

# Εργασία 7η

Διονύσης Κοτζαΐτης

Μελετήστε την εργασία

Bowers, Alex J. (2010) "Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis," *Practical Assessment, Research, and Evaluation*: Vol. 15 , Article 7. DOI: <https://doi.org/10.7275/r4zq-9c31>. Available at: <https://scholarworks.umass.edu/pare/vol15/iss1/7>

και απαντήστε τις παρακάτω ερωτήσεις. (Αποφεύγετε να αντιγράφετε σε μεγάλο βαθμό αυτούσια τμήματα της εργασίας ως απαντήσεις στις ερωτήσεις. Οι απαντήσεις να είναι ακριβείς, περιεκτικές. Αποφεύγετε αυτόματες μεταφράσεις όπως από Google translate).

## 1. Διατυπώστε το σκοπό της εργασίας

Σκοπός της εργασίας ήταν η εφαρμογή της ιεραρχικής ανάλυσης cluster και της δημιουργίας heatmaps σε δεδομένα από μαθητές για να γίνει έγκαιρη λήψη αποφάσεων, πάνω στους μαθητές, για την βελτίωση και «διάσωση» της ακαδημαϊκής τους πορείας. Τα δεδομένα που χρησιμοποιούνται είναι 3DM και είναι δεδομένα ενός σχετικά μικρού dataset. Ο τελικός σκοπός της μελέτης είναι σε τι βαθμό μπορούν οι βαθμοί των μαθητών να επηρεάσουν την μετέπειτα ακαδημαϊκή τους καριέρα και κατά πόσο η ταχεία βελτίωση αυτών μπορεί να αλλάξει τα πράγματα.

## Method

2. Τι πληροφορίες μας δίνουν για τη μελέτη που διενήργησαν. Η απάντησή σας να είναι επιγραμματική (απαντήστε στα παρακάτω σημεία)

- **Ιδρυμα/βαθμίδα:** Νηπιαγωγείο, Δημοτικό, Γυμνάσιο, Λύκειο
- **Χρονική διάρκεια:** Όλες οι τάξεις της σχολικής χρονιάς που θα τελείωνε τον Ιούνιο 2016.
- **Πληθυσμός:** Οι επιστήμονες έλαβαν δείγμα από 188 μαθητές από έναν πληθυσμό 6000 ατόμων. Τα άτομα προήλθαν από 2 περιοχές και οι μαθητές είχαν ως «κλάση» αποφοίτησης την χρονία 2006, χωρίς να λαμβάνεται υποψιν το dropout.
- **Συμμετέχοντες (πλήθος, προφίλ συμμετεχόντων):** 188 μαθητές από 2 περιοχές της Αμερικής (Α και Β). Η περιοχή Α είναι κεντρική πόλη μεσαίου μεγέθους. Οι κάτοικοι είναι κατά 70% σε χαμηλότερα οικονομικά στρώματα, 50% ισπανόφωνοι, 30% λευκοί και 15% αφροαμερικανοί. Σ' αυτή, έχουμε 2 δημοτικά σχολεία, 1 γυμνάσιο και 1 λύκειο, με μία τάξη νηπιαγωγείου. Η περιοχή Β έχει 50% ατόμων σε χαμηλότερα οικονομικά στρώματα, 50% λευκούς, 20% ισπανόφωνους και 15% αφροαμερικάνους. Εδώ υπάεχουν 3 δημοτικά, ένα γυμνάσιο, ένα λύκειο και μία τάξη νηπίων.

3. Δώστε τις πηγές δεδομένων, δομή στιγμιοτύπου των δεδομένων. Μέγεθος δείγματος (γραμμές, στήλες).

- **Πηγές δεδομένων:** Από κάθε περιοχή λήφθηκαν οι κάρτες αναφοράς των μαθητών που ανήκαν στην class 2006, είτε αν αποφοίτησε, είτε όχι. Τα δεδομένα ήταν οι βαθμοί του μαθητή σε κάθε τάξη, με τα μαθήματα να είναι κατηγοριοποιημένα ανάλογα με τα σχολεία της κάθε περιοχής και το πρόγραμμα σπουδών (δηλαδή μαθηματικά, αγγλικά, ομιλία, γραφή, ανάγνωση, ορθογραφία, γραφικό χαρακτήρα, επιστήμη, κοινωνικές σπουδές, ξένη γλώσσα, κυβέρνηση, οικονομικά, μουσική, φυσική αγωγή, υγεία, υπολογιστές, τέχνη, οικιακή οικονομία)

- *Δεδομένα σαν πίνακας με στιγμιότυπα:* Έχουμε 188 γραμμές για τους μαθητές και 169 στήλες για τα μαθήματα. Η αξιολόγηση είναι με το σύστημα A, B, C, D, E, F (με +/- και ακριβώς για κάθε βαθμό πλήν από τα E και F) σε σχέση με το GPA του κάθε μαθητή (από το 0 στο 4 ανα 0.333) . Οι μη αθροιστικές βαθμολογίες GPA για κάθε τάξη βγήκαν από τον μ.ο. των GPA για όλα τα μαθήματα, με τις βαθμολογίες να χωρίζονται και να καταγράφονται ανα εξάμηνο με τρόπο S1 και S2. Επιπλέον μεταβλητές για κάθε μαθητή είναι το φύλο, αν ο μαθητής μετακινείται για να φτάσει στο σχολείο, αν έλαβε μέρος στις ACT εξετάσεις για να μπει στο πανεπιστήμιο και αν αποφοίτησε εγκαίρως ή εγκατέλειψε.

#### **4. Περιγράψτε τον αλγόριθμο *Hierarchical cluster analysis* τον οποίο χρησιμοποίησαν οι συγγραφείς με τη μορφή βημάτων**

Οι συγγραφείς, ακολούθησαν τον αλγόριθμο με τα εξής βήματα:

1. Κατηγοριοποίηση των μαθημάτων ανα τάξη, εξάμηνο και αντικείμενο.
2. Οι βαθμοί έγιναν από string σε int [0,4]
3. Δημιουργείται ο πίνακας δεδομένων που περιέχει τις βαθμολογίες για τους μαθητές για τους μαθητές. Οι μαθητές χωρίζονται, όπως αναφέραμε, σε δυο κοόρτες. Προφανώς κάθε γραμμή αντιστοιχεί σε κάθε μαθητή και κάθε στήλη σε μάθημα.
4. Γίνεται κανονικοποίηση του πίνακα μέσω z-scores, δηλαδή κάθε μάθημα αντικαθίσταται με τα z-scores με βάση τις μέσες τιμές από ένα μάθημα. Αυτό αποκλείει τις περιπτώσεις δεδομένων όπου δίνεται μεγαλύτερη βαρύτητα σε cluster αλγορίθμους.
5. Κάνουμε πίνακα similarity αποστάσεων, με την χρήση uncentered correlation ως μετρική απόστασης.
6. Για ιεραρχική συσταδοποίηση χρησιμοποιούν μέση σύνδεση επαναληπτικά στον πίνακα similarity. Η επιλογή αυτή γίνεται λόγω της αποτελεσματικότητας του σε pattern recognition μεθόδους. Ξεκινάμε κάθε μάθημα ως ξεχωριστό cluster, και ενώνουμε επαναληπτικά τις συστάδες ανάλογα με την ομοιότητα συστάδων. Τα clusters που παρουσιάζουν τη μεγαλύτερη ομοιότητα ενώνονται μεταξύ τους. Η διαδικασία αυτή συνεχίζεται επαναληπτικά με την ένωση παρόμοιων ανά δύο clusters, μέχρι ο αλγόριθμος να ορίσει ολόκληρο το σύνολο δεδομένων ως ένα cluster.

#### **5. Ποιες ήταν οι αιτίες για ελλιπή δεδομένα και πως τα αντιμετώπισαν;**

Οι λόγοι που παρουσιάζονται για τα ελλιπή δεδομένα είναι:

- Αλλαγές σχολείων μαθητή.
- Εγκατάλειψη μαθητή.
- Διαφορετικά μαθήματα επιλογής ανάλογα με τον μαθητή.

Το πρόβλημα αυτό αντιμετωπίστηκε η μέθοδος που αναφέραμε πάνω, δηλαδή average linkage, χρησιμοποιούμε δηλαδή τις μέσες αποστάσεις ανάμεσα στα στοιχεία που πρέπει να συσταδοποιηθούν. Έτσι αν υπάρχουν ελλιπείς τιμές για κάθε μαθητή, χρησιμοποιείται την μέση απόσταση από τους βαθμούς του σε σχέση με τον επόμενο, για να κοιτάξει αν βρίσκεται στο ίδιο cluster. Γι αυτό και όπως αναφέραμε μαθητές είναι ιδανικός για pattern recognition. Πχ μαθητές που εγκατέληψαν στην ίδια τάξη θα έχουν κοινό μοτίβο συσταδοποίησης, ενώ θα δωθεί μεγαλύτερη βαρύτητα στις τάξεις που παρακολούθησε.

#### **6. Τι αναπαριστούν στην *Figure 2*. Περιγράψτε (δώστε κάθε πληροφορία που αφορά την οπτικοποίηση των δεδομένων και πληροφορίας). Με τα *heatmaps* ποια δεδομένα/πληροφορία αναπαριστούν; Ο συνδυασμός *heatmaps* με *cluster tree* τι επιτυγχάνει;**

- Heatmaps: Οπτικοποιήσεις οι οποίες δείχνουν τα δεδομένα συσταδοποιημένα και με συγκεκριμένους χρωματισμούς δείχνουν στοιχεία των δεδομένων, όπως συχνότητες και ποσοότητες.

- Clustergram: Συνδυασμός heatmaps και cluster tree. Το cluster tree αποτελείται από τις ιεραρχικές συσταδοποιήσεις από τον αλγόριθμο και δείχνουν τον τρόπο με τον οποίο έγιναν. Ο συνδυασμός αυτός δείχνει την συσταδοποίηση σε σχέση με τα πραγματικά δεδομένα. Σ αυτά, τα δεδομένα είναι z-scored όπως αναφέραμε, με την διαφορά ότι έχουμε και την οπτικοποίηση της συσταδοποίησης.

Στο figure 2 έχουμε το clustergram όπου κάθε γραμμή αντιπροσωπεύει έναν μαθητή. Οι στήλες δείχνουν ένα επαναλαμβανόμενο μοτίβο μαθημάτων για κάθε τάξη, τα οποία ταξινομούνται από αριστερά στα δεξιά από πιο σημαντικά στα λιγότερα. Έτσι έχουμε συνδυασμούς τάξεις και μαθήματα:

- Νηπιαγωγείο με μαθηματικά, ομιλία, γραφή, ανάγνωση.
- Τάξεις 1 έως 4 με μαθηματικά, ανάγνωση, γραφή, ορθογραφία, τρόπος γραφής, επιστήμη, κοινωνικές επιστήμες.
- Τάξη 5 με μαθηματικά, ανάγνωση, Αγγλικά, ορθογραφία, τρόπος γραφής, επιστήμη, κοινωνικές επιστήμες.
- Τάξη 6 με μαθηματικά, ανάγνωση, Αγγλικά, ορθογραφία, τρόπος γραφής, επιστήμη, κοινωνικές επιστήμες, μουσική, φυσική αγωγή, τέχνη.
- Τάξη 7 με μαθηματικά, Αγγλικά, επιστήμη, κοινωνικές επιστήμες, μουσική, φυσική αγωγή, υγεία, τέχνη. Τάξη 8 με μαθηματικά, Αγγλικά, επιστήμη, κοινωνικές επιστήμες, μουσική, γυμναστική, οικιακή οικονομία, τέχνη;
- Τάξη 9 πρώτο εξάμηνο με μαθηματικά, Αγγλικά, επιστήμη, κοινωνικές επιστήμες, ξένη γλώσσα, κυβέρνηση, οικονομικά, μουσική, φυσική αγωγή, υπολογιστές, τέχνη, οικιακή οικονομία.
- Τάξη 9 δεύτερο εξάμηνο μέχρι την τάξη 12 δεύτερο εξάμηνο με μαθηματικά, Αγγλικά, επιστήμη, κοινωνικές επιστήμες, ξένη γλώσσα, κυβέρνηση, οικονομικά, μουσική, φυσική αγωγή, υπολογιστές, τέχνη, οικιακή οικονομία.

Οπού κάθε στήλη έχει και τον βαθμό του κάθε μαθητή. Το heatmap δείχνει σαν χρώματα την απόκλιση τριών μονάδων από τον μέσο όρο του μαθήματος με το μπλέ να είναι -3 το κόκκινο +3 και το γκρι 0. Λευκές τιμές δείχνουν έλλειψη δεδομένων. Στις τελευταίες στήλες του clustergram έχουμε τις κατηγορικές «ειδικές» μεταβλητές που αναφέρθηκαν παραπάνω.

## **Findings**

**7. Τι αναπαριστούν στην Figure 3; Περιγράψτε (δώστε κάθε πληροφορία που αφορά στην οπτικοποίηση των δεδομένων και εν γένει της πληροφορίας). Μπορείτε να απαντήσετε συνδυάζοντας μαζί τις ερωτήσεις 7 και 8**

Το διάγραμμα 7 δείχνει το clustergram της έρευνας. Τα δεδομένα παρουσιάζονται σε 3 τμήματα.

- Ιεραρχική συσταδοποίηση αριστερά, όπως γίνεται με τον αλγόριθμο που αναφέραμε και average linking.
- Το κομμάτι του heatmap με z-scored βαθμολογίες και τυπική απόκλιση, όπως αναφέρθηκε στο 6.
- Αριστερά οι καταγεγραμμένες κατηγορικές μεταβλητές για κάθε μαθητή.

**8.Ο συγγραφέας ισχυρίζεται ότι η αναπαράσταση είναι εξαιρετικά ενημερωτική όσον αφορά τους μαθητές που αποτυγχάνουν. Επίσης κάνει κάποιες διαπιστώσεις σχετικά με τα 2 μεγαλύτερα clusters που σχηματίζονται. Περιγράψτε.**

Οι συγγραφείς αναφέρουν ότι χρησιμοποιώντας την ιεραρχική συσταδοποίηση και την οπτικοποίηση που αναφέραμε, προκύπτουν ενδιαφέροντα συμπεράσματα όσον αφορά τους μαθητές που αποτυγχάνουν στις πρώτες τάξεις και στην μετέπειτα ακαδημαϊκή τους πορεία. Στο figure 3 βλέπουμε δύο μεγάλες ομάδες που οι συγγραφείς κατέδειξαν με την διακεκομμένη γραμμή. Στην πρώτη συστάδα, οι μαθητές λαμβάνουν κυρίως υψηλές βαθμολογίες κατά τη σχολική τους πορεία και αποφοιτούν εγκαίρως. Στη δεύτερη συστάδα, οι μαθητές λαμβάνουν χαμηλότερες βαθμολογίες κοντά στο μέσο όρο και βλέπουμε οι μαθητές της δεύτερης συστάδας να εγκαταλείπουν πιο συχνά το σχολείο σε σχέση με τους μαθητές της πρώτης συστάδας. Επιπλέον, υπάρχει η παρατήρηση ότι το 38% (88.6% των dropout) των μαθητών της δεύτερης συστάδας εγκαταλείπει το σχολείο, ενώ για την πρώτη συστάδα έχουμε μόνο 6%, όπου οι περισσότεροι μαθητές δώσαν εισαγωγικές εξετάσεις για πανεπιστήμιο.

**9. Διαπιστώνονται κάποιες μικρές διαφορές στα 2 clusters σχετικά με τις επιδόσεις των μαθητών. Αναφέρετε σε ποιες κατηγορίες μαθητών παρατηρήθηκαν στατιστικά σημαντικές διαφορές και ποιον έλεγχο διενήργησαν και ποια τα αποτελέσματα;**

Η διαφορά ανάμεσα στα cluster είναι για το αν οι μαθητές προέρχονται από την Α ή την Β περιοχή. Για αυτό το λόγο κάνουμε έλεγχο σημαντικότητας  $\chi^2$ , για να ελέγξουμε αν η διαφορά αυτή είναι στατιστικά σημαντική. Τα αποτελέσματα του στατιστικού ελέγχου ισούνται με 3.97 έχει 1 και 18 βαθμούς ελευθέριας και η p τιμή του ελέγχου είναι 0.045, δηλαδή μικρότερη από τη στάθμη σημαντικότητας 5%. Με p-value 0.05 έχουμε τιμή μικρότερη, άρα δεν αποδεχόμαστε την  $H_0$  άρα οι δυο συστάδες δεν διαφέρουν ανάμεσα στην καταγωγή των μαθητών.

**10. Τι αναπαριστούν στην Figure 4; Περιγράψτε (δώστε κάθε πληροφορία που αφορά στην οπτικοποίηση των δεδομένων και εν γένει της πληροφορίας).**

Στο figure 4 βλέπουμε την ιεραρχική ανάλυση συστάδων για τον εντοπισμό πιθανών μαθητών που θα εγκαταλείψουν το σχολείο, αλλά αυτή την φορά σπάμε το σύνολο μαθητών σε δυο υποσύνολα:

1. Το πρώτο υποσύνολο είναι μαθητές από νηπιαγωγείο μέχρι 8<sup>η</sup> τάξη, με ακρίβεια συσταδοποίησης 93.9%.
2. Το δεύτερο υποσύνολο είναι από νηπιαγωγείο μέχρι 6<sup>η</sup> τάξη, με ακρίβεια ότι θα εγκαταλήψουν το σχολείο 63%.

Έτσι, βλέπουμε 2 clustergrams για τα δύο υποσύνολα που σχηματίζονται από την ανάλυση ιεραρχικής συσταδοποίησης σ' αυτά τα δύο υποσύνολα.

**11. Στην Figure 3 οπτικοποιούν «four individual course grade clusters». Πως τα αναπαριστούν και τι πληροφορία δίνουν;**

Στο figure 3 έχουμε την οπτικοποίηση αυτή, η οποία γίνεται ανάμεσα στο heatmap και στο cluster tree. Αυτές οι 4 συστάδες αναπαρίστανται με κάθετες μπάρες στα αριστερά του γραφήματος 3 με 4 διαφορετικά χρώματα για τις 4 διαφορετικές συστάδες:

- Συστάδα high-high → πορτοκαλί χρώμα
- Συστάδα low-high → κίτρινο χρώμα
- Συστάδα high-low → πράσινο χρώμα
- Συστάδα low-low → μώβ χρώμα

Αυτά τα grade clusters δείχνουν τις αποδόσεις των μαθητών από τις μικρές τάξεις μέχρι 12<sup>η</sup> τάξη και δίνουν πληροφορίες για το αν ένας μαθητής πρόκειται να εγκαταλήψει το σχολείο σε μετέπειτα πορεία. Για παράδειγμα στην high high κλάση έχουμε μαθητές που ξεκινούν με υψηλές βαθμολογίες και συνεχίζουν με αυτές μέχρι το τέλος, ενώ έχουν υψηλό ποσοστό ότι θα δώσουν εισαγωγικές για το κολλέγιο.

**12. Στην Figure 5 τι αναπαριστούν και με ποιο τρόπο; Η αναπαράσταση αυτή πιο συμπέρασμα βοηθά να εξαχθεί;**

Εδώ έχουμε την κατανομή των GPA σε σχέση με τα grade clusters που δημιουργούνται και αναφέρονται παραπάνω.

- *High high cluster*: Οι μαθητές έχουν μ.ο. 3.5 gpa με το 97.7% των μαθητών να έχει αποφοιτήσει στην ώρα του.
- *Low low cluster*: έχουν χαμηλό GPA εξ αρχής, με το GPA να παραμένει χαμηλό, με 40% να εγκαταλείπουν το σχολείο πριν την αποφοίτηση.
- *High low cluster*: Αρχικά υψηλό GPA με αυτό να πέφτει όσο περνάνε οι τάξεις, το 45% των μαθητών εγκατέλειψε.
- *Low high cluster*: Χαμηλό GPA αρχικά, που ανέβηκε ο βαθμός σταδιακά, με 100% των μαθητών να αποφοιτά.

Το συμπέρασμα που εξαγάγουμε από το figure είναι ότι για αυτό το σύνολο δεδομένων, τα μοτίβα των συστάδων δείχνουν ότι οι βαθμολογίες που μπαίνουν από τους καθηγητές δεν έχουν σταθερή τάση μέχρι την 4<sup>η</sup> τάξη. Μετά την 4<sup>η</sup> τάξη, φαίνεται ότι οι συστάδες να παρουσιάζουν συγκεκριμένες τάσεις όσον αφορά τον βαθμό και την απόδοση των μαθητών στην εισαγωγή στο πανεπιστήμιο.

### **13. Η μέθοδος που χρησιμοποιήθηκε για να εντοπιστεί η “flag” γιατί είναι καλύτερη από προηγούμενες που περιγράφονται στη βιβλιογραφία;**

Ο βασικός στόχος ήταν να εντοπιστεί ένα flag ώστε να μπορέσουμε να εντοπίσουμε με ακρίβεια αν ένας μαθητής εγκαταλείπει το σχολείο. Μεσα από αυτά τα flags, οι εκπαιδευτικοί θα μπορέσουν να βγάλουν συμπεράσματα και να λάβουν σύντομα αποφάσεις για να εντοπίσουν το ρίσκο του μαθητή ως προς την εγκατάλειψη, αλλά και να λάβουν αποφάσεις πριν να είναι αργά. Παλιότερες βιβλιογραφικές έρευνες χρησιμοποιούσαν διάφορες μεθόδους, όπως:

- Παλαιότερα, για τον εντοπισμό των flags ως μεταβλητές χρησιμοποιούσαν γραμμική και λογιστική παλινδρόμησης σε συνδυασμό με έλεγχο σημαντικότητας και υπολόγιζαν το ποσοστό των μαθητών που εγκατέλειψαν το σχολείο και ταυτόχρονα παρουσίαζαν την καθορισμένη ‘flag’ ή το συνδυασμό από ‘flags’.
  - Gleason & Dynarski με πολλαπλή παλινδρόμηση είχαν τη δυνατότητα να εντοπίζουν το 43% των μαθητών που εγκαταλείπουν το σχολείο. Για το λύκειο χρησιμοποιήσαν:
    - Αν η οικογένεια που δέχεται δημόσια βοήθεια.
    - Εγκατάλειψη του σχολείου από τα αδέρφια τους.
    - Αριθμός απουσιών.
    - External locus of control, κλπ.

Σε επίπεδο γυμνασίου, εντόπισαν μόνο το 23% των μαθητών που τελικά εγκατέλειψαν το σχολείο.

- Οι Balfanz χρησιμοποίησαν λογιστική παλινδρόμηση και εντόπισαν flags για την 6<sup>η</sup> τάξη, με το 60% των μαθητών που τελικά εγκατέλειψαν το σχολείο πριν αποφοιτήσουν από το λύκειο στο δείγμα που μελέτησαν. Τα flags ήταν:
  - Χαμηλή παρακολούθηση.
  - Μη ικανοποιητική συμπεριφορά.
  - Αποτυχίες στα μαθηματικά και τα αγγλικά.
- Η ιεραρχική ανάλυση συνιστωσών που παρουσιάζεται εντοπίζει μαθητές που εγκαταλείπουν το σχολείο χρησιμοποιώντας μόνο ένα τύπο δεδομένων που έχουν ήδη συλλεχτεί από το σχολείο, τα οποία είναι οι βαθμοί και φαίνεται ότι η μέθοδος αυτή παρουσιάζει πλεονεκτήματα σε σχέση με τις προηγούμενες μεθόδους.

### **14. Ο συγγραφέας συγκρίνει τη regression analysis με τη μέθοδο που χρησιμοποίησε στην εργασία του. Δώστε τους ισχυρισμούς του.**

Κατά τη σύγκριση έχουμε:

- Η μέθοδος HCA κρατάει τα δεδομένα για κάθε μαθητή, και κάνει ομαδοποίηση ανάλογα με τα χαρακτηριστικά, δημιουργώντας κοινά patterns. Έτσι, μπορούμε να παρατηρήσουμε ότι οι τάσεις των μαθητών φαίνονται μετά την 4<sup>η</sup> τάξη καλύτερα.
- Η μέθοδος είναι το ίδιο αποτελεσματική με προηγούμενες μεθόδους για την 6<sup>η</sup> τάξη και τα αποτελέσματα της μεθόδου HCA δίνουν την ένδειξη ότι η μέθοδος είναι μάλλον καλύτερη για δεδομένα μεγαλύτερων τάξεων σε σχέση με τις προηγούμενες μεθόδους.
- Χρησιμοποιείται ένας τύπος δεδομένων, οι βαθμοί του μαθητή που είναι προσβάσιμοι σε όλα τα σχολεία σχεδόν.

- Γίνεται συνδυασμός μεθόδου ανάλυσης των δεδομένων, αλλά και οπτικοποίησης.

**15. Τα ευρήματα της βιβλιογραφίας αναφορικά με τη χρήση των βαθμών για τη διαπίστωση της εγκατάλειψης του σχολείου δείχνουν ότι είναι προβληματική η χρήση τους ή όχι; Δώστε τους 4 λόγους και τι ισχύει για τη δική τους έρευνα.**

Ο συγγραφέας αναφέρει ότι όντως η χρήση των βαθμών είναι προβληματική, σε αντίθεση με προηγούμενες έρευνες. Οι λόγοι είναι οι εξής:

- Δίνεται έμφαση στις αποτυχίες του μαθητή σε μαθήτα γενικής παιδείας (πχ μαθηματικά ή αγγλικά), μαθήματα τα οποία οι μαθητές έχουν αρνητικές εμπειρίες από νωρίς στην ακαδημαϊκή πορεία. Στην έτευνά τους οι συγγραφείς λαμβάνουν υπόψιν όλα τα μαθήματα, όπως και τα επιλογής.
- Προηγούμενες μελέτες απαιτούν, επιπλέον μεταβλητές, όπως μία μεταβλητή για την παρακολούθηση και μία μεταβλητή για την μη ικανοποιητική συμπεριφορά προκειμένου να κάνουν αναλύσεις. Αντίθετα, σύμφωνα με τη μελέτη που γίνεται στην παρούσα εργασία για το σύνολο δεδομένων προτείνεται ότι μόνο οι βαθμολογίες αρκούν για να γίνει πρόβλεψη.
- Οι προηγούμενες μελέτες δίνουν βάρος σε μεμονωμένους βαθμούς και τάξεις, σε αντίθεση με την τωρινή έρευνά που χρησιμοποιεί το σύνολο των βαθμών.
- Στις προηγούμενες μελέτες μελετώνται ως προς την ακαδημαϊκή πορεία τα μαθήματα κορμού και η πορεία των μαθητών σε αυτά.

**16. Εξετάζοντας την Figure 3 ο συγγραφέας αξιολογεί τους τύπους των μαθημάτων στα οποία οι μαθητές εγγράφονται καθ' όλη τη διάρκεια της φοίτησης τους και αναλύει τα πρότυπα επιλογής των μαθημάτων για διαφορετικές ομάδες μαθητών. Τι διαπιστώνει;**

Στο clustergram του διαγράμματος 3 γίνεται και μια επιπλέον αξιολόγηση των μαθημάτων, όσον αφορά το βάρος τους και την επιτυχία του μαθητή στην μετέπειτα ακαδημαϊκή του πορεία. Επειδή τα μαθήματα εμφανίζονται για κάθε τάξη/semester από αριστερά προς τα δεξιά με φθίνουσα σημαντικότητα, παρατηρούμε ότι τα μαθήματα κορμού εμφανίζουν αυξημένη σημαντικότητα σε σχέση με τα υπόλοιπα μαθήματα, χωρίς αυτό να σημαίνει ότι τα υπόλοιπα δεν παίζουν ρόλο στην εκπαίδευση του μαθητή. Επιπλέον όσο μεγαλύτερη είναι η βαθμίδα του μαθητή σε συστάδες βαθμών, τόσο τείνει ο ίδιος να επιλέγει μαθήματα κορμού και να είναι καλός σ' αυτά. Δηλαδή, ένας μαθητής της τάξης high high τείνει να έχει μεγαλύτερους βαθμούς στα μαθήματα γενικής παιδείας από τους μαθητές των άλλων cluster, ενώ η επιλογή των μαθημάτων αυτών φθίνει όσο μειώνουμε την απόδοση των μαθητών.

