡΟΓΡ. ΕΠΙΣΤΗΜΟΝ. ΕΦ.-Ε	Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών - Ε
ΑΡΙΘΜ ΑΝΑΛΥΣΗ & ΠΡΟΓΡ. ΕΠΙΣΤΗΜΟΝ. ΕΦ.-Θ	Αριθ. Ανάλυση και Προγρ/μός Επιστ. Εφαρμογών - Θ
Ασύρματα και Κινητά Δίκτυα Επικοινωνιών	Δίκτυα Ασύρματων και Κινητών Επικοινωνιών
ΑΣΦΑΛΕΙΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ-Ε	Ασφάλεια Πληροφοριακών Συστημάτων - Ε
ΑΣΦΑΛΕΙΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ-Θ	Ασφάλεια Πληροφοριακών Συστημάτων - Θ
Ασφάλεια Πληροφορικών Συστημάτων - Ε	Ασφάλεια Πληροφορικών Συστημάτων - Ε
Ασφάλεια Πληροφορικών Συστημάτων - Θ	Ασφάλεια Πληροφορικών Συστημάτων - Θ
ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ Ι-Ε	Συστήματα Διαχείρισης Βάσεων Δεδομένων - Ε
ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ Ι-Θ	Συστήματα Διαχείρισης Βάσεων Δεδομένων - Θ
ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΙΙ-Ε	Τεχνολογία Βάσεων Δεδομένων - Ε
ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΙΙ-Θ	Τεχνολογία Βάσεων Δεδομένων - Θ
ΓΛΩΣΣΕΣ ΚΑΙ ΜΕΤΑΓΛΩΤΤΙΣΤΕΣ-Ε	‘N/A
ΓΛΩΣΣΕΣ ΚΑΙ ΜΕΤΑΓΛΩΤΤΙΣΤΕΣ-Θ	‘N/A
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
Γλώσσες και Τεχνολογίες Ιστού - Ε	Γλώσσες και Τεχνολογίες Ιστού - Ε
Γλώσσες και Τεχνολογίες Ιστού - Θ	Γλώσσες και Τεχνολογίες Ιστού - Θ
Γραφικά Υπολογιστών	Γραφικά Υπολογιστών
ΓΡΑΦΙΚΑ ΥΠΟΛΟΓΙΣΤΩΝ-Ε	Γραφικά Υπολογιστών
ΓΡΑΦΙΚΑ ΥΠΟΛΟΓΙΣΤΩΝ-Θ	Γραφικά Υπολογιστών
Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Ε	Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Ε
Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Θ	Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Θ
ΔΕΞΙΟΤΗΤΕΣ ΕΠΙΚΟΙΝΩΝΙΑΣ-Ε	Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Ε
ΔΕΞΙΟΤΗΤΕΣ ΕΠΙΚΟΙΝΩΝΙΑΣ-Θ	Δεξιότητες Επικοινωνίας/Κοινωνικά Δίκτυα - Θ
Διαδικτυακές Υπηρεσίες Προστιθέμενης Αξίας (ε-εομ/γο/ε-λεαρν)	Διαδικτυακές Υπηρεσίες Προστιθέμενης Αξίας
ΔΙΑΚΡΙΤΑ ΜΑΘΗΜΑΤΙΚΑ	Διακριτά Μαθηματικά
Διακριτά Μαθηματικά	Διακριτά Μαθηματικά
ΔΙΔΑΚΤΙΚΕΣ ΜΕΘΟΔΟΙ & ΔΕΟΝΤΟΛΟΓΙΑ ΕΠΑΓΓΕΛ	*N/A
Δίκτυα Η/Υ - Ε	Δίκτυα Η/Υ - Ε
Δίκτυα Η/Υ - Θ	Δίκτυα Η/Υ - Θ
ΔΙΚΤΥΑ Η/Υ-Ε	Δίκτυα Η/Υ - Ε
ΔΙΚΤΥΑ Η/Υ-Θ	Δίκτυα Η/Υ - Θ
Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Ε	Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Ε
Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Θ	Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Θ
ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ-Ε	Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Ε
ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ-Θ	Δομές Δεδομένων και Ανάλυση Αλγορίθμων - Θ
Ειδικά θέματα δικτύων I - Ε	Ειδικά Θέματα Δικτύων I - Ε
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
Ειδικά θέματα δικτύων I - Θ	Ειδικά Θέματα Δικτύων I - Θ
Ειδικά θέματα δικτύων II - Ε	Ειδικά Θέματα Δικτύων II - Ε
Ειδικά θέματα δικτύων II - Θ	Ειδικά Θέματα Δικτύων II - Θ
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΔΙΚΤΥΩΝ I-E	Ειδικά Θέματα Δικτύων I - Ε
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΔΙΚΤΥΩΝ I-Θ	Ειδικά Θέματα Δικτύων II - Θ
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΔΙΚΤΥΩΝ I-Θ	Ειδικά Θέματα Δικτύων I - Θ
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΔΙΚΤΥΩΝ II-E	Ειδικά Θέματα Δικτύων II - Ε
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΔΙΚΤΥΩΝ II-Θ	Ειδικά Θέματα Δικτύων II - Θ
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣ.-Ε	“N/A
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣ.-Θ	“N/A
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜ/ΝΗΣ-Ε	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Ε
ΕΙΔΙΚΑ ΘΕΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜ/ΝΗΣ-Θ	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Θ
Εισαγωγή στα Λειτουργικά Συστήματα - Ε	Εισαγωγή στα Λειτουργικά Συστηματα - Ε
Εισαγωγή στα Λειτουργικά Συστηματα - Θ	Εισαγωγή στα Λειτουργικά Συστηματα - Θ
Εισαγωγή στην Πληροφορική - Ε	Εισαγωγή στην Πληροφορική - Ε
Εισαγωγή στην Πληροφορική - Θ	Εισαγωγή στην Πληροφορική - Θ
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ-Ε	Εισαγωγή στην Πληροφορική - Ε
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ-Θ	Εισαγωγή στην Πληροφορική - Θ
ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ & ΕΙΚΟΝΑΣ-Ε	“N/A
ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ & ΕΙΚΟΝΑΣ-Θ	“N/A
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ I	“N/A
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ II	“N/A
ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ	Επιχειρησιακή Έρευνα
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
Επιχειρησιακή Έρευνα	Επιχειρησιακή Έρευνα
Ευφυή Συστήματα	Ευφυή Συστήματα
ΕΥΦΥΗ ΣΥΣΤΗΜΑΤΑ-Ε	Ευφυή Συστήματα
ΕΥΦΥΗ ΣΥΣΤΗΜΑΤΑ-Θ	Ευφυή Συστήματα
ΗΛΕΚΤΡΟΝΙΚΗ ΜΑΘΗΣΗ-Ε	‘N/A
ΗΛΕΚΤΡΟΝΙΚΗ ΜΑΘΗΣΗ-Θ	‘N/A
ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ-Ε	Διαδικτυακές Υπηρεσίες Προστιθέμενης Αξίας
ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ-Θ	Διαδικτυακές Υπηρεσίες Προστιθέμενης Αξίας
Θεωρία Λειτουργικών Συστημάτων	Αρχές Σχεδίασης Λειτουργικών Συστημάτων
ΘΕΩΡΙΑ ΠΙΘΑΝ/ΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ-Ε	Θεωρία Πιθανοτήτων και Στατιστική - Ε
ΘΕΩΡΙΑ ΠΙΘΑΝ/ΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ-Θ	Θεωρία Πιθανοτήτων και Στατιστική - Θ
Θεωρία Πιθανοτήτων και Στατιστική - Ε	Θεωρία Πιθανοτήτων και Στατιστική - Ε
Θεωρία Πιθανοτήτων και Στατιστική - Θ	Θεωρία Πιθανοτήτων και Στατιστική - Θ
ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ Ι	Αρχές Σχεδίασης Λειτουργικών Συστημάτων
ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ ΙΙ-Ε	Εισαγωγή στα Λειτουργικά Συστήματα - Ε
ΛΕΙΤΟΥΡΓΙΚΑ ΣΥΣΤΗΜΑΤΑ ΙΙ-Θ	Εισαγωγή στα Λειτουργικά Συστήματα - Θ
ΜΑΘΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ	Μαθηματική Ανάλυση και Γραμμική Άλγεβρα
Μαθηματική Ανάλυση	Μαθηματική Ανάλυση και Γραμμική Άλγεβρα
ΜΕΘΟΔΟΛΟΓΙΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ Ι-Ε	Μεθοδολογίες Προγραμματισμού - Ε
ΜΕΘΟΔΟΛΟΓΙΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ Ι-Θ	Μεθοδολογίες Προγραμματισμού - Θ
Μεθοδολογίες Προγραμματισμού - Ε	Μεθοδολογίες Προγραμματισμού - Ε
Μεθοδολογίες Προγραμματισμού - Θ	Μεθοδολογίες Προγραμματισμού - Θ
ΜΕΘΟΔΟΛΟΓΙΕΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΙΙ-Ε	Μεθοδολογίες Προγραμματισμού - Ε
ΜΕΘΟΔΟΛΟΓΙΕΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΙΙ-Θ	Μεθοδολογίες Προγραμματισμού - Ε
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
ΜΕΘΟΔΟΛΟΓΙΕΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΙΙ-Θ	Μεθοδολογίες Προγραμματισμού - Θ
ΜΕΝΤΟΡΕΣ	*N/A
ΜΗΧΑΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ	Μηχανική Λογισμικού Ι - Θ
Μηχανική Λογισμικού ΙΙ	Μηχανική Λογισμικού ΙΙ
Μηχανική Λογισμικού Ι - Ε	Μηχανική Λογισμικού Ι - Ε
Μηχανική Λογισμικού Ι - Θ	Μηχανική Λογισμικού Ι - Θ
Μηχανική Μάθηση - Ε	Μηχανική Μάθηση - Ε
Μηχανική Μάθηση - Θ	Μηχανική Μάθηση - Θ
ΝΕΥΡΩΝΙΚΑ (ΝΕΥΡΟΜΟΡΦΙΚΑ) ΔΙΚΤΥΑ-Ε	Μηχανική Μάθηση - Ε
ΝΕΥΡΩΝΙΚΑ (ΝΕΥΡΟΜΟΡΦΙΚΑ) ΔΙΚΤΥΑ-Θ	Μηχανική Μάθηση - Θ
ΞΕΝΗ ΓΛΩΣΣΑ Ι	*N/A
ΟΙΚΟΝΟΜΙΑ ΤΩΝ ΕΠΙΧ/ΩΝ & ΟΡΓ/ΣΗ Δ/ΣΗ ΕΠΙΧ	*N/A
ΟΡΓ/ΣΗ & ΑΡΧ/ΚΗ ΥΠΟΛ/ΚΩΝ ΣΥΣ/ΤΩΝ-Ε	Οργάνωση και Αρχιτεκτονική Υπολ. Συστημάτων - Ε
ΟΡΓ/ΣΗ & ΑΡΧ/ΚΗ ΥΠΟΛ/ΚΩΝ ΣΥΣ/ΤΩΝ-Θ	Οργάνωση και Αρχιτεκτονική Υπολ. Συστημάτων - Θ
Οργάνωση & Αρχιτεκτονική Υπολ. Συστημάτων - Ε	Οργάνωση και Αρχιτεκτονική Υπολ. Συστημάτων - Ε
Οργάνωση & Αρχιτεκτονική Υπολ. Συστημάτων - Θ	Οργάνωση και Αρχιτεκτονική Υπολ. Συστημάτων - Θ
Οργάνωση Δεδομένων & Εξόρυξη Πληροφορίας	Οργάνωση Δεδομένων και Εξόρυξη Πληροφορίας
ΠΑΡΑΛΛΗΛΑ ΚΑΙ ΚΑΤΑΝΕΜΗΜΕΝΑ ΣΥΣ/ΤΑ-Ε	Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα
ΠΑΡΑΛΛΗΛΑ ΚΑΙ ΚΑΤΑΝΕΜΗΜΕΝΑ ΣΥΣ/ΤΑ-Θ	Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα
ΠΛΗΡ/ΚΗ ΚΑΙ Κ/ΝΙΑ & ΤΕΧ/ΚΕΣ ΠΡ/ΣΗΣ ΑΓΟΡΑΣ	*N/A
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΔΙΟΙΚΗΣΗΣ-Ε	Διοίκηση και Διαχείριση Έργων Πληροφορικής
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΔΙΟΙΚΗΣΗΣ-Θ	Διοίκηση και Διαχείριση Έργων Πληροφορικής
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ Ι	Πληροφοριακά Συστήματα Ι
Πληροφοριακά Συστήματα Ι	Πληροφοριακά Συστήματα Ι
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΙΙ	Πληροφοριακά Συστήματα ΙΙ
Πληροφοριακά Συστήματα ΙΙ - Θ	Πληροφοριακά Συστήματα ΙΙ
ΠΟΙΟΤΗΤΑ & ΑΞΙΟΠΙΣΤΙΑ ΛΟΓΙΣΜΙΚΟΥ-Ε	‘N/A
ΠΟΙΟΤΗΤΑ & ΑΞΙΟΠΙΣΤΙΑ ΛΟΓΙΣΜΙΚΟΥ-Θ	‘N/A
ΠΡΟΓ/ΣΜΟΣ ΔΙΑΔΙΚΤΥΑΚΩΝ ΕΦΑΡ/ΓΩΝ-Ε	Ανάπτυξη Διαδικτυακών Συστ. και Εφαρμογών - Ε
ΠΡΟΓ/ΣΜΟΣ ΔΙΑΔΙΚΤΥΑΚΩΝ ΕΦΑΡ/ΓΩΝ-Θ	Ανάπτυξη Διαδικτυακών Συστ. και Εφαρμογών - Θ
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΥΠΟΛΟΓΙΣΤΩΝ Ι-Ε	Αλγοριθμική και Προγραμματισμός - Ε
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΥΠΟΛΟΓΙΣΤΩΝ Ι-Θ	Αλγοριθμική και Προγραμματισμός - Θ
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΥΠΟΛΟΓΙΣΤΩΝ ΙΙ-Ε	Αντικειμενοστραφής Προγραμματισμός - Ε
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΥΠΟΛΟΓΙΣΤΩΝ ΙΙ-Θ	Αντικειμενοστραφής Προγραμματισμός - Θ
ΠΡΟΗΓΜΕΝΕΣ ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ Η/Υ-Ε	Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα
ΠΡΟΗΓΜΕΝΕΣ ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ Η/Υ-Θ	Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα
Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα	Προηγμένες Αρχιτεκτονικές Υπολογιστών και Παράλληλα Συστήματα
ΣΥΣΤΗΜΑΤΑ ΓΕΩΓΡΑΦΙΚΩΝ ΠΛΗΡΟΦΟΡΙΩΝ-Ε	‘N/A
ΣΥΣΤΗΜΑΤΑ ΓΕΩΓΡΑΦΙΚΩΝ ΠΛΗΡΟΦΟΡΙΩΝ-Θ	‘N/A
Συστήματα Διαχείρισης Βάσεων Δεδομένων - Ε	Συστήματα Διαχείρισης Βάσεων Δεδομένων - Ε
Συστήματα Διαχείρισης Βάσεων Δεδομένων - Θ	Συστήματα Διαχείρισης Βάσεων Δεδομένων - Θ
Συνέχεια στην επόμενη σελίδα	

Πίνακας 7 – Συνέχεια από προηγούμενη σελίδα

Τίτλος Μαθήματος (Π3,Π4)	Τίτλος Μαθήματος (Π5)
Τεχνητή Νοημοσύνη (Γλώσσες & Τεχνικές) - Ε	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Ε
Τεχνητή Νοημοσύνη (Γλώσσες & Τεχνικές) - Θ	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Θ
ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ-Ε	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Ε
ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ-Θ	Τεχνητή Νοημοσύνη: Γλώσσες και Τεχνικές - Θ
Τεχνολογία Βάσεων Δεδομένων - Ε	Τεχνολογία Βάσεων Δεδομένων - Ε
Τεχνολογία Βάσεων Δεδομένων - Θ	Τεχνολογία Βάσεων Δεδομένων - Θ
Τεχνολογία Πολυμέσων - Ε	Τεχνολογία Πολυμέσων - Ε
Τεχνολογία Πολυμέσων - Θ	Τεχνολογία Πολυμέσων - Θ
ΤΕΧΝΟΛΟΓΙΑ ΠΟΛΥΜΕΣΩΝ-Ε	Τεχνολογία Πολυμέσων - Ε
ΤΕΧΝΟΛΟΓΙΑ ΠΟΛΥΜΕΣΩΝ-Θ	Τεχνολογία Πολυμέσων - Θ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ & ΔΙΚΤΥΑ-Ε	Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Ε
ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ & ΔΙΚΤΥΑ-Θ	Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Θ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΕΣ ΚΑΙ ΔΙΚΤΥΑ Η/Υ	Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Θ
Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Ε	Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Ε
Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Θ	Τηλεπικοινωνίες και Δίκτυα Υπολογιστών - Θ
Ψηφιακά Συστήματα	Ψηφιακά Συστήματα
ΨΗΦΙΑΚΑ ΣΥΣΤΗΜΑΤΑ-Ε	Ψηφιακά Συστήματα
ΨΗΦΙΑΚΑ ΣΥΣΤΗΜΑΤΑ-Θ	Ψηφιακά Συστήματα

Αναφορές

References

- [Cha91] Hastie Chambers. *Statistical Models in S*. Pacific Grove, 1991.
- [AIS93] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. “Mining Association Rules Between Sets of Items in Large Databases”. In: *SIGMOD Rec.* 22.2 (June 1993), pp. 207–216. ISSN: 0163-5808. DOI: 10.1145/170036.170072. URL: <http://doi.acm.org/10.1145/170036.170072>.
- [FPS96] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. “The KDD Process for Extracting Useful Knowledge from Volumes of Data”. In: *Commun. ACM* 39.11 (Nov. 1996), pp. 27–34. ISSN: 0001-0782. DOI: 10.1145/240455.240464. URL: <http://doi.acm.org/10.1145/240455.240464>.
- [Ber02] Pavel Berkhin. “Survey Of Clustering Data Mining Techniques”. In: *A Survey of Clustering Data Mining Techniques. Grouping Multidimensional Data: Recent Advances in Clustering*. 10 (Aug. 2002). DOI: 10.1007/3-540-28349-8_2.
- [Heg03] Markus Hegland. “Advanced Lectures on Machine Learning”. In: ed. by Shahar Mendelson and Alexander J. Smola. New York, NY, USA: Springer-Verlag New York, Inc., 2003. Chap. Algorithms for Association Rules, pp. 226–234. ISBN: 3-540-00529-3. URL: <http://dl.acm.org/citation.cfm?id=863714.863722>.
- [DBP05] Naeimeh Delavari, M. Beikzadeh, and Somnuk Phon-Amnuaisuk. “Application of enhanced analysis model for data mining processes in higher educational system”. In: *2005 6th International Conference on Information Technology Based Higher Education and Training* (2005), F4B/1–F4B/6.
- [KK05] Sotiris Kotsiantis and D Kanellopoulos. “Association Rules Mining: A Recent Overview”. In: *GESTS International Transactions on Computer Science and Engineering* 32 (Nov. 2005), pp. 71–82.
- [RV07] C. Romero and S. Ventura. “Educational Data Mining: A Survey from 1995 to 2005”. In: *Expert Syst. Appl.* 33.1 (July 2007), pp. 135–146. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2006.04.005. URL: <http://dx.doi.org/10.1016/j.eswa.2006.04.005>.

- [CC08] Chen-Fu Chien and Li-Fei Chen. “Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry”. In: *Expert Systems with Applications* 34 (Jan. 2008), pp. 280–290. DOI: 10.1016/j.eswa.2006.09.003.
- [ElH09] Alaa El-Halees. “Mining Students Data to Analyze Learning Behavior: A Case Study”. In: (Jan. 2009).
- [NXC09] E.W.T. Ngai, Li Xiu, and D.C.K. Chau. “Application of data mining techniques in customer relationship management: A literature review and classification”. In: *Expert Systems with Applications* 36.2, Part 2 (2009), pp. 2592–2602. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2008.02.021>. URL: <http://www.sciencedirect.com/science/article/pii/S0957417408001243>.
- [Wic09] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.
- [SKC10] Rob Schumaker, Osama K. Solieman, and Hsiu-chin Chen. “Sports Knowledge Management and Data Mining”. In: *ARIST* 44 (Jan. 2010), pp. 115–157. DOI: 10.1002/aris.2010.1440440110.
- [Hah+11] Michael Hahsler et al. “The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets”. In: *Journal of Machine Learning Research* 12 (2011), pp. 1977–1981. URL: <http://jmlr.csail.mit.edu/papers/v12/hahsler11a.html>.
- [Wic11] Hadley Wickham. “The Split-Apply-Combine Strategy for Data Analysis”. In: *Journal of Statistical Software* 40.1 (2011), pp. 1–29. URL: <http://www.jstatsoft.org/v40/i01/>.
- [HKP12] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Jan. 2012. DOI: 10.1016/C2009-0-61819-5.
- [OP12] Stefanos Ougiaroglou and Giorgos Paschalis. “Association Rules Mining from the Educational Data of ESOG Web-Based Application”. In: *AIAI*. 2012.
- [Cha14] Winston Chang. *extrafont: Tools for using fonts*. R package version 0.17. 2014. URL: <https://CRAN.R-project.org/package=extrafont>.

- [Dra14] Adrian A. Dragulescu. *xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files*. R package version 0.5.7. 2014. URL: <https://CRAN.R-project.org/package=xlsx>.
- [HD15] Pooja R. Harne and Prof.S.D. Deshpande. “Mining of Association Rules A Review Paper”. In: *International Journal of Enfgineering Technology, Management and Applied Sciences* (2015).
- [RSt15] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2015. URL: <http://www.rstudio.com/>.
- [SOC16] Maximilian Schleich, Dan Olteanu, and Radu Ciucanu. “Learning Linear Regression Models over Factorized Joins”. In: *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. 2016, pp. 3–18. DOI: 10.1145/2882903.2882939. URL: <https://doi.org/10.1145/2882903.2882939>.
- [R C17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: <https://www.R-project.org/>.
- [Wic17a] Hadley Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.2.0. 2017. URL: <https://CRAN.R-project.org/package=stringr>.
- [Wic17b] Hadley Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. 2017. URL: <https://CRAN.R-project.org/package=tidyverse>.
- [FZR18] Ari Fadli, Mulki Zulfa, and Yogi Ramadhani. “Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness”. In: *Jurnal Teknologi dan Sistem Komputer* 6 (Oct. 2018), p. 158. DOI: 10.14710/jtsiskom.6.4.2018.158-163.
- [Hah18] Michael Hahsler. *arulesViz: Visualizing Association Rules and Frequent Itemsets*. R package version 1.3-1. 2018. URL: <https://CRAN.R-project.org/package=arulesViz>.
- [Sta18] Laerd Statistics. *Descriptive and Inferential Statistics*. 2018. URL: <https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php>.