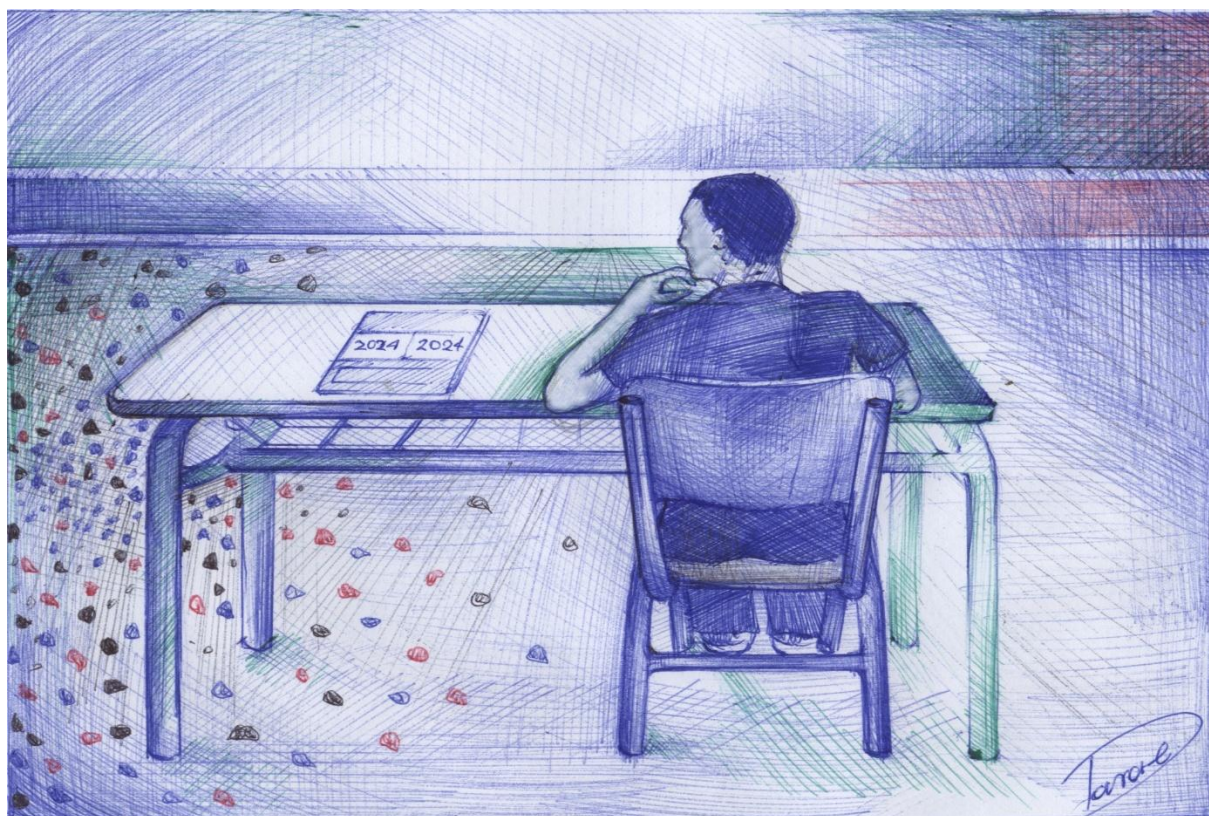




ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
«ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΕΞΟΡΥΞΗ
ΔΕΔΟΜΕΝΩΝ ΠΑΝΕΛΛΑΔΙΚΩΝ ΕΞΕΤΑΣΕΩΝ»



Της φοιτήτριας
Παρασκευής Δημητριάδου
Αρ. Μητρώου: 9/2024

Επιβλέπων
Στέφανος Ουγιάρογλου
Επίκουρος Καθηγητής

Φεβρουάριος 2026

Τίτλος Δ.Ε. Στατιστική Ανάλυση και Εξόρυξη Δεδομένων Πανελλαδικών Εξετάσεων
Κωδικός Δ.Ε. 25321

Όνοματεπώνυμο φοιτητή: Παρασκευή Δημητριάδου

Όνοματεπώνυμο εισηγητή: Στέφανος Ουγιάρογλου

Ημερομηνία ανάληψης Δ.Ε.: 05/10/2025

Ημερομηνία περάτωσης Δ.Ε.: 11/02/2026

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Δημητριάδου Παρασκευής που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιοδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

*«Αφιερώνω στα παιδιά μου,
Κωνσταντίνα & Δανάη»*

Εικόνα εξώφυλλου: Δανάη Παναγιώτου

Περίληψη

Η παρούσα διπλωματική εργασία εστιάζει στη στατιστική ανάλυση και την εξόρυξη δεδομένων των Πανελλαδικών Εξετάσεων των ετών 2023 και 2024, αξιοποιώντας ανοικτά δεδομένα που διατίθενται από το Υπουργείο Παιδείας, Θρησκευμάτων και Αθλητισμού. Τα δεδομένα περιλαμβάνουν δημογραφικά χαρακτηριστικά υποψηφίων, ομάδες προσανατολισμού, βαθμολογικές επιδόσεις, τις προτιμήσεις των υποψηφίων στο μηχανογραφικό δελτίο καθώς και τη σχολή επιτυχίας (εφόσον φυσικά έχει πετύχει) του κάθε υποψήφιου.

Στόχος της εργασίας είναι η ανάδειξη προτύπων, τάσεων και συσχετίσεων που χαρακτηρίζουν τόσο τις εκπαιδευτικές επιλογές όσο και τις επιδόσεις των μαθητών, μέσω της εφαρμογής στατιστικών μεθόδων και τεχνικών εξόρυξης γνώσης. Ιδιαίτερη έμφαση δίνεται στη διερεύνηση διαφορών μεταξύ των Ομάδων Προσανατολισμού, καθώς και στη μελέτη της επίδρασης θεσμικών αλλαγών στο σύστημα των Πανελλαδικών Εξετάσεων, όπως η καθιέρωση της Ελάχιστης Βάσης Εισαγωγής και η διαφοροποίηση των συντελεστών βαρύτητας ανά Τμήμα.

Η μεθοδολογία που ακολουθήθηκε περιλαμβάνει αρχικά εκτενή προεπεξεργασία και ενοποίηση των δεδομένων, ώστε να διασφαλιστεί η ποιότητά τους και η καταλληλότητά τους για ανάλυση. Στη συνέχεια πραγματοποιείται στατιστική ανάλυση με τη χρήση της γλώσσας προγραμματισμού Python στο περιβάλλον Google Colab, με σκοπό την αποτύπωση βασικών χαρακτηριστικών και την εξαγωγή περιγραφικών συμπερασμάτων.

Σε επόμενο στάδιο εφαρμόζονται τεχνικές εξόρυξης δεδομένων, και συγκεκριμένα αλγόριθμοι συσταδοποίησης k-means και DBSCAN, κατηγοριοποίησης Random Forest, Random Forest Regression & Naïve Bayes και ο αλγόριθμος Apriori για την εξαγωγή κανόνων συσχέτισης με τη χρήση της γλώσσας προγραμματισμού Python. Μέσω των τεχνικών αυτών επιδιώκεται η αποκάλυψη κρυφών ή μη προφανών μοτίβων μέσα στα δεδομένα, η ομαδοποίηση υποψηφίων με παρόμοια χαρακτηριστικά και η αναγνώριση συσχετίσεων μεταξύ προτιμήσεων, επιδόσεων και αποτελεσμάτων εισαγωγής.

Τέλος, τα αποτελέσματα της ανάλυσης συγκρίνονται, όπου είναι δυνατό, με αντίστοιχα δεδομένα προηγούμενων ετών, προκειμένου να αναδειχθούν διαχρονικές τάσεις και μεταβολές στη συμπεριφορά των υποψηφίων και στο πλαίσιο των αλλαγών που έχουν επέλθει κατά καιρούς στο σύστημα εισαγωγής στην τριτοβάθμια εκπαίδευση.

Λέξεις Κλειδιά: Πανελλαδικές εξετάσεις, Ανοικτά Δεδομένα, Υπουργικές αποφάσεις, Εξόρυξη δεδομένων, Αλγόριθμοι Συσταδοποίησης, Κατηγοριοποίηση, Κανόνες συσχέτισης

«STATISTICAL ANALYSIS AND DATA MINING OF NATIONAL EXAMS»

«Paraskevi Dimitriadou»

Abstract

This thesis focuses on the statistical analysis and data mining of the of the 2023 and 2024 Panhellenic Examinations, utilizing open data provided by the Ministry of Education, Religious Affairs and Sports. The data includes candidates' demographic characteristics, orientation groups, grade performance, preferences of candidates in the computerized application, as well as the school of admission for each successful candidate.

The aim of this work is to highlight patterns, trends, and correlations that characterize both educational choices and student performance through the application of statistical methods and knowledge mining techniques. Particular emphasis is placed on examining differences among the Orientation Groups, as well as on studying the impact of institutional changes in the Panhellenic Examination system, such as the establishment of the Minimum Entry Base and the differentiation of Weighting Factors per Department.

The methodology initially includes extensive data pre-processing and integration to ensure data quality and suitability for analysis. Subsequently, statistical analysis is carried out using the Python programming language in the Google Colab environment, in order to capture key features and extract descriptive conclusions.

In a subsequent stage, data mining techniques are applied, including clustering algorithms k-means and DBSCAN, classification using Random Forest, Random Forest Regression, Naive Bayes and association rule mining through the Apriori algorithm using the Python programming language. These techniques are employed to uncover hidden or non-obvious patterns within the data, group candidates with similar characteristics, and identify correlations between preferences, performance, and admission results.

Finally, the results of the analysis are compared, where possible, with corresponding data from previous years in order to highlight trends and changes in candidate behavior over time, as well as in the context of changes that have occurred over time in the higher education admission system.

Keywords: National exams, Open Data, Ministerial decisions, Data Mining, Clustering algorithms, Association rules, Categorization

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες προς τον επιβλέποντα καθηγητή μου για την πολύτιμη καθοδήγηση, την επιστημονική υποστήριξη και τη συνεχή ενθάρρυνση που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της εργασίας. Η ουσιαστική συμβολή και οι εποικοδομητικές του παρατηρήσεις συνέβαλαν καθοριστικά στη βελτίωση και την ποιότητα της παρούσας διπλωματικής εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω τον σύζυγό μου και τα παιδιά μου για την κατανόηση, την υπομονή και τη στήριξή τους σε όλη αυτή τη διαδρομή. Ο χρόνος που αφιέρωσα στην έρευνα και τη συγγραφή σήμαινε λιγότερες στιγμές μαζί τους, ωστόσο η αγάπη και η αφοσίωσή τους αποτέλεσαν πηγή δύναμης και μου έδωσαν το κουράγιο να συνεχίσω.

Περιεχόμενα

Περίληψη.....	vi
Abstract	viii
Ευχαριστίες	iii
Περιεχόμενα	ix
Κατάλογος Σχημάτων	xii
Κατάλογος Πινάκων.....	xii
Κατάλογος Γραφημάτων.....	xii
Κατάλογος Εικόνων	xiii

Κεφάλαιο 1^ο: Εισαγωγή

1.1. Εισαγωγή στα Πανεπιστήμια της Ελλάδας	1
1.2. Υπουργικές Αποφάσεις και τρέχων σύστημα των Πανελληνίων.....	1
1.3. Εισαγωγή νέων μηχανισμών αξιολόγησης.....	3
1.4. Ελάχιστη Βάση Εισαγωγής – 2021	3
1.5. Στόχοι και θεσμικό πλαίσιο της Ελάχιστης Βάσης Εισαγωγής.....	4
1.6. Προϋποθέσεις εισαγωγής	4
1.7. Συντελεστές Βαρύτητας – 2022	4
1.8. Τρόπος υπολογισμού μορίων	5
1.9. Κίνητρο.....	6
1.10. Συνεισφορά.....	6
1.11. Σχετικές εργασίες.....	7
1.12. Οργάνωση Διπλωματικής Εργασίας	8

Κεφάλαιο 2^ο: Αλγόριθμοι Εξόρυξης Δεδομένων

2.1. Εξόρυξη γνώσης	9
2.2. Συσταδοποίηση (Clustering).....	11
2.2.1. Αλγόριθμος k-means	11
2.2.2. Elbow	12
2.2.3. Αλγόριθμος DBSCAN	12
2.3. Εξαγωγή κανόνων συσχέτισης.....	12
2.3.1. Ο αλγόριθμος Apriori.....	12
2.4. Κατηγοριοποίηση.....	13

2.4.1. Random Forest	13
2.4.2. Random Forest Regression	14
2.4.3. Naive Bayes και Gaussian Naive Bayes	15
Κεφάλαιο 3^ο: Συλλογή Δεδομένων	
3.1. Ανοιχτά Δεδομένα.....	17
3.2. Πρωτογενή δεδομένα Πανελλαδικών εξετάσεων	17
3.3. Προεπεξεργασία και Μετασχηματισμός των δεδομένων για στατιστική ανάλυση.....	18
3.4. Τελικά Σύνολα Δεδομένων	19
Κεφάλαιο 4^ο: Στατιστική Ανάλυση	
4.1 Στατιστική Ανάλυση Υποψηφίων 2023-2024	21
4.2 Επιτυχόντες και Αποτυχόντες Υποψήφιοι ανά Είδος Λυκείου και ανά Ομάδα Προσανατολισμού	23
4.3 Υποψήφιοι ανά Φύλο	28
4.4. Επιτυχόντες ανά Φύλο	30
4.5. Ποσοστά Επιτυχόντων & Αποτυχόντων ανά Διεύθυνση 2023 και 2024	35
4.6. Στατιστικά Βαθμών ανά Μάθημα, ανά Ομάδα και ανά Έτος.....	38
4.7. Ποσοστά αριστούχων και μαθητών κάτω της βάσης ανά Μάθημα - 2024	41
4.8. Πλήθος επιτυχόντων ανά σειρά προτίμησης στο μηχανογραφικό του 2024	42
4.9. Πλήθος επιτυχόντων ανά σειρά προτίμησης και ανά Φύλο	44
4.10. Ποιες σχολές είναι υψηλά στις προτιμήσεις των υποψηφίων το 2024	45
4.11. Ποσοστά Δηλώσεων και Επιτυχιών σε Σχολές Πληροφορικής Θεσσαλονίκης	46
4.12. Σύγκριση διετιών 2016-2017 και 2023-2024	48
Κεφάλαιο 5^ο: Συσταδοποίηση	
5.1. Προεργασία δεδομένων	51
5.2. Αλγόριθμος k-means / Elbow	52
5.3. Αλγόριθμος DBSCAN	61
Κεφάλαιο 6^ο: Κατηγοριοποίηση	
6.1. Random Forest Regression.....	63
6.2. Gaussian Naive Bayes	70
6.3. Random Forest.....	76
6.4. Συγκριτική ανάλυση των αλγορίθμων Gaussian Naive Bayes και Random Forest .	82
Κεφάλαιο 7^ο: Εξαγωγή Κανόνων συσχέτισης	

7.1.	Κανόνες συσχέτισης σχετικά με τις Προτιμήσεις των Υποψηφίων	84
7.2.	Κανόνες συσχέτισης που αφορούν το Φύλο και τη Σειρά Επιτυχίας των Υποψηφίων ως προς την προτίμηση τους	95
7.3.	Κανόνες Συσχέτισης σχετικές με την Επίδοση των Μαθημάτων	97
Κεφάλαιο 8ο: Συμπεράσματα και Μελλοντική Έρευνα		
8.1.	Συμπεράσματα	103
8.2.	Μελλοντική Έρευνα	104
ΒΙΒΛΙΟΓΡΑΦΙΑ		105
ΠΑΡΑΡΤΗΜΑ: ΤΙΤΛΟΣ ΠΑΡΑΡΤΗΜΑΤΟΣ.....		107

Κατάλογος Σχημάτων

Σχήμα 2.1: Στάδια της διαδικασίας Ανακάλυψης Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD)	10
Σχήμα 3.1: Συγχώνευση Πρωτογενών Αρχείων και Δημιουργία Νέων	20
Σχήμα 5.1.: Δημιουργία αρχείων κατάλληλων για Συσταδοποίηση	52
Σχήμα 7.1: Δημιουργία αρχείων κατάλληλων για τον εξαγωγή Κανόνων Συσχέτισης	85
Σχήμα 7.2: Δημιουργία αρχείων κατάλληλων για τον εξαγωγή Κανόνων Συσχέτισης	87

Κατάλογος Πινάκων

Πίνακας 4.1.1: Πλήθος Υποψηφίων ανά Είδος Λυκείου	21
Πίνακας 4.1.2: Πλήθος Υποψηφίων ανά Ομάδα Προσανατολισμού - ΓΕΛ	22
Πίνακας 4.2.1 Επιτυχόντες / Αποτυχόντες ανά Έτος ΕΠΑΛ	23
Πίνακας 4.2.2: Αποτυχόντες ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ	24
Πίνακας 4.2.3: Επιτυχόντες ανά Ομάδα Προσανατολισμού & ανά Έτος - ΓΕΛ	24
Πίνακας 4.3.1: Υποψήφιοι ανά Φύλο, ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ	28
Πίνακας 4.4.1: Επιτυχόντες ανά Φύλο, ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ	31
Πίνακας 4.4.2: Ποσοστά Μεταβολών ανά Ομάδα	33
Πίνακας 4.11.1: Πλήθος Δηλώσεων και Επιτυχιών σε Σχολών Πληροφορικής Θεσ/νίκης	46
Πίνακας 4.12.1: Ποσοστά επιτυχίας ανά Ομάδα Προσανατολισμού και ανά Έτος	49
Πίνακας 5.2.1: Μέσοι όροι χαρακτηριστικών ανά cluster (k-Means – Ανθρωπιστικών Σπουδών	53
Πίνακας 5.2.2: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)- Θετικών Σπουδών	55
Πίνακας 5.2.3: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)- Σπουδών Υγείας	57
Πίνακας 5.2.4: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)-Σπουδών Οικονομίας & Πληροφορικής	59
Πίνακας 6.1: Συγκριτικός πίνακας Μαθημάτων	70

Κατάλογος Γραφημάτων

Γράφημα 4.1.1: Ποσοστά Υποψηφίων 2023	21
Γράφημα 4.1.2: Ποσοστά Υποψηφίων 2024	21
Γράφημα 4.1.3 & 4.1.4: Υποψήφιοι ανά Ομάδα Προσανατολισμού 2023 - 2024	23
Γράφημα 4.2.1: Επιτυχόντες και Αποτυχόντες από ΓΕΛ 2023-2024	25
Γράφημα 4.2.2: Επιτυχόντες και Αποτυχόντες από ΕΠΑΛ 2023-2024	25
Γράφημα 4.2.3: Επιτυχόντες και Αποτυχόντες ανά Ομάδα Προσανατολισμού από ΓΕΛ 2023	26
Γράφημα 4.2.4: Ποσοστά Επιτυχόντων & Αποτυχόντων Ανά Ομάδα Προσανατολισμού 2024	26
Γράφημα 4.2.5 Ποσοστά Επιτυχόντων Ανά Ομάδα Προσανατολισμού και ανά Έτος ως προς το σύνολο	27
υποψηφίων της ομάδας	27
Γράφημα 4.3.1: Υποψήφιοι ΓΕΛ ανά Φύλο 2023	29
Γράφημα 4.3.2: Υποψήφιοι ΓΕΛ ανά Φύλο 2024	29
Γράφημα 4.3.3: Υποψήφιοι ανά Φύλο & ανά Ομάδα Προσανατολισμού 2023	29
Γράφημα 4.3.4: Υποψήφιοι ανά Φύλο & ανά Ομάδα Προσανατολισμού 2024	30
Γραφήματα 4.4.1 και 4.4.2: Ποσοστά Επιτυχόντων ανά Φύλο 2023 & 2024	31
Γράφημα 4.4.3: Ποσοστά Επιτυχόντων/Αποτυχόντων ανά Φύλο ως προς το σύνολο των αγοριών και των κοριτσιών αντίστοιχα - 2024	32
Γράφημα 4.4.4: Ποσοστά Επιτυχόντων ανά Φύλο & ανά Ομάδα Προσανατολισμού 2023	32
Γράφημα 4.4.5: Ποσοστά Επιτυχόντων ανά Φύλο & ανά Ομάδα Προσανατολισμού 2024	33
Γράφημα 4.4.6: Μεταβολή Ποσοστού Επιτυχίας 2024 σε σχέση με 2023	34
Γράφημα 4.5.1: Ποσοστά Επιτυχόντων ανά Διεύθυνση 2023	35
Γράφημα 4.5.2: Ποσοστά Επιτυχόντων ανά Διεύθυνση 2024	36

Γραφήματα 4.6.1 & 4.6.2: Κατανομή Βαθμών Ανθρωπιστικών Σπουδών 2023-2024.....	38
Γραφήματα 4.6.3 & 4.6.4: Κατανομή Βαθμών Θετικών Σπουδών 2023-2024.....	39
Γραφήματα 4.6.5 & 4.6.6: Κατανομή Βαθμών Σπουδών Υγείας 2023-2024	39
Γραφήματα 4.6.7 & 4.6.8 : Κατανομή Βαθμών Σπουδών Οικονομίας & Πληροφορικής 2023-2024.....	40
Γράφημα 4.7.1: Ποσοστό Αριστούχων (18-20) ανά Μάθημα 2024	41
Γράφημα 4.7.2: Ποσοστό Μαθητών κάτω από τη βάση (0-9.9) ανά Μάθημα 2024	42
Γράφημα 4.8.1: Πλήθος Επιτυχόντων ανά Σειρά Προτίμησης (Top 20 προτιμήσεις).....	43
Γράφημα 4.8.2: Ποσοστό Επιτυχόντων ανά Σειρά Προτίμησης (Top 20).....	43
Γράφημα 4.9.1: Πλήθος Επιτυχόντων Αγοριών ανά Σειρά Προτίμησης (Top 20 επιλογές)	44
Γράφημα 4.9.2: Πλήθος Επιτυχόντων Κοριτσιών ανά Σειρά Προτίμησης (Top 20 επιλογές)	44
Γράφημα 4.9.3: Κατανομή Επιτυχόντων Αγοριών-Κοριτσιών ανά Σειρά Προτίμησης(1η-20 ^η).....	45
Γράφημα 4.10.1: Κατανομή Σχολών με βάση τις προτιμήσεις των υποψηφίων 2024.....	46
Γράφημα 4.11.1: Ποσοστά 1 ^{ης} προτίμησης σε Σχολές Πληροφορικής Θεσσαλονίκης.....	47
Γράφημα 4.11.2: Δηλώσεις και Επιτυχίες σε Σχολές Πληροφορικής Θεσσαλονίκης	47
Γράφημα 4.12.1: Ποσοστά Επιτυχόντων Υποψηφίων ΓΕΛ έτη 2016-2017.....	48
Γράφημα 4.12.2: Ποσοστά Επιτυχόντων Υποψηφίων ΓΕΛ έτη 2023-2024.....	49
Γράφημα 4.12.3: Εξέλιξη Ποσοστών ανά Ομάδα Προσανατολισμού (2016-2024)	50
Γράφημα 5.2.1: Elbow k-means (Ανθρωπιστικών Σπουδών)	53
Γράφημα 5.2.2: Clusters k-means with Parallel Coordinates (Ανθρωπιστικών Σπουδών)	54
Γράφημα 5.2.3: Elbow/k-means (Θετικών Σπουδών).....	55
Γράφημα 5.2.4: Clusters k-means with Parallel Coordinates (Θετικών Σπουδών).....	56
Γράφημα 5.2.5: Elbow/k-means (Σπουδών Υγείας)	57
Γράφημα 5.2.6: Clusters k-means with Parallel Coordinates (Σπουδών Υγείας)	58
Γράφημα 5.2.7: Elbow/k-means (Σπουδών Οικονομίας & Πληροφορικής)	59
Γράφημα 5.2.8: Clusters k-means with Parallel Coordinates (Σπουδών Οικονομίας & Πληροφορικής)	60
Γράφημα 5.3.1: k-distance με αυτόματο εύρος eps=0.897 και min_samples=10 και Clusters.....	61
Γράφημα 5.3.2: k-distance με αυτόματο εύρος eps=0.987 και min_samples=10 και Clusters.....	61
Γράφημα 5.3.3: k-distance με αυτόματο εύρος eps=1.347 και min_samples=10 και Clusters.....	62
Γράφημα 5.3.4: k-distance με αυτόματο εύρος eps=0.564 και min_samples=10 και Clusters.....	62

Κατάλογος Εικόνων

Εικόνα 6.1: Random Forest - Αποτελέσματα 1 –Ανθρωπιστικών Σπουδών.....	63
Εικόνα 6.2: Random Forest - Αποτελέσματα 2 –Ανθρωπιστικών Σπουδών.....	64
Εικόνα 6.3: Random Forest - Αποτελέσματα 1 – Θετικών Σπουδών.....	65
Εικόνα 6.4: Random Forest - Αποτελέσματα 2 – Θετικών Σπουδών.....	65
Εικόνα 6.5: Random Forest - Αποτελέσματα 1 - Σπουδών Υγείας.....	66
Εικόνα 6.6: Random Forest - Αποτελέσματα 2 - Σπουδών Υγείας.....	67
Εικόνα 6.7: Random Forest - Αποτελέσματα 1 -Σπουδών Οικονομίας & Πληροφορικής	68
Εικόνα 6.8: Random Forest - Αποτελέσματα 2- Σπουδών Οικονομίας & Πληροφορικής	69
Εικόνα 6.9: Gaussian Naive Bayes - Αποτελέσματα Ανθρωπιστικών Σπουδών	71
Εικόνα 6.10: Gaussian Naive Bayes -Αποτελέσματα Θετικών Σπουδών	72
Εικόνα 6.11: Gaussian Naive Bayes - Σπουδών Υγείας	73
Εικόνα 6.12: Gaussian Naive Bayes -Σπουδών Οικονομίας & Πληροφορικής	74
Εικόνα 7.1: Στιγμιότυπο αρχείου ΦΥΛΟ_ΠΡΟΤΙΜΗΣΕΙΣ_ΑΠΟΤΕΛΕΣΜΑΤΑ.....	84
Εικόνα 7.2: Στιγμιότυπο αρχείου ΦΥΛΟ_ΣΧΟΛΗ_ΕΠΙΤΥΧΙΑΣ_ΒΑΘΜΟΙ.....	85
Εικόνα 7.3: Στιγμιότυπο αρχείου ΠΡΟΤΙΜΗΣΕΙΣ.....	85
Εικόνα 7.4: Στιγμιότυπο αρχείου ΠΡΟΤΙΜΗΣΕΙΣ 1,2,3,4.....	86
Εικόνα 7.5: Στιγμιότυπο αρχείου MERGE_VATHMOI_PROT1234_ANTHROP_2024.csv	86
Εικόνα 7.6: Στιγμιότυπο αρχείου ΒΑΘΜΟΙ_ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ.....	95

Κεφάλαιο 1^ο: Εισαγωγή

1.1. Εισαγωγή στα Πανεπιστήμια της Ελλάδας

Οι πανελλαδικές εξετάσεις εισαγωγής στα δημόσια πανεπιστήμια της Ελλάδας, γνωστές ως Πανελλήνιες Εξετάσεις, αποτελούν τον βασικό μηχανισμό πρόσβασης στην τριτοβάθμια εκπαίδευση και διεξάγονται κάθε χρόνο στα τέλη Μαΐου και κατά τη διάρκεια του Ιουνίου.

Η διαδικασία αυτή συνοδεύεται από έντονη ψυχολογική πίεση, καθώς οι μαθητές προετοιμάζονται εντατικά για μια εξεταστική περίοδο περιορισμένης διάρκειας, συχνά επί μακρό χρονικό διάστημα. Η εισαγωγή στην τριτοβάθμια εκπαίδευση καθορίζεται αποκλειστικά από τις επιδόσεις στα τέσσερα πανελλαδικώς εξεταζόμενα μαθήματα, χωρίς να λαμβάνεται υπόψη ο βαθμός του απολυτηρίου και ο βαθμός των υπόλοιπων τάξεων του λυκείου. Σε συνδυασμό με την κοινωνική αντίληψη περί της σημασίας της πανεπιστημιακής εκπαίδευσης, οι Πανελλήνιες Εξετάσεις αποκτούν ιδιαίτερη βαρύτητα για τους υποψηφίους.[1]

Η προετοιμασία των μαθητών υπερβαίνει συχνά τα όρια της τυπικής σχολικής εκπαίδευσης και περιλαμβάνει ιδιωτικά μαθήματα ή φοίτηση σε φροντιστήρια, τα οποία αποτελούν διαδεδομένη πρακτική στην Ελλάδα. Η επιλογή αυτή συνεπάγεται σημαντικό οικονομικό κόστος για τις οικογένειες, το οποίο διαφοροποιείται ανάλογα με το γνωστικό αντικείμενο, τη διάρκεια και τη συχνότητα της εκπαιδευτικής υποστήριξης.

Παρότι η γενική εισαγωγή στην τριτοβάθμια εκπαίδευση επιτυγχάνεται από υψηλό ποσοστό υποψηφίων η εισαγωγή σε συγκεκριμένα πανεπιστήμια και τμήματα υψηλής ζήτησης εξαρτάται από τη γεωγραφική τους θέση και το γνωστικό αντικείμενο. Πανεπιστήμια μεγάλων αστικών κέντρων και τμήματα όπως η Νομική, η Ιατρική και οι Πολυτεχνικές Σχολές είναι υψηλά στις προτιμήσεις των υποψηφίων παρά τις ιδιαίτερα υψηλές βαθμολογικές επιδόσεις που απαιτούν και το αυξημένο κόστος διαβίωσης των φοιτητών. Αντίθετα, ενώ η εισαγωγή σε λιγότερο δημοφιλή τμήματα, σε αστικά ή επαρχιακά κέντρα, είναι ευκολότερη, τα τμήματα αυτά δεν προσελκύουν αντίστοιχο ενδιαφέρον και συχνά εμφανίζουν κενές θέσεις. Πράγμα που εντείνεται και με την Ελάχιστη Βάση Εισαγωγής που θεσπίστηκε το 2021.[1]

Κάθε εξέταση βαθμολογείται στην κλίμακα 0-20, με το 20 να είναι το υψηλότερο και το 10 να θεωρείται (η βάση). Ωστόσο, ανάλογα με τη μέση βαθμολογία του, ένας μαθητής μπορεί να καταφέρει να μπει στο πανεπιστήμιο ακόμα κι αν, για παράδειγμα, σημείωσε βαθμό κάτω του 10 σε κάποιο από τα εξεταζόμενα μαθήματα, ανάλογα πάντα και με την Ελάχιστη Βάση Εισαγωγής του μαθήματος στην εκάστοτε σχολή.

Οι Πανελλήνιες Εξετάσεις διαρθρώνονται σε δύο φάσεις: την εξέταση των βασικών πανελλαδικώς εξεταζόμενων μαθημάτων και, όπου απαιτείται, την εξέταση ειδικών μαθημάτων ή πρακτικών δοκιμασιών, όπως ξένες γλώσσες, μουσική και σχέδιο, που αφορούν συγκεκριμένες κατηγορίες σπουδών.

1.2. Υπουργικές Αποφάσεις και τρέχων σύστημα των Πανελληνίων

2020 - Οι πρώτες τροποποιήσεις στο σύστημα των Πανελλαδικών Εξετάσεων που ισχύει μέχρι και σήμερα ξεκίνησαν το έτος 2020. Σύμφωνα με την Υπουργική Απόφαση Φ.253/11812/Α5 (ΦΕΚ Β' 345/07.02.2020), στο άρθρο 4Α του ν. 4186/2013, όπως αυτό προστέθηκε με την παρ. 5 του άρθρου 100 του ν. 4610/2019, οι Σχολές, τα Τμήματα και οι Εισαγωγικές Κατευθύνσεις Τμημάτων κατατάσσονται σε τέσσερα (4) Επιστημονικά Πεδία και τέσσερις Ομάδες Προσανατολισμού, τα οποία είναι τα εξής:

Επιστημονικά Πεδία

- 1ο Επιστημονικό Πεδίο: Ανθρωπιστικές, Νομικές και Κοινωνικές Επιστήμες
- 2ο Επιστημονικό Πεδίο: Θετικές και Τεχνολογικές Επιστήμες
- 3ο Επιστημονικό Πεδίο: Επιστήμες Υγείας και Ζωής
- 4ο Επιστημονικό Πεδίο: Επιστήμες Οικονομίας και Πληροφορικής

Ομάδες Προσανατολισμού

- Ομάδα Προσανατολισμού Ανθρωπιστικών Σπουδών
- Ομάδα Προσανατολισμού Θετικών Σπουδών
- Ομάδα Προσανατολισμού Σπουδών Υγείας
- Ομάδα Προσανατολισμού Σπουδών Οικονομίας και Πληροφορικής

Κάθε Ομάδα Προσανατολισμού επιτρέπει την πρόσβαση σε 1 ΜΟΝΟ συγκεκριμένο Επιστημονικό Πεδίο κατόπιν συμμετοχής σε εξετάσεις που διεξάγονται σε πανελλαδικό επίπεδο με θέματα από την εξεταστέα ύλη της τάξης αυτής που προκύπτουν αποκλειστικά από Κεντρική Επιτροπή Εξετάσεων. Επομένως, η πρόσβαση σε συγκεκριμένο Επιστημονικό Πεδίο γίνεται ανάλογα με την Ομάδα Προσανατολισμού που έχει επιλέξει ο κάθε υποψήφιος/α.

Οι μαθητές/τριες της τελευταίας τάξης Γενικού Λυκείου από το σχολικό έτος 2019-2020 και μέχρι σήμερα εξετάζονται και παρακολουθούν τέσσερα (4) μαθήματα και είναι υποψήφιοι/ες για ένα (1) και μόνο Επιστημονικό Πεδίο ανάλογα με την Ομάδα Προσανατολισμού όπου ανήκει ο/η κάθε υποψήφιος/α. Συγκεκριμένα:

Ομάδα Προσανατολισμού ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ

Οι υποψήφιοι/ες της Ομάδας Προσανατολισμού Ανθρωπιστικών Σπουδών, έχουν πρόσβαση στις Σχολές και τα Τμήματα του 1ου Επιστημονικού Πεδίου και τα μαθήματα στα οποία εξετάζονται υποχρεωτικά είναι τα εξής τέσσερα (4):

1. Νεοελληνική Γλώσσα και Λογοτεχνία Γενικής Παιδείας
2. Αρχαία Ελληνικά Ομάδας Προσανατολισμού
3. Ιστορία Ομάδας Προσανατολισμού
4. Κοινωνιολογία που έγινε το 2021 Λατινικά Ομάδας Προσανατολισμού

Ομάδα Προσανατολισμού ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ

Οι υποψήφιοι/ες της Ομάδας Προσανατολισμού Θετικών Σπουδών, έχουν πρόσβαση στις Σχολές και τα Τμήματα του 2ου Επιστημονικού Πεδίου και τα μαθήματα στα οποία εξετάζονται υποχρεωτικά είναι τα εξής τέσσερα (4):

1. Νεοελληνική Γλώσσα και Λογοτεχνία Γενικής Παιδείας
2. Φυσική Ομάδας Προσανατολισμού
3. Χημεία Ομάδας Προσανατολισμού
4. Μαθηματικά Ομάδας Προσανατολισμού

Ομάδα Προσανατολισμού ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ

Οι υποψήφιοι/ες της Ομάδας Προσανατολισμού Σπουδών Υγείας, έχουν πρόσβαση στις Σχολές και τα Τμήματα του 3ου Επιστημονικού Πεδίου και τα μαθήματα στα οποία εξετάζονται υποχρεωτικά είναι τα εξής τέσσερα (4):

1. Νεοελληνική Γλώσσα και Λογοτεχνία Γενικής Παιδείας
2. Φυσική Ομάδας Προσανατολισμού
3. Χημεία Ομάδας Προσανατολισμού
4. Βιολογία Ομάδας Προσανατολισμού

Ομάδα Προσανατολισμού **ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ**

Οι υποψήφιοι/ες της Ομάδας Προσανατολισμού Σπουδών Οικονομίας και Πληροφορικής, έχουν πρόσβαση στις Σχολές και τα Τμήματα του 4ου Επιστημονικού Πεδίου και τα μαθήματα στα οποία εξετάζονται υποχρεωτικά είναι τα εξής τέσσερα (4):

1. Νεοελληνική Γλώσσα και Λογοτεχνία Γενικής Παιδείας
2. Μαθηματικά Ομάδας Προσανατολισμού
3. Οικονομία Ομάδας Προσανατολισμού
4. Πληροφορική Ομάδας Προσανατολισμού

Μέχρι και το σχολικό έτος 2020–2021, οι υποψήφιοι για την εισαγωγή στην τριτοβάθμια εκπαίδευση της Ομάδας Προσανατολισμού Ανθρωπιστικών Σπουδών εξετάζονταν υποχρεωτικά στα μαθήματα της Νεοελληνικής Γλώσσας και Λογοτεχνίας, των Αρχαίων Ελληνικών, της Ιστορίας και της Κοινωνιολογίας. Από το σχολικό έτος 2021–2022, το εξεταζόμενο μάθημα της Κοινωνιολογίας αντικαταστάθηκε εκ νέου από το μάθημα των Λατινικών.

Επιπλέον, στα βασικά χαρακτηριστικά του νέου συστήματος εισαγωγής συγκαταλέγονται η καθιέρωση του μαθήματος της Νεοελληνικής Γλώσσας και Λογοτεχνίας Γενικής Παιδείας ως κοινού εξεταζόμενου μαθήματος για όλους τους υποψηφίους, καθώς και η δυνατότητα επιλογής αποκλειστικά ενός (1) Επιστημονικού Πεδίου. Αντιθέτως, στο προηγούμενο σύστημα προβλεπόταν η δυνατότητα επιλογής ενός (1) ή δύο (2) Επιστημονικών Πεδίων, με αντίστοιχη εξέταση σε τέσσερα (4) ή πέντε (5) μαθήματα. Τέλος, τα δύο συστήματα (παλαιό και νέο) διαφοροποιούνται ως προς την εξεταστέα ύλη, το ποσοστό κατανομής των θέσεων εισακτέων ανά σύστημα, καθώς και ως προς τον τρόπο υπολογισμού των μορίων.

1.3. Εισαγωγή νέων μηχανισμών αξιολόγησης

Στο πλαίσιο αυτών των θεσμικών και δομικών αλλαγών, από το έτος 2021 εισήχθησαν στον μηχανισμό υπολογισμού των μορίων οι **Συντελεστές Βαρύτητας** και η **Ελάχιστη Βάση Εισαγωγής (ΕΒΕ)**, οι οποίοι συνιστούν διακριτούς μηχανισμούς του συστήματος των Πανελλαδικών Εξετάσεων, με διαφορετικό ρόλο και λειτουργία στη διαδικασία εισαγωγής στην τριτοβάθμια εκπαίδευση.

1.4. Ελάχιστη Βάση Εισαγωγής - 2021

Σύμφωνα με τον Νόμο υπ' αριθμ. 4777 (ΦΕΚ Α' 25/17.02.2021), που αφορά διατάξεις περί «Εισαγωγής στην Τριτοβάθμια Εκπαίδευση», θεσπίστηκε η Ελάχιστη Βάση Εισαγωγής (ΕΒΕ) για την εισαγωγή αποφοίτων Γενικού Λυκείου και Επαγγελματικών Λυκείων στην τριτοβάθμια εκπαίδευση και εφαρμόστηκε για πρώτη φορά στις Πανελλαδικές Εξετάσεις του Ιουνίου 2021[2].

Η ελάχιστη και η μέγιστη τιμή του συντελεστή της ΕΒΕ καθορίζονται με απόφαση του Υπουργού Παιδείας πριν από το τέλος Ιουνίου κάθε έτους, κατόπιν εισηγήσεων των ίδιων των ΑΕΙ.

Η ΕΒΕ: α) καθορίζει το ελάχιστο απαιτούμενο επίπεδο επίδοσης, β) λειτουργεί αποκλειστικά ως προϋπόθεση εισαγωγής, καθώς, εάν δεν καλυφθεί η ΕΒΕ, ο υποψήφιος δεν δύναται να δηλώσει το αντίστοιχο Τμήμα, και γ) δεν επηρεάζει τη σειρά κατάταξης των υποψηφίων, αλλά το δικαίωμα συμμετοχής τους στη διαδικασία επιλογής.

1.5. Στόχοι και θεσμικό πλαίσιο της Ελάχιστης Βάσης Εισαγωγής

Σύμφωνα με τη νομοθεσία, η καθιέρωση της Ελάχιστης Βάσης Εισαγωγής αποσκοπεί στη διασφάλιση των απαραίτητων ακαδημαϊκών προϋποθέσεων για την επιτυχή φοίτηση των υποψηφίων φοιτητών στην ανώτατη τριτοβάθμια εκπαίδευση, καθώς και στην έγκαιρη ολοκλήρωση των σπουδών τους. Παράλληλα, εκτιμήθηκε ότι μέσω της θεσμοθέτησής της θα επιτευχθεί η ποιοτική αναβάθμιση των παρεχόμενων σπουδών στα ελληνικά Ανώτατα Εκπαιδευτικά Ιδρύματα (ΑΕΙ).

Γενικά, η ΕΒΕ διαμορφώνεται ενιαία μέσω συσχέτισης των μέσων όρων των βαθμολογικών επιδόσεων των υποψηφίων. Οι μέσοι όροι αφορούν τις επιδόσεις στα τέσσερα πανελλαδικώς εξεταζόμενα μαθήματα κάθε επιστημονικού πεδίου. Ο προκύπτων μέσος όρος πολλαπλασιάζεται με έναν συντελεστή, ο οποίος κυμαίνεται συνήθως μεταξύ 0,80, 1,00 και 1,20 και αποφασίζεται από κάθε Τμήμα ΑΕΙ. Το αποτέλεσμα στρογγυλοποιείται στο δεύτερο δεκαδικό ψηφίο. Ο εν λόγω συντελεστής οφείλει να βρίσκεται εντός του εύρους ελάχιστης και μέγιστης τιμής που ορίζεται θεσμικά.[2]

1.6. Προϋποθέσεις εισαγωγής

Για να επιτύχει ένας υποψήφιος, απαιτείται στις Πανελλαδικές Εξετάσεις να επιτύχει βαθμολογική επίδοση ίση ή μεγαλύτερη της ΕΒΕ ανά Σχολή, Τμήμα ή Εισαγωγική Κατεύθυνση (π.χ. $\geq 8,26$ με άριστα το 20).

Στα Τμήματα των ΑΕΙ που εντάσσονται σε περισσότερα του ενός επιστημονικά πεδία, για τον υπολογισμό της ΕΒΕ λαμβάνεται υπόψη ο μικρότερος από τους μέσους όρους των πανελλαδικώς εξεταζόμενων μαθημάτων των επιστημονικών πεδίων στα οποία ανήκουν τα συγκεκριμένα Τμήματα.[2]

1.7. Συντελεστές Βαρύτητας - 2022

Οι συντελεστές βαρύτητας ανά πανελλαδικώς εξεταζόμενο μάθημα καθορίζονται με απόφαση της Συγκλήτου του οικείου ΑΕΙ για κάθε Σχολή, Τμήμα ή Εισαγωγική Κατεύθυνση Τμήματος και ανακοινώθηκαν για πρώτη φορά με την Υπουργική Απόφαση Φ.253.1/148469/Α5/18-11-2021 (ΦΕΚ Β΄ 5399), η οποία εκδόθηκε βάσει του Ν. 4777/2021.

Κατ' αυτόν τον τρόπο, τα ΑΕΙ συμμετέχουν ενεργά στη διαμόρφωση της ακαδημαϊκής φυσιογνωμίας των Σχολών και Τμημάτων τους, καθορίζοντας τον συντελεστή βαρύτητας κάθε πανελλαδικώς εξεταζόμενου μαθήματος και κοινοποιώντας τον στην αρχή του ημερολογιακού έτους.

Ο υπολογισμός του συνολικού αριθμού μορίων κάθε υποψηφίου που συμμετέχει στις Πανελλαδικές Εξετάσεις από το έτος 2022 και εφεξής πραγματοποιείται ως εξής: ο γραπτός βαθμός σε καθένα από τα τέσσερα πανελλαδικώς εξεταζόμενα μαθήματα, τα οποία προβλέπονται από την Ομάδα Προσανατολισμού του υποψηφίου για το αντίστοιχο Επιστημονικό Πεδίο, πολλαπλασιάζεται με τον αντίστοιχο συντελεστή βαρύτητας. Τα τέσσερα γινόμενα αθροίζονται και το τελικό άθροισμα πολλαπλασιάζεται επί χίλια (1000), ώστε να προκύψει ο συνολικός αριθμός μορίων.

Δεδομένου ότι οι συντελεστές βαρύτητας των μαθημάτων δύνανται να διαφοροποιούνται μεταξύ Σχολών, Τμημάτων ή Εισαγωγικών Κατευθύνσεων ακόμη και του ίδιου γνωστικού αντικειμένου, ενδέχεται να προκύπτουν διαφοροποιήσεις στον αριθμό μορίων των υποψηφίων.

Σύμφωνα με τον Νόμο 4777/2021 (ΦΕΚ Α΄ 25), ο συντελεστής βαρύτητας κάθε πανελλαδικώς εξεταζόμενου μαθήματος εκφράζεται σε ποσοστό (%) και δεν μπορεί να είναι μικρότερος του 20%, ενώ το άθροισμα των συντελεστών ισούται με 100%.

Στις περιπτώσεις Σχολών, Τμημάτων ή Εισαγωγικών Κατευθύνσεων για τις οποίες απαιτείται εξέταση σε ειδικό μάθημα ή πρακτικές δοκιμασίες, ο συντελεστής βαρύτητας που αποδίδεται στο ειδικό μάθημα ή στις δοκιμασίες αυτές είναι είτε 10% είτε 20%. Στις περιπτώσεις εξέτασης δύο ειδικών μαθημάτων, το άθροισμα των συντελεστών τους ανέρχεται σε 20%, ενώ ο συντελεστής κάθε ειδικού μαθήματος δεν μπορεί να είναι μικρότερος του 8%. Τα ποσοστά των ειδικών μαθημάτων προστίθενται στο 100% των τεσσάρων πανελλαδικώς εξεταζόμενων μαθημάτων [3].

1.8. Τρόπος υπολογισμού μορίων

Σχετικά με τον υπολογισμό των μορίων για την εισαγωγή στην Τριτοβάθμια Εκπαίδευση ισχύουν τα εξής: α) Ο γραπτός βαθμός σε καθένα από τα τέσσερα (4) πανελλαδικώς εξεταζόμενα μαθήματα πολλαπλασιάζεται με τον αντίστοιχο συντελεστή βαρύτητας. Τα γινόμενα αθροίζονται και το τελικό άθροισμα πολλαπλασιάζεται επί χίλια (1000), ώστε να προκύψει ο συνολικός αριθμός μορίων. Ο βαθμός κάθε μαθήματος εκφράζεται στην εικοσάβαθμη κλίμακα με προσέγγιση δεκάτου.

Ανάλογα με το Επιστημονικό Πεδίο, κάθε υποψήφιος Γενικού Λυκείου δύναται να συγκεντρώνει από 144 έως 247 διαφορετικές βαθμολογίες.

Ενδεικτικά, υποψήφιος του 1ου Επιστημονικού Πεδίου μπορεί να συγκεντρώνει διαφορετικό αριθμό μορίων για τη Νομική Σχολή Θεσσαλονίκης (ΑΠΘ) και τη Νομική Σχολή Κομοτηνής (ΔΠΘ), λόγω διαφοροποίησης των συντελεστών βαρύτητας.

Για παράδειγμα:

- Υποψήφιος για το 1^ο Επιστημονικό Πεδίο που έλαβε την εξής βαθμολογία:

Νεοελληνική Γλώσσα & Λογοτεχνία	Αρχαία Ελληνικά	Ιστορία	Λατινικά
17	18,6	20	20

συγκεντρώνει για τη Νομική Θεσ/νίκης (Α.Π.Θ.) **18.680** μόρια, ενώ για τη Νομική Κομοτηνής (Δ.Π.Θ.) **18.900** μόρια, σύμφωνα με τους συντελεστές βαρύτητας μαθημάτων των αντίστοιχων Σχολών:

	Νεοελληνική Γλώσσα & Λογοτεχνία	Αρχαία Ελληνικά	Ιστορία	Λατινικά
Νομική Θεσ/ νίκης-ΑΠΘ	30%	30%	20%	20%
Νομική Κομοτηνής- Δ.Π.Θ.	25%	25%	25%	25%

β) Στα Τμήματα όπου απαιτείται εξέταση ειδικού μαθήματος ή πρακτικής δοκιμασίας, η εξέταση πραγματοποιείται σε πανελλαδικό επίπεδο. Για τον υπολογισμό της συνολικής βαθμολογίας:

1. Ο βαθμός κάθε ειδικού μαθήματος πολλαπλασιάζεται με τον αντίστοιχο συντελεστή βαρύτητας.
2. Τα γινόμενα προστίθενται στο άθροισμα των τεσσάρων βασικών μαθημάτων.
3. Το συνολικό άθροισμα πολλαπλασιάζεται επί χίλια (1000), ώστε να προκύψουν τα συνολικά μόρια.

Παράδειγμα:

Υποψήφιος με βαθμό 18 σε ειδικό μάθημα με συντελεστή βαρύτητας 20% συγκεντρώνει:
 $18 \times 0,20 = 3,6 \rightarrow 3,6 \times 1000 = 3.600$ μόρια.

Τα μόρια αυτά προστίθενται στα μόρια που προκύπτουν από τα τέσσερα βασικά μαθήματα [3].

1.9. Κίνητρο

Κατά την τελευταία δεκαετία έχουν υλοποιηθεί εκτεταμένες μεταρρυθμίσεις τόσο στο εκπαιδευτικό σύστημα όσο και στο σύστημα εισαγωγής στην Τριτοβάθμια Εκπαίδευση, οι οποίες αναμφίβολα έχουν επηρεάσει σε σημαντικό βαθμό τις εκπαιδευτικές επιλογές και τις δυνατότητες πρόσβασης των υποψηφίων. Στο πλαίσιο αυτό, καθίσταται ιδιαίτερα σημαντική η εκ νέου μελέτη των διαθέσιμων δεδομένων, προκειμένου να αποτιμηθούν οι επιπτώσεις των μεταρρυθμίσεων και να αναδειχθούν πιθανές μεταβολές σε πρότυπα συμμετοχής και πρόσβασης των υποψηφίων.

Κίνητρο λοιπόν για την εκπόνηση της παρούσας διπλωματικής εργασίας αποτέλεσε αφενός η ύπαρξη μεγάλου όγκου δεδομένων που αφορούν τις Πανελλαδικές Εξετάσεις, τα οποία δεν έχουν τύχει συστηματικής ανάλυσης από το 2017, και αφετέρου η ανάγκη κατανόησης και ερμηνείας αυτού του εκτεταμένου και ανεκμετάλλευτου εκπαιδευτικού αποθέματος. Στόχος είναι η εξαγωγή συνδυαστικών και ουσιαστικών πληροφοριών, καθώς και η παραγωγή πρόσθετης γνώσης, η οποία δύναται να αξιοποιηθεί μελλοντικά από τους αρμόδιους φορείς για χάραξη νέας εκπαιδευτικής πολιτικής, περισσότερο δίκαιης και αποτελεσματικής για τους μελλοντικούς υποψηφίους.

Επιπλέον, καθοριστικό ρόλο στην επιλογή του θέματος διαδραμάτισε το επαγγελματικό μου ενδιαφέρον, καθώς υπηρετώ επί σειρά ετών ως Εκπαιδευτικός Πληροφορικής στη δευτεροβάθμια εκπαίδευση και, τα τελευταία χρόνια, στην πρωτοβάθμια. Παράλληλα, η προσωπική μου εμπλοκή ως γονέας ενίσχυσε περαιτέρω το ενδιαφέρον αυτό, δεδομένου ότι η πρώτη μου κόρη συμμετείχε στις Πανελλαδικές Εξετάσεις το 2024, ενώ το 2026 υποψήφια θα είναι και η δεύτερη κόρη μου.

Συνοψίζοντας, η παρούσα διπλωματική εργασία προκύπτει από το συνδυασμό επιστημονικού ενδιαφέροντος, επαγγελματικής εμπειρίας και προσωπικού προβληματισμού, σε ένα εκπαιδευτικό περιβάλλον που μεταβάλλεται διαρκώς. Η αξιοποίηση σύγχρονων μεθόδων ανάλυσης και εξόρυξης δεδομένων φιλοδοξεί να συμβάλει στην ουσιαστική κατανόηση των εξελίξεων που έχουν επέλθει στο σύστημα των Πανελλαδικών Εξετάσεων και συνεχίζουν να μεταβάλλονται, αναδεικνύοντας πρότυπα, τάσεις και κρίσιμους παράγοντες που επηρεάζουν την πρόσβαση στην Τριτοβάθμια Εκπαίδευση. Μέσα από αυτή τη διαδικασία επιδιώκεται η παραγωγή τεκμηριωμένων συμπερασμάτων, τα οποία δύναται να αποτελέσουν χρήσιμο εργαλείο υποστήριξης μελλοντικών εκπαιδευτικών παρεμβάσεων και πολιτικών αποφάσεων.

1.10. Συνεισφορά

Η παρούσα διπλωματική εργασία διαφοροποιείται από προηγούμενες προσεγγίσεις, καθώς πραγματοποιεί εκτεταμένη στατιστική ανάλυση των δεδομένων των Πανελλαδικών Εξετάσεων σε συνδυασμό με την εφαρμογή τεχνικών εξόρυξης γνώσης, οι οποίες, σύμφωνα με τη διαθέσιμη

βιβλιογραφία και τα δημόσια διαθέσιμα δεδομένα, δεν έχουν χρησιμοποιηθεί έως σήμερα με αυτόν τον τρόπο και σε αυτή την κλίμακα για τα συγκεκριμένα δεδομένα.

Συγκεκριμένα, υλοποιείται ένα ολοκληρωμένο πλαίσιο ανάλυσης βασισμένο στη διαδικασία KDD, το οποίο περιλαμβάνει προεπεξεργασία, στατιστική διερεύνηση, συσταδοποίηση, κατηγοριοποίηση και εξαγωγή κανόνων συσχέτισης, μέσω των αλγορίθμων k-means, Random Forest, Naïve Bayes και Apriori. Αυτά τα ευρήματα είναι ιδιαίτερα χρήσιμα για την ερμηνεία της συμπεριφοράς επιλογών, τη συνύπαρξη ακαδημαϊκών και κοινωνικών παραμέτρων και μπορούν να αξιοποιηθούν τόσο στην εκπαιδευτική συμβουλευτική όσο και στη μελέτη των προτιμήσεων εισαγωγής ανά Ομάδα Προσανατολισμού.

Η συνδυαστική αυτή προσέγγιση επιτρέπει την ανάδειξη προτύπων, τάσεων και συσχετίσεων που δεν είναι άμεσα ορατές με παραδοσιακές μεθόδους ανάλυσης, προσφέροντας νέα γνώση σχετικά με τις εκπαιδευτικές επιλογές και τις επιδόσεις των υποψηφίων. Παράλληλα, η εργασία προτείνει ένα σαφές και οργανωμένο πλαίσιο ανάλυσης εκπαιδευτικών δεδομένων, το οποίο μπορεί να αξιοποιηθεί σε μελλοντικές μελέτες.

1.11. Σχετικές εργασίες

Κατά την διαδικασία της βιβλιογραφικής αναζήτησης πηγών, για την ανακάλυψη θεματικών περιοχών συναφή με το αντικείμενο της παρούσης εργασίας, εντοπίστηκαν και μελετήθηκαν οι παρακάτω εργασίες οι οποίες αφορούν την στατική ανάλυση και την πρόγνωση της επίδοσης μαθητών.

Η διπλωματική εργασία Στατιστική Ανάλυση Αποτελεσμάτων των Πανελληνίων Εξετάσεων 2001 (Κατωπόδης, 2003), αφορά στο να εντοπιστούν και να παρουσιαστούν οι ποσοτικές και ποιοτικές διαφορές που σχετίζονται με την επιτυχία ή την αποτυχία των αποφοίτων Λυκείου του έτους 2000-2001, στις εξετάσεις για την εισαγωγή στα Πανεπιστήμια & Τ.Ε.Ι. της χώρας. Η ανάλυσή βασίζεται σε διάφορες κατηγοριοποιήσεις των Λυκείων και διερευνώνται οι διάφοροι προσδιοριστικοί παράγοντες της επιτυχίας και της αποτυχίας. Τέλος, γίνεται μια αξιολόγηση των ποσοτικών αλλά και των ποιοτικών διαφορών με βάση ορισμένα κριτήρια.

Η εργασία Association Rules Mining from the Educational Data of ESOG Web-Based Application (Ougiarioglou & Paschalis, 2012), αφορά την παρουσίαση της διαδικασίας KDD και την εφαρμογή του αλγορίθμου Apriori σε εκπαιδευτικά δεδομένα που αφορούν το ενδιαφέρον που επιδεικνύουν μαθητές Λυκείου για μαθήματα που παρακολουθούν στο Λύκειο. Απ' τη διαδικασία αυτή προέκυψε ένα σύνολο από κανόνες συσχέτισης, οι οποίοι είναι ικανοί να βοηθήσουν και να καθοδηγήσουν τους Έλληνες Εκπαιδευτικούς και τους Διευθυντές Σχολείων, στη λήψη εκπαιδευτικών αποφάσεων, στο σχεδιασμό μαθησιακών δραστηριοτήτων σύμφωνα με τα ενδιαφέροντα των μαθητών τους και την αποτελεσματική διαχείριση μιας τάξης.

Επίσης εντοπίστηκε η διπλωματική εργασία της Κυπρίλλης Αναστασίας με τίτλο «Ανάλυση Δεδομένων Πανελλαδικών Εξετάσεων και Εξόρυξη Γνώσης», η οποία αφορούσε την ανάλυση δεδομένων Πανελλαδικών Εξετάσεων για τα έτη 2016 και 2017. Η συγκεκριμένη εργασία κρίθηκε ιδιαίτερα σημαντική, καθώς μας έδωσε τη δυνατότητα να πραγματοποιήσουμε συγκριτική ανάλυση των δεδομένων τόσο μεταξύ διαφορετικών διετιών (2016–2017 και 2023–2024) όσο και μεταξύ διαφορετικών συστημάτων εξετάσεων, επιτρέποντάς μας παράλληλα να παρατηρήσουμε την εξέλιξη των ποσοστών επιτυχίας κατά την τελευταία δεκαετία, γεγονός που προσέφερε χρήσιμα συμπεράσματα και πολύτιμη μεθοδολογική καθοδήγηση για την παρούσα έρευνα.

1.12. Οργάνωση Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία είναι δομημένη στα ακόλουθα κεφάλαια:

Στο Κεφάλαιο 2 παρουσιάζεται αρχικά το αντικείμενο της εξόρυξης γνώσης, με έμφαση στον ορισμό της, στα στάδια της διαδικασίας Εξόρυξης Γνώσης από Δεδομένα (Knowledge Discovery in Databases – KDD), καθώς και στις βασικές κατηγορίες αλγορίθμων εξόρυξης γνώσης, συνοδευόμενες από σύντομη αναφορά στα αντίστοιχα πεδία εφαρμογής τους. Επιπλέον, περιγράφονται αναλυτικά οι αλγόριθμοι εξόρυξης γνώσης που εφαρμόστηκαν στο πλαίσιο της παρούσας εργασίας.

Στο Κεφάλαιο 3 αναλύεται η διαδικασία συλλογής των δεδομένων, καθώς και τα στάδια της προεπεξεργασίας και του μετασχηματισμού τους, με σκοπό τη μετέπειτα ανάλυσή τους.

Στο Κεφάλαιο 4 πραγματοποιείται η στατιστική ανάλυση των δεδομένων, αρχικά με τη χρήση του λογισμικού Microsoft Excel και στη συνέχεια με τη γλώσσα προγραμματισμού Python, στο περιβάλλον Google Colab δημιουργούνται τα κατάλληλα γραφήματα, με στόχο τη διευκόλυνση της εξαγωγής συμπερασμάτων.

Στο Κεφάλαιο 5 περιγράφεται η περαιτέρω προεπεξεργασία των δεδομένων και με τη χρήση της γλώσσας προγραμματισμού Python παρουσιάζονται τα αποτελέσματα της εφαρμογής των αλγορίθμων συσταδοποίησης k-means και DBSCAN μέσω πινάκων και γραφημάτων.

Στο Κεφάλαιο 6 παρουσιάζεται η πρακτική εφαρμογή του αλγορίθμου κατηγοριοποίησης Random Forest Regression, στα ίδια σύνολα δεδομένων που χρησιμοποιήθηκαν και στη συσταδοποίηση. Επίσης παρουσιάζονται τα αποτελέσματα της εφαρμογής του αλγορίθμου Gaussian Naive Bayes και Random Forest σε ίδια δεδομένα και συγκριτική ανάλυση αποτελεσμάτων.

Στο Κεφάλαιο 7, κατόπιν κατάλληλης προεπεξεργασίας των δεδομένων, παρουσιάζονται οι κανόνες συσχέτισης που προκύπτουν από την εφαρμογή του αλγορίθμου Apriori, ξεχωριστά για κάθε Ομάδα Προσανατολισμού.

Τέλος, στο Κεφάλαιο 8 συνοψίζονται τα κύρια συμπεράσματα της εργασίας και διατυπώνονται προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2^ο: Αλγόριθμοι Εξόρυξης Γνώσης

2.1. Εξόρυξη γνώσης

Πλήθος ερευνητικών προσπαθειών έχει στραφεί προς την εξόρυξη γνώσης από εκπαιδευτικά δεδομένα που προέρχονται από εκπαιδευτικό λογισμικό και Συστήματα Διαχείρισης Μάθησης. Βασικός σκοπός αυτών των μελετών είναι η αξιοποίηση των διαθέσιμων πληροφοριών για την υποστήριξη των εκπαιδευτικών, τόσο στη βελτιστοποίηση της λειτουργίας των εκπαιδευτικών δομών όσο και στον επανασχεδιασμό των μαθησιακών δραστηριοτήτων, με απώτερο στόχο την ενίσχυση των μαθησιακών αποτελεσμάτων [4].

Στο πλαίσιο της Ανακάλυψης Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD), η εξόρυξη δεδομένων συνιστά το κεντρικό και πιο κρίσιμο στάδιο της συνολικής διαδικασίας. Σε αυτό το στάδιο εφαρμόζονται εξειδικευμένοι αλγόριθμοι και υπολογιστικές τεχνικές με στόχο τον εντοπισμό έγκυρων, πρωτότυπων και δυνητικά χρήσιμων προτύπων, βασισμένων σε δεδομένα που έχουν προηγουμένως υποστεί επιλογή, καθαρισμό και κατάλληλο μετασχηματισμό [5].

Η εξόρυξη δεδομένων (data mining) αποτελεί θεμελιώδη τομέα της επιστήμης δεδομένων και της τεχνητής νοημοσύνης, καθώς επικεντρώνεται στην ανακάλυψη γνώσης, συσχετίσεων και προτύπων από μεγάλης κλίμακας σύνολα δεδομένων[6]. Μέσω της αξιοποίησης υπολογιστικών μεθόδων, επιδιώκεται η εξαγωγή πληροφορίας που δεν είναι άμεσα εμφανής, αλλά μπορεί να υποστηρίξει την κατανόηση σύνθετων φαινομένων και τη λήψη τεκμηριωμένων αποφάσεων[7].

Η διαδικασία της εξόρυξης δεδομένων περιλαμβάνει ένα σύνολο τεχνικών και αλγορίθμων που καθιστούν δυνατή την αποκάλυψη λανθάνοντων προτύπων, την πρόβλεψη μελλοντικών συμπεριφορών ή τάσεων και την ενίσχυση της διαδικασίας λήψης αποφάσεων. Παράλληλα, εντάσσεται ως επιμέρους στάδιο στη συνολική διαδικασία KDD, η οποία εκτείνεται από την αρχική συλλογή των δεδομένων έως την τελική εξαγωγή και ερμηνεία της γνώσης [5].

Η εξόρυξη δεδομένων μπορεί να οριστεί ως η διαδικασία εντοπισμού ενδιαφερόντων και ουσιαστικών μοτίβων μέσα από πολύ μεγάλες ποσότητες δεδομένων. Ως μέρος της ανακάλυψης γνώσης, περιλαμβάνει συνήθως μια σειρά από βήματα, όπως ο καθαρισμός και η ενοποίηση δεδομένων, η επιλογή και ο μετασχηματισμός τους, η ανακάλυψη και αξιολόγηση προτύπων, καθώς και η παρουσίαση της παραγόμενης γνώσης [6]. Οι βασικές διαστάσεις της εξόρυξης δεδομένων αφορούν τα ίδια τα δεδομένα, τη γνώση που παράγεται, τις τεχνολογίες που χρησιμοποιούνται και τις εφαρμογές στις οποίες στοχεύει.

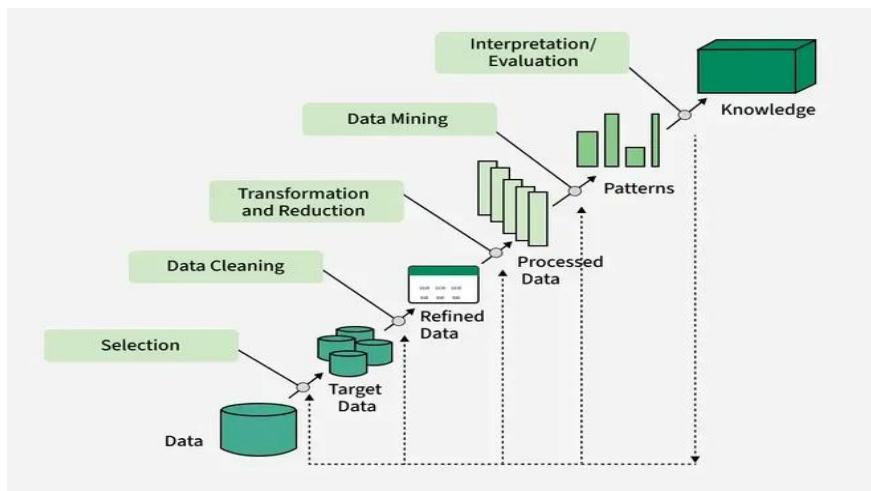
Η συνεχώς αυξανόμενη ζήτηση για εξόρυξη δεδομένων αποτελεί φυσική συνέπεια της εξέλιξης των τεχνολογιών πληροφορίας και της ραγδαίας αύξησης του όγκου των διαθέσιμων δεδομένων [9]. Η εξόρυξη δεδομένων εξετάζεται από πολλαπλές οπτικές γωνίες, όπως τα είδη των δεδομένων που μπορούν να αναλυθούν, οι μορφές γνώσης που είναι δυνατόν να εξαχθούν, οι τεχνολογίες που εφαρμόζονται και οι τομείς εφαρμογής, προσφέροντας έτσι μια πολυδιάστατη προσέγγιση του αντικειμένου. Η διαδικασία αυτή μπορεί να εφαρμοστεί σε κάθε τύπο δεδομένων που είναι σχετικός με μια συγκεκριμένη εφαρμογή, όπως δεδομένα βάσεων δεδομένων, αποθηκών δεδομένων, δεδομένα συναλλαγών, καθώς και πιο σύνθετους ή προηγμένους τύπους δεδομένων [10].

Η γνώση που αφορά τα χαρακτηριστικά και τη δομή των δεδομένων διαδραματίζει καθοριστικό ρόλο στην προεπεξεργασία, η οποία αποτελεί το πρώτο και ιδιαίτερα σημαντικό στάδιο της εξόρυξης δεδομένων. Παράλληλα, ο τομέας της οπτικοποίησης δεδομένων προσφέρει πλήθος τεχνικών γραφικής

αναπαράστασης, οι οποίες διευκολύνουν την ανίχνευση συσχετίσεων, τάσεων και προκαταλήψεων που συχνά παραμένουν κρυμμένες σε μεγάλα ή μη δομημένα σύνολα δεδομένων.

Οι βασικές έννοιες και μέθοδοι της προεπεξεργασίας δεδομένων κατηγοριοποιούνται σε καθαρισμό, ενοποίηση, μείωση και μετασχηματισμό δεδομένων. Η ποιότητα των δεδομένων θεωρείται ικανοποιητική όταν ανταποκρίνεται στις απαιτήσεις της προβλεπόμενης χρήσης και επηρεάζεται από παράγοντες όπως η ακρίβεια, η πληρότητα, η συνέπεια, η επικαιρότητα, η αξιοπιστία και η δυνατότητα ερμηνείας [11]. Ο καθαρισμός δεδομένων στοχεύει στην απομάκρυνση θορύβου και στη διόρθωση ασυνεπειών, ενώ η ενοποίηση δεδομένων συνδυάζει πληροφορία από πολλαπλές πηγές σε ένα ενιαίο σύνολο, όπως μια αποθήκη δεδομένων. Η μείωση δεδομένων επιδιώκει τον περιορισμό του όγκου τους μέσω τεχνικών όπως η συγκέντρωση, η αφαίρεση πλεοναζόντων χαρακτηριστικών ή η ομαδοποίηση.

Επιπλέον, οι μετασχηματισμοί δεδομένων, όπως η κανονικοποίηση, επιτρέπουν την κλιμάκωση των τιμών σε μικρότερα και ομοιόμορφα εύρη, γεγονός που μπορεί να βελτιώσει τόσο την ακρίβεια όσο και την αποδοτικότητα των αλγορίθμων εξόρυξης που βασίζονται σε μετρήσεις απόστασης [12]. Οι τεχνικές προεπεξεργασίας δεν λειτουργούν ανεξάρτητα, αλλά συχνά συνδυάζονται, καθώς ο καθαρισμός μπορεί να περιλαμβάνει μετασχηματισμούς για τη διόρθωση εσφαλμένων τιμών, όπως η ενοποίηση διαφορετικών μορφών ημερομηνιών. Η κατανόηση των τύπων και των χαρακτηριστικών των δεδομένων συμβάλλει στον εντοπισμό ακραίων ή λανθασμένων τιμών, ενισχύοντας την αποτελεσματικότητα των σταδίων καθαρισμού και ενοποίησης. Συνολικά, η εφαρμογή τεχνικών προεπεξεργασίας πριν από την εξόρυξη δεδομένων μπορεί να βελτιώσει σημαντικά τόσο την ποιότητα των προτύπων που προκύπτουν όσο και τον χρόνο που απαιτείται για τη διαδικασία εξόρυξης [6].



Σχήμα 2.1: Στάδια της διαδικασίας Ανακάλυψης Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD) [5]

Κύρια στάδια της διαδικασίας (Σχήμα 2.1)

1. Συλλογή δεδομένων – από βάσεις δεδομένων, αρχεία, αισθητήρες, web.
2. Προεπεξεργασία – καθαρισμός, αντιμετώπιση ελλειπών τιμών, κανονικοποίηση.
3. Μετασχηματισμός – επιλογή και μείωση χαρακτηριστικών.
4. Εξόρυξη δεδομένων – εφαρμογή αλγορίθμων.
5. Αξιολόγηση & ερμηνεία – έλεγχος χρησιμότητας και εγκυρότητας αποτελεσμάτων.

Βασικές τεχνικές εξόρυξης δεδομένων

- Ομαδοποίηση (Clustering)
- Κατηγοριοποίηση (Classification)
- Ανάλυση συσχετίσεων (Association Rules)
- Παλινδρόμηση (Regression)
- Ανίχνευση ανωμαλιών (Anomaly Detection)

Η παρούσα εργασία θα μελετήσει τις τέσσερις πρώτες τεχνικές εξόρυξης δεδομένων και κατόπιν θα τις εφαρμόσει στα δεδομένα.

2.2. Συσταδοποίηση (Clustering)

Πολλές εφαρμογές απαιτούν την διαμέριση σημείων δεδομένων σε διαισθητικά παρόμοιες ομάδες. Η διαμέριση ενός μεγάλου αριθμού σημείων δεδομένων σε μικρότερο αριθμό ομάδων βοηθά σημαντικά στη σύνοψη των δεδομένων και στην κατανόησή τους για μια ποικιλία εφαρμογών εξόρυξης δεδομένων.

Η βασική αρχή της συσταδοποίησης έγκειται στη μεγιστοποίηση της ομοιότητας μεταξύ των στοιχείων της ίδιας συστάδας και, ταυτόχρονα, στη μεγιστοποίηση της ανομοιότητας μεταξύ στοιχείων διαφορετικών συστάδων. Η ομοιότητα αυτή εκφράζεται συνήθως μέσω μετρικών απόστασης, με συνηθέστερη την Ευκλείδεια απόσταση, αν και σε ορισμένες περιπτώσεις χρησιμοποιούνται και άλλες μετρικές, όπως η Manhattan ή η cosine distance [13].

Η συσταδοποίηση εφαρμόζεται ευρέως σε πλήθος επιστημονικών και εφαρμοσμένων πεδίων, όπως η ανάλυση δεδομένων πελατών, η αναγνώριση προτύπων, η επεξεργασία εικόνας, η βιοπληροφορική και η ανίχνευση ανωμαλιών. Η σημασία της έγκειται στη δυνατότητά της να αποκαλύπτει υποκείμενες δομές στα δεδομένα, οι οποίες δεν είναι άμεσα ορατές μέσω απλών στατιστικών μεθόδων[14].

2.2.1. Αλγόριθμος k-means

Ο αλγόριθμος k-means συγκαταλέγεται στους πλέον ευρέως χρησιμοποιούμενους αλγορίθμους συσταδοποίησης, λόγω της απλότητας και της υπολογιστικής του αποδοτικότητας. Σκοπός του είναι η κατάτμηση ενός συνόλου δεδομένων σε k διακριτές συστάδες, όπου ο αριθμός των συστάδων καθορίζεται εκ των προτέρων από τον ερευνητή ή τον χρήστη (MacQueen, 1967). Κάθε συστάδα περιγράφεται από ένα αντιπροσωπευτικό κέντρο, γνωστό ως κεντροειδές (centroid), το οποίο υπολογίζεται ως ο μέσος όρος των σημείων που την απαρτίζουν.

Η διαδικασία λειτουργίας του αλγορίθμου είναι επαναληπτική. Σε αρχικό στάδιο, επιλέγονται τυχαία k κεντροειδή, ενώ στη συνέχεια κάθε παρατήρηση ανατίθεται στη συστάδα της οποίας το κέντρο βρίσκεται στη μικρότερη απόσταση, βάσει της επιλεγμένης μετρικής. Κατόπιν, τα κεντροειδή επαναπροσδιορίζονται ως ο μέσος όρος των δεδομένων που έχουν αποδοθεί σε κάθε συστάδα. Η παραπάνω διαδικασία επαναλαμβάνεται έως ότου επιτευχθεί σύγκλιση, δηλαδή έως τη σταθεροποίηση των αναθέσεων των σημείων στις συστάδες.

Ο k-means στοχεύει στην ελαχιστοποίηση της συνάρτησης κόστους που είναι γνωστή ως Άθροισμα Τετραγώνων των Αποστάσεων εντός των συστάδων (Within-Cluster Sum of Squares – WCSS). Παρότι χαρακτηρίζεται από υψηλή αποδοτικότητα, παρουσιάζει ορισμένους περιορισμούς, όπως η απαίτηση προκαθορισμού του αριθμού των συστάδων, η ευαισθησία στις αρχικές τιμές των κέντρων και η περιορισμένη ικανότητά του να διαχειρίζεται συστάδες μη σφαιρικού σχήματος ή την παρουσία

θορύβου στα δεδομένα [15]. Επίσης ένα θεμελιώδες πρόβλημα του αλγορίθμου k-means είναι η αδυναμία του να χειριστεί διάφορους τύπους δεδομένων.

2.2.2. Elbow

Η μέθοδος Elbow εφαρμόζεται για τον προσδιορισμό του κατάλληλου αριθμού συστάδων k κατά την υλοποίηση του αλγορίθμου k-means. Η προσέγγιση αυτή στηρίζεται στη μελέτη της συμπεριφοράς του Αθροίσματος Τετραγώνων των Αποστάσεων εντός των συστάδων (WCSS) σε σχέση με τον αριθμό των συστάδων που εξετάζονται [16]. Συγκεκριμένα, για διαδοχικές τιμές του k υπολογίζεται το WCSS και αποτυπώνεται σε γραφική παράσταση.

Η τιμή του k που αντιστοιχεί στο σημείο όπου ο ρυθμός μείωσης του WCSS μειώνεται αισθητά, δημιουργώντας ένα χαρακτηριστικό «γόνατο» στην καμπύλη, θεωρείται ότι αντιπροσωπεύει τη βέλτιστη επιλογή. Παρά τα πλεονεκτήματά της, η μέθοδος Elbow εμφανίζει ορισμένους περιορισμούς, καθώς το σημείο αυτό δεν είναι πάντοτε εύκολα αναγνωρίσιμο και η επιλογή του μπορεί να επηρεάζεται από υποκειμενική κρίση, ιδιαίτερα σε σύνολα δεδομένων με σύνθετη ή ασαφή δομή [13].

2.2.3. Αλγόριθμος DBSCAN

Ο DBSCAN (Density-Based Spatial Clustering of Applications with Noise) αποτελεί αλγόριθμο συσταδοποίησης που στηρίζεται στην έννοια της πυκνότητας των δεδομένων και εισήχθη από τους Ester et al. (1996) [17]. Σε αντίθεση με τον αλγόριθμο k-means, δεν προϋποθέτει τον προκαθορισμό του αριθμού των συστάδων, ενώ παρέχει τη δυνατότητα εντοπισμού συστάδων με μη προκαθορισμένο ή αυθαίρετο γεωμετρικό σχήμα.

Η λειτουργία του DBSCAN καθορίζεται από δύο βασικές παραμέτρους: την ακτίνα γειτονιάς ϵ (epsilon) και τον ελάχιστο αριθμό σημείων MinPts. Ένα σημείο θεωρείται σημείο πυρήνα όταν εντός της ϵ -γειτονιάς του περιλαμβάνεται τουλάχιστον ο αριθμός MinPts παρατηρήσεων. Σημεία που γειτνιάζουν με σημεία πυρήνα αλλά δεν πληρούν τα κριτήρια πυκνότητας χαρακτηρίζονται ως οριακά, ενώ όσα δεν εντάσσονται σε καμία συστάδα κατατάσσονται ως σημεία θορύβου.

Μεταξύ των κύριων πλεονεκτημάτων του DBSCAN συγκαταλέγονται η αυξημένη ανθεκτικότητά του στην παρουσία θορύβου και η ικανότητά του να αναγνωρίζει συστάδες αυθαίρετου σχήματος. Παρ' όλα αυτά, η επιλογή κατάλληλων τιμών για τις παραμέτρους ϵ και MinPts αποτελεί καθοριστικό παράγοντα για την απόδοσή του, ιδίως σε σύνολα δεδομένων με ανομοιογενή ή μεταβαλλόμενη πυκνότητα, γεγονός που μπορεί να επηρεάσει αρνητικά την αποτελεσματικότητά του [17].

2.3. Εξαγωγή κανόνων συσχέτισης

2.3.1. Αλγόριθμος Apriori

Η εξαγωγή κανόνων συσχέτισης αποτελεί βασική τεχνική της εξόρυξης δεδομένων και εντάσσεται στο ευρύτερο πλαίσιο της Εξαγωγής Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD). Στόχος της είναι η ανακάλυψη συσχετίσεων μεταξύ αντικειμένων σε μεγάλα σύνολα συναλλαγών [18].

Μέτρα Ενδιαφέροντος Κανόνων

Έστω $I = \{i_1, i_2, \dots, i_n\}$ το σύνολο αντικειμένων και $D = \{T_1, T_2, \dots, T_m\}$ το σύνολο συναλλαγών. Ένας κανόνας συσχέτισης έχει τη μορφή $X \rightarrow Y$, όπου σύμφωνα με τη σχέση (2.1)

$$\mathbf{X \subseteq I, Y \subseteq I \text{ και } X \cap Y = \emptyset} \quad (2.1)$$

Για την αξιολόγηση ενός κανόνα χρησιμοποιούνται κυρίως τα εξής μέτρα:

Η Υποστήριξη (Support) μετρά τη συχνότητα εμφάνισης σύμφωνα με τη σχέση (2.2):

$$\mathbf{support(X)} = \frac{|T \in D: X \subseteq T|}{|D|} \quad (2.2)$$

Η Εμπιστοσύνη (Confidence) μετρά την αξιοπιστία του κανόνα σύμφωνα με τη σχέση (2.3):

$$\mathbf{confidence(X \Rightarrow Y)} = \frac{support(X \cup Y)}{support(X)} \quad (2.3)$$

Ενώ η Άρση (Lift) εκφράζει τον βαθμό στατιστικής εξάρτησης μεταξύ των συνόλων σύμφωνα με τη σχέση (2.4):

$$\mathbf{lift(X \Rightarrow Y)} = \frac{confidence(X \Rightarrow Y)}{support(Y)} \quad (2.4)$$

Συχνά Σύνολα Αντικειμένων

Ένα σύνολο αντικειμένων X χαρακτηρίζεται ως συχνό όταν η υποστήριξή του υπερβαίνει ένα προκαθορισμένο κατώφλι $minsup$. Η εύρεση όλων των συχνών συνόλων αποτελεί υπολογιστικά απαιτητικό πρόβλημα λόγω του εκθετικού πλήθους πιθανών συνδυασμών.

Από κάθε συχνό σύνολο X , παράγονται κανόνες της μορφής σύμφωνα με τη σχέση (2.5):

$$\mathbf{A \Rightarrow X \setminus A, A \subset X} \quad (2.5)$$

Οι κανόνες διατηρούνται μόνο αν ικανοποιούν το κατώφλι ελάχιστης εμπιστοσύνης ($minconf$).

Ο αλγόριθμος Apriori βασίζεται στην ιδιότητα ότι κάθε υποσύνολο ενός συχνού συνόλου αντικειμένων είναι επίσης συχνό. Η ιδιότητα αυτή επιτρέπει την αποτελεσματική μείωση του χώρου αναζήτησης μέσω κλαδέματος. [18].

Πλεονεκτήματα και Περιορισμοί

Ο Apriori είναι εννοιολογικά απλός και ιδιαίτερα χρήσιμος για εκπαιδευτικούς σκοπούς. Ωστόσο, παρουσιάζει περιορισμούς λόγω των πολλαπλών σαρώσεων της βάσης δεδομένων και του μεγάλου αριθμού υποψηφίων συνόλων. Η όλη διαδικασία καθίσταται χρονοβόρα και απαιτεί πολύ μνήμη με αποτέλεσμα η απόδοση να είναι πολύ χαμηλή. Έτσι λοιπόν κρίνεται ακατάλληλος για μεγάλους όγκους δεδομένων [6].

2.4 Κατηγοριοποίηση

2.4.1. Random Forest

Ο Random Forest αποτελεί έναν αλγόριθμο ensemble learning βασισμένο σε πολλαπλά δέντρα απόφασης, ο οποίος μπορεί να εφαρμοστεί τόσο σε προβλήματα ταξινόμησης όσο και παλινδρόμησης. Η μέθοδος εισήχθη από τον Leo Breiman και βασίζεται στη δημιουργία πολλών ανεξάρτητων δέντρων

μέσω bootstrap sampling και τυχαίας επιλογής χαρακτηριστικών σε κάθε κόμβο, με στόχο τη μείωση της διακύμανσης και τη βελτίωση της γενίκευσης [19].

Στην περίπτωση της Random Forest Classification, η τελική πρόβλεψη προκύπτει μέσω πλειοψηφικής ψήφου των επιμέρους δέντρων σύμφωνα με τη σχέση (2.6), ενώ στη Random Forest Regression υπολογίζεται ως ο μέσος όρος των προβλέψεων σύμφωνα με τη σχέση (2.7). Η διαφορά μεταξύ των δύο έγκειται αποκλειστικά στον τρόπο συνδυασμού των αποτελεσμάτων και στη συνάρτηση απώλειας, ενώ η δομή του αλγορίθμου και οι υπερπαραμέτροι παραμένουν ίδιες [20].

Η Random Forest Regression χρησιμοποιείται ευρέως σε προβλήματα πρόβλεψης συνεχών μεταβλητών, προσφέροντας υψηλή ακρίβεια, ανθεκτικότητα στο overfitting και δυνατότητα εκτίμησης της σημαντικότητας των χαρακτηριστικών, γεγονός που την καθιστά ιδιαίτερα ελκυστική σε εφαρμογές μηχανικής μάθησης [21].

2.4.2. Random Forest Regression

Ο **Random Forest Regression** είναι ένας αλγόριθμος **επιβλεπόμενης μηχανικής μάθησης** που χρησιμοποιείται για την **πρόβλεψη συνεχών αριθμητικών τιμών**. Ανήκει στην οικογένεια μεθόδων μάθησης συνόλων και εισήχθη αρχικά από τον Breiman [19] ως επέκταση των μοντέλων δέντρων αποφάσεων, με στόχο τη βελτίωση της προγνωστικής ακρίβειας και της γενίκευσης του μοντέλου.

Ο Random Forest Regression βασίζεται στη δημιουργία και τον συνδυασμό **πολλών δέντρων απόφασης παλινδρόμησης**. Κάθε δέντρο εκπαιδεύεται σε διαφορετικό τυχαίο υποσύνολο των δεδομένων και των χαρακτηριστικών, ενώ η τελική πρόβλεψη προκύπτει ως **ο μέσος όρος των προβλέψεων όλων των δέντρων**. Αυτά τα υποσύνολα δημιουργούνται χρησιμοποιώντας δειγματοληψία bootstrap, όπου οι παρατηρήσεις αντλούνται με αντικατάσταση από το αρχικό σύνολο δεδομένων.

Εκτός από τη δειγματοληψία δεδομένων, εισάγεται επίσης τυχειότητα κατά την κατασκευή του δέντρου. Σε κάθε διαίρεση ενός δέντρου αποφάσεων, λαμβάνεται υπόψη μόνο ένα τυχαία επιλεγμένο υποσύνολο των διαθέσιμων χαρακτηριστικών. Κάθε δέντρο παλινδρόμησης παράγει μια ανεξάρτητη πρόβλεψη για τη μεταβλητή-στόχο και η τελική έξοδος του μοντέλου Random Forest Regression υπολογίζεται ως ο αριθμητικός μέσος όρος όλων των προβλέψεων των μεμονωμένων δέντρων[2].

Αυτή η στρατηγική συνόλου επιτρέπει στην Random Forest Regression να μοντελοποιεί αποτελεσματικά πολύπλοκες και μη γραμμικές σχέσεις μεταξύ των μεταβλητών εισόδου και της μεταβλητής-στόχου. Επιπλέον, ο αλγόριθμος δεν απαιτεί κλιμάκωση ή κανονικοποίηση χαρακτηριστικών και μπορεί να χειριστεί αποτελεσματικά μεγάλα σύνολα δεδομένων [19].

Παρά τα πλεονεκτήματά της, η Παλινδρόμηση Τυχαίου Δάσους έχει ορισμένους περιορισμούς. Το υπολογιστικό κόστος είναι υψηλότερο σε σύγκριση με τα απλούστερα μοντέλα παλινδρόμησης, ειδικά όταν χρησιμοποιείται μεγάλος αριθμός δέντρων. Επιπλέον, ενώ είναι πιο ερμηνεύσιμη από πολλά μοντέλα μαύρου κουτιού, παραμένει λιγότερο διαφανής από ένα μόνο δέντρο αποφάσεων. Παρ' όλα αυτά, αυτά τα μειονεκτήματα συχνά αντισταθμίζονται από την ανώτερη ακρίβεια και ανθεκτικότητα του μοντέλου [21].

Συμπερασματικά, η Τυχαία Δασική Παλινδρόμηση αποτελεί μια ισχυρή και αξιόπιστη προσέγγιση για προβλήματα πρόβλεψης συνεχών τιμών και εφαρμόζεται ευρέως σε διάφορους επιστημονικούς τομείς, συμπεριλαμβανομένης της εκπαίδευσης, των οικονομικών και της επιστήμης δεδομένων [19].

Classification

$$\hat{y} = \text{mode}(h_1(x), \dots, h_T(x)) \quad (2.6)$$

Regression

$$\hat{y} = 1/T \sum_{t=1}^T h_t(x) \quad (2.7)$$

2.4.3. Naive Bayes και Gaussian Naive Bayes

Οι πιθανοτικοί ταξινομητές αποτελούν βασική κατηγορία αλγορίθμων Μηχανικής Μάθησης και βασίζονται στη θεωρία πιθανοτήτων για τη λήψη αποφάσεων ταξινόμησης. Ένας από τους πιο διαδεδομένους αλγορίθμους αυτής της κατηγορίας είναι ο Naive Bayes, ο οποίος χαρακτηρίζεται από απλότητα, υπολογιστική αποδοτικότητα και υψηλή ερμηνευσιμότητα [6].

Ο αλγόριθμος **Naive Bayes** βασίζεται στο θεώρημα του Bayes, το οποίο ορίζεται σύμφωνα με τη σχέση (2.8):

$$P(C | X) = (P(X | C) \cdot P(C)) / P(X) \quad (2.8)$$

όπου C είναι η κλάση και $X = (x_1, x_2, \dots, x_n)$ το διάνυσμα χαρακτηριστικών.

Η βασική απλοποιητική υπόθεση του Naive Bayes είναι ότι τα χαρακτηριστικά είναι υπό συνθήκη ανεξάρτητα μεταξύ τους δεδομένης της κλάσης, δηλαδή σύμφωνα με τη σχέση (2.9):

$$P(X | C) = \prod P(x_i | C) \quad (2.9)$$

Η υπόθεση αυτή σπάνια ισχύει πλήρως στην πράξη, ωστόσο ο αλγόριθμος παρουσιάζει αξιοσημείωτη ανθεκτικότητα σε παραβιάσεις της ανεξαρτησίας [22].

Ο **Gaussian Naive Bayes** αποτελεί εξειδίκευση του Naive Bayes για περιπτώσεις όπου τα χαρακτηριστικά είναι συνεχείς αριθμητικές μεταβλητές. Υποθέτει ότι κάθε χαρακτηριστικό ακολουθεί κανονική (Gaussian) κατανομή εντός κάθε κλάσης.

Η πιθανότητα ενός χαρακτηριστικού x_i δεδομένης της κλάσης C δίνεται από τη συνάρτηση πυκνότητας της κανονικής κατανομής σύμφωνα με τη σχέση (2.10):

$$P(x_i | C) = (1 / \sqrt{2\pi\sigma_C^2}) \cdot \exp(- (x_i - \mu_C)^2 / (2\sigma_C^2)) \quad (2.10)$$

όπου μ_C είναι ο μέσος όρος και σ_C^2 η διακύμανση του χαρακτηριστικού για την κλάση C .

Κατά τη φάση της εκπαίδευσης, το μοντέλο υπολογίζει τις εκ των προτέρων πιθανότητες των κλάσεων καθώς και τις παραμέτρους μ και σ^2 για κάθε χαρακτηριστικό. Κατά την πρόβλεψη, το στιγμιότυπο κατατάσσεται στην κλάση που μεγιστοποιεί την εκ των υστέρων πιθανότητα $P(C | X)$.

Ο Gaussian Naive Bayes παρουσιάζει σημαντικά πλεονεκτήματα, όπως χαμηλή υπολογιστική πολυπλοκότητα, ταχεία εκπαίδευση και καλή ερμηνευσιμότητα. Ωστόσο, η απόδοσή του εξαρτάται από την εγκυρότητα της υπόθεσης κανονικότητας των δεδομένων και από την ανεξαρτησία των χαρακτηριστικών [21].

Ο αλγόριθμος έχει εφαρμοστεί εκτενώς στην ανάλυση εκπαιδευτικών δεδομένων, όπως στην πρόβλεψη ακαδημαϊκής επίδοσης και επιτυχίας σε εξετάσεις, όπου η διαφάνεια των προβλέψεων αποτελεί κρίσιμο παράγοντα [23].

Κεφάλαιο 3^ο: Συλλογή Δεδομένων

3.1. Ανοιχτά Δεδομένα

Με τον όρο *Ανοιχτά (κρατικά) δεδομένα* ορίζονται οι πληροφορίες που συλλέγονται ή παράγονται από δημόσιους φορείς και διατίθενται ελεύθερα προς περαιτέρω χρήση για οποιονδήποτε σκοπό, υπό συγκεκριμένους όρους άδειας. Τα δεδομένα αυτά αναφέρονται συχνά και ως Πληροφορίες Δημόσιου Τομέα (Public Sector Information – PSI) και αποτελούν βασικό πυλώνα της ψηφιακής διακυβέρνησης και της ανοικτής καινοτομίας (European Commission, 2019).

Σε ευρωπαϊκό επίπεδο, το κοινό νομικό πλαίσιο καθορίζεται από την Οδηγία (ΕΕ) 2019/1024 του European Union για τα ανοικτά δεδομένα και την περαιτέρω χρήση πληροφοριών του δημόσιου τομέα, η οποία οικοδομείται γύρω από τους βασικούς άξονες της εσωτερικής αγοράς: την ελεύθερη ροή δεδομένων, τη διαφάνεια και τον θεμιτό ανταγωνισμό. Ωστόσο, δεν υπάγονται όλες οι πληροφορίες του δημόσιου τομέα αυτομάτως στο καθεστώς των ανοικτών δεδομένων, καθώς υφίστανται περιορισμοί που σχετίζονται με την προστασία προσωπικών δεδομένων, την εθνική ασφάλεια και τα πνευματικά δικαιώματα.

Στην Ελλάδα, το θεσμικό πλαίσιο για τη διάθεση και επαναχρησιμοποίηση δεδομένων του δημόσιου τομέα καθορίζεται από τον Ν.4727/2020 (ΦΕΚ 184/τ. Α’/23-09-2020), ο οποίος ενσωματώνει την ευρωπαϊκή οδηγία στο εθνικό δίκαιο. Στο πλαίσιο εφαρμογής του Κεφαλαίου Ι’ του νόμου αυτού πραγματοποιήθηκε η καταγραφή, ταξινόμηση και αξιολόγηση των συνόλων εγγράφων, πληροφοριών και δεδομένων των Γενικών Γραμματειών και Γενικών Διευθύνσεων του Υπουργείου Παιδείας, Θρησκευμάτων και Αθλητισμού, με στόχο τη συστηματική διάθεση ανοικτών δεδομένων και την ενίσχυση της διαφάνειας και της αξιοποίησής τους για ερευνητικούς και αναπτυξιακούς σκοπούς.

Η αξιοποίηση των ανοικτών κρατικών δεδομένων θεωρείται πλέον κρίσιμος παράγοντας για την ανάπτυξη εφαρμογών επιστήμης δεδομένων, τη χάραξη τεκμηριωμένων πολιτικών και την προώθηση της καινοτομίας, ιδιαίτερα σε τομείς όπως η εκπαίδευση, η οικονομία και η δημόσια διοίκηση [24].

3.2. Πρωτογενή δεδομένα Πανελλαδικών εξετάσεων

Στο πλαίσιο της αξιοποίησης ανοικτών δεδομένων, όπως αυτά ορίζονται από το ισχύον θεσμικό πλαίσιο για την περαιτέρω χρήση πληροφοριών του δημόσιου τομέα, ιδιαίτερης σημασίας υπήρξε η συνεισφορά του Υπουργείου Παιδείας, Θρησκευμάτων και Αθλητισμού μέσω της διάθεσης των δεδομένων των Πανελλαδικών Εξετάσεων. Τα δεδομένα αυτά είναι διαθέσιμα στον επίσημο διαδικτυακό τόπο του Υπουργείου, στην ενότητα «Ανοιχτά Δεδομένα», και αποτέλεσαν τη βασική πηγή πληροφόρησης για την παρούσα μελέτη. Επιπλέον, κατόπιν σχετικού αιτήματος που υποβλήθηκε μέσω ηλεκτρονικού ταχυδρομείου προς τη Διεύθυνση Ανάπτυξης Πληροφοριακών Συστημάτων (Τμήμα Ε’ Εφαρμογών Εξετάσεων), παρασχέθηκαν συμπληρωματικά αρχεία που αφορούν το φύλο των υποψηφίων για τα έτη 2023 και 2024, δεδομένου ότι η συγκεκριμένη πληροφορία δεν περιλαμβανόταν στα δημοσιευμένα ανοικτά σύνολα δεδομένων.

Η παρούσα εργασία βασίστηκε στη μελέτη των ανοικτών συνόλων δεδομένων με έτη αναφοράς το 2023 και το 2024. Ειδικότερα, αξιοποιήθηκαν τα αρχεία Αποτελεσμάτων, Βαθμολογιών και Προτιμήσεων, καθώς και τα αρχεία που αφορούν το φύλο των υποψηφίων, για κάθε έτος ξεχωριστά.

Το αρχείο Αποτελεσμάτων περιλάμβανε τις ακόλουθες στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ - Ο Κωδικός Υποψηφίου κρυπτογραφημένος.

2. Δ/ΝΣΗ Δ.Ε. - Η Διεύθυνση Δευτεροβάθμιας Εκπαίδευσης στη οποία ανήκει η Σχολική Μονάδα του μαθητή.
3. ΕΙΔΟΣ ΛΥΚΕΙΟΥ - ΓΕΝΙΚΟ ή ΕΠΑΓΓΕΛΜΑΤΙΚΟ
4. ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ - Αν ο υποψήφιος ανήκει σε ημερήσιο ή εσπερινό Λύκειο.
5. ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ-ΕΙΔΙΚΟΤΗΤΑ - Η Ομάδα προσανατολισμού στην οποία ανήκει ο υποψήφιος.
6. ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ – Η Σχολή που έχει πετύχει ο υποψήφιος εάν έχει πετύχει.
7. ΕΙΔΟΣ ΕΠΙΛΟΓΗΣ – ΓΕΛ Ή ΕΠΑΛ (συντομογραφία σχολείου)

Το αρχείο Βαθμών περιλάμβανε τις εξής στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ - Ο Κωδικός Υποψηφίου κρυπτογραφημένος.
2. ΕΙΔΟΣ ΣΧΟΛΕΙΟΥ - ΓΕΛ ή ΕΠΑΛ
3. ΜΑΘΗΜΑ – Το εξεταζόμενο μάθημα.
4. ΒΑΘΜΟΣ – Ο βαθμός που πήρε στο συγκεκριμένο μάθημα.

Το αρχείο Προτιμήσεων περιλάμβανε τις στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ - Ο Κωδικός Υποψηφίου κρυπτογραφημένος.
2. ΣΕΙΡΑ ΕΠΙΛΟΓΗΣ – Η σειρά προτίμησης
3. ΟΝΟΜΑ ΣΧΟΛΗΣ – Η σχολή που δήλωσε στη συγκεκριμένη σειρά προτίμησης

Το αρχείο Φύλο περιλάμβανε τις στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ - Ο Κωδικός Υποψηφίου κρυπτογραφημένος.
2. ΦΥΛΟ – Το φύλο του Υποψηφίου (Άρρεν ή Θήλυ).

3.3. Προεπεξεργασία και Μετασχηματισμός των δεδομένων για στατιστική ανάλυση

Αρχικά, στο αρχείο **ΑΠΟΤΕΛΕΣΜΑΤΑ** προστέθηκε νέα μεταβλητή με την ονομασία *ΕΠΙΤΥΧΩΝ*, η οποία δημιουργήθηκε βάσει συνθήκης και απέδιδε την τιμή «ΔΕΝ ΠΕΡΑΣΕ» όταν το πεδίο *ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ* ήταν κενό και την τιμή «ΠΕΡΑΣΕ» όταν περιείχε καταχώριση σχολής. Η διαδικασία αυτή επέτρεψε τη δυαδική κωδικοποίηση του αποτελέσματος των εξετάσεων, διευκολύνοντας την περαιτέρω ανάλυση.

Στη συνέχεια πραγματοποιήθηκε συγχώνευση του αρχείου **ΦΥΛΟ** με το αρχείο **ΑΠΟΤΕΛΕΣΜΑΤΑ**, με κοινό αναγνωριστικό τον *ΚΩΔΙΚΟ ΥΠΟΨΗΦΙΟΥ*. Ακολούθως αφαιρέθηκε η μεταβλητή *ΕΙΔΟΣ ΕΠΙΛΟΓΗΣ* και, μέσω φιλτραρίσματος, εξήχθησαν βασικά στατιστικά στοιχεία σχετικά με το πλήθος των υποψηφίων ανά τύπο σχολείου, φύλο, ομάδα προσανατολισμού και τελικό αποτέλεσμα στις εξετάσεις.

Στο αρχείο **ΒΑΘΜΟΙ**, όπου οι επιδόσεις των μαθημάτων ήταν αποθηκευμένες ως δεκαδικές τιμές, πραγματοποιήθηκε μετατροπή του διαχωριστικού από κόμμα σε τελεία, ώστε όλα τα πεδία να αναγνωριστούν ορθά ως αριθμητικές μεταβλητές.

Ακολούθως, τα επεξεργασμένα αρχεία, μαζί με τα πρωτογενή σύνολα δεδομένων, μετατράπηκαν από μορφή Excel σε CSV και μεταφορτώθηκαν στο Google Drive, προκειμένου να καταστεί δυνατή η περαιτέρω ανάλυσή τους με κώδικα Python στο περιβάλλον του Google Colab.

Στο περιβάλλον αυτό πραγματοποιήθηκε, μέσω Python, η συγχώνευση των αρχείων **ΦΥΛΟ**, **ΑΠΟΤΕΛΕΣΜΑΤΑ** και **ΒΑΘΜΟΙ** για κάθε έτος ξεχωριστά. Παράλληλα, δημιουργήθηκε δεύτερο ενοποιημένο σύνολο δεδομένων μέσω συγχώνευσης των αρχείων **ΦΥΛΟ**, **ΑΠΟΤΕΛΕΣΜΑΤΑ** και **ΠΡΟΤΙΜΗΣΕΙΣ**, με σκοπό τη διερεύνηση της σχέσης μεταξύ δημογραφικών χαρακτηριστικών, επίδοσης και εκπαιδευτικών επιλογών των υποψηφίων.

3.4. Τελικά Σύνολα Δεδομένων

Στην πρώτη φάση της στατιστικής ανάλυσης και κατόπιν των απαραίτητων συγχωνεύσεων προέκυψαν τα εξής σύνολα δεδομένων:

Το αρχείο **ΦΥΛΟ_ΑΠΟΤΕΛΕΣΜΑΤΑ_ΒΑΘΜΟΙ** με τις εξής στήλες:

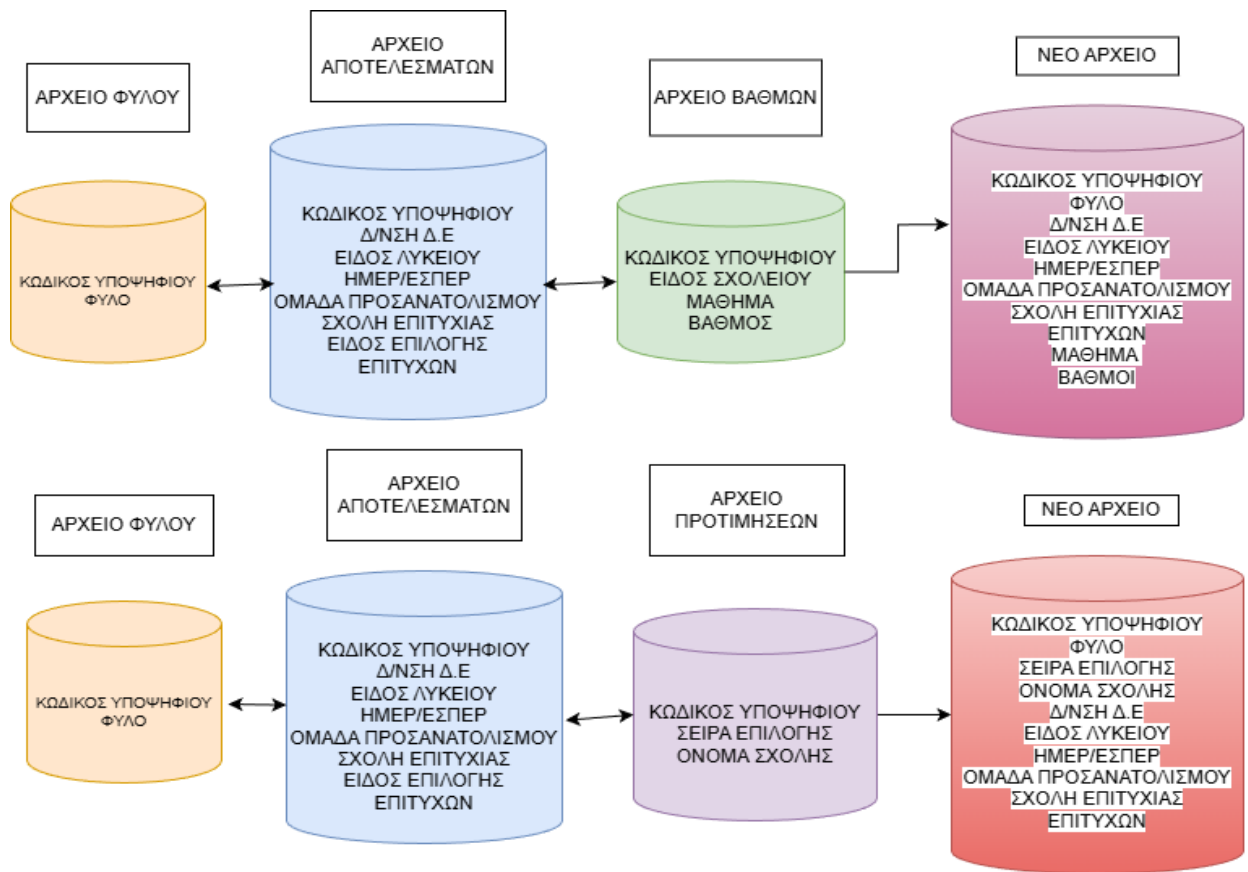
1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΦΥΛΟ
3. Δ/ΝΣΗ Δ.Ε.
4. ΕΙΔΟΣ ΛΥΚΕΙΟΥ
5. ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ
6. ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ-ΕΙΔΙΚΟΤΗΤΑ
7. ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ
8. ΕΠΙΤΥΧΩΝ
9. ΜΑΘΗΜΑ
10. ΒΑΘΜΟΣ

Στο αρχείο αυτό για κάθε ΚΩΔΙΚΟ ΥΠΟΨΗΦΙΟΥ δημιουργήθηκαν τόσες γραμμές όσα και τα μαθήματα εξέτασης του υποψηφίου. Συνολικά **415.352** γραμμές.

Και το αρχείο **ΦΥΛΟ_ΠΡΟΤΙΜΗΣΕΙΣ_ΑΠΟΤΕΛΕΣΜΑΤΑ** με τις εξής στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΦΥΛΟ
3. ΣΕΙΡΑ ΕΠΙΛΟΓΗΣ
4. ΟΝΟΜΑ ΣΧΟΛΗΣ
5. Δ/ΝΣΗ Δ.Ε.
6. ΕΙΔΟΣ ΛΥΚΕΙΟΥ
7. ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ
8. ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ-ΕΙΔΙΚΟΤΗΤΑ
9. ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ
10. ΕΠΙΤΥΧΩΝ

Όπου και εδώ για κάθε ΚΩΔΙΚΟ ΥΠΟΨΗΦΙΟΥ δημιουργήθηκαν τόσες γραμμές όσες και οι προτιμήσεις που δήλωσε ο υποψήφιος στο μηχανογραφικό. Συνολικά **1.048.576** γραμμές.



Σχήμα 3.1: Συγχώνευση Πρωτογενών Αρχείων και Δημιουργία Νέων

Κεφάλαιο 4^ο: Στατιστική Ανάλυση

4.1. Στατιστική Ανάλυση Υποψηφίων 2023-2024

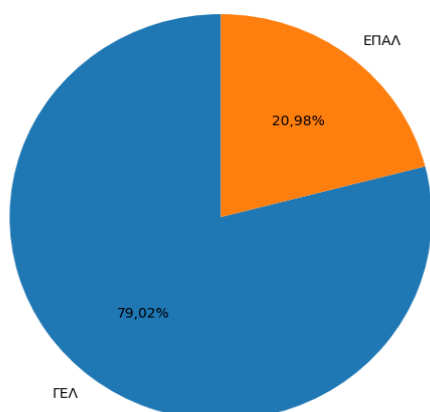
Για την πληρέστερη και αξιόπιστη διερεύνηση των δεδομένων πραγματοποιήθηκε συνολική στατιστική ανάλυση, αφού προηγουμένως ολοκληρώθηκαν οι απαραίτητες διαδικασίες προεπεξεργασίας και συγχώνευσης των επιμέρους αρχείων σε ενιαίο σύνολο δεδομένων.

Στον Πίνακα 4.1.1. παρουσιάζεται το πλήθος των υποψηφίων ανά έτος, ανά τύπο σχολικής μονάδας και ανά κατηγορία Ημερήσιων και Εσπερινών Λυκείων. Από τα στοιχεία προκύπτει ότι το έτος 2024 καταγράφηκε ελαφρώς αυξημένος αριθμός υποψηφίων (101.372) σε σύγκριση με το έτος 2023 (98.266). Επιπλέον, παρατηρείται ότι ο αριθμός των υποψηφίων των Γενικών Λυκείων (ΓΕΛ) είναι περίπου τετραπλάσιος σε σχέση με τον αντίστοιχο αριθμό των υποψηφίων των Επαγγελματικών Λυκείων (ΕΠΑΛ), γεγονός που αποτυπώνεται γραφικά στα κυκλικά Γραφήματα 4.1.1. και 4.1.2.

Πίνακας 4.1.1: Πλήθος Υποψηφίων ανά Είδος Λυκείου

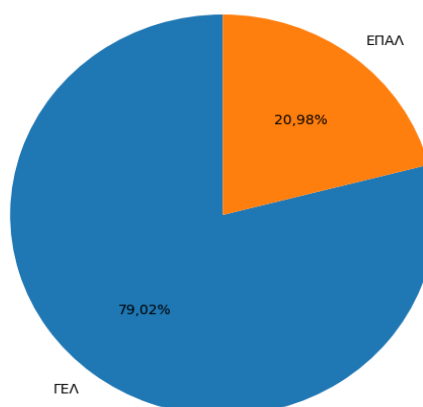
Πλήθος Υποψηφίων			
	ΓΕΛ	2023	2024
	ΗΜΕΡΗΣΙΑ	77146	79634
	ΕΣΠΕΡΙΝΑ	503	474
ΣΥΝΟΛΑ		77649	80108
	ΕΠΑΛ		
	ΗΜΕΡΗΣΙΑ	19189	19881
	ΕΣΠΕΡΙΝΑ	1428	1383
ΣΥΝΟΛΑ		20617	21264

Ποσοστά Υποψηφίων Ανά Είδος Λυκείου 2023



Γράφημα 4.1.1: Ποσοστά Υποψηφίων 2023

Ποσοστά Υποψηφίων Ανά Είδος Λυκείου 2024



Γράφημα 4.1.2: Ποσοστά Υποψηφίων 2024

Στον Πίνακα 4.1.2 παρουσιάζεται η κατανομή του πλήθους των υποψηφίων των Γενικών Λυκείων της χώρας ανά Ομάδα Προσανατολισμού και ανά έτος. Από την ανάλυση των δεδομένων προκύπτει ότι και για τα δύο εξεταζόμενα έτη η Ομάδα Σπουδών Οικονομίας και Πληροφορικής συγκεντρώνει τον μεγαλύτερο αριθμό υποψηφίων. Ειδικότερα, για το έτος 2024, οι υποψήφιοι της συγκεκριμένης Ομάδας αντιστοιχούν στο 37,33% του συνολικού πλήθους. Ακολουθεί η Ομάδα Ανθρωπιστικών Σπουδών με ποσοστό 27,19%, ενώ οι Ομάδες Θετικών Σπουδών και Σπουδών Υγείας παρουσιάζουν ισοδύναμη συμμετοχή, με ποσοστό που προσεγγίζει το 17,5% εκάστη επί του συνόλου των υποψηφίων.

Πίνακας 4.1.2: Πλήθος Υποψηφίων ανά Ομάδα Προσανατολισμού - ΓΕΛ

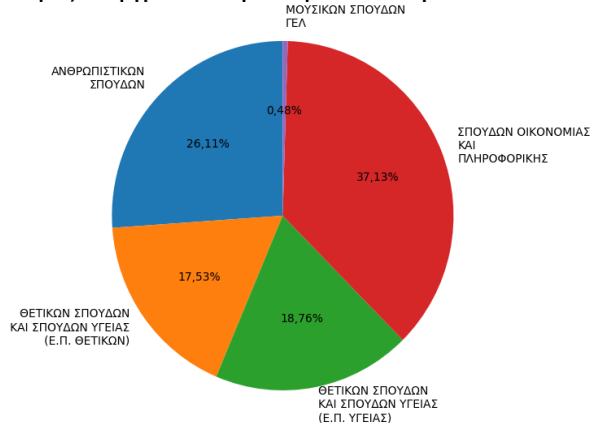
ΠΛΗΘΟΣ ΥΠΟΨΗΦΙΩΝ ΑΝΑ ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ - ΓΕΛ						
	2023			2024		
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	20050	224	20274	21561	218	21779
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ)	13567	45	13612	14255	79	14334
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ)	14477	87	14564	14064	27	14091
ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ	28689	141	28830	29754	150	29904
ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ ΓΕΛ	363	6	369			
ΣΥΝΟΛΑ ΥΠΟΨΗΦΙΩΝ	77146	503	77649	79634	474	80108

Επίσης σύμφωνα με τον Πίνακα 4.1.2. παρατηρείται ότι κατά το έτος 2023 υπήρχε η Ομάδα Μουσικών Σπουδών, η οποία καταργήθηκε το 2024. Οι υποψήφιοι της συγκεκριμένης Ομάδας ενσωματώθηκαν στις υπόλοιπες Ομάδες Προσανατολισμού, εξεταζόμενοι πλέον στα βασικά μαθήματα της αντίστοιχης Ομάδας τους, ενώ τα μαθήματα μουσικής εντάχθηκαν στη κατηγορία των μαθημάτων επιλογής.

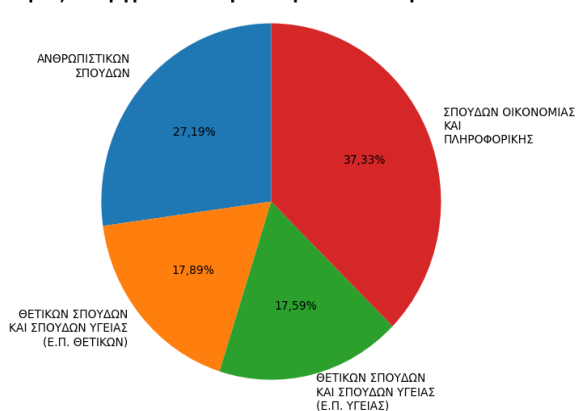
Η σύγκριση της κατανομής των υποψηφίων μεταξύ των ετών 2023 και 2024 καταδεικνύει γενική σταθερότητα στις εκπαιδευτικές προτιμήσεις των μαθητών, με το εντονότερο ενδιαφέρον να διατηρείται στις Ομάδες Σπουδών Οικονομίας και Πληροφορικής καθώς και Ανθρωπιστικών Σπουδών. Παρά τη συνολική αυτή σταθερότητα, παρατηρούνται μικρές αλλά αξιοσημείωτες μεταβολές μεταξύ των επιμέρους επιστημονικών πεδίων.

Στα Γραφήματα 4.1.3. και 4.1.4. η Ομάδα Σπουδών Οικονομίας και Πληροφορικής παραμένει η δημοφιλέστερη επιλογή και για τα δύο έτη, συγκεντρώνοντας ποσοστό άνω του ενός τρίτου του συνολικού αριθμού των υποψηφίων (37,13% το 2023 και 37,33% το 2024). Η Ομάδα Ανθρωπιστικών Σπουδών εμφανίζει τη μεγαλύτερη ποσοστιαία αύξηση, σημειώνοντας άνοδο κατά 1,08 ποσοστιαίες μονάδες, από 26,11% το 2023 σε 27,19% το 2024. Η αύξηση αυτή ενδέχεται να αντανακλά ενίσχυση του ενδιαφέροντος των μαθητών για σπουδές κλασικού, φιλολογικού και νομικού προσανατολισμού.

Πλήθος Υποψηφίων Ανά Ομάδα Προσανατολισμού ΓΕΛ - 2023



Πλήθος Υποψηφίων Ανά Ομάδα Προσανατολισμού ΓΕΛ - 2024



Γράφημα 4.1.3 & 4.1.4: Υποψήφιοι ανά Ομάδα Προσανατολισμού 2023 - 2024

Αντίθετα, στις Ομάδες Θετικών Σπουδών και Σπουδών Υγείας καταγράφεται διαφοροποιημένη τάση, η οποία υποδηλώνει μια ήπια εσωτερική ανακατανομή των προτιμήσεων. Συγκεκριμένα, η Ομάδα Σπουδών Υγείας παρουσιάζει τη μεγαλύτερη μείωση, της τάξης των 1,17 ποσοστιαίων μονάδων, με το ποσοστό συμμετοχής να υποχωρεί από 18,76% το 2023 σε 17,59% το 2024. Αντιθέτως, η Ομάδα Θετικών Σπουδών εμφανίζει μικρή αύξηση κατά 0,36 ποσοστιαίες μονάδες, από 17,53% σε 17,89%.

Η παρατηρούμενη αυτή τάση ενδέχεται να συνδέεται με τον αυξημένο ανταγωνισμό και τις υψηλές απαιτήσεις εισαγωγής στις Σχολές Υγείας, γεγονός που πιθανώς οδηγεί μέρος των υποψηφίων στην επιλογή των Θετικών Σπουδών, προσφέροντας μεγαλύτερη ευελιξία ως προς τις διαθέσιμες ακαδημαϊκές κατευθύνσεις.

Συμπερασματικά, η κατανομή των υποψηφίων το 2024 σε σύγκριση με το 2023 χαρακτηρίζεται από σταθερότητα στη δημοτικότητα των κυρίαρχων επιστημονικών πεδίων (Οικονομία & Πληροφορική), ενίσχυση του ενδιαφέροντος για τις Ανθρωπιστικές Σπουδές και ήπια αποδυνάμωση του ενδιαφέροντος για τις Σπουδές Υγείας, με τις σχετικές μεταβολές να κατευθύνονται κυρίως προς τις Ανθρωπιστικές και, σε μικρότερο βαθμό, προς τις Θετικές Σπουδές.

4.2. Επιτυχόντες και Αποτυχόντες Υποψήφιοι ανά Είδος Λυκείου και ανά Ομάδα Προσανατολισμού

Στον παρακάτω Πίνακα 4.2.1. παρουσιάζονται τα πλήθη των επιτυχόντων και αποτυχόντων των Επαγγελματικών Λυκείων ενώ στον Πίνακα 4.2.2. παρουσιάζεται το πλήθος των μη επιτυχόντων υποψηφίων των Γενικών Λυκείων ανά Ομάδα Προσανατολισμού και ανά Έτος, ενώ στον επόμενο Πίνακα 4.2.3. αποτυπώνεται το αντίστοιχο πλήθος των επιτυχόντων υποψηφίων των Γενικών Λυκείων.

Πίνακας 4.2.1 Επιτυχόντες / Αποτυχόντες ανά Έτος ΕΠΑΛ

ΕΠΙΤΥΧΟΝΤΕΣ / ΑΠΟΤΥΧΟΝΤΕΣ ΕΠΑΛ			
ΕΠΑΛ	ΠΕΡΑΣΕ	ΔΕΝ ΠΕΡΑΣΕ	ΣΥΝΟΛΑ
2023	5939	14678	20617
2024	6306	14958	21264

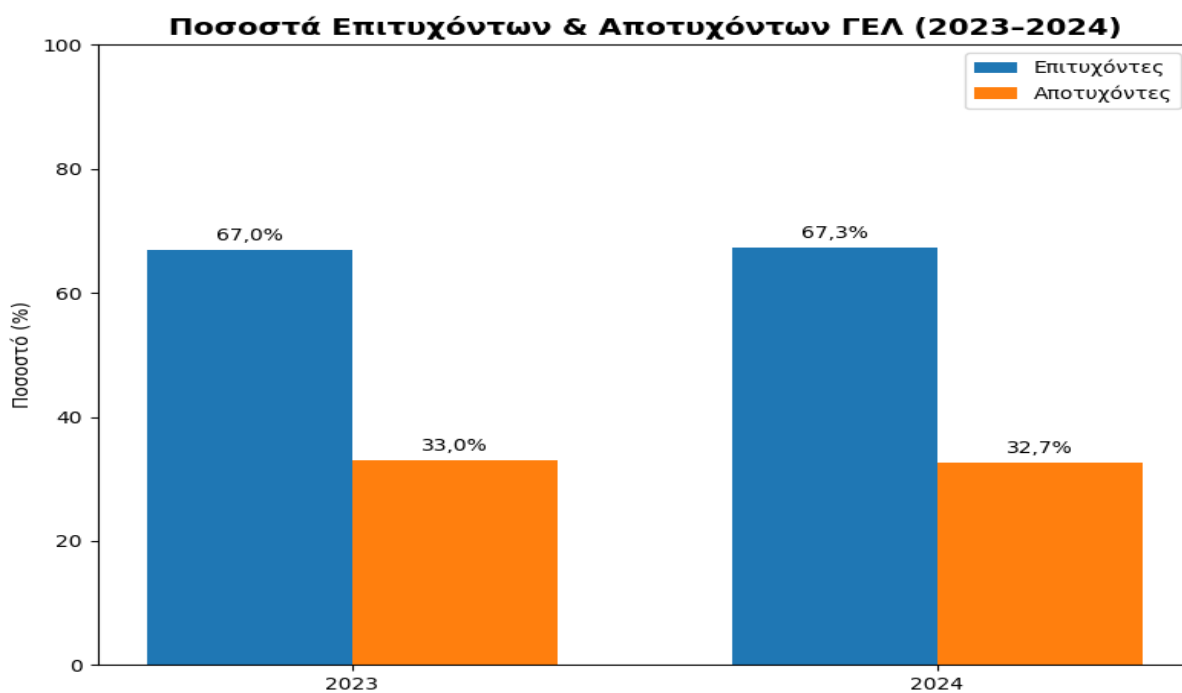
Πίνακας 4.2.2: Αποτυχόντες ανά Ομάδα Προσανατολισμού & ανά Έτος - ΓΕΛ

ΜΗ ΕΠΙΤΥΧΟΝΤΕΣ ΑΝΑ ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ ΓΕΛ						
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	2023			2024		
	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	6490	186	6676	7368	193	7561
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ)	2904	39	2943	2707	21	2728
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ)	4540	72	4612	3973	68	4041
ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ	11177	119	11296	11695	136	11831
ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ	124	5	129			
ΣΥΝΟΛΑ	25235	421	25656	25743	418	26161

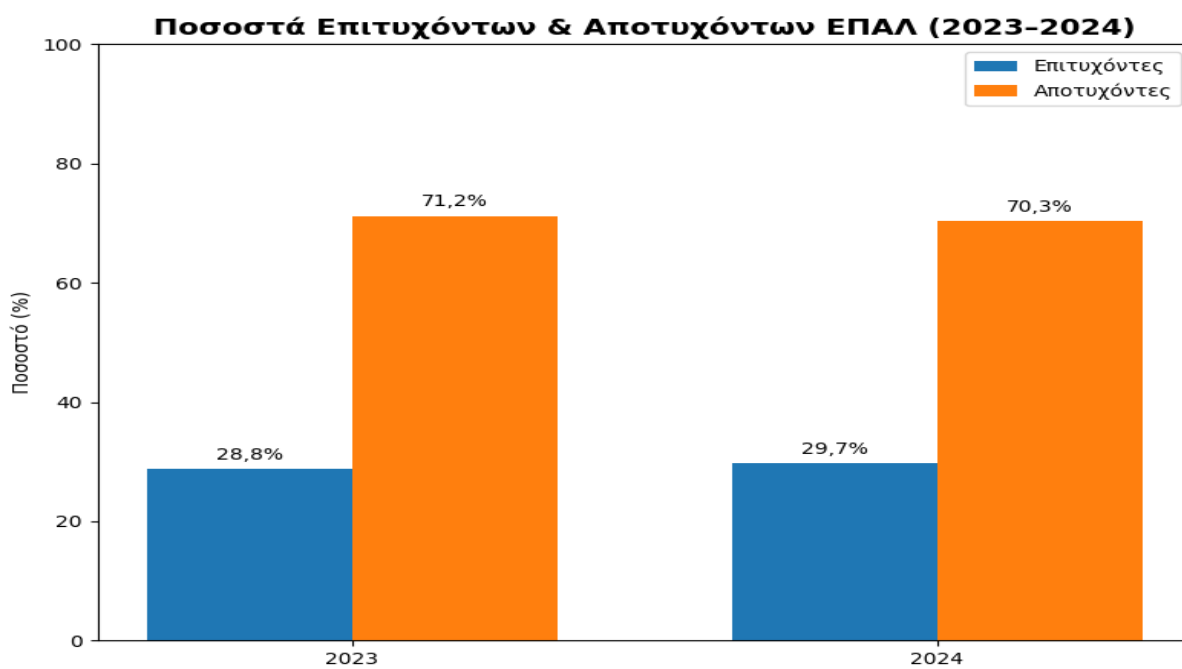
Πίνακας 4.2.3: Επιτυχόντες ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ

ΕΠΙΤΥΧΟΝΤΕΣ ΑΝΑ ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ ΓΕΛ						
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	2023			2024		
	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ	ΗΜΕΡΗΣΙΑ	ΕΣΠΕΡΙΝΑ	ΣΥΝΟΛΑ
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	13560	38	13598	14193	25	14218
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ)	10663	6	10669	11357	6	11363
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ)	9937	15	9952	10282	11	10293
ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ	17512	22	17534	18059	14	18073
ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ	239	1	240			
ΣΥΝΟΛΑ	51911	82	51993	53891	56	53947

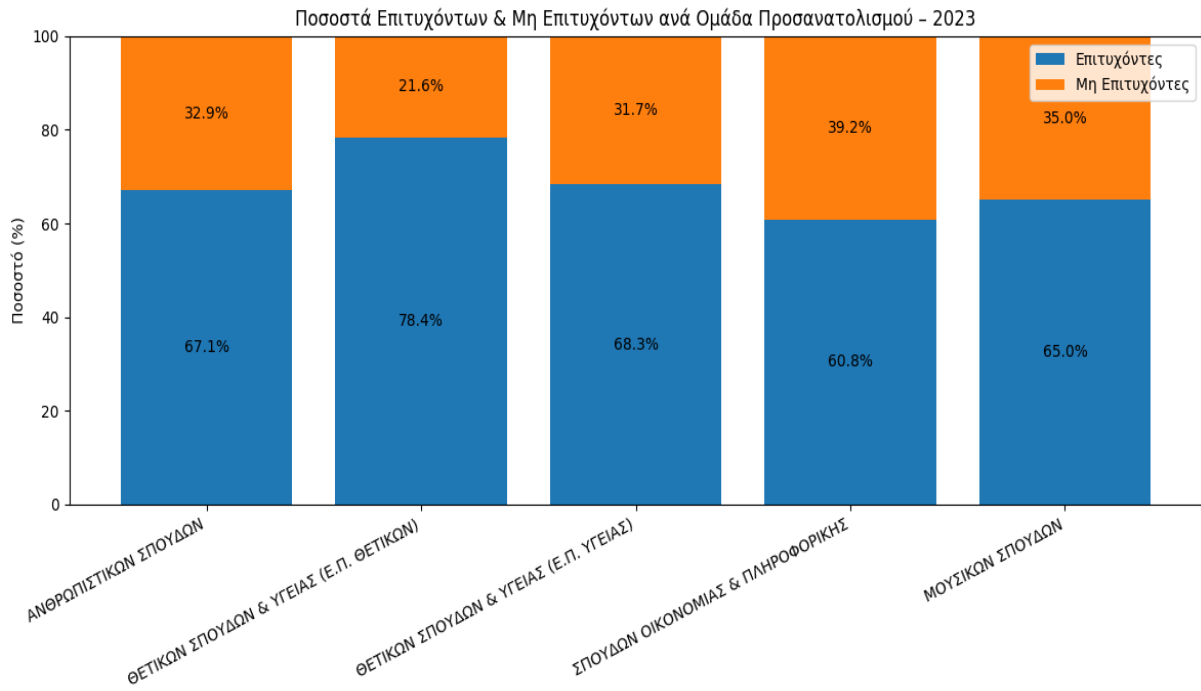
Βάσει των δεδομένων των Πινάκων 4.2.2 και 4.2.3. και του γραφήματος 4.2.1. διαπιστώνεται ότι σε όλες τις Ομάδες Προσανατολισμού του Γενικού Λυκείου το ποσοστό των επιτυχόντων υποψηφίων υπερβαίνει το αντίστοιχο ποσοστό των μη επιτυχόντων για τα έτη 2023 και 2024, γεγονός που υποδηλώνει συνολικά ικανοποιητικό επίπεδο ακαδημαϊκής επίδοσης του μαθητικού πληθυσμού. Σε αντίθεση με τα ΕΠΑΛ που το ποσοστό των επιτυχόντων είναι αντιστρόφως ανάλογο με τα ΓΕΛ και φαίνεται ξεκάθαρα στο Γράφημα 4.2.2.



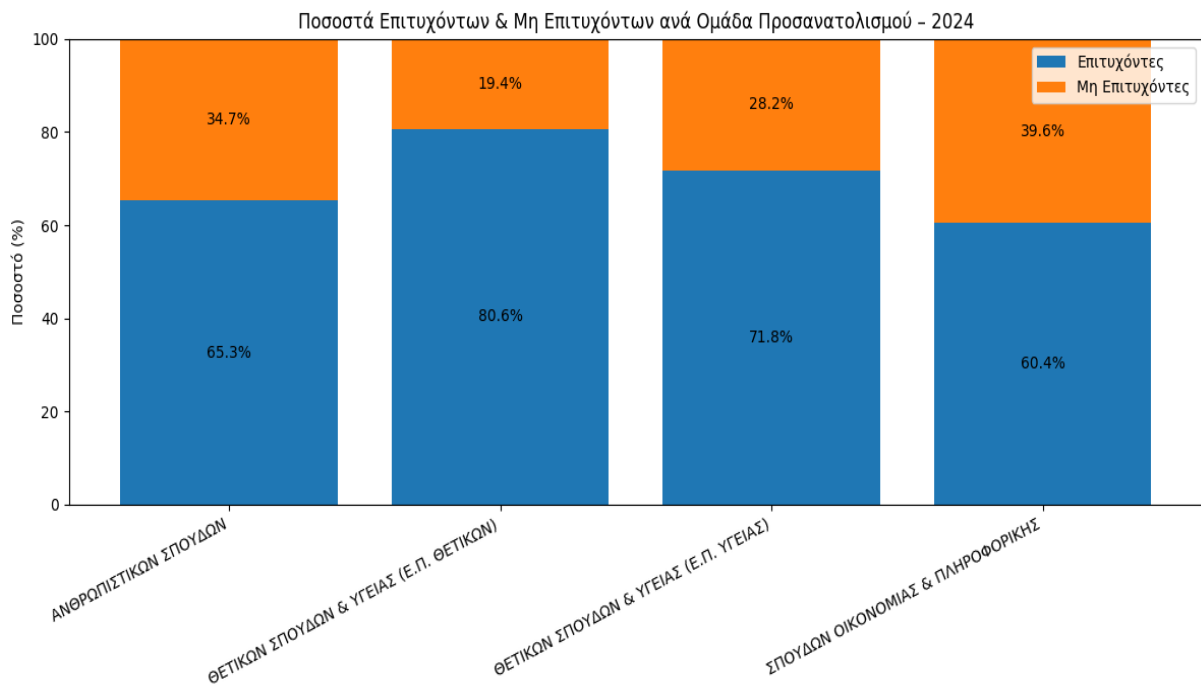
Γράφημα 4.2.1 : Επιτυχόντες και Αποτυχόντες από ΓΕΛ 2023-2024



Γράφημα 4.2.2: Επιτυχόντες και Αποτυχόντες από ΕΠΑΛ 2023-2024



Γράφημα 4.2.3: Επιτυχόντες και Αποτυχόντες ανά Ομάδα Προσανατολισμού από ΓΕΛ 2023



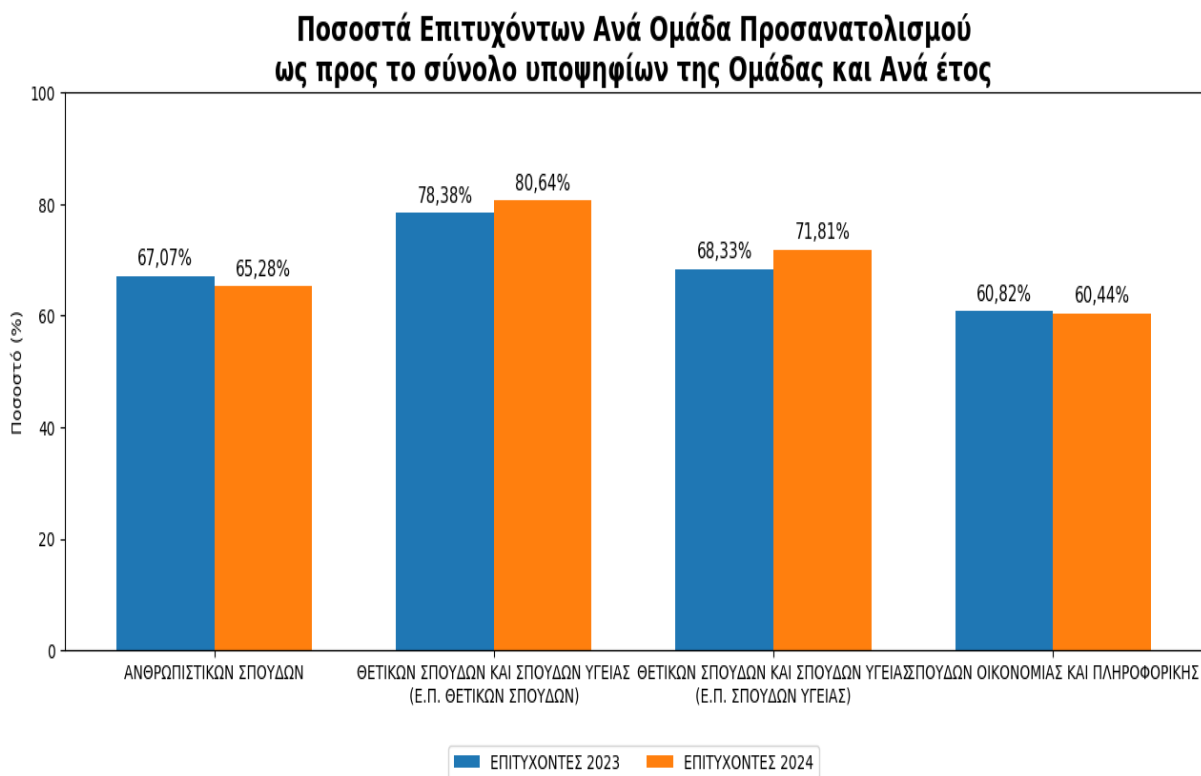
Γράφημα 4.2.4: Ποσοστά Επιτυχόντων & Αποτυχόντων Ανά Ομάδα Προσανατολισμού 2024

Ειδικότερα, για το έτος 2024 (Γράφημα 4.2.4), η Ομάδα Θετικών Σπουδών εμφανίζει το υψηλότερο ποσοστό επιτυχίας, το οποίο ανέρχεται στο 80,64%, με αντίστοιχα χαμηλό ποσοστό αποτυχίας (19,36%). Το εύρημα αυτό ενδέχεται να αντανακλά υψηλό βαθμό προετοιμασίας των υποψηφίων ή αποτελεσματική ευθυγράμμιση της εξεταστικής διαδικασίας με το προβλεπόμενο αναλυτικό πρόγραμμα σπουδών.

Αντιθέτως, η Ομάδα Σπουδών Οικονομίας και Πληροφορικής καταγράφει το χαμηλότερο ποσοστό επιτυχίας (60,44%) και το υψηλότερο ποσοστό αποτυχίας (39,56%). Η διαφοροποίηση αυτή πιθανώς συνδέεται με τις αυξημένες γνωστικές απαιτήσεις της Ομάδας, τον συνδυασμό θεωρητικών και ποσοτικών δεξιοτήτων που απαιτούνται, καθώς και με το ετερογενές μαθησιακό υπόβαθρο των υποψηφίων που την επιλέγουν.

Οι Ανθρωπιστικές Σπουδές και η Ομάδα Σπουδών Υγείας παρουσιάζουν ενδιάμεσα ποσοστά επιτυχίας, συγκεκριμένα 65,28% και 71,81% αντίστοιχα, αναδεικνύοντας διαφοροποιήσεις που ενδέχεται να αποδίδονται στη φύση των γνωστικών αντικειμένων και στο επίπεδο δυσκολίας των εξεταζόμενων μαθημάτων.

Συνολικά, τα αποτελέσματα της ανάλυσης υπογραμμίζουν την ανάγκη για στοχευμένες παιδαγωγικές παρεμβάσεις, ιδίως στις Ομάδες Προσανατολισμού με χαμηλότερα ποσοστά επιτυχίας, με στόχο τη βελτίωση των μαθησιακών αποτελεσμάτων και τη μείωση των εκπαιδευτικών ανισοτήτων.



Γράφημα 4.2.5 Ποσοστά Επιτυχόντων Ανά Ομάδα Προσανατολισμού και ανά Έτος ως προς το σύνολο υποψηφίων της ομάδας

Σύμφωνα με το συγκριτικό Γράφημα 4.2.5 για τα έτη 2023-2024 παρατηρείται για το έτος 2024 γενικευμένη τάση αύξησης του ποσοστού επιτυχίας σχεδόν σε όλες τις Ομάδες Προσανατολισμού σε σύγκριση με το 2023. Αυτό σημαίνει ότι ένα μεγαλύτερο ποσοστό υποψηφίων από κάθε ομάδα κατάφερε να εισαχθεί στην τριτοβάθμια εκπαίδευση το 2024.

Τα υψηλότερα ποσοστά επιτυχίας καταγράφονται στις Ομάδες Θετικών Σπουδών και Σπουδών Υγείας. Ειδικότερα, στις Θετικές Σπουδές περισσότεροι από οκτώ στους δέκα υποψηφίους πέτυχαν την εισαγωγή τους σε πανεπιστημιακές σχολές, γεγονός που ενδέχεται να σχετίζεται είτε με την αναλογία των διαθέσιμων θέσεων σε σχέση με τον αριθμό των υποψηφίων είτε με υψηλότερο επίπεδο ακαδημαϊκής επίδοσης των εξεταζόμενων.

Αντιθέτως, η Ομάδα Ανθρωπιστικών Σπουδών αποτελεί τη μοναδική Ομάδα που εμφανίζει μείωση στο ποσοστό επιτυχίας, της τάξης των 1,79 ποσοστιαίων μονάδων. Η μείωση αυτή, σε συνδυασμό με την αύξηση του συνολικού αριθμού των υποψηφίων στην εν λόγω Ομάδα το 2024, υποδηλώνει ενίσχυση του ανταγωνισμού, με αποτέλεσμα μικρότερο ποσοστό υποψηφίων να επιτυγχάνει τελικά την εισαγωγή στην τριτοβάθμια εκπαίδευση.

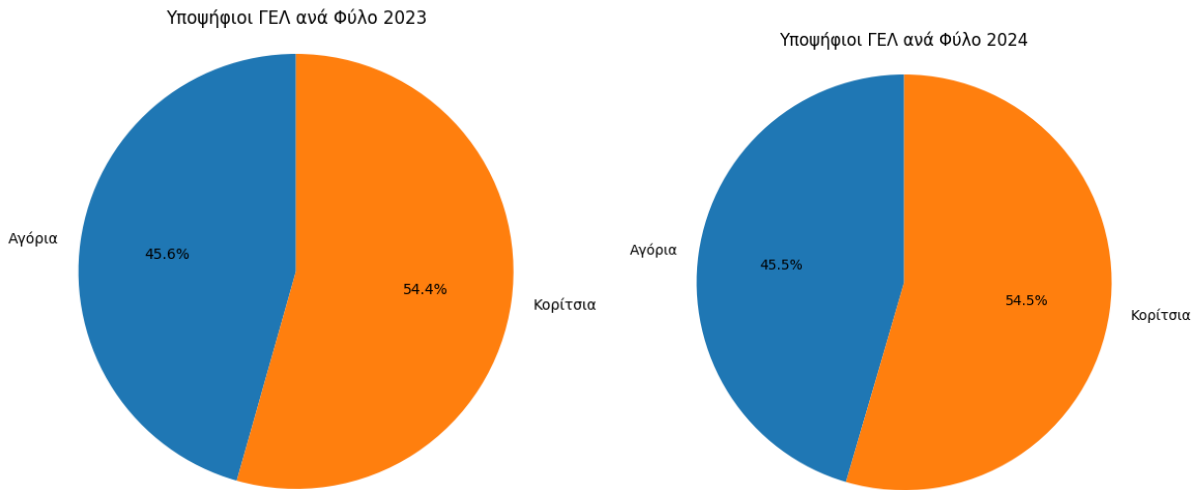
Η Ομάδα Σπουδών Οικονομίας και Πληροφορικής παρουσιάζει τη μικρότερη μεταβολή στο ποσοστό επιτυχίας (-0,38 ποσοστιαίες μονάδες), διατηρώντας ωστόσο το χαμηλότερο συνολικό ποσοστό επιτυχίας μεταξύ των Ομάδων Προσανατολισμού. Συγκεκριμένα, λιγότεροι από έξι στους δέκα υποψηφίους της Ομάδας καταφέρνουν να εισαχθούν σε πανεπιστημιακές σχολές, γεγονός που αναδεικνύει την ανάγκη περαιτέρω διερεύνησης των παραγόντων που επηρεάζουν την επίδοση των υποψηφίων της συγκεκριμένης κατεύθυνσης.

4.3. Υποψήφιοι ανά Φύλο

Πίνακας 4.3.1: Υποψήφιοι ανά Φύλο, ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ

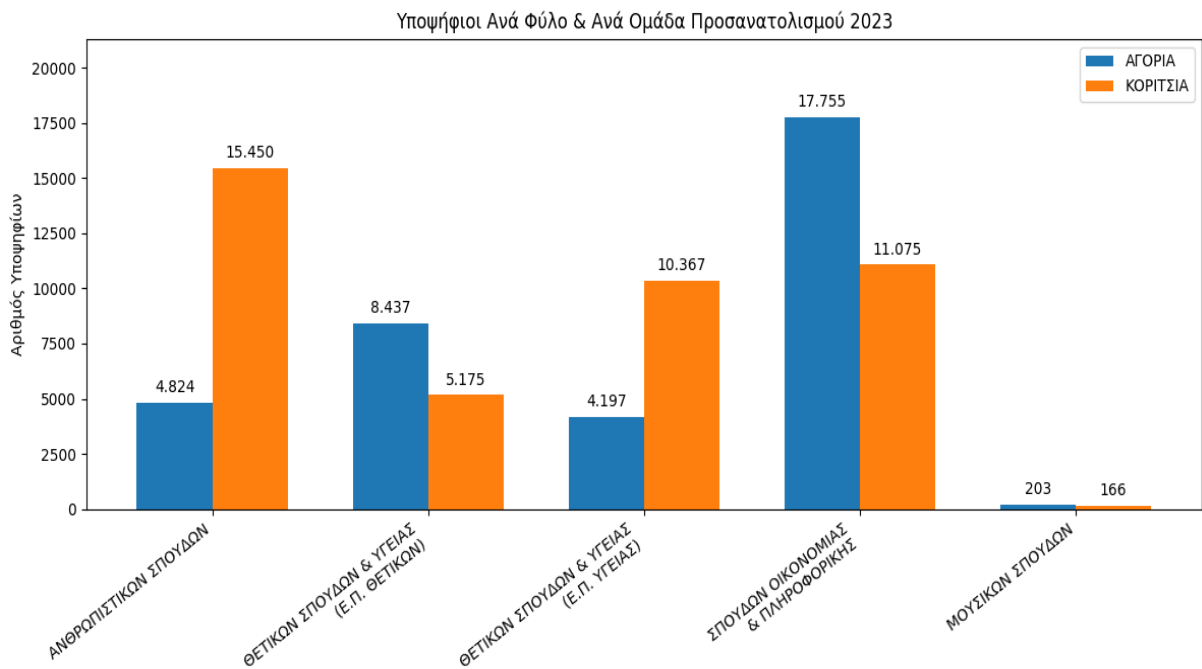
ΥΠΟΨΗΦΙΟΙ ΑΝΑ ΦΥΛΟ & ΑΝΑ ΕΤΟΣ						
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	2023			2024		
	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΣΥΝΟΛΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΣΥΝΟΛΑ
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	4.824	15.450	20.274	5.151	16.628	21.779
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ)	8.437	5.175	13.612	8.658	5.433	14.091
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ)	4.197	10.367	14.564	4.027	10.307	14.334
ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ	17.755	11.075	28.830	18.597	11.307	29.904
ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ	203	166	369	-	-	-
ΣΥΝΟΛΑ	35.416	42.233	77.649	36.433	43.675	80.108

Στο πλαίσιο των υποψηφίων των Γενικών Λυκείων, τα κυκλικά Γραφήματα 4.3.3 και 4.3.4 καταδεικνύουν ότι και για τα δύο εξεταζόμενα έτη περίπου το 55% των υποψηφίων είναι μαθήτριες, ενώ το υπόλοιπο 45% αντιστοιχεί σε μαθητές. Η συγκεκριμένη κατανομή υποδηλώνει ότι οι μαθήτριες αποτελούν την πλειονότητα του υποψηφίου πληθυσμού των Πανελλαδικών Εξετάσεων του Γενικού Λυκείου. Παρότι η διαφορά μεταξύ των δύο φύλων δεν εμφανίζεται ιδιαίτερα έντονη, είναι σαφής και συνεπής με τη γενικότερη τάση αυξημένης συμμετοχής των κοριτσιών στη δευτεροβάθμια εκπαίδευση και ειδικότερα στο Γενικό Λύκειο.

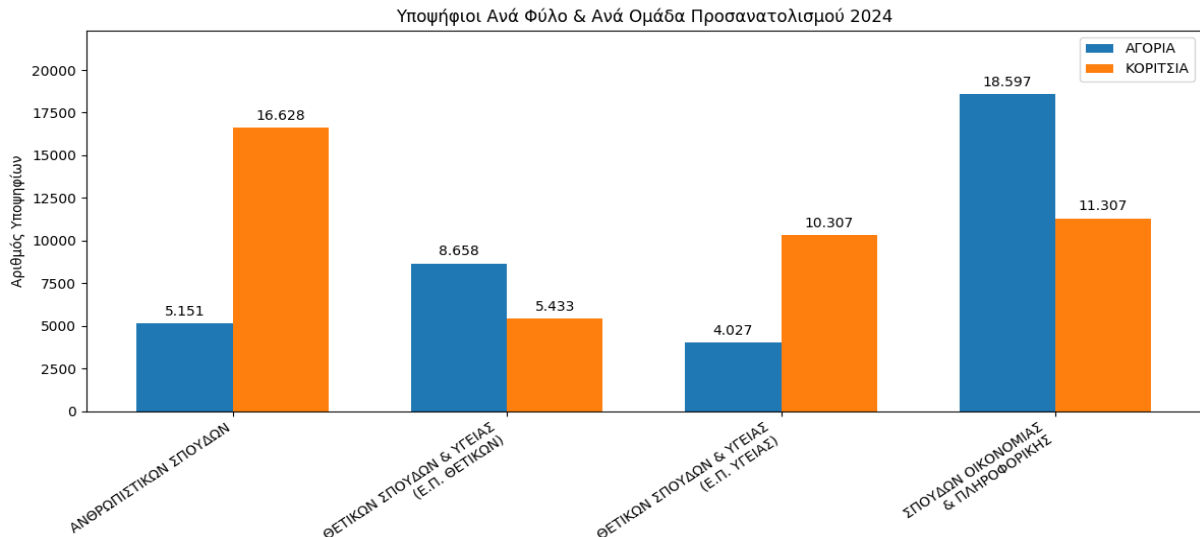


Γράφημα 4.3.1: Υποψήφιοι ΓΕΛ ανά Φύλο 2023

Γράφημα 4.3.2: Υποψήφιοι ΓΕΛ ανά Φύλο 2024



Γράφημα 4.3.3: Υποψήφιοι ανά Φύλο & ανά Ομάδα Προσανατολισμού 2023



Γράφημα 4.3.4: Υποψήφιοι ανά Φύλο & ανά Ομάδα Προσανατολισμού 2024

Στα Γραφήματα 4.3.3 και 4.3.4 αποτυπώνονται σαφείς διαφοροποιήσεις μεταξύ μαθητών και μαθητριών ως προς την επιλογή Ομάδων Προσανατολισμού και για τα δύο εξεταζόμενα έτη. Οι μαθήτριες εμφανίζουν ιδιαίτερα αυξημένη συμμετοχή στις Ανθρωπιστικές Σπουδές, υπερτερώντας αριθμητικά σε σημαντικό βαθμό έναντι των μαθητών, γεγονός που συνάδει με τη διαχρονικά καταγεγραμμένη τάση ισχυρότερης παρουσίας του γυναικείου φύλου σε θεωρητικά και ανθρωποκεντρικά επιστημονικά πεδία.

Αντιθέτως, οι μαθητές παρουσιάζουν σαφή προτίμηση προς τις Σπουδές Οικονομίας και Πληροφορικής, οι οποίες συγκεντρώνουν το μεγαλύτερο ποσοστό υποψηφίων του ανδρικού φύλου, αν και η συμμετοχή των μαθητριών παραμένει αξιόλογη. Στις Θετικές Σπουδές καταγράφεται μικρότερη απόκλιση μεταξύ των δύο φύλων, με εναλλαγή υπεροχής ανά επιμέρους κατηγορία, στοιχείο που υποδηλώνει τάση σύγκλισης των φύλων σε επιστημονικά πεδία υψηλής γνωστικής απαίτησης.

Συνολικά, τα ευρήματα καταδεικνύουν ότι, παρά τη διατήρηση διαφοροποιήσεων στις εκπαιδευτικές επιλογές μεταξύ αγοριών και κοριτσιών, σε ορισμένες Ομάδες Προσανατολισμού παρατηρείται σταδιακή εξισορρόπηση και σύγκλιση ως προς τη συμμετοχή των δύο φύλων.

4.4. Επιτυχόντες ανά Φύλο

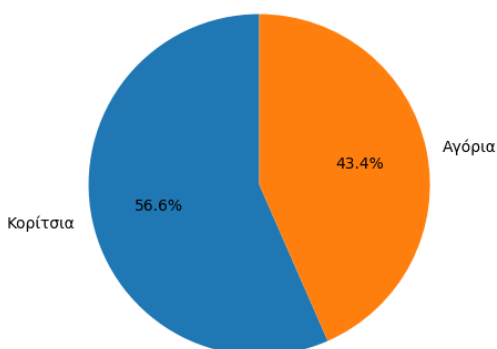
Στον Πίνακα 4.4.1 παρουσιάζονται τα πλήθη των επιτυχόντων υποψηφίων των Γενικών Λυκείων ανά Φύλο και ανά Ομάδα Προσανατολισμού.

Ενώ τα κυκλικά Γραφήματα 4.4.1 και 4.4.2 απεικονίζουν την κατανομή των επιτυχόντων υποψηφίων των Γενικών Λυκείων με βάση το φύλο. Σύμφωνα με τα στοιχεία και για τα δύο εξεταζόμενα έτη, οι μαθήτριες αντιπροσωπεύουν περίπου το 56,5% του συνόλου των επιτυχόντων, ενώ οι μαθητές το υπόλοιπο 43,5%.

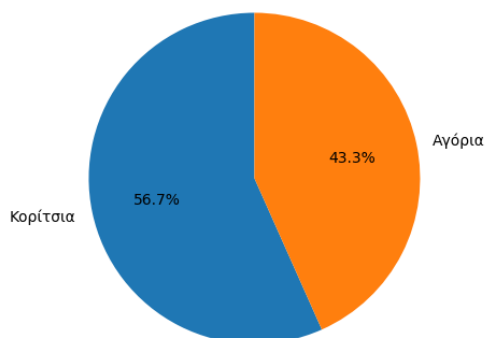
Πίνακας 4.4.1: Επιτυγχόντες ανά Φύλο, ανά Ομάδα Προσανατολισμού & ανά Έτος – ΓΕΛ

ΕΠΙΤΥΧΟΝΤΕΣ ΓΕΛ ΑΝΑ ΦΥΛΟ						
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	2023			2024		
	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΣΥΝΟΛΑ	ΑΓΟΡΙΑ	ΚΟΡΙΤΣΙΑ	ΣΥΝΟΛΑ
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	2.813	10.785	13.598	2.907	11.311	14.218
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ)	6.563	4.106	10.669	6.880	4.483	11.363
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ)	2.904	7.048	9.952	2.903	7.390	10.293
ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ	10.178	7.356	17.534	10.685	7.388	18.073
ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ	123	117	240	-	-	-
ΣΥΝΟΛΑ	22.581	29.412	51.993	23.375	30.572	53.947

Επιτυγχόντες ανά φύλο 2023

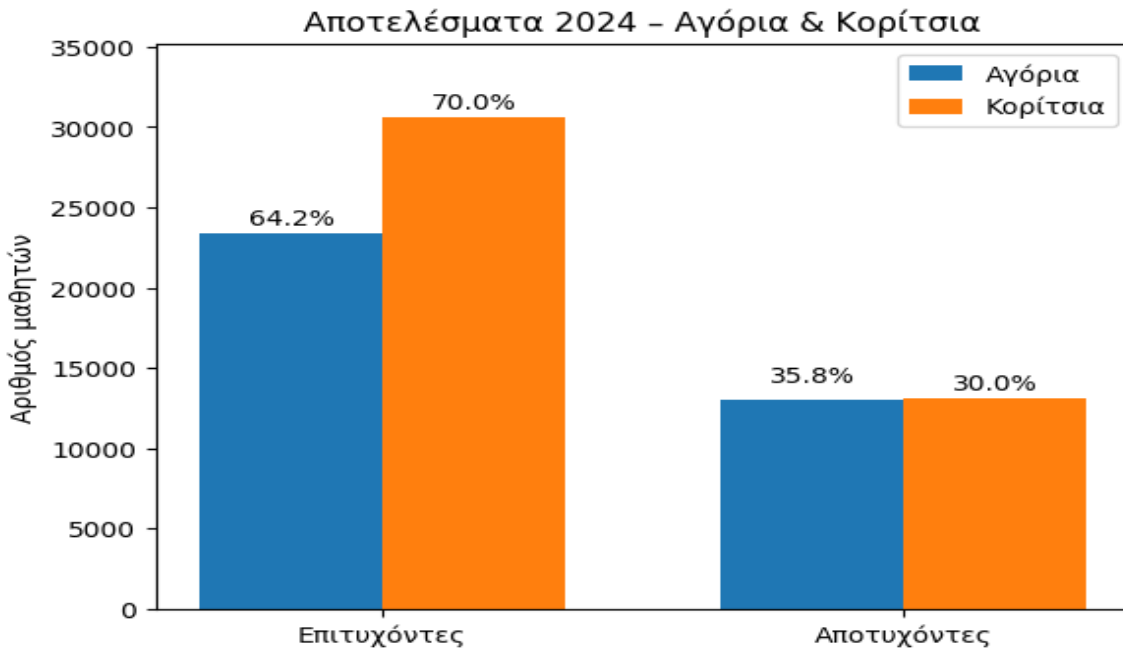


Επιτυγχόντες ανά φύλο 2024



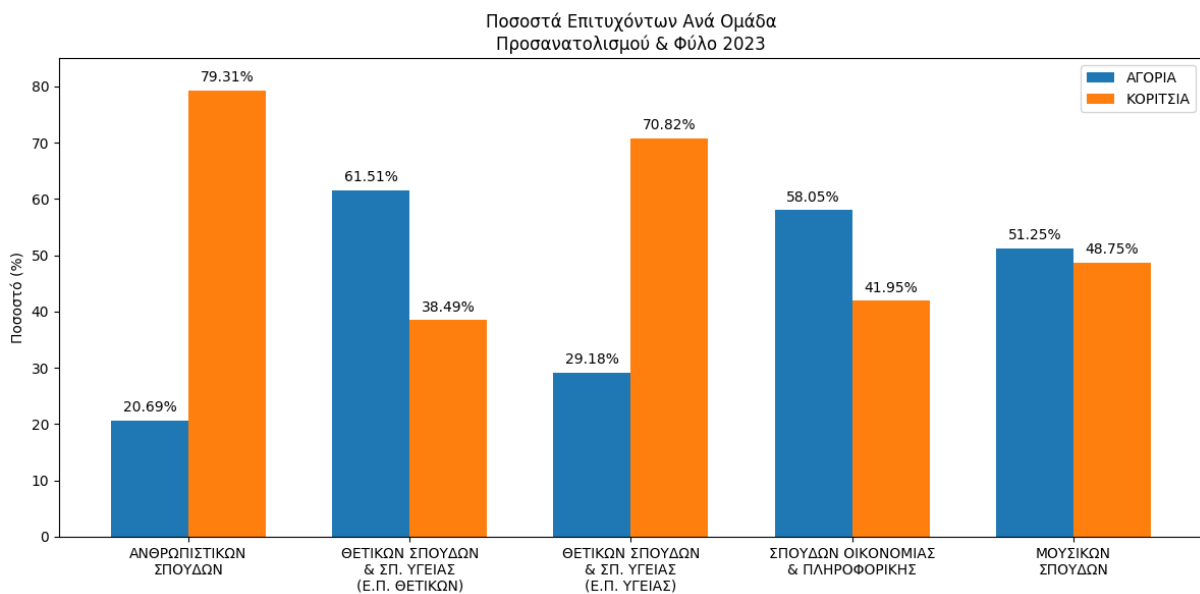
Γραφήματα 4.4.1 και 4.4.2: Ποσοστά Επιτυχόντων ανά Φύλο 2023 & 2024

Επίσης στο Γράφημα 4.4.3 παρατηρείται ότι το 2024, τα κορίτσια όχι μόνο αποτελούν την πλειονότητα των επιτυχόντων (56,7%), αλλά εμφανίζουν και υψηλότερο ποσοστό επιτυχίας σε σχέση με τα αγόρια (70% έναντι 64,2%) . Τα ίδια αποτελέσματα ισχύουν και για το 2023 (69,6% έναντι 63,8%). Αυτό δείχνει ότι η υπεροχή τους στους επιτυχόντες είναι τόσο ποσοτική όσο και αναλογική.

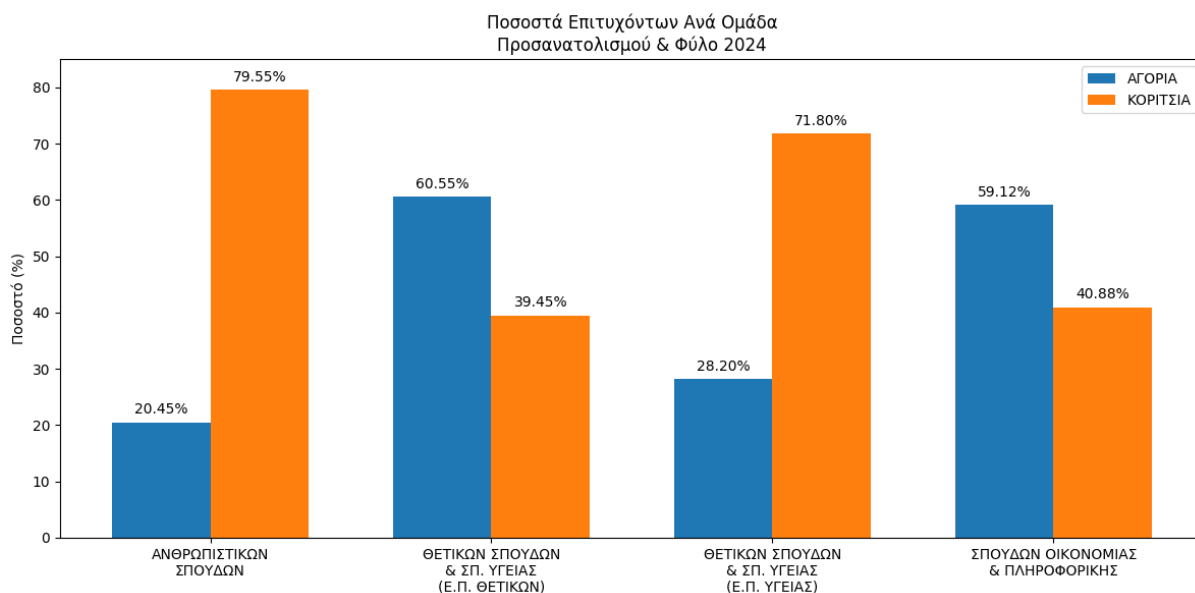


Γράφημα 4.4.3: Ποσοστά Επιτυχόντων/Αποτυχόντων ανά Φύλο ως προ το σύνολο των αγοριών και των κοριτσιών αντίστοιχα - 2024

Η συγκεκριμένη εικόνα συνάδει με τη μεγαλύτερη συμμετοχή των κοριτσιών στο σύνολο των υποψηφίων, ωστόσο παράλληλα υποδηλώνει και υψηλότερη αναλογία επιτυχίας σε σύγκριση με τους μαθητές. Συνολικά, τα δεδομένα αναδεικνύουν ότι οι μαθήτριες των Γενικών Λυκείων όχι μόνο συμμετέχουν σε μεγαλύτερο βαθμό στις Πανελλαδικές Εξετάσεις, αλλά και εμφανίζουν αυξημένα ποσοστά επιτυχίας για τα έτη 2023 και 2024.



Γράφημα 4.4.4: Ποσοστά Επιτυχόντων ανά Φύλο & ανά Ομάδα Προσανατολισμού 2023



Γράφημα 4.4.5: Ποσοστά Επιτυχόντων ανά Φύλο & ανά Ομάδα Προσανατολισμού 2024

Η σύγκριση των ποσοστών επιτυχόντων ανά Ομάδα Προσανατολισμού και φύλο μεταξύ των ετών 2023 και 2024 αναδεικνύει όχι ιδιαίτερες διαφοροποιήσεις, γεγονός που υποδηλώνει σχετική σταθερότητα στη διάρθρωση των επιτυχόντων μεταξύ των δύο φύλων, τόσο ως προς τις δηλώσεις προτίμησης όσο και ως προς την ακαδημαϊκή τους επίδοση.

Πίνακας 4.4.2: Ποσοστά Μεταβολών ανά Ομάδα

Ομάδα	Ποσοστό 2023	Ποσοστό 2024	Μεταβολή
Ανθρωπιστικών Σπουδών	67,0%	65,3%	-1,7%
Θετικών (Ε.Π. Θετικών)	78,4%	80,6%	+2,2%
Θετικών (Ε.Π. Υγείας)	68,3%	71,8%	+3,5%
Οικονομίας & Πληροφορικής	60,8%	60,9%	+0,1%

Ομάδες Προσανατολισμού και κυρίαρχο φύλο

Από την παραπάνω ανάλυση και τον συνολικό Πίνακα 4.4.2. προκύπτει η επικράτηση συγκεκριμένου φύλου σε ορισμένα επιστημονικά πεδία. Ειδικότερα:

Ανθρωπιστικές Σπουδές: Οι μαθήτριες διατηρούν συντριπτική υπεροχή, καταλαμβάνοντας περίπου το 80% του συνόλου των επιτυχόντων και κατά τα δύο εξεταζόμενα έτη.

Θετικές Σπουδές: Οι μαθητές εμφανίζουν σαφή υπεροχή, με ποσοστά που κυμαίνονται περίπου μεταξύ 60% και 61,5% των επιτυχόντων.

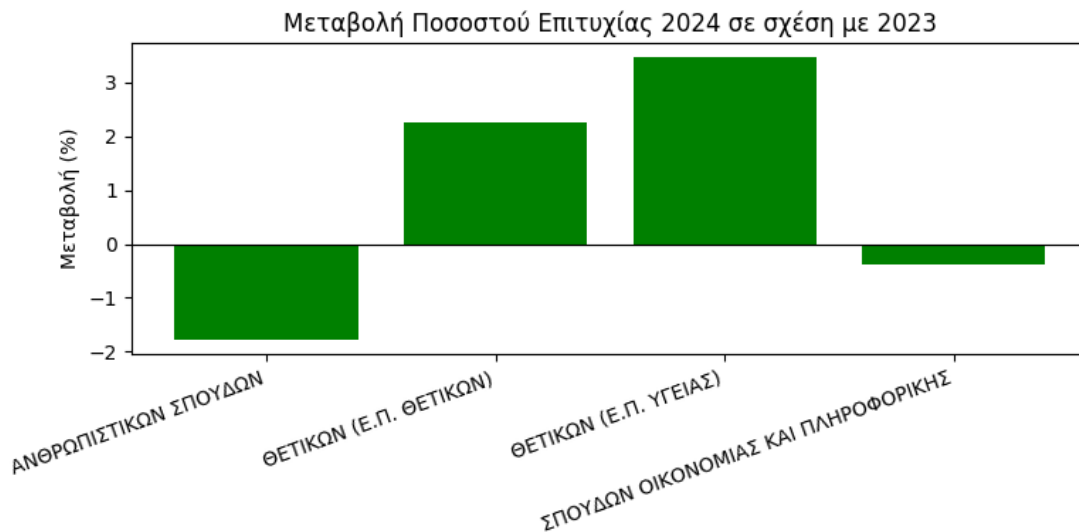
Σπουδές Υγείας: Και σε αυτό το πεδίο οι μαθήτριες κυριαρχούν, συγκεντρώνοντας ποσοστά της τάξης του 71%–72% των επιτυχόντων. Αξιοσημείωτη είναι η αύξηση της συμμετοχής τους το 2024 κατά 0,98 ποσοστιαίες μονάδες, η οποία συνιστά τη δεύτερη μεγαλύτερη μεταβολή που καταγράφεται.

Σπουδές Οικονομίας και Πληροφορικής: Οι μαθητές υπερτερούν οριακά, με ποσοστά που κυμαίνονται περίπου μεταξύ 58% και 59% του συνόλου των επιτυχόντων.

Συνοψίζοντας, τα αποτελέσματα καταδεικνύουν ότι η κατανομή των επιτυχόντων ανά φύλο παραμένει ιδιαίτερα σταθερή μεταξύ των ετών 2023 και 2024. Οι παραδοσιακές διαφοροποιήσεις στις επιλογές σπουδών βάσει φύλου—όπως η υπεροχή των μαθητριών στις Ανθρωπιστικές Σπουδές και στις Σπουδές Υγείας και των μαθητών στις Θετικές Σπουδές και στις Σπουδές Οικονομίας και Πληροφορικής—εξακολουθούν να διατηρούνται σε μεγάλο βαθμό.

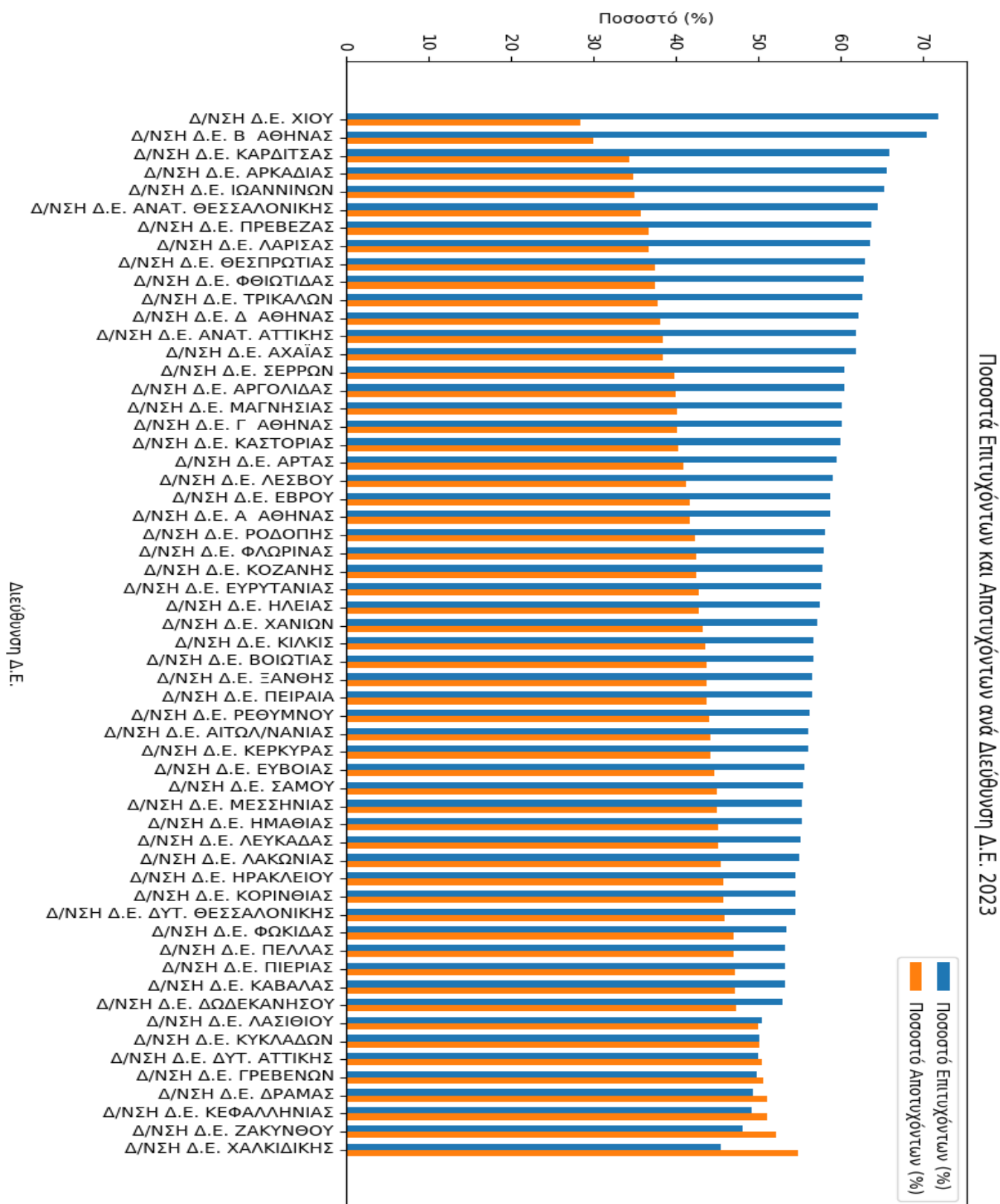
Το 2024 χαρακτηρίζεται από βελτίωση των αποτελεσμάτων στις Θετικές Σπουδές και σταθερότητα στα υπόλοιπα πεδία, γεγονός που υποδηλώνει ομαλή και ελαφρώς ανοδική τάση συνολικής επίδοσης των μαθητών ΓΕΛ.

Με βάση τα πλήρη στατιστικά στοιχεία του Υπουργείου Παιδείας για τις Πανελλαδικές Εξετάσεις του 2024, η ομάδα από την οποία προέρχονται οι περισσότεροι αποτυχόντες (υποψήφιοι που δεν εισήχθησαν σε σχολή) είναι και πάλι η Ομάδα Σπουδών Οικονομίας & Πληροφορικής. Στο επόμενο Γράφημα 4.4.6. βλέπουμε τα ποσοστά μεταβολών ως προς την επιτυχία ανά Ομάδα Προσανατολισμού.

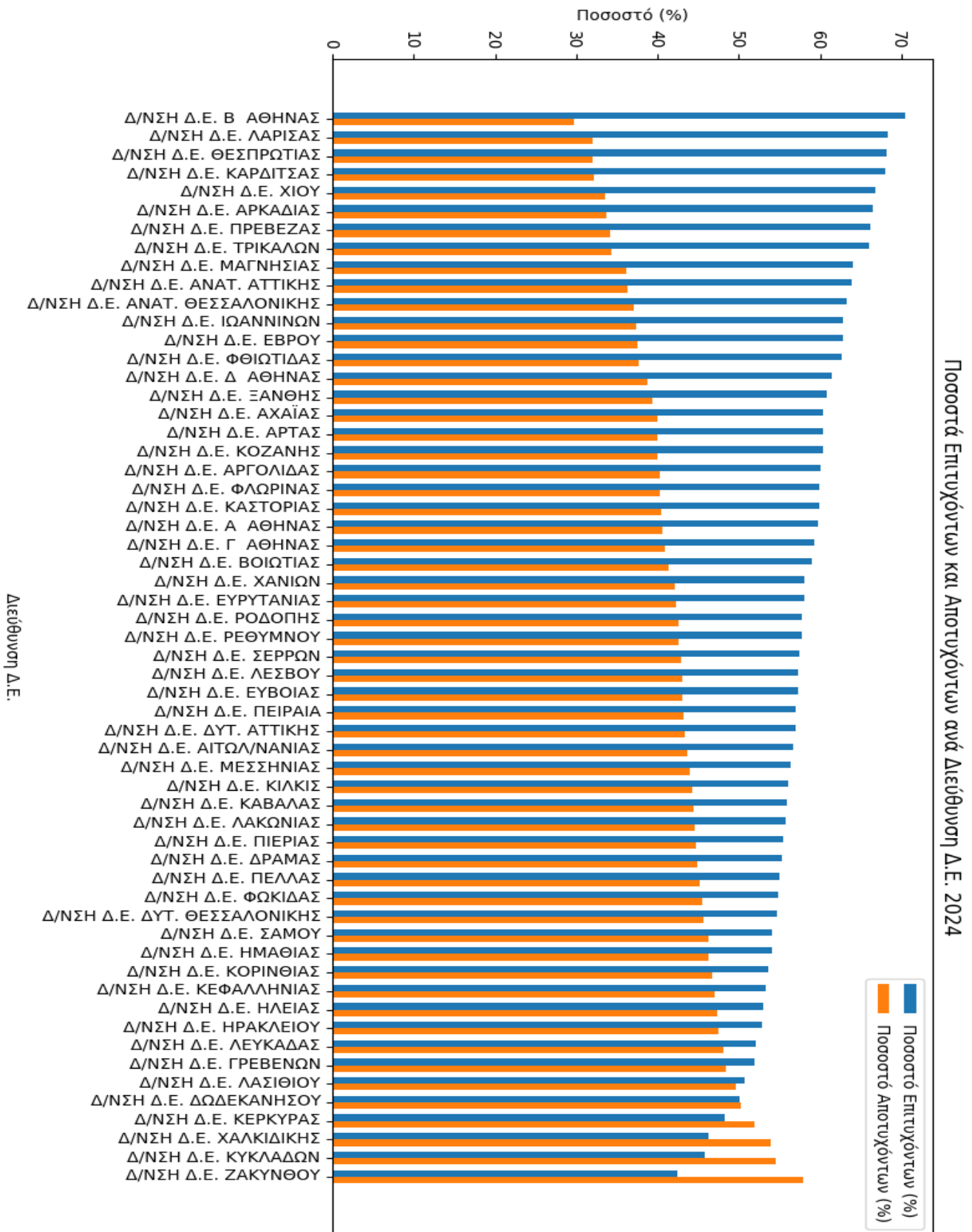


Γράφημα 4.4.6: Μεταβολή Ποσοστού Επιτυχίας 2024 σε σχέση με 2023

4.5. Ποσοστά Επιτυχόντων & Αποτυχόντων ανά Διεύθυνση 2023 και 2024



Γράφημα 4.5.1: Ποσοστά Επιτυχόντων ανά Διεύθυνση 2023



Γράφημα 4.5.2: Ποσοστά Επιτυχόντων ανά Διεύθυνση 2024

Τα Γραφήματα 4.5.1. και 4.5.2. απεικονίζονται τα ποσοστά επιτυχίας και αποτυχίας των Διευθύνσεων Δευτεροβάθμιας Εκπαίδευσης για τα έτη 2023 και 2024, με στόχο τη διερεύνηση της χωρικής διαφοροποίησης της εκπαιδευτικής επίδοσης και την ανάδειξη ενδεχόμενων ανισοτήτων μεταξύ των επιμέρους Διευθύνσεων.

Διευθύνσεις με υψηλά ποσοστά επιτυχίας

Σύμφωνα με τα διαθέσιμα δεδομένα, στις ανώτερες θέσεις της κατανομής για το έτος 2023 καταγράφονται οι Διευθύνσεις Δευτεροβάθμιας Εκπαίδευσης Χίου, Β' Αθήνας, Καρδίτσας, Αρκαδίας και Ιωαννίνων, ενώ για το 2024 αντίστοιχα οι Διευθύνσεις Β' Αθήνας, Λάρισας, Θεσπρωτίας, Καρδίτσας και Χίου. Οι εν λόγω Διευθύνσεις εμφανίζουν ποσοστά επιτυχίας που κυμαίνονται μεταξύ 65% και 70%, επίπεδα σημαντικά υψηλότερα σε σύγκριση με τον συνολικό μέσο όρο. Το εύρημα αυτό μπορεί να ερμηνευθεί είτε ως ένδειξη αυξημένης αποτελεσματικότητας του εκπαιδευτικού συστήματος στις εν λόγω περιοχές είτε ως αποτέλεσμα υψηλής στοχοπροσήλωσης των υποψηφίων, η οποία, σε συνδυασμό με τα κοινωνικοοικονομικά χαρακτηριστικά των τοπικών κοινωνιών, συμβάλλουν στην επίτευξη υψηλότερων ποσοστών επιτυχίας.

Διευθύνσεις με χαμηλά ποσοστά επιτυχίας

Αντιθέτως, στο κατώτερο άκρο της κατανομής και για τα δύο έτη εντοπίζονται οι Διευθύνσεις Δευτεροβάθμιας Εκπαίδευσης Κεφαλληνίας, Ζακύνθου, Κυκλάδων, Χαλκιδικής, Κέρκυρας και Δωδεκανήσου. Στις περιοχές αυτές τα ποσοστά επιτυχίας δεν υπερβαίνουν το 35%, γεγονός που συνεπάγεται αυξημένα επίπεδα αποτυχίας και υποδεικνύουν χαμηλότερη συνολική εκπαιδευτική επίδοση.

Εύρος τιμών και βαθμός διαφοροποίησης

Η διαφορά μεταξύ της Διεύθυνσης με την υψηλότερη και εκείνης με τη χαμηλότερη επίδοση υπερβαίνει τις 35 ποσοστιαίες μονάδες. Η έκταση αυτής της απόκλισης αναδεικνύει την ύπαρξη έντονων ανισοτήτων στην εκπαιδευτική επίδοση σε περιφερειακό επίπεδο. Παράλληλα, η πλειονότητα των Διευθύνσεων συγκεντρώνεται σε ενδιάμεσες τιμές επιτυχίας, οι οποίες κυμαίνονται περίπου μεταξύ 40% και 55%, στοιχείο που υποδηλώνει μια μέση συνολική εικόνα με σημαντικές, ωστόσο, εσωτερικές διαφοροποιήσεις.

Ερμηνευτικοί παράγοντες

Η παρατηρούμενη διαφοροποίηση στα ποσοστά επιτυχίας ενδέχεται να ερμηνευθεί ως αποτέλεσμα συνδυασμού παραγόντων, όπως η σύνθεση του μαθητικού πληθυσμού (αστικές έναντι αγροτικών ή νησιωτικών περιοχών), η ποιότητα και η επάρκεια των εκπαιδευτικών υποδομών και του ανθρώπινου δυναμικού, καθώς και τα κοινωνικοοικονομικά χαρακτηριστικά των τοπικών κοινωνιών. Επιπλέον, σε ορισμένες περιοχές — κυρίως παραθαλάσσιες ή νησιωτικές — πολύ πιθανόν η επαγγελματική στόχευση των υποψηφίων να κατευθύνεται προς τουριστικά επαγγέλματα με ή χωρίς εξειδίκευση, γεγονός που ενδέχεται να επηρεάζει αρνητικά τις επιδόσεις τους. Τέλος, το μέγεθος του πληθυσμού των υποψηφίων συνιστά κρίσιμο παράγοντα, δεδομένου ότι στις μικρότερες Διευθύνσεις το περιορισμένο δείγμα μαθητών μπορεί να οδηγήσει στην εμφάνιση στατιστικά ακραίων τιμών.

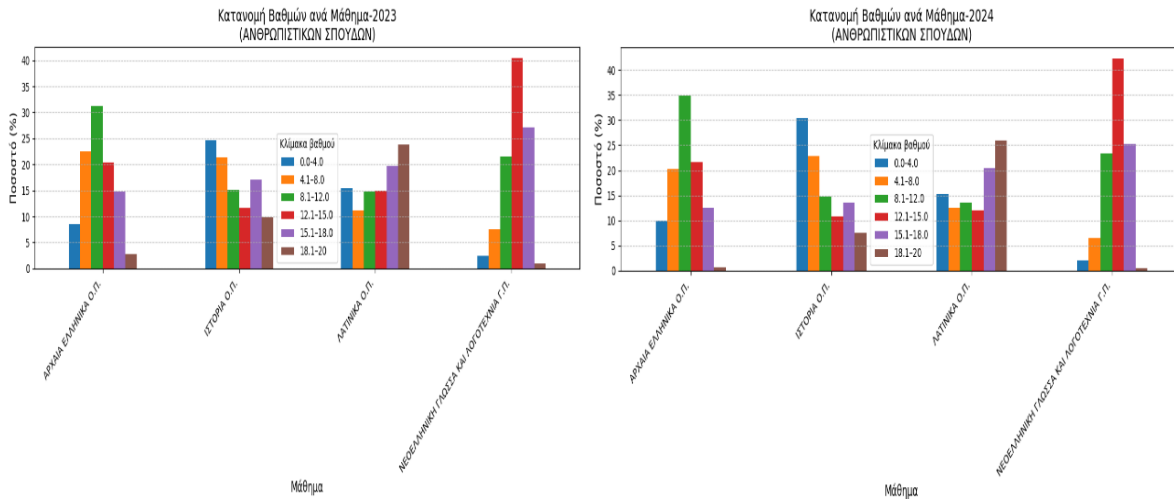
Συζήτηση και συμπεράσματα

Συνολικά, τα αποτελέσματα καταδεικνύουν ότι ορισμένες Διευθύνσεις εμφανίζουν συστηματικά υψηλές επιδόσεις, κυρίως σε αστικά κέντρα και σε ορισμένες περιοχές της ηπειρωτικής χώρας, ενώ άλλες, κυρίως νησιωτικές ή γεωγραφικά απομακρυσμένες, παρουσιάζουν σημαντικά χαμηλότερα ποσοστά επιτυχίας. Η γενική εικόνα αποτυπώνει έναν μέσο όρο επιτυχίας κοντά στο 50%, με ωστόσο

έντονες αποκλίσεις, γεγονός που αναδεικνύει την ανάγκη στοχευμένων εκπαιδευτικών πολιτικών για τη μείωση των περιφερειακών ανισοτήτων.

4.6. Στατιστικά Βαθμών ανά Μάθημα, ανά Ομάδα και ανά Έτος

Τα Γραφήματα 4.6.1 και 4.6.2 παρουσιάζουν την κατανομή των βαθμών ανά Μάθημα και ανά Έτος, για την Ομάδα Ανθρωπιστικών Σπουδών με την εξής διακριτοποίηση: (0, 4, 8, 12, 15, 18, 20).



Γραφήματα 4.6.1 & 4.6.2: Κατανομή Βαθμών Ανθρωπιστικών Σπουδών 2023-2024

Αρχαία Ελληνικά: Η πλειονότητα των μαθητών συγκεντρώνεται στη μεσαία βαθμολογική κλίμακα (8.1–12.0), ενώ οι υψηλές βαθμολογίες (18.1–20) είναι εξαιρετικά λίγες. Δηλώνει μάθημα με μέτρια επίδοση συνολικά.

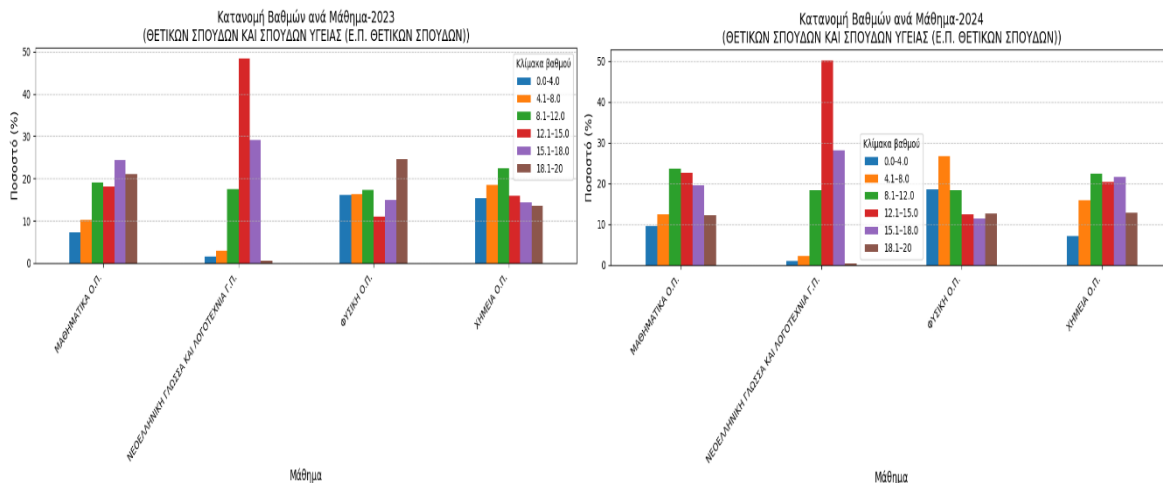
Ιστορία: Εμφανίζει αρκετά χαμηλούς βαθμούς: πάνω από 30% των μαθητών έχει 0.0–4.0, ενώ λίγοι φτάνουν πάνω από 15.0. Πρόκειται για μάθημα με αυξημένη δυσκολία.

Λατινικά: Διαφαίνεται πιο ισορροπημένη κατανομή, με σημαντικό ποσοστό άνω του 15.0 και υψηλότερα ποσοστά στο εύρος 18.1–20. Δείχνει καλύτερη επίδοση σε σχέση με τα υπόλοιπα της Ο.Π.

Νεοελληνική Γλώσσα & Λογοτεχνία: Έχει το υψηλότερο ποσοστό μαθητών στη βαθμολογική ζώνη 12.1–15.0, με λίγες πολύ χαμηλές ή πολύ υψηλές επιδόσεις. Συνεπώς, είναι μάθημα μέσης δυσκολίας.

Συνολικά: Η Ομάδα Ανθρωπιστικών Σπουδών χαρακτηρίζεται από μέτριες προς χαμηλές επιδόσεις, με τα Λατινικά να ξεχωρίζουν θετικά.

Τα Γραφήματα 4.6.3 και 4.6.4 παρουσιάζουν την κατανομή των βαθμών ανά Μάθημα και ανά έτος, για την Ομάδα Θετικών Σπουδών, με την εξής διακριτοποίηση: (0, 4, 8, 12, 15, 18, 20).



Γραφήματα 4.6.3 & 4.6.4: Κατανομή Βαθμών Θετικών Σπουδών 2023-2024

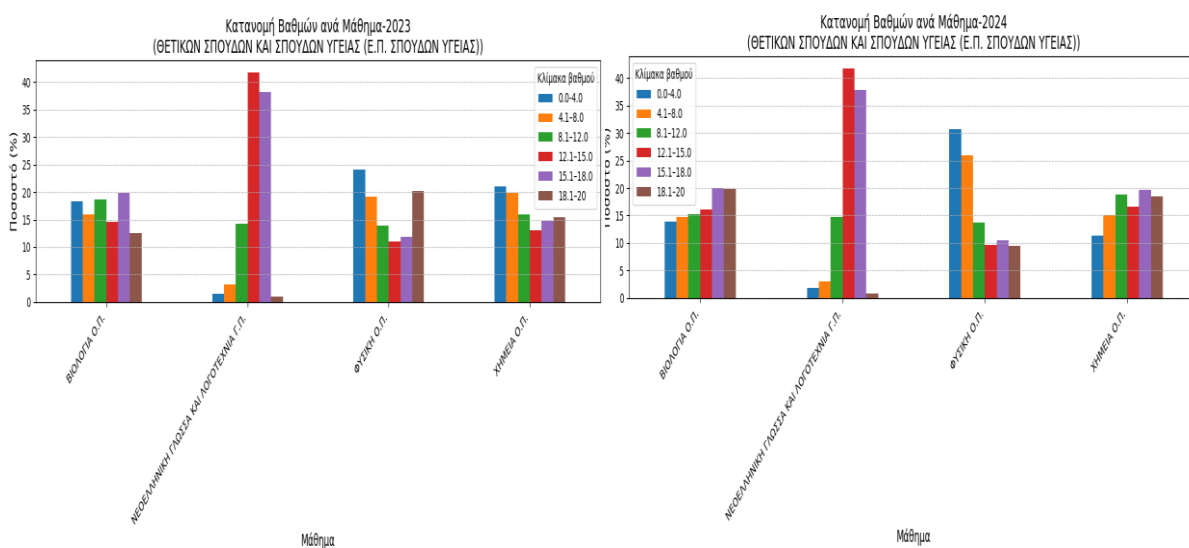
Μαθηματικά: Κυριαρχεί η μεσαία βαθμολογική ζώνη (8.1–12.0), αλλά με αρκετούς κάτω του 8.0, δείγμα σχετικά απαιτητικού μαθήματος.

Νεοελληνική Γλώσσα & Λογοτεχνία: Εμφανίζει πολύ μεγάλη συγκέντρωση στη ζώνη 12.1–15.0 (σχεδόν 50%). Είναι το πιο “σταθερό” μάθημα με περιορισμένες ακραίες τιμές.

Φυσική: Παρατηρείται χαμηλή επίδοση, με το μεγαλύτερο ποσοστό κάτω από 8.0 (κυρίως στη ζώνη 4.1–8.0). **Χημεία:** Η κατανομή είναι σχετικά ομοιόμορφη, αλλά με τάση προς τη μεσαία βαθμολογική περιοχή (8.1–15.0).

Συνολικά: Στην Ομάδα Θετικών Σπουδών, τα Μαθηματικά και η Φυσική αποτελούν “δύσκολα” μαθήματα, ενώ η Χημεία και η Νεοελληνική Γλώσσα δείχνουν πιο “βατή” συμπεριφορά.

Τα Γραφήματα 4.6.5 και 4.6.6 παρουσιάζουν την κατανομή των βαθμών ανά Μάθημα και ανά έτος, για την Ομάδα Σπουδών Υγείας, με την εξής διακριτοποίηση: (0, 4, 8, 12, 15, 18, 20).



Γραφήματα 4.6.5 & 4.6.6: Κατανομή Βαθμών Σπουδών Υγείας 2023-2024

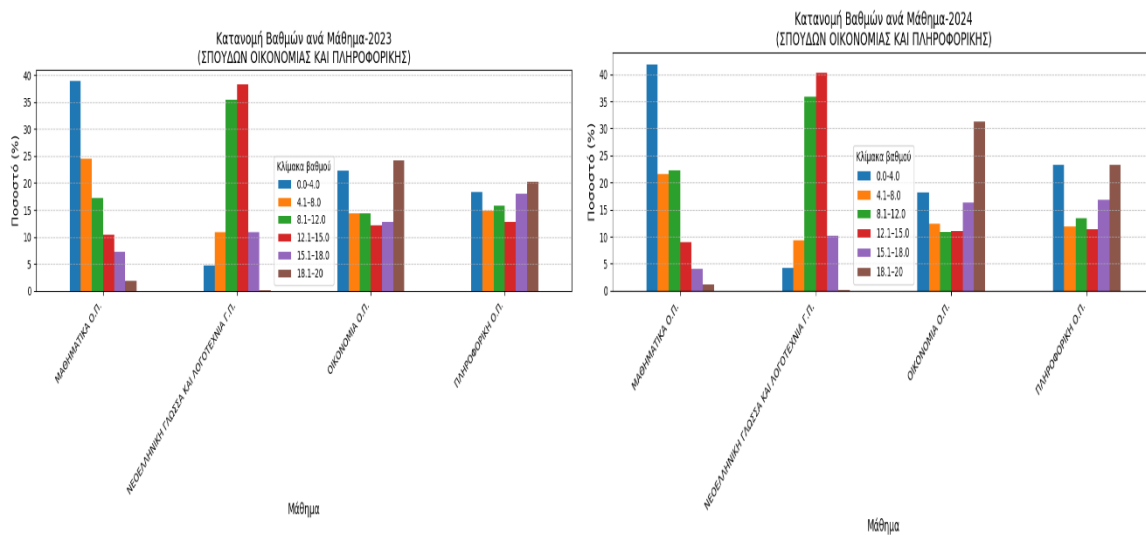
Βιολογία: Παρουσιάζει ομοιόμορφη κατανομή με σχετικά υψηλές επιδόσεις. Οι βαθμοί άνω του 15.0 αποτελούν σημαντικό ποσοστό. Είναι το πιο “δυνατό” μάθημα της Ομάδας.

Νεοελληνική Γλώσσα & Λογοτεχνία: Όπως και στις άλλες Ομάδες, η κατανομή κορυφώνεται στο 12.1–15.0. Φυσική: Εδώ εμφανίζεται έντονη συγκέντρωση χαμηλών βαθμών (0.0–8.0), ιδιαίτερα το 2024. Είναι το δυσκολότερο μάθημα της Ο.Π. Σπουδών Υγείας.

Χημεία: Σχετικά ισορροπημένη με αξιοπρόσεκτο ποσοστό μαθητών άνω του 15.0.

Συνολικά: Οι μαθητές των Σπουδών Υγείας τα πήγαν καλύτερα στη Βιολογία και Χημεία, ενώ η Φυσική παραμένει το “δύσκολο” μάθημα της ομάδας.

Και τέλος τα Γραφήματα 4.6.7 και 4.6.8 παρουσιάζουν την κατανομή των βαθμών ανά Μάθημα και ανά έτος, για την Ομάδα Σπουδών Οικονομίας και Πληροφορικής, με την εξής διακριτοποίηση: (0, 4, 8, 12, 15, 18, 20).



Γραφήματα 4.6.7 & 4.6.8 : Κατανομή Βαθμών Σπουδών Οικονομίας & Πληροφορικής 2023-2024

Μαθηματικά: Παρατηρείται έντονη συγκέντρωση χαμηλών βαθμών (0.0–8.0), η οποία αποτελεί το μεγαλύτερο ποσοστό. Οι βαθμοί άνω του 15.0 είναι ιδιαίτερα περιορισμένοι, γεγονός που δείχνει ότι το μάθημα παρουσιάζει αυξημένη δυσκολία για τους υποψηφίους.

Νεοελληνική Γλώσσα & Λογοτεχνία Γ.Π.: Η κατανομή εμφανίζει κορύφωση στις μεσαίες βαθμολογικές κλίμακες (12.1–15.0 και 15.1–18.0). Οι πολύ χαμηλοί βαθμοί είναι λιγότεροι, γεγονός που υποδηλώνει πιο σταθερές και γενικά καλύτερες επιδόσεις σε σύγκριση με τα Μαθηματικά.

Οικονομία.: Οι βαθμολογίες κατανέμονται σχετικά ισορροπημένα, με αισθητό ποσοστό στις υψηλές κλίμακες (15.1–18.0 και 18.1–20). Το μάθημα παρουσιάζει καλή επίδοση για σημαντικό μέρος των μαθητών και συγκαταλέγεται στα πιο «δυνατά» της ομάδας.

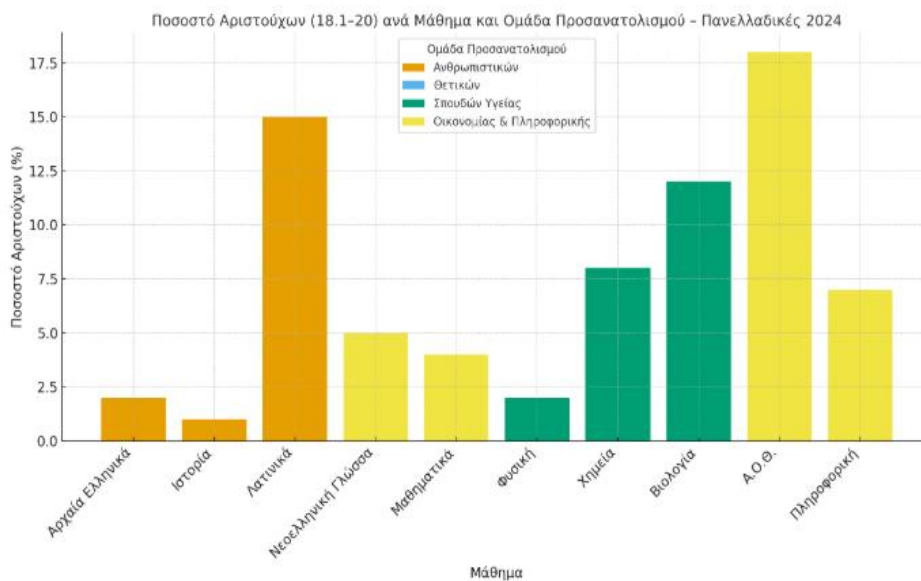
Πληροφορική: Εμφανίζει ιδιαίτερα θετική εικόνα, με υψηλό ποσοστό βαθμών άνω του 15.0 και κάτω του 4.0 αλλά και περιορισμένη συγκέντρωση στις μεσαίες βαθμολογίες. Πρόκειται για μάθημα στο οποίο γενικά οι υποψήφιοι φαίνεται να ανταποκρίνονται αποτελεσματικά.

Συνολικά: Οι υποψήφιοι της Ομάδας Προσανατολισμού Οικονομίας και Πληροφορικής παρουσιάζουν καλύτερες επιδόσεις στην Πληροφορική και την Οικονομία, ενώ τα Μαθηματικά αποτελούν το πιο απαιτητικό μάθημα της ομάδας. Η Νεοελληνική Γλώσσα & Λογοτεχνία διατηρεί ενδιάμεση και πιο ισορροπημένη εικόνα επιδόσεων.

4.7. Ποσοστά αριστούχων και μαθητών κάτω της βάσης ανά Μάθημα - 2024

Οι Ανθρωπιστικές Σπουδές και οι Θετικές Σπουδές εμφανίζουν τη μεγαλύτερη δυσκολία (πολλοί κάτω από τη βάση). Οι Σπουδές Υγείας παρουσιάζουν καλύτερη συνολική επίδοση. Η Οικονομία & Πληροφορική δείχνει πιο ισορροπημένη κατανομή, με θετικά αποτελέσματα στην Οικονομία και στην Πληροφορική.

Στο Γράφημα 4.7.1. αποτυπώνονται τα ποσοστά αριστούχων υποψηφίων (βαθμολογία 18,1–20) ανά εξεταζόμενο μάθημα και ανά Ομάδα Προσανατολισμού για τις Πανελλαδικές Εξετάσεις του έτους 2024. Η ανάλυση των δεδομένων αναδεικνύει σημαντικές διαφοροποιήσεις τόσο μεταξύ των μαθημάτων όσο και μεταξύ των Ομάδων Προσανατολισμού.



Γράφημα 4.7.1: Ποσοστό Αριστούχων (18-20) ανά Μάθημα 2024

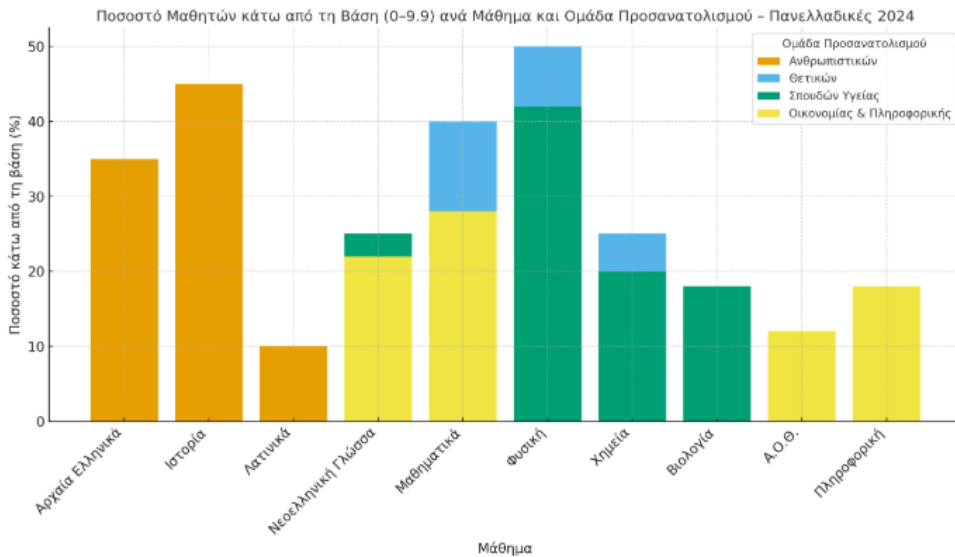
Ειδικότερα, το μάθημα των Αρχών Οικονομικής Θεωρίας (Α.Ο.Θ.) εμφανίζει το υψηλότερο ποσοστό αριστούχων, το οποίο προσεγγίζει το 18%, γεγονός που υποδηλώνει αυξημένο επίπεδο επίδοσης των υποψηφίων ή συγκριτικά χαμηλότερο βαθμό δυσκολίας του εξεταστικού δοκιμίου. Ακολουθούν τα Λατινικά και η Βιολογία, τα οποία παρουσιάζουν επίσης αυξημένα ποσοστά υψηλών βαθμολογιών, αντανακλώντας ικανοποιητική προετοιμασία των υποψηφίων στις αντίστοιχες Ομάδες Προσανατολισμού.

Αντιθέτως, τα Μαθηματικά και η Φυσική καταγράφουν διαχρονικά χαμηλά ποσοστά αριστούχων σε όλες τις Ομάδες Προσανατολισμού, γεγονός που επιβεβαιώνει τον υψηλό βαθμό απαιτήσεων των συγκεκριμένων μαθημάτων. Παράλληλα, τα μαθήματα της Πληροφορικής και της Χημείας εμφανίζουν γενικά ικανοποιητικές επιδόσεις, με ποσοστά αριστείας που κινούνται σε ενδιάμεσα επίπεδα.

Η Νεοελληνική Γλώσσα και Λογοτεχνία παρουσιάζει σχετικά σταθερά αλλά μέτρια ποσοστά αριστούχων, της τάξης του 4%–5%, στοιχείο που υποδηλώνει περιορισμένη συγκέντρωση πολύ υψηλών βαθμολογιών, παρά τη γενικευμένη συμμετοχή όλων των υποψηφίων στο μάθημα.

Όσον αφορά τα ποσοστά χαμηλών επιδόσεων (Γράφημα 4.7.2.), τα μαθήματα της Ιστορίας (Ομάδα Ανθρωπιστικών Σπουδών), των Αρχαίων Ελληνικών και της Φυσικής (Ομάδες Θετικών Σπουδών και Σπουδών Υγείας) καταγράφουν τα υψηλότερα ποσοστά αποτυχίας, με σχεδόν έναν στους δύο

υποψηφίους να βαθμολογείται κάτω από τη βάση. Αντίστοιχα, τα Μαθηματικά στις Ομάδες Θετικών Σπουδών και Σπουδών Οικονομίας και Πληροφορικής εμφανίζουν ιδιαίτερα αυξημένα ποσοστά αποτυχίας, τα οποία κυμαίνονται περίπου μεταξύ 28% και 40%.



Γράφημα 4.7.2: Ποσοστό Μαθητών κάτω από τη βάση (0-9.9) ανά Μάθημα 2024

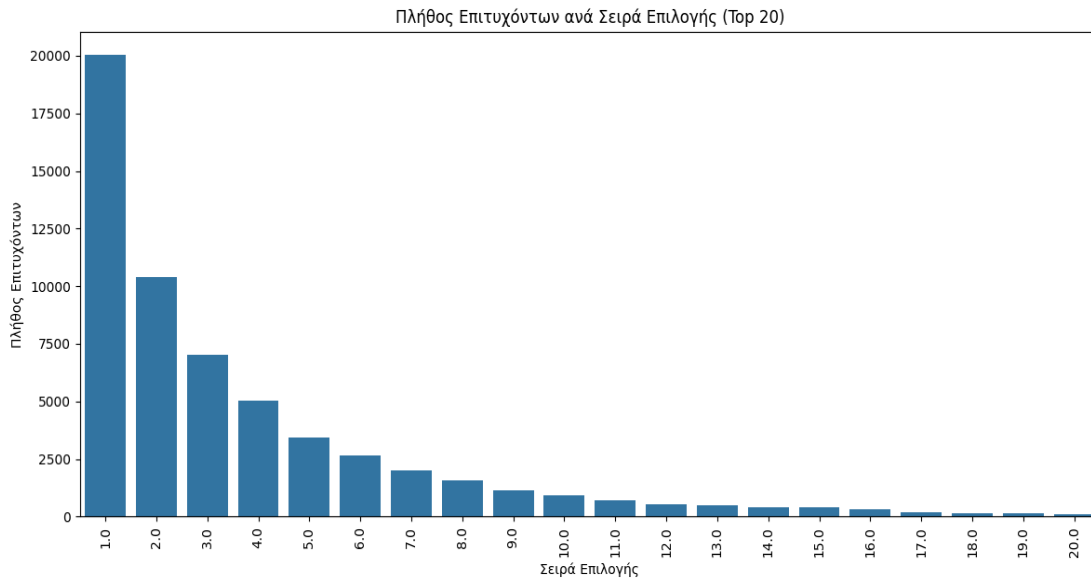
Αντίθετα, τα μαθήματα των Λατινικών και των Αρχών Οικονομικής Θεωρίας παρουσιάζουν τα υψηλότερα ποσοστά επιτυχίας, γεγονός που υποδηλώνει ότι αξιολογούνται ως συγκριτικά «βατά» από τους υποψηφίους, τόσο ως προς τη δομή των θεμάτων όσο και ως προς τις απαιτούμενες γνωστικές δεξιότητες.

4.8. Πλήθος επιτυχόντων ανά σειρά προτίμησης στο μηχανογραφικό του 2024

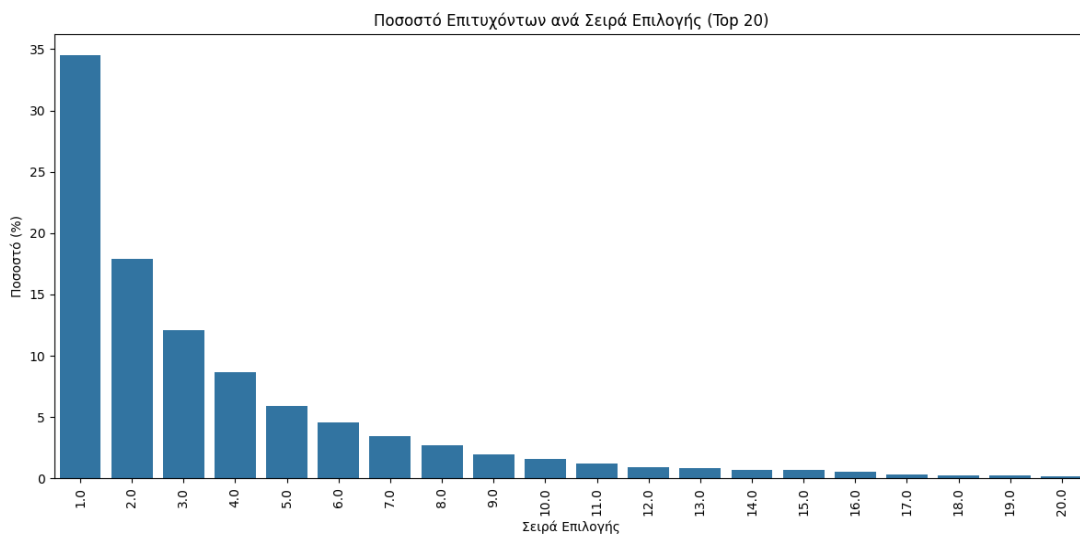
Στα Γραφήματα 4.8.1 και 4.8.2 παρουσιάζεται η κατανομή των επιτυχόντων υποψηφίων των Πανελλαδικών Εξετάσεων ως προς τη σειρά προτίμησης με την οποία δηλώθηκαν οι σχολές εισαγωγής στο μηχανογραφικό δελτίο (πρώτες 20 σειρές προτίμησης). Η κατανομή αυτή αναδεικνύει σαφή συγκέντρωση της επιτυχίας στις ανώτερες δηλωθείσες προτιμήσεις.

Ειδικότερα, παρατηρείται ότι ιδιαίτερα υψηλό ποσοστό υποψηφίων εισάγεται στη σχολή πρώτης προτίμησης, ενώ ήδη από τη δεύτερη και τρίτη επιλογή καταγράφεται απότομη μείωση των ποσοστών εισαγωγής. Καθώς η σειρά προτίμησης αυξάνεται, τα ποσοστά επιτυχίας μειώνονται σταδιακά και τείνουν σε οριακές τιμές, γεγονός που υποδηλώνει ότι η εισαγωγή σε σχολές χαμηλότερης προτεραιότητας αποτελεί σαφώς λιγότερο συχνό ενδεχόμενο.

Η παρατηρούμενη κατανομή υποδηλώνει ότι σημαντικό ποσοστό των υποψηφίων επιτυγχάνει την εισαγωγή του σε σχολές που ανταποκρίνονται στους αρχικούς ακαδημαϊκούς στόχους και τις προσωπικές του προτιμήσεις. Παράλληλα, τα ευρήματα αναδεικνύουν τον έντονα ανταγωνιστικό χαρακτήρα του συστήματος εισαγωγής στην τριτοβάθμια εκπαίδευση, στο οποίο η επιτυχία συνδέεται στενά με την επίτευξη υψηλών βαθμολογικών επιδόσεων που επιτρέπουν την κάλυψη των πρώτων επιλογών στο μηχανογραφικό δελτίο.



Γράφημα 4.8.1: Πλήθος Επιτυχόντων ανά Σειρά Προτίμησης (Top 20 προτιμήσεις)

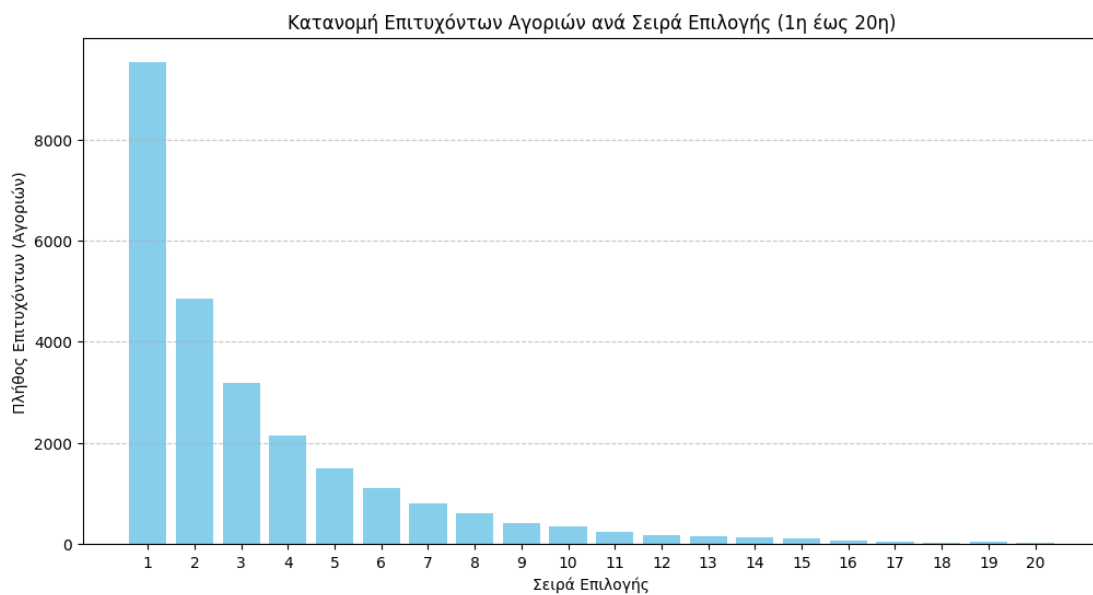


Γράφημα 4.8.2: Ποσοστό Επιτυχόντων ανά Σειρά Προτίμησης (Top 20)

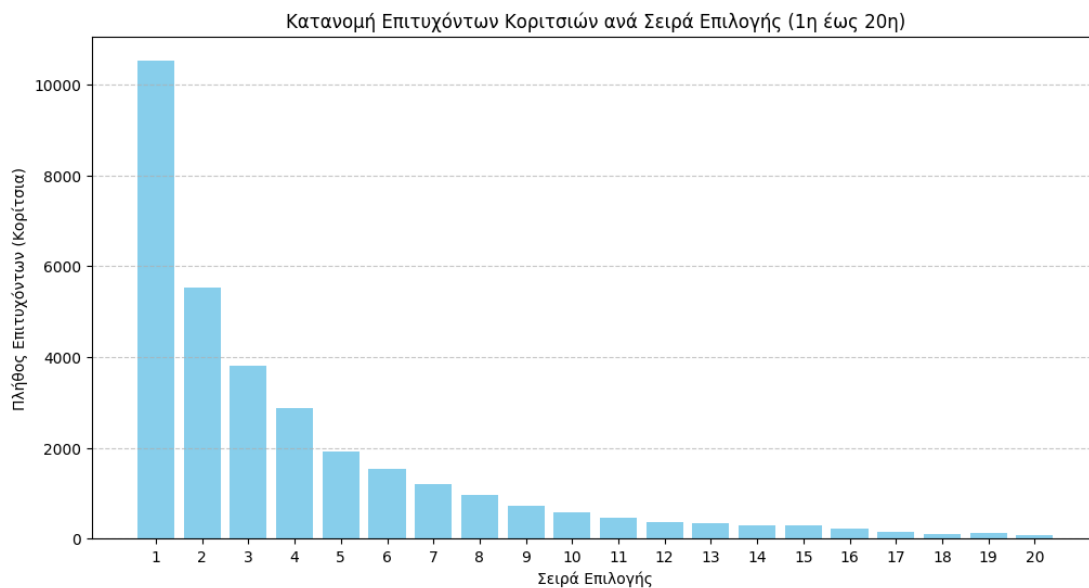
4.9. Πλήθος επιτυχόντων ανά σειρά προτίμησης και ανά Φύλο

Από την ανάλυση των Γραφημάτων 4.9.1 και 4.9.2, προκύπτει ότι οι ράβδοι που αντιστοιχούν στις μαθήτριες εμφανίζονται συστηματικά υψηλότερες στις πρώτες σειρές προτίμησης, γεγονός που υποδηλώνει ότι αναλογικά μεγαλύτερο ποσοστό κοριτσιών επιτυγχάνει την εισαγωγή του στις αρχικές επιλογές του μηχανογραφικού δελτίου. Επιπλέον, μπορεί να υποδηλώνει πιο ρεαλιστική και στοχευμένη στρατηγική επιλογών εκ μέρους των μαθητριών.

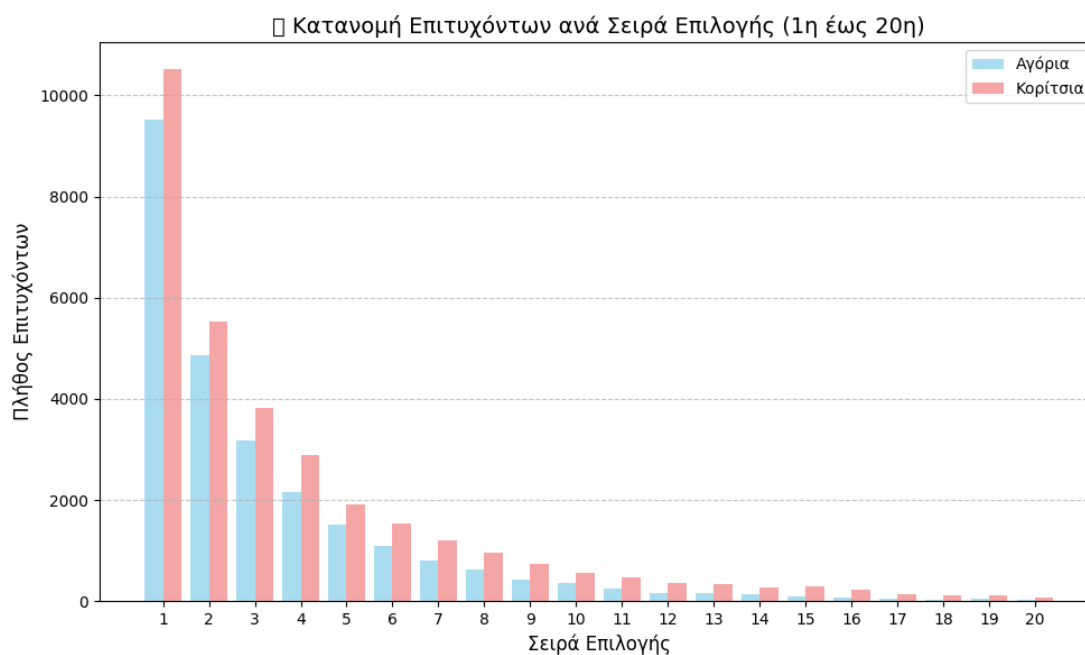
Με βάση τα στατιστικά δεδομένα της ανάλυσης και υπό την παραδοχή ότι οι ίδιες συμπεριφορές επιλογών και επιδόσεων θα διατηρηθούν και τα επόμενα έτη, εκτιμάται ότι το ποσοστό επιτυχίας των κοριτσιών θα συνεχίσει να αυξάνεται με ήπιο ρυθμό, της τάξης του 2% ετησίως. Σε αυτό το πλαίσιο, αναμένεται η διατήρηση και ενίσχυση του προβαδίσματος των μαθητριών, με τις διαφορές στις πρώτες σειρές προτίμησης (1η–3η επιλογή) να διευρύνονται ελαφρώς.



Γράφημα 4.9.1: Πλήθος Επιτυχόντων Αγοριών ανά Σειρά Προτίμησης (Τop 20 επιλογές)



Γράφημα 4.9.2: Πλήθος Επιτυχόντων Κοριτσιών ανά Σειρά Προτίμησης (Τop 20 επιλογές)



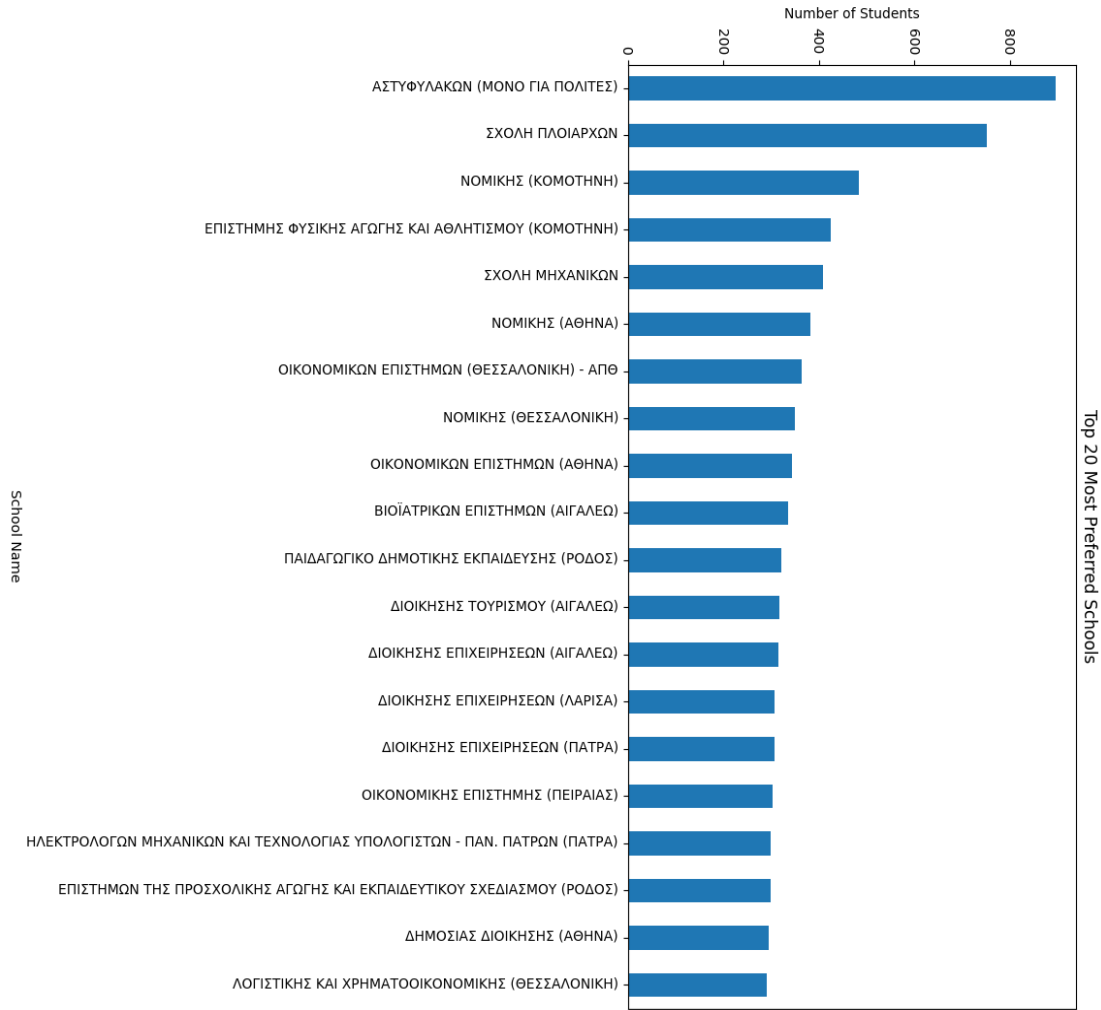
Γράφημα 4.9.3: Κατανομή Επιτυχόντων Αγοριών-Κοριτσιών ανά Σειρά Προτίμησης(1η-20^η)

Συνολικά, προβλέπεται ότι το αθροιστικό ποσοστό επιτυχίας των κοριτσιών στις πρώτες είκοσι σειρές προτίμησης θα αυξηθεί κατά περίπου 6% εντός της επόμενης τριετίας, γεγονός που υποδηλώνει ότι το χάσμα επιτυχιών υπέρ του γυναικείου φύλου ενδέχεται να συνεχίσει να διευρύνεται με την πάροδο του χρόνου, εφόσον δεν μεταβληθούν οι υφιστάμενες συνθήκες και πρακτικές επιλογής.

4.10. Ποιες σχολές είναι υψηλά στις προτιμήσεις των υποψηφίων το 2024

Στο Γράφημα 4.10.1. παρουσιάζονται οι 20 σχολές που συγκεντρώνουν τον μεγαλύτερο αριθμό δηλώσεων προτίμησης από τους υποψηφίους των Πανελλαδικών Εξετάσεων και αποτυπώνονται σαφείς τάσεις ως προς τις εκπαιδευτικές και επαγγελματικές επιδιώξεις τους. Στην κορυφή των προτιμήσεων εμφανίζεται η Σχολή Αστυφυλάκων (μόνο για πολίτες), πιθανώς λόγω της άμεσης επαγγελματικής αποκατάστασης και της εργασιακής σταθερότητας. Ακολουθεί η Σχολή Πλοιάρχων, επιβεβαιώνοντας τη διαχρονική σημασία του ναυτικού επαγγέλματος στην ελληνική κοινωνία αλλά και εδώ την άμεση επαγγελματική αποκατάσταση και την εργασιακή σταθερότητα.

Ιδιαίτερα υψηλή θέση κατέχουν επίσης σχολές Νομικής (Κομοτηνή, Αθήνα, Θεσσαλονίκη), καθώς και σχολές Οικονομικών Επιστημών, γεγονός που αναδεικνύει τη διαχρονική απήχηση επαγγελματιών κύρους και ευρύτερων επαγγελματικών διεξόδων. Παράλληλα, η παρουσία σχολών Διοίκησης Επιχειρήσεων, Δημόσιας Διοίκησης και Τουρισμού υποδηλώνει τη σύνδεση των προτιμήσεων με τις ανάγκες της σύγχρονης οικονομίας και της ελληνικής οικονομικής δραστηριότητας. Τέλος, η συμμετοχή σχολών παιδαγωγικού και επιστημών αγωγής, καθώς και πολυτεχνικών τμημάτων, καταδεικνύει τη διαφοροποίηση των ενδιαφερόντων των υποψηφίων, με το σύνολο των προτιμήσεων να αντανακλά τόσο την αναζήτηση επαγγελματικής ασφάλειας όσο και τη διατήρηση παραδοσιακών ακαδημαϊκών επιλογών.



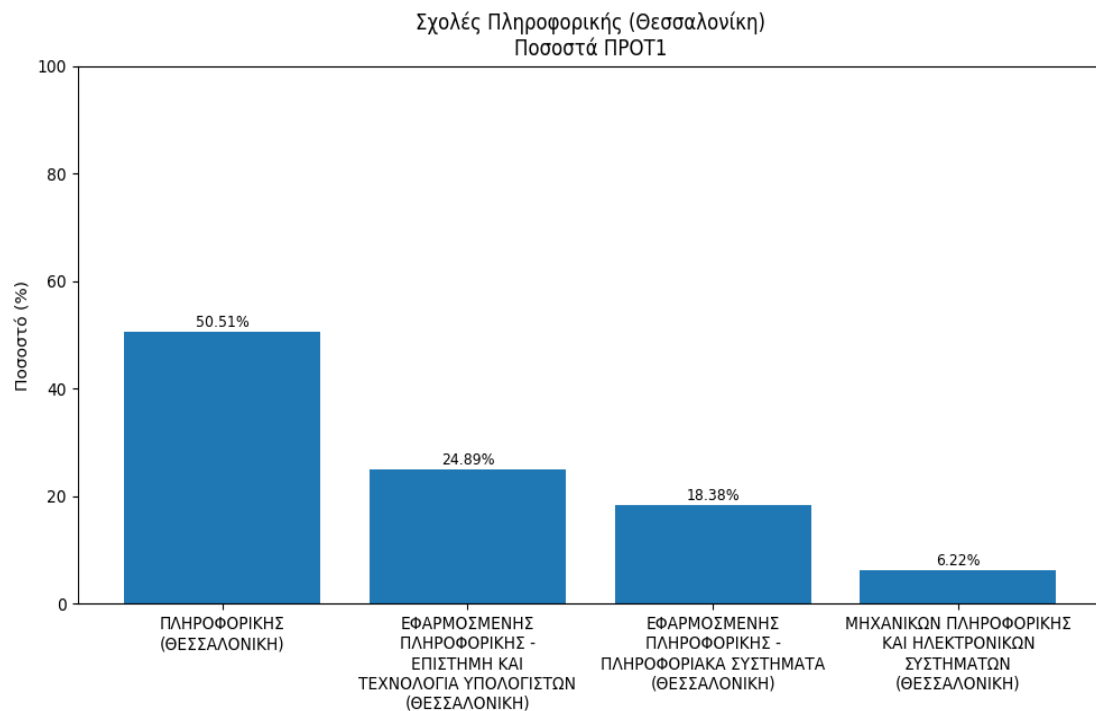
Γράφημα 4.10.1: Κατανομή Σχολών με βάση τις προτιμήσεις των υποψηφίων 2024

4.11. Ποσοστά Δηλώσεων και Επιτυχιών σε Σχολές Πληροφορικής Θεσσαλονίκης

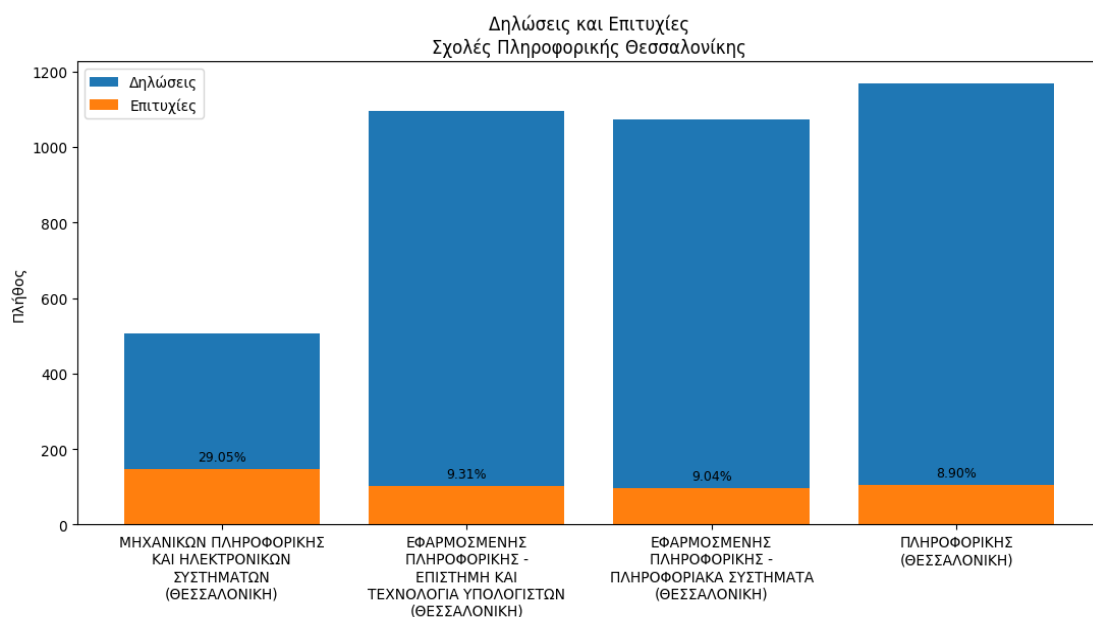
Πίνακας 4.11.1: Πλήθος Δηλώσεων και Επιτυχιών σε Σχολών Πληροφορικής Θεσ/νίκης

ΣΧΟΛΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΘΕΣΣΑΛΟΝΙΚΗΣ	Δηλώσεις	Επιτυχίες
ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑ -ΔΙΠΑΕ	506	147
ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛ. - ΠΑΜΑΚ	1096	102
ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜ. - ΠΑΜΑΚ	1073	97
ΠΛΗΡΟΦΟΡΙΚΗΣ (ΑΠΘ ΘΕΣΣΑΛΟΝΙΚΗ)	1168	104

Σύμφωνα με τον πίνακα 4.11.1 και τα παρουσιαζόμενα Γραφήματα 4.11.1 και 4.11.2 καταδεικνύεται έντονη αναντιστοιχία μεταξύ των δηλωθεισών προτιμήσεων των υποψηφίων και της τελικής επιτυχίας εισαγωγής στις σχολές Πληροφορικής της Θεσσαλονίκης. Ειδικότερα, το Τμήμα Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης συγκεντρώνει το 50,51% των δηλώσεων πρώτης προτίμησης, γεγονός που αποτυπώνει την ιδιαίτερα υψηλή ελκυστικότητα του. Ωστόσο, το ποσοστό επιτυχίας περιορίζεται μόλις στο 8,90%, στοιχείο που υποδηλώνει εξαιρετικά αυξημένο επίπεδο ανταγωνισμού και σαφή αναντιστοιχία μεταξύ ζήτησης και διαθέσιμων θέσεων.



Γράφημα 4.11.1: Ποσοστά 1^{ης} προτίμησης σε Σχολές Πληροφορικής Θεσσαλονίκης



Γράφημα 4.11.2: Δηλώσεις και Επιτυχίες σε Σχολές Πληροφορικής Θεσσαλονίκης

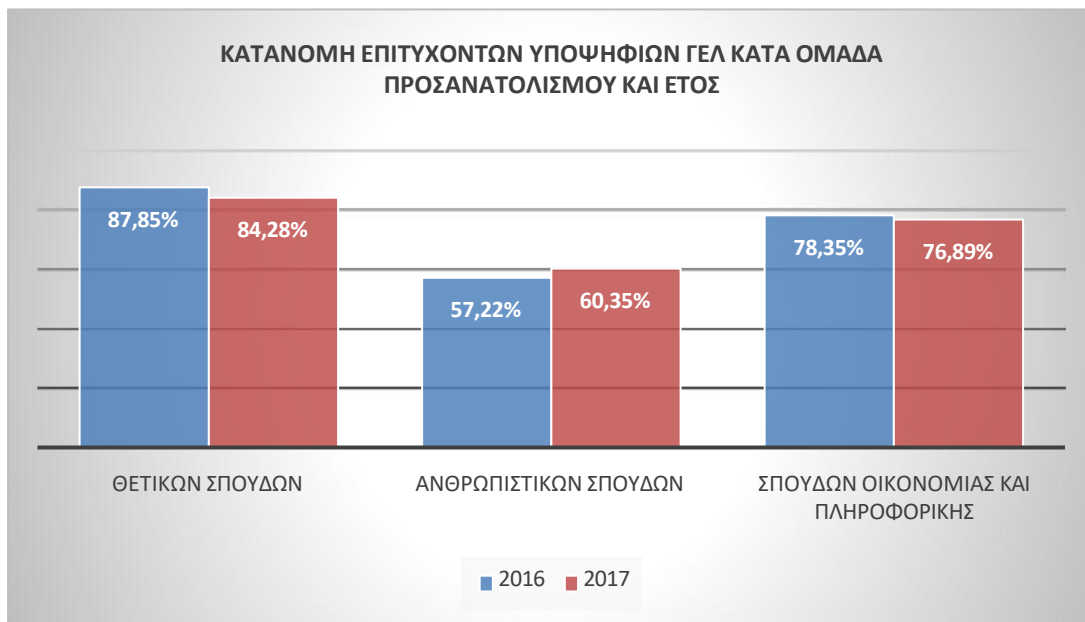
Αντίστοιχα χαμηλά ποσοστά επιτυχίας καταγράφονται και στα Τμήματα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας, με ποσοστά που προσεγγίζουν το 9%, παρά τον υψηλό αριθμό δηλώσεων προτίμησης. Η εικόνα αυτή ενισχύει την υπόθεση έντονου κορεσμού των συγκεκριμένων τμημάτων από πλευράς υποψηφίων.

Αντιθέτως, το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, αν και δηλώνεται ως πρώτη προτίμηση από μόλις το 6,22% των υποψηφίων, παρουσιάζει σημαντικά υψηλότερο ποσοστό επιτυχίας, της τάξης του 29,05%. Το εύρημα αυτό αναδεικνύει τη σημασία της στρατηγικής δήλωσης σχολών στο μηχανογραφικό δελτίο και υποδηλώνει ότι επιλογές με χαμηλότερη δημοφιλία μπορούν να προσφέρουν αυξημένες πιθανότητες εισαγωγής σε συναφές επιστημονικό αντικείμενο.

Συνολικά, τα αποτελέσματα υπογραμμίζουν την ανάγκη ενημερωμένης και ρεαλιστικής προσέγγισης κατά τη διαδικασία δήλωσης προτιμήσεων, λαμβάνοντας υπόψη όχι μόνο το κύρος ή τη δημοφιλία των σχολών, αλλά και τις πραγματικές πιθανότητες επιτυχίας που αυτές συνεπάγονται.

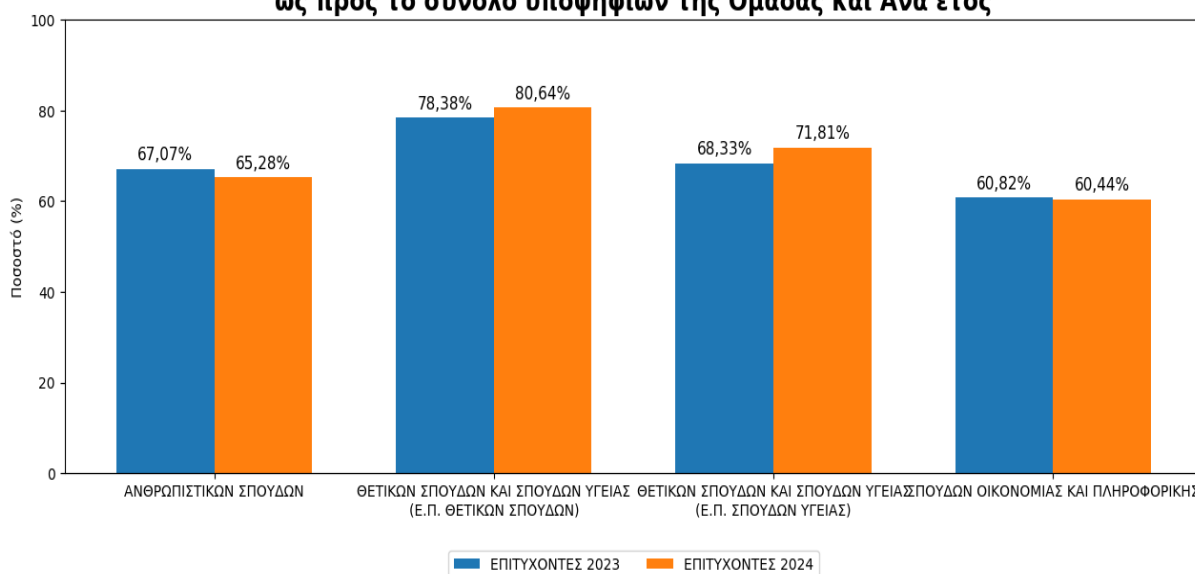
4.12. Σύγκριση διετιών 2016-2017 και 2023-2024

Εξετάζοντας τα στατιστικά στοιχεία (Γράφημα 4.12.1.) της διπλωματικής εργασίας της φοιτήτριας Κυπρίλλης Αναστασίας, του Μεταπτυχιακού Προγράμματος Σπουδών «Ευφυείς Τεχνολογίες στο Διαδίκτυο», με θέμα «Ανάλυση Δεδομένων Πανελλαδικών Εξετάσεων και Εξόρυξη Γνώσης» για τα έτη 2016–2017 και συγκρίνοντας τα με τα στατιστικά στοιχεία των ετών 2023–2024 (Γράφημα 4.12.2.), προκύπτουν τα ακόλουθα συγκριτικά αποτελέσματα:



Γράφημα 4.12.1: Ποσοστά Επιτυχόντων Υποψηφίων ΓΕΛ έτη 2016-2017

**Ποσοστά Επιτυχόντων Ανά Ομάδα Προσανατολισμού
ως προς το σύνολο υποψηφίων της Ομάδας και Ανά έτος**



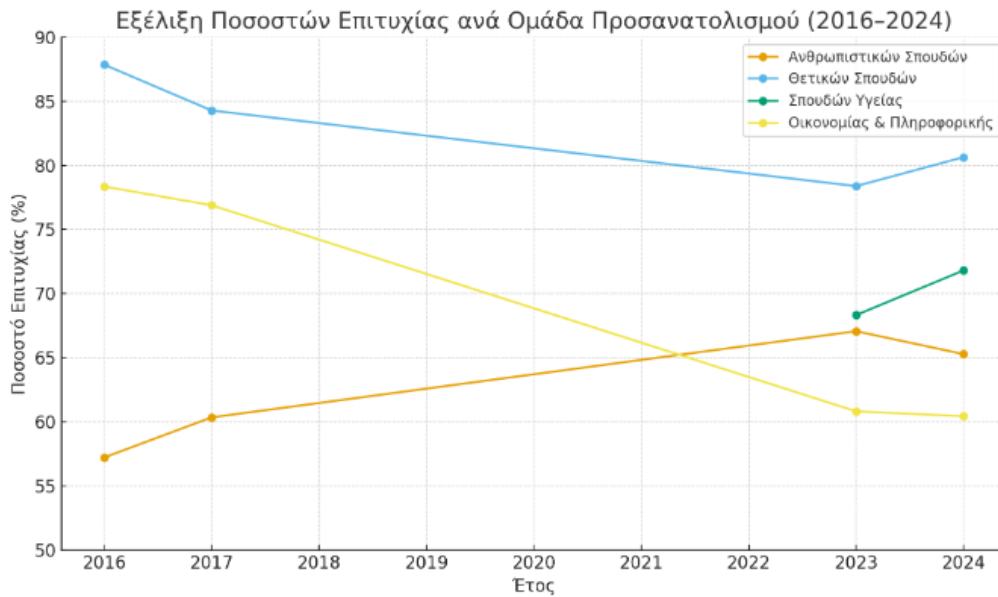
Γράφημα 4.12.2: Ποσοστά Επιτυχόντων Υποψηφίων ΓΕΛ έτη 2023-2024

Η περίοδος 2016–2017 χαρακτηρίζεται γενικά από υψηλότερα ποσοστά επιτυχίας, ειδικά στις Θετικές Σπουδές (όπου ανήκουν και οι υποψήφιοι των Σπουδών Υγείας). Η περίοδος 2023–2024 δείχνει διαφοροποίηση των επιστημονικών πεδίων, με μικρότερες διαφορές μεταξύ ομάδων και τάση σταθεροποίησης.

Οι Θετικές Σπουδές παραμένουν η πιο επιτυχημένη κατηγορία, ενώ οι Οικονομικές και Πληροφορικής δείχνουν σαφή πτωτική πορεία από ~78% σε ~60%. Οι Ανθρωπιστικές Σπουδές έχουν βελτιωθεί σημαντικά σε σχέση με το 2016, πλησιάζοντας τις άλλες ομάδες.

Πίνακας 4.12.1: Ποσοστά επιτυχίας ανά Ομάδα Προσανατολισμού και ανά Έτος

ΜΕΣΟΙ ΟΡΟΙ ΠΟΣΟΣΤΩΝ ΕΠΙΤΥΧΙΑΣ ΑΝΑ ΕΤΟΣ				
ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ	2016	2017	2023	2024
ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ	57,22	60,35	67,07	65,28
ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ	78,35	76,89	60,82	60,44
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ	87,85	84,28	78,38	80,64
ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ			68,33	71,81
ΜΕΣΟΙ ΟΡΟΙ ΠΟΣΟΣΤΩΝ	74,47	73,84	68,65	69,54



Γράφημα 4.12.3: Εξέλιξη Ποσοστών ανά Ομάδα Προσανατολισμού (2016-2024)

Από την εξέταση του Πίνακα 4.12.1 και του Γραφήματος 4.12.3 διαπιστώνεται μια αξιοσημείωτη πτωτική τάση των σχετικών ποσοστών σε βάθος χρόνου, με μείωση από περίπου 74% το έτος 2016 σε περίπου 70% κατά μέσο όρο το 2024. Η παρατηρούμενη μείωση των ποσοστών επιτυχίας μετά το 2017 και κυρίως κατά την περίοδο 2023–2024 μπορεί να αποδοθεί σε συνδυασμό θεσμικών και εκπαιδευτικών παραγόντων.

Καταρχάς, η εισαγωγή της **Ελάχιστης Βάσης Εισαγωγής (ΕΒΕ)** λειτούργησε ως μηχανισμός αποκλεισμού σημαντικού αριθμού υποψηφίων, ανεξάρτητα από τη σχετική κατάταξή τους, καθώς απαιτείται πλέον η επίτευξη ελάχιστου βαθμολογικού ορίου για κάθε επιστημονικό πεδίο και τμήμα. Αυτό είχε ως άμεση συνέπεια τη μείωση των εισακτέων και, κατ' επέκταση, των συνολικών ποσοστών επιτυχίας.

Παράλληλα, η εφαρμογή διαφορετικών **συντελεστών βαρύτητας** στα εξεταζόμενα μαθήματα ενίσχυσε τις επιπτώσεις χαμηλών επιδόσεων σε συγκεκριμένα γνωστικά αντικείμενα, ιδίως στις Ομάδες Προσανατολισμού Οικονομίας & Πληροφορικής και Θετικών Σπουδών, όπου παρατηρείται η εντονότερη πτώση. Έτσι, ακόμη και μικρές διαφοροποιήσεις στις επιδόσεις οδηγούν πλέον σε δυσανάλογες απώλειες μορίων.

Επιπλέον, οι συνεχείς αλλαγές στο σύστημα εισαγωγής και στη δομή των εξετάσεων δημιούργησαν συνθήκες αβεβαιότητας τόσο για τους μαθητές όσο και για το εκπαιδευτικό προσωπικό, επηρεάζοντας αρνητικά τη στρατηγική προετοιμασίας και τις εκπαιδευτικές επιλογές των υποψηφίων.

Τα δεδομένα καταδεικνύουν επίσης διαφοροποιημένες επιπτώσεις ανά Ομάδα Προσανατολισμού: ενώ στις Ανθρωπιστικές Σπουδές παρατηρείται σχετική σταθερότητα με ήπιες διακυμάνσεις, στις Ομάδες Οικονομίας & Πληροφορικής και Θετικών Σπουδών καταγράφεται σαφής πτωτική τάση, γεγονός που υποδηλώνει ότι οι θεσμικές αλλαγές δεν επηρέασαν ομοιόμορφα όλα τα επιστημονικά πεδία.

Συνεπώς, η συνολική πτωτική πορεία των ποσοστών επιτυχίας δεν φαίνεται να οφείλεται αποκλειστικά στη μεταβολή των επιδόσεων των υποψηφίων, αλλά κυρίως στη μετατόπιση του πλαισίου αξιολόγησης και επιλογής, με την ΕΒΕ και τους συντελεστές βαρύτητας να λειτουργούν ως καθοριστικοί παράγοντες αναδιαμόρφωσης των αποτελεσμάτων.

Κεφάλαιο 5^ο: Συσταδοποίηση

5.1. Προεργασία δεδομένων

Για την καλύτερη εφαρμογή των αλγορίθμων συσταδοποίησης στο ήδη δημιουργημένο αρχείο ΦΥΛΟ_ΑΠΟΤΕΛΕΣΜΑΤΑ_ΒΑΘΜΟΙ με **415.351** γραμμές και με τις εξής στήλες:

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΦΥΛΟ
3. Δ/ΝΣΗ Δ.Ε.
4. ΕΙΔΟΣ ΛΥΚΕΙΟΥ
5. ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ
6. ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ-ΕΙΔΙΚΟΤΗΤΑ
7. ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ
8. ΕΠΙΤΥΧΩΝ
9. ΜΑΘΗΜΑ
10. ΒΑΘΜΟΣ

Εφαρμόστηκαν περεταίρω επεξεργασίες όπως α) με τη βοήθεια του φίλτρου στο excel επιλέχθηκαν μόνο οι υποψήφιοι του ΓΕΛ και στη συνέχεια διαγράφηκαν οι στήλες: ΦΥΛΟ, Δ/ΝΣΗ, ΕΙΔΟΣ ΛΥΚΕΙΟΥ, ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ, ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ, ΕΠΙΤΥΧΩΝ που δεν ήταν απαραίτητες για την εφαρμογή του clustering. Στη συνέχεια φιλτραρίστηκαν οι Ομάδες Προσανατολισμού και δημιουργήθηκαν 4 διαφορετικά αρχεία. Με κώδικα rython σε κάθε ένα αρχείο δημιουργήθηκε μία στήλη για το κάθε ένα από τα 4 βασικά μαθήματα της Ομάδας και μία γραμμή για κάθε κωδικό υποψηφίου. Με αποτέλεσμα για κάθε Ομάδα Προσανατολισμού να δημιουργηθούν τα παρακάτω αρχεία (βλέπε Σχήμα 5.1.) με τις εξής στήλες:

ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ
3. ΙΣΤΟΡΙΑ
4. ΛΑΤΙΝΙΚΑ
5. ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ

ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ

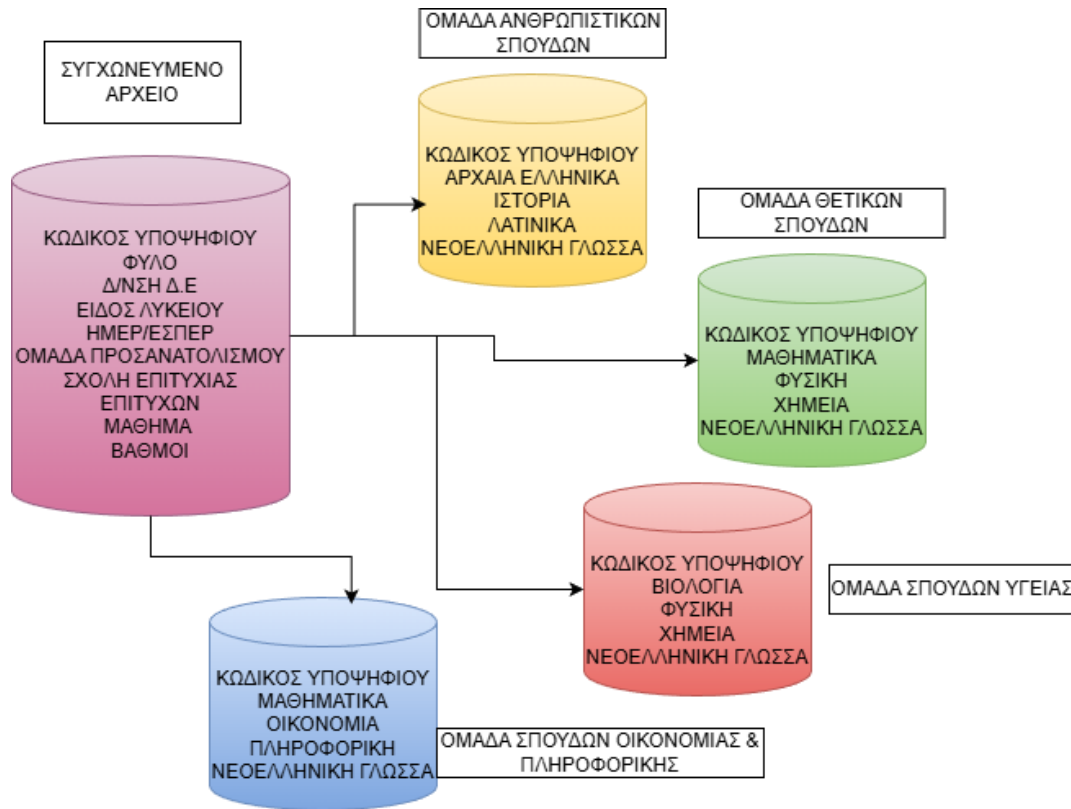
1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΜΑΘΗΜΑΤΙΚΑ
3. ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ
4. ΦΥΣΙΚΗ
5. ΧΗΜΕΙΑ

ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΒΙΟΛΟΓΙΑ
3. ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ
4. ΦΥΣΙΚΗ
5. ΧΗΜΕΙΑ

ΣΠΟΥΔΕΣ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ

1. ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ
2. ΜΑΘΗΜΑΤΙΚΑ
3. ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ
4. ΟΙΟΝΟΜΙΑ
5. ΠΛΗΡΟΦΟΡΙΚΗ



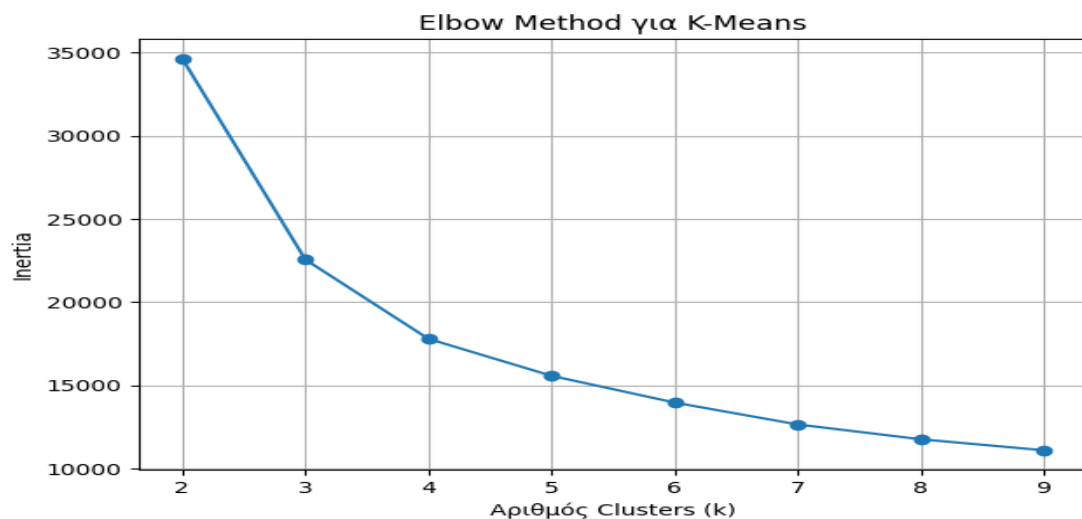
Σχήμα 5.1.: Δημιουργία αρχείων κατάλληλων για Συσταδοποίηση

Πολύ σημαντικό να επισημανθεί ότι πριν εφαρμοστούν οι αλγόριθμοι συσταδοποίησης καθαρίστηκαν τα δεδομένα από εγγραφές υποψηφίων που δεν είχαν κανένα βαθμό και στα 4 μαθήματα. Στη συνέχεια με Γλώσσα Προγραμματισμού Python εφαρμόστηκαν οι αλγόριθμοι k-means και BDSCAN.

5.2. Αλγόριθμος k-means / Elbow

ΟΜΑΔΑ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ

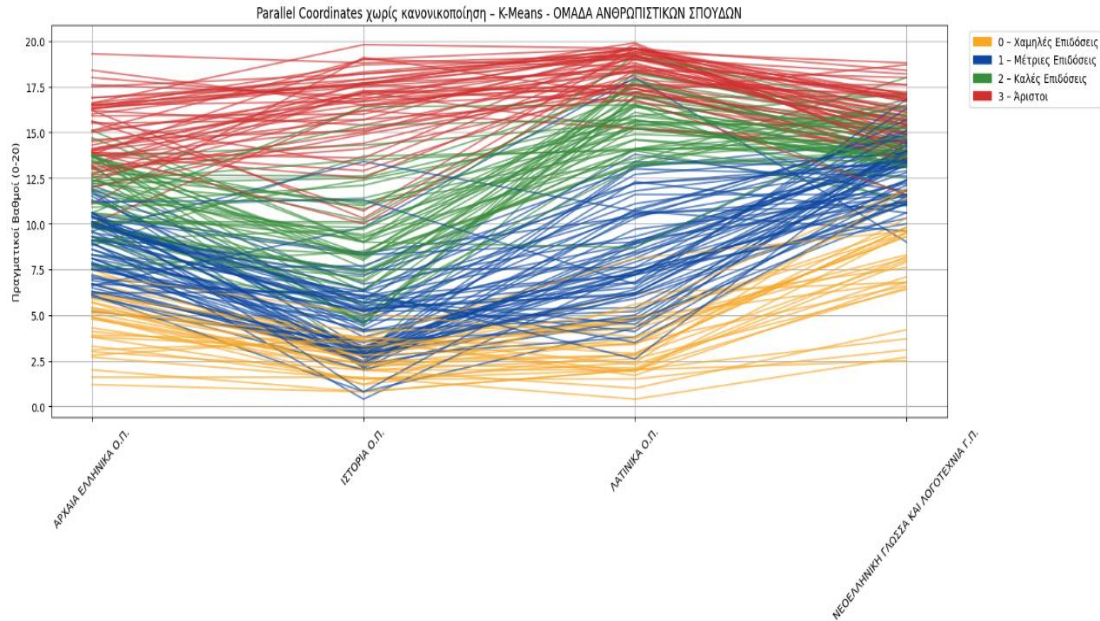
Στα δεδομένα της Ομάδας Ανθρωπιστικών Σπουδών ο αλγόριθμος k-means εντόπισε 4 διακριτές ομάδες με τη μέθοδο elbow για k-means όπως φαίνεται και στο Γράφημα 5.2.1.



Γράφημα 5.2.1: Elbow k-means (Ανθρωπιστικών Σπουδών)

Πίνακας 5.2.1: Μέσοι όροι χαρακτηριστικών ανά cluster (k-Means –Ανθρωπιστικών Σπουδών)

KMeans_ cluster	ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.	ΙΣΤΟΡΙΑ Ο.Π.	ΛΑΤΙΝΙΚΑ Ο.Π.	ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.	ΠΛΗΘΟΣ	Χαρακτη- ρισμός
0	4.45	2.56	3.36	8.18	3026	Χαμηλές επιδόσεις
1	8.55	4.59	8.30	12.63	5750	Μέτριοι
2	11.77	9.03	15.04	14.44	6080	Καλοί
3	15.24	16.73	18.: 53	15.93	5771	Άριστοι



Γράφημα 5.2.2: Clusters k-means with Parallel Coordinates (Ανθρωπιστικών Σπουδών)

Σύμφωνα με τον Πίνακα 5.2.1 η κατανομή είναι σχετικά ισορροπημένη (εκτός από το cluster 0 που είναι μικρότερο), άρα ο k-means κατάφερε να εντοπίσει **τέσσερις σταθερές υπο-ομάδες επιδόσεων**, οι οποίες διακρίνονται και χρωματικά στο Γράφημα 5.2.2 του **Parallel Coordinates**.

Προφίλ των clusters:

Cluster 0 (3026 μαθητές) – Χαμηλές επιδόσεις: Το πιο αδύναμο cluster, εξαιρετικά χαμηλές επιδόσεις σε Αρχαία–Ιστορία–Λατινικά, και μέτρια προς χαμηλή Νεοελληνική. Αυτό είναι και το μικρότερο cluster — λογικό, λίγοι γράφουν τόσο χαμηλά.

Cluster 1 (5750 μαθητές) – Μέτριοι: Αυτοί είναι οι μαθητές με **μέτριες επιδόσεις**, με σχετικά καλή Νεοελληνική αλλά αδύναμη Ιστορία. Μοιάζουν με «μέσο επίπεδο» μαθητών.

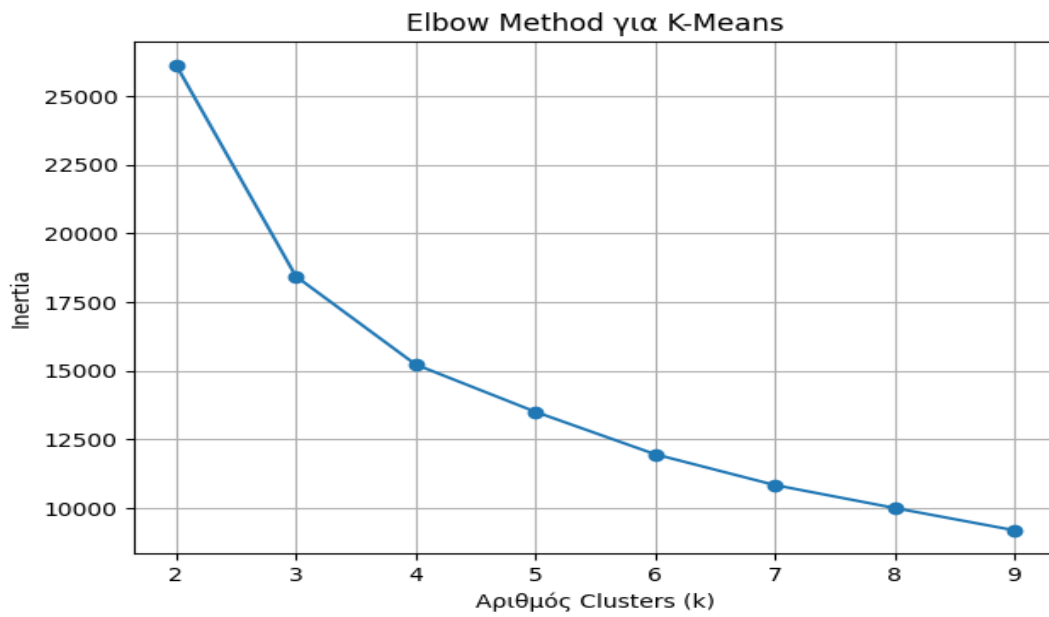
Cluster 2 (6080 μαθητές) – Καλοί, με δύναμη στα Λατινικά: Καλοί μαθητές συνολικά, με **ξεκάθαρη ισχύ στα Λατινικά και Νεοελληνική**, και μέτριες επιδόσεις στην Ιστορία. Αυτό είναι το μεγαλύτερο cluster.

Cluster 3 (5771 μαθητές)– Άριστοι / Πάρα πολύ καλοί: Πρόκειται για **τους άριστους μαθητές** του πεδίου. Εξαιρετικοί στα Λατινικά και Ιστορία, αλλά και στα Αρχαία & Γλώσσα

ΟΜΑΔΑ ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ

Στα δεδομένα της Ομάδας Θετικών Σπουδών ο αλγόριθμος k-means εντόπισε 4 διακριτές ομάδες με τη μέθοδο elbow για k-means όπως φαίνεται και στο Γράφημα 5.2.3.

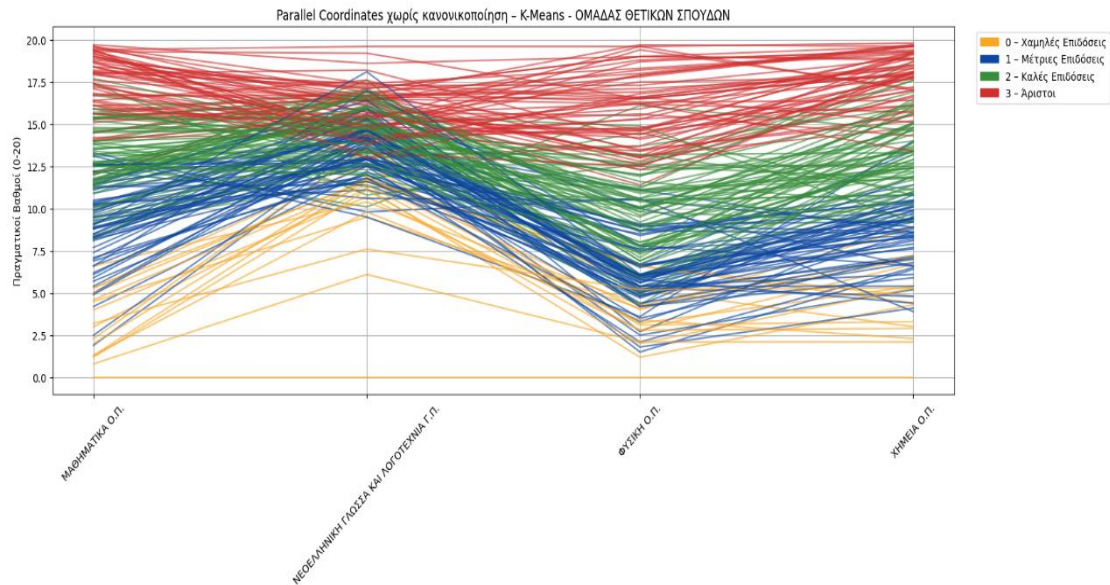
Βέλτιστο k για K-Means: 4



Γράφημα 5.2.3: Elbow/k-means (Θετικών Σπουδών)

Πίνακας 5.2.2: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)- Θετικών Σπουδών

KMeans_ cluster	ΜΑΘΗΜΑΤΙΚΑ Ο.Π.	ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.	ΦΥΣΙΚΗ Ο.Π.	ΧΗΜΕΙΑ Ο.Π.	ΠΛΗΘΟΣ	ΧΑΡΑΚΤΗ- ΡΙΣΜΟΣ
0	3.59	9.63	3.30	4.44	1391	Χαμηλές επιδόσεις
1	8.86	13.54	5.53	8.57	3930	Μέτριοι / καλοί στη Γλώσσα
2	13.31	14.42	9.73	13.80	4524	Πολύ Καλοί
3	17.22	15.63	16.19	17.84	4036	Άριστοι



Γράφημα 5.2.4: Clusters k-means with Parallel Coordinates (Θετικών Σπουδών)

Σύμφωνα με τον Πίνακα 5.2.2 η κατανομή είναι ισορροπημένη (εκτός από το cluster 0 που είναι μικρότερο), άρα ο k-means κατάφερε να εντοπίσει **τέσσερις σταθερές υπο-ομάδες επιδόσεων**, οι οποίες διακρίνονται και χρωματικά στο Γράφημα 5.2.4 του **Parallel Coordinates**.

Προφίλ των clusters:

Cluster 0 (1391 μαθητές)– Χαμηλές επιδόσεις: Η ομάδα με τις χαμηλότερες επιδόσεις, ιδιαίτερα στα θετικά μαθήματα. Ωστόσο στη Νεοελληνική έχουν κάπως καλύτερες επιδόσεις.

Cluster 1 (3930 μαθητές) – Μέτριο/Καλοί: Μέτριες επιδόσεις στα θετικά μαθήματα, σαφώς καλύτερες στη Νεοελληνική Γλώσσα. Θα μπορούσε να θεωρηθεί ομάδα μαθητών με σαφή προτίμηση/ικανότητα σε θεωρητικά μαθήματα.

Cluster 2 (4524 μαθητές) – Πολύ καλοί: Καλοί έως πολύ καλοί μαθητές. Δυνατοί στα θεωρητικά και χημεία, μέτριοι προς καλούς στη φυσική.

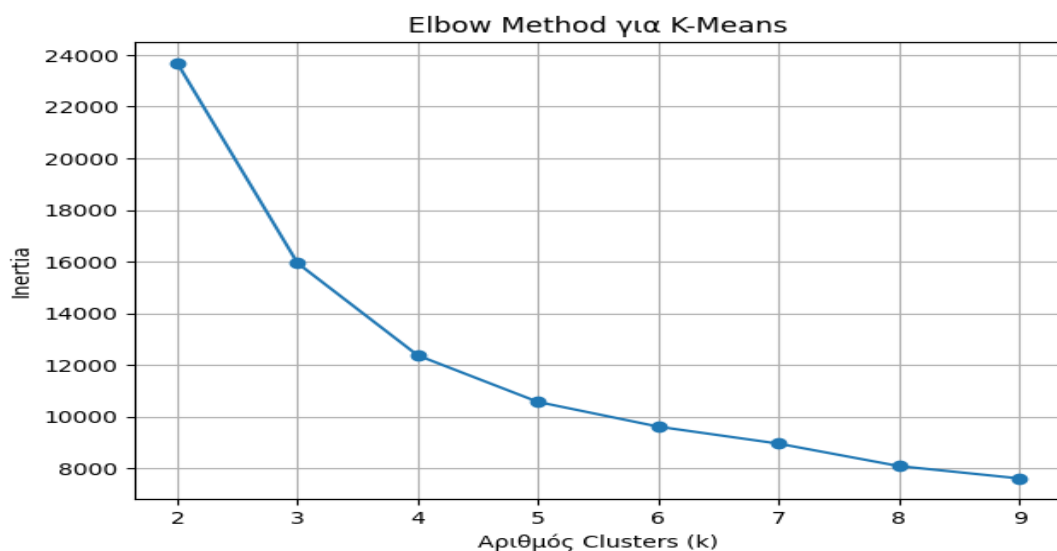
Cluster 3 (4036 μαθητές)– Άριστοι: Πρόκειται για τους αριστούχους, με υψηλές επιδόσεις σε όλα τα αντικείμενα. Πολύ ισορροπημένα και υψηλά σκορ.

Η κατανομή μοιάζει πολύ ρεαλιστική για πανελλαδικές: μεγάλα μεσαία clusters, μικρότερο το cluster 0 των χαμηλών επιδόσεων.

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ

Στα δεδομένα της Ομάδας Σπουδών Υγείας ο αλγόριθμος k-means εντόπισε 4 διακριτές ομάδες με τη μέθοδο elbow για k-means όπως φαίνεται και στο Γράφημα 5.2.5.

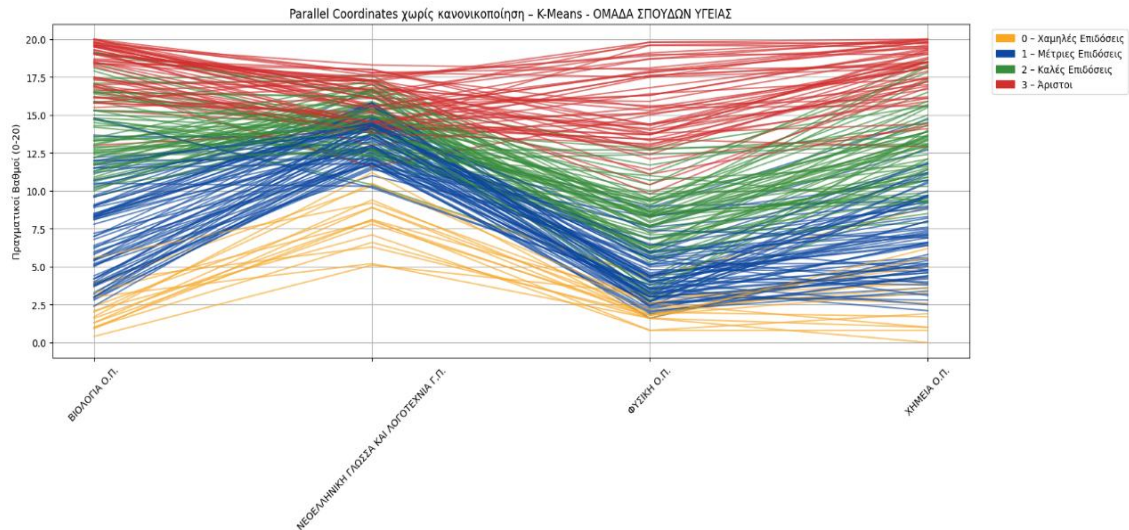
Βέλτιστο K για K-Means: 4



Γράφημα 5.2.5: Elbow/k-means (Σπουδών Υγείας)

Πίνακας 5.2.3: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)- Σπουδών Υγείας

KMeans_ cluster	ΒΙΟΛΟΓΙΑ Ο.Π.	ΝΕΟΕΛΛΗΝ ΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝ ΙΑ Γ.Π.	ΦΥΣΙΚΗ Ο.Π.	ΧΗΜΕΙΑ Ο.Π.	Πλήθος	Χαρακτη- ρισμός
0	3.38	8.11	2.68	3.32	1251	Πολύ χαμηλοί
1	7.53	13.45	4.28	7.46	3735	Μέτριοι/Προς Χαμηλοί
2	14.27	14.96	7.76	13.34	4491	Καλοί
3	18.17	16.10	15.40	18.09	4164	Άριστοι



Γράφημα 5.2.6: Clusters k-means with Parallel Coordinates (Σπουδών Υγείας)

Σύμφωνα με τον Πίνακα 5.2.3 η κατανομή είναι ισορροπημένη (εκτός από το cluster 0 που είναι μικρότερο), άρα ο k-means κατάφερε να εντοπίσει **τέσσερις σταθερές υπο-ομάδες επιδόσεων**, οι οποίες διακρίνονται και χρωματικά στο Γράφημα 5.2.6 του **Parallel Coordinates**.

Προφίλ των clusters:

Cluster 0 (1251 μαθητές)— **Πολύ χαμηλές επιδόσεις (το αδύναμο cluster)**: Πρόκειται για μαθητές που γράφουν **πολύ χαμηλά στα θετικά μαθήματα** και μέτρια στη Γλώσσα. Είναι ξεκάθαρα το "αδύναμο άκρο".

Cluster 1 (3735 μαθητές) – **Μέτριοι / προς χαμηλοί**: Μαθητές με μέτρια θετική επίδοση, αρκετά καλύτεροι στη Γλώσσα παρά στα επιστημονικά μαθήματα.

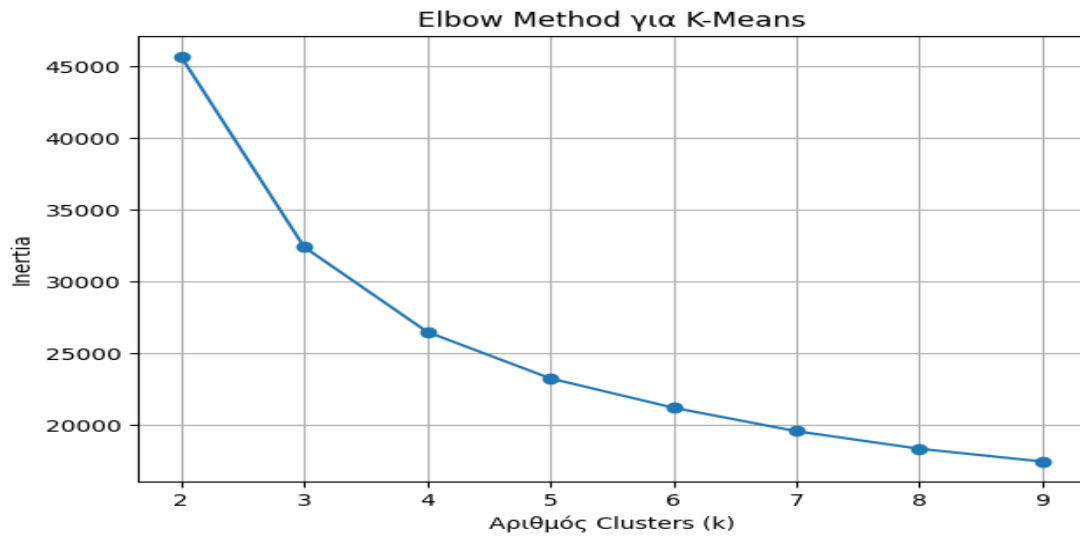
Cluster 2 (4491 μαθητές) – **Καλοί μαθητές**: Πολύ καλοί στη Βιολογία & Χημεία, μέτριοι στη Φυσική, πολύ καλοί στη Νεοελληνική. Αυτό είναι το «μεσαίο-προς-καλό» επίπεδο.

Cluster 3 (4164 μαθητές)— **Άριστοι**: Το **ισχυρό cluster**. Άριστοι σε όλα τα επιστημονικά μαθήματα και υψηλοί στη Γλώσσα. Το πιο ισορροπημένο υψηλής επίδοσης προφίλ.

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Στα δεδομένα της Ομάδας Σπουδών Οικονομίας & Πληροφορικής ο αλγόριθμος k-means εντόπισε 4 διακριτές ομάδες με τη μέθοδο elbow για k-means όπως φαίνεται και στο Γράφημα 5.2.7.

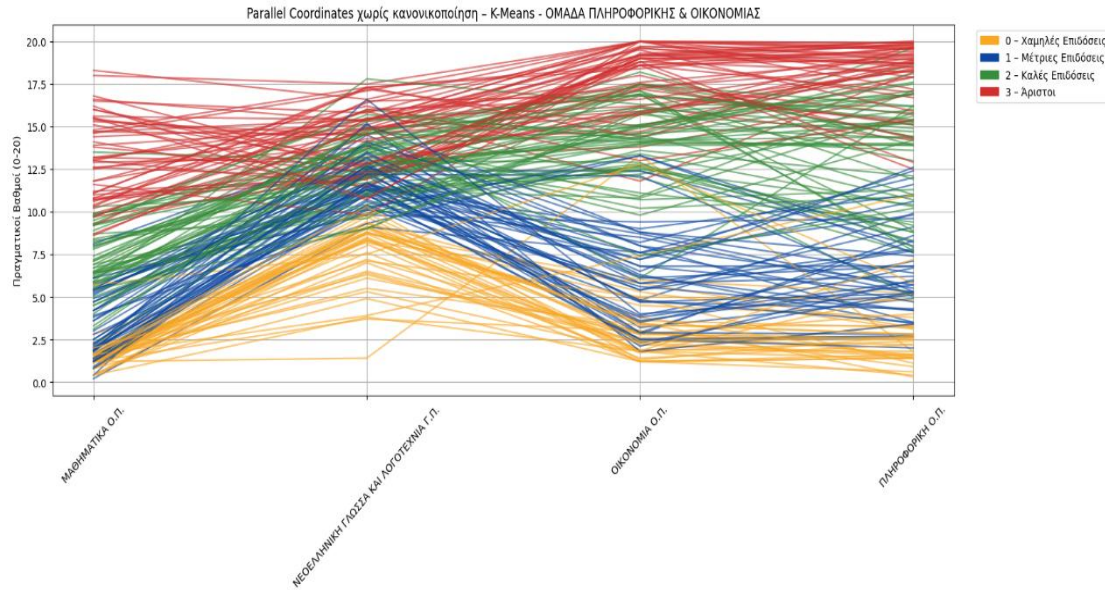
Βέλτιστο k για K-Means: 4



Γράφημα 5.2.7: Elbow/k-means (Σπουδών Οικονομίας & Πληροφορικής)

Πίνακας 5.2.4: Μέσοι όροι χαρακτηριστικών ανά cluster (K-Means)-Σπουδών Οικονομίας & Πληροφορικής

KMeans_ cluster	ΜΑΘΗΜΑΤΙ ΚΑ Ο.Π.	ΝΕΟΕΛΛΗΝΙ ΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙ Α Γ.Π.	ΟΙΚΟΝΟΜ ΙΑ Ο.Π.	ΠΛΗΡΟΦΟΡΙ ΚΗ Ο.Π.	Πλήθος	Χαρακτη- ρισμός
0	1.40	6.94	2.78	2.56	4566	Πολύ αδύναμοι
1	2.96	11.73	6.23	6.41	6742	Μέτριοι
2	6.78	12.82	13.66	13.56	8548	Καλοί
3	12.70	14.56	18.08	18.13	7735	Άριστοι



Γράφημα 5.2.8: Clusters k-means with Parallel Coordinates (Σπουδών Οικονομίας & Πληροφορικής)

Σύμφωνα με τον Πίνακα 5.2.4 η κατανομή είναι ισορροπημένη (εκτός από το cluster 0 που είναι μικρότερο), άρα ο k-means κατάφερε να εντοπίσει **τέσσερις σταθερές υπο-ομάδες επιδόσεων**, οι οποίες διακρίνονται και χρωματικά στο Γράφημα 5.2.8 του **Parallel Coordinates**.

Προφίλ των clusters:

Cluster 0 (4566 μαθητές) – Πολύ χαμηλές επιδόσεις (το αδύναμο cluster): Πρόκειται για μαθητές που γράφουν **πολύ χαμηλά σε όλα τα μαθήματα**.

Cluster 1 (6742 μαθητές) – Μέτριοι / προς χαμηλοί: Μαθητές με χαμηλή επίδοση σε όλα τα μαθήματα εκτός τη Νεοελληνική Γλώσσα που είναι μέτρια.

Cluster 2 (8548 μαθητές) – Καλοί μαθητές: Μέτρια έως καλή επίδοση σε όλα τα μαθήματα εκτός τα Μαθηματικά.

Cluster 3 (7735 μαθητές) – Άριστοι: Άριστοι στην Οικονομία και στη Πληροφορική αλλά καλοί στη Γλώσσα και μέτριοι στα Μαθηματικά.

Η ομάδα αυτή παρουσιάζει τη μεγαλύτερη διασπορά (από 1.4 έως 18), δείχνοντας ότι οι μαθητές της είναι πιο ετερογενείς και όχι τελείως ξεκάθαρα τα clusters π.χ. το cluster 3, οι Άριστοι, δεν είναι άριστοι σε όλα τα μαθήματα.

Συνολικά, τα αποτελέσματα της ανάλυσης καταδεικνύουν ότι σε όλες τις ομάδες αναδύονται τέσσερα διακριτά φυσικά clusters επιδόσεων, τα οποία αντιστοιχούν σε μεγάλο βαθμό στις κλασικές βαθμίδες αξιολόγησης (αδύναμοι, μέτριοι, καλοί, άριστοι), αναδεικνύοντας μια **σταθερή καθοριζόμενη ταξινόμηση** των μαθητών. Η μέθοδος K-Means αποτυπώνει με σαφήνεια ποιοτικά επίπεδα επίδοσης, επιτυγχάνοντας ιδιαίτερα καθαρό διαχωρισμό μεταξύ των ομάδων. Επιπλέον, οι ομάδες Θετικών Επιστημών και Υγείας παρουσιάζουν πιο ευδιάκριτα και «καθαρά» clusters, στοιχείο που υποδηλώνει ότι τα αντίστοιχα μαθήματα διαφοροποιούν εντονότερα τις μαθητικές επιδόσεις, σε αντίθεση με τις Ανθρωπιστικές Σπουδές, όπου παρατηρείται υψηλός βαθμός ομοιογένειας στα ανώτερα επίπεδα, με σημαντική συγκέντρωση άριστων επιδόσεων, ιδίως στα Λατινικά. Τέλος, η ομάδα Οικονομικών–Πληροφορικής εμφανίζει τη μεγαλύτερη ετερογένεια, καθώς χαρακτηρίζεται από ευρύ φάσμα βαθμολογιών.

5.3. Αλγόριθμος DBSCAN

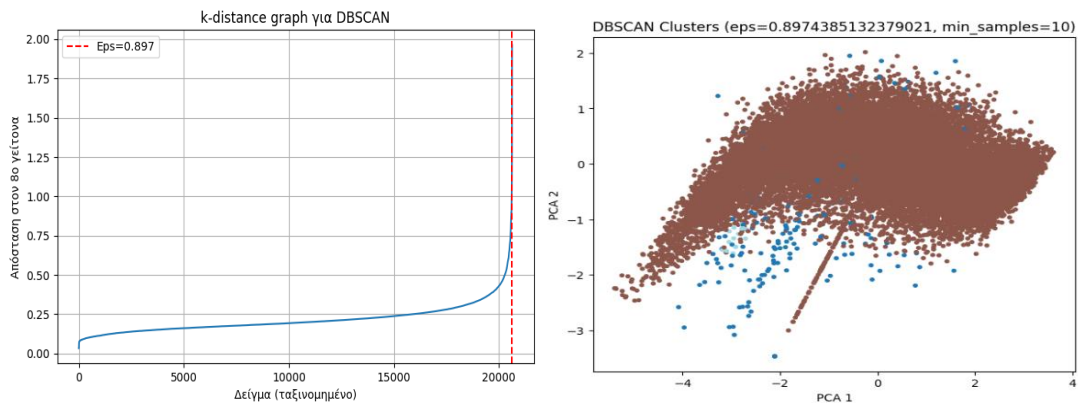
Παρακάτω παρουσιάζονται τα αποτελέσματα από την εφαρμογή του αλγόριθμου DBSCAN στα δεδομένα των βαθμολογιών ανά κωδικό υποψηφίου και ανά Ομάδα Προσανατολισμού προέκυψαν τα εξής δεδομένα:

ΟΜΑΔΑ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ

DBSCAN_cluster

0 20618

-1 9



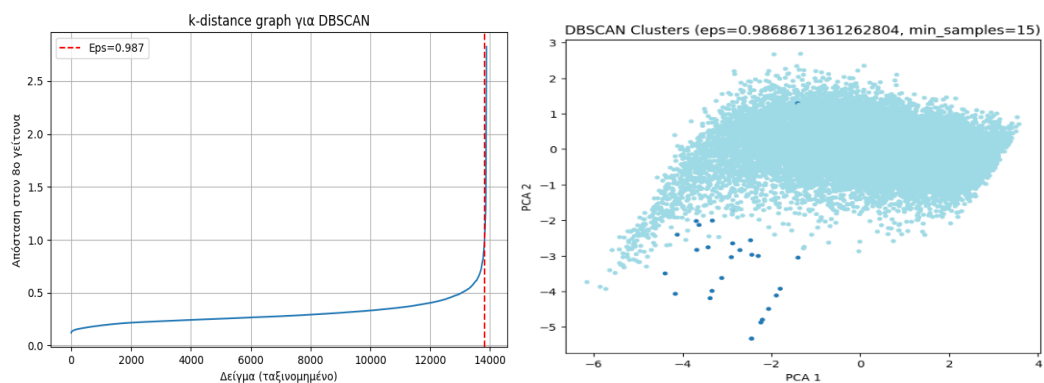
Γράφημα 5.3.1: k-distance με αυτόματο εύρος $\epsilon=0.897$ και $\min_samples=10$ και Clusters

ΟΜΑΔΑ ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ

DBSCAN_cluster:

0 13853,

-1 28



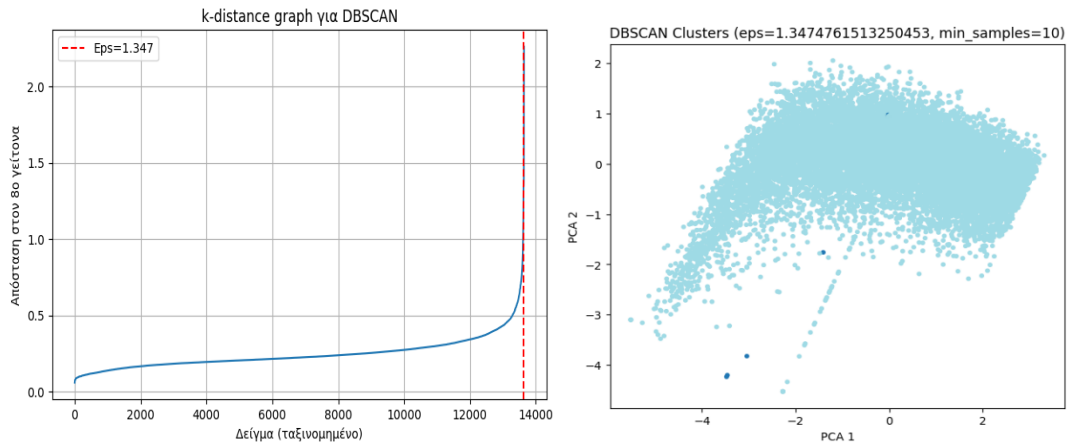
Γράφημα 5.3.2: k-distance με αυτόματο εύρος $\epsilon=0.987$ και $\min_samples=10$ και Clusters

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ

DBSCAN_cluster

0 13636

-1 5



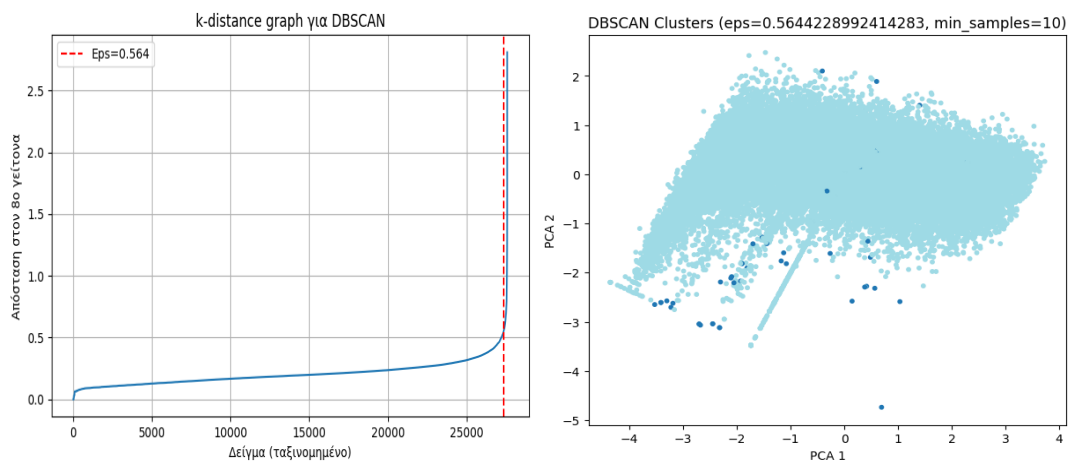
Γράφημα 5.3.3: k-distance με αυτόματο εύρος $\text{eps}=1.347$ και $\text{min_samples}=10$ και Clusters

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ

DBSCAN_cluster

0 27541

-1 50



Γράφημα 5.3.4: k-distance με αυτόματο εύρος $\text{eps}=0.564$ και $\text{min_samples}=10$ και Clusters

Σε όλες τις ομάδες έγιναν τροποποιήσεις στο min_samples π.χ. 15 και 20, χωρίς όμως καμία ιδιαίτερη μεταβολή στα αποτελέσματα, συνέχιζε ουσιαστικά ο DBSCAN να εντοπίζει μόνο **μια "κύρια" ομάδα** και τους υπόλοιπους ως **εξαιρέσεις**.

Η εφαρμογή του αλγορίθμου DBSCAN στις βαθμολογίες έδειξε ότι τα δεδομένα σχηματίζουν ένα κυρίως συνεχές σύνολο χωρίς σαφή διαχωρισμό σε πολλαπλές πυκνές ομάδες, με την ύπαρξη περιορισμένου αριθμού ακραίων περιπτώσεων που εμφανίζονται εδώ ως noise (-1) δηλαδή θόρυβο. Ο DBSCAN βρίσκει ένα ενιαίο σώμα μαθητών και λίγες ακραίες περιπτώσεις (outliers).

Κεφάλαιο 6^ο: Κατηγοριοποίηση

6.1. Random Forest Regression

Ο αλγόριθμος Random Forest Regression εφαρμόστηκε στα ίδια δεδομένα που χρησιμοποιήθηκαν και στην Συσταδοποίηση. Για κάθε Ομάδα Προσανατολισμού ένα αρχείο με δεδομένα: τον κωδικό υποψηφίου και τους βαθμούς στα τέσσερα βασικά μαθήματα.

Περίπτωση Ομάδα Ανθρωπιστικών Σπουδών

Καθότι το μάθημα της Ιστορίας έχει τις χαμηλότερες επιδόσεις μπήκε ως στόχος στην εφαρμογή του αλγόριθμου Random Forest Regression στην εν λόγω Ομάδα Προσανατολισμού. Από το σύνολο των υποψηφίων **20.628**, οι **240 εγγραφές** απορρίφθηκαν λόγω ελλιπών τιμών στον στόχο (*Ιστορία Ο.Π.*). Το τελικό σύνολο δεδομένων θεωρείται **επαρκές και αντιπροσωπευτικό**, καθώς ο αριθμός παρατηρήσεων μετά τον καθαρισμό παραμένει μεγάλος. Πράγμα που αποτρέπει μεροληψία (bias) και διασφαλίζει την εγκυρότητα των δεικτών απόδοσης.

Τα αποτελέσματα της παλινδρόμησης για την πρόβλεψη του βαθμού Ιστορίας καταδεικνύουν την υψηλή αποδοτικότητα του μοντέλου, με συντελεστή προσδιορισμού $R^2 \approx 0,72$, τιμή που θεωρείται ιδιαίτερα υψηλή για εκπαιδευτικά δεδομένα. Η χαμηλή διακύμανση (standard deviation) που παρατηρήθηκε κατά τη διαδικασία του cross-validation υποδηλώνει απουσία φαινομένων υπερπροσαρμογής (overfitting) και ισχυρή ικανότητα γενίκευσης του μοντέλου σε νέα δεδομένα. Συνεπώς, το μοντέλο εμφανίζεται αξιόπιστο ως προς την πρόβλεψη της επίδοσης των μαθητών στο μάθημα της Ιστορίας, γεγονός που ενισχύει τη στατιστική και παιδαγωγική εγκυρότητα των αποτελεσμάτων.

```
Αφαιρέθηκαν 240 γραμμές με NaN στον στόχο για Regression.

=== REGRESSION (Predict: ΙΣΤΟΡΙΑ Ο.Π.) ===
MAE : 2.302
RMSE: 3.130
R^2 : 0.718
CV R^2 (5-fold): mean=0.700, std=0.013

Feature importances (Regression):
ΛΑΤΙΝΙΚΑ Ο.Π.                0.797564
ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.        0.111145
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.091291
dtype: float64

Αφαιρέθηκαν 240 γραμμές με NaN στον στόχο για Classification.

=== CLASSIFICATION (Pass/Fail for: ΙΣΤΟΡΙΑ Ο.Π.) ===
Threshold: 10.0
Accuracy: 0.846
ROC AUC : 0.927
```

Εικόνα 6.1: Random Forest - Αποτελέσματα 1 –Ανθρωπιστικών Σπουδών

```

Confusion matrix [ [TN FP], [FN TP] ]:
[[2116  324]
 [ 303 1335]]

Classification report:
              precision    recall  f1-score   support

   fail         0.87         0.87         0.87         2440
   pass         0.80         0.82         0.81         1638

 accuracy
macro avg         0.84         0.84         0.84         4078
weighted avg         0.85         0.85         0.85         4078

Feature importances (Classification):
ΛΑΤΙΝΙΚΑ Ο.Π.                0.443948
ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.        0.335375
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.220677
dtype: float64

```

Εικόνα 6.2: Random Forest - Αποτελέσματα 2 –Ανθρωπιστικών Σπουδών

Η ανάλυση των χαρακτηριστικών (feature importances) στο μοντέλο παλινδρόμησης αναδεικνύει τα Λατινικά Ομάδας Προσανατολισμού ως τον κυρίαρχο προβλεπτικό παράγοντα, με ποσοστό συνεισφοράς 0,79, ενώ τα Αρχαία Ελληνικά και η Νεοελληνική Γλώσσα & Λογοτεχνία εμφανίζουν σαφώς μικρότερη αλλά συμπληρωματική επίδραση. Το εύρημα αυτό υποδηλώνει την ύπαρξη κοινής γνωστικής βάσης μεταξύ των μαθημάτων, κυρίως σε δεξιότητες κατανόησης κειμένου, ερμηνείας ιστορικών πηγών και απομνημόνευσης πληροφοριών. Επομένως, η υψηλή επίδοση στα Λατινικά συνδέεται άμεσα με αυξημένη πιθανότητα υψηλής επίδοσης στην Ιστορία, ενώ τα υπόλοιπα φιλολογικά μαθήματα λειτουργούν ενισχυτικά στο συνολικό γνωστικό προφίλ του μαθητή.

Όσον αφορά το μοντέλο ταξινόμησης για τη διάκριση μεταξύ επιτυχίας και αποτυχίας με όριο βαθμολογίας το 10, τα αποτελέσματα κρίνονται ιδιαίτερα ικανοποιητικά, με ακρίβεια (Accuracy) 0,846 και δείκτη ROC AUC 0,927, τιμές που υποδηλώνουν εξαιρετική διαχωριστική ικανότητα. Η ανάλυση του πίνακα σύγκρισης δείχνει περιορισμένο αριθμό ψευδώς αρνητικών προβλέψεων, γεγονός παιδαγωγικά σημαντικό, καθώς ελαχιστοποιείται ο κίνδυνος να χαρακτηριστούν λανθασμένα μαθητές ως αποτυγχόντες ενώ θα μπορούσαν να επιτύχουν. Παράλληλα, το classification report καταδεικνύει ισορροπία μεταξύ precision και recall και για τις δύο κατηγορίες, αποφεύγοντας τόσο την υπερβολική αυστηρότητα όσο και την υπεραισιοδοξία.

Τέλος, τα Λατινικά παραμένουν ο σημαντικότερος δείκτης επιτυχίας, ενώ τα Αρχαία Ελληνικά αποκτούν αυξημένο ρόλο στη διάκριση pass/fail και η Νεοελληνική Γλώσσα λειτουργεί ως παράγοντας εξισορρόπησης, επιβεβαιώνοντας τη συνολική σημασία των φιλολογικών μαθημάτων στην πρόβλεψη της σχολικής επίδοσης.

Περίπτωση Ομάδα Θετικών Σπουδών

Όπως παρατηρήθηκε στην Ομάδα Θετικών Σπουδών το μάθημα της Φυσικής είχε τις χαμηλότερες επιδόσεις και γι' αυτό το λόγο επιλέχθηκε ως target μάθημα στην εφαρμογή του αλγορίθμου Random Forest.

Από το σύνολο των υποψηφίων **13882, οι 68 εγγραφές** αφαιρέθηκαν λόγω ελλিপών τιμών στον στόχο (Φυσική Ο.Π.). Το τελικό δείγμα παραμένει **μεγάλο και στατιστικά επαρκές**, επιτρέποντας ασφαλή εκπαίδευση των μοντέλων.

```

Αφαιρέθηκαν 68 γραμμές με NaN στον στόχο για Regression.

=== REGRESSION (Predict: ΦΥΣΙΚΗ Ο.Π.) ===
MAE : 1.912
RMSE: 2.472
R^2 : 0.771
CV R^2 (5-fold): mean=0.762, std=0.011

Feature importances (Regression):
ΧΗΜΕΙΑ Ο.Π.                0.498828
ΜΑΘΗΜΑΤΙΚΑ Ο.Π.           0.431890
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.069282
dtype: float64

Αφαιρέθηκαν 68 γραμμές με NaN στον στόχο για Classification.

=== CLASSIFICATION (Pass/Fail for: ΦΥΣΙΚΗ Ο.Π.) ===
Threshold: 10.0
Accuracy: 0.856
ROC AUC : 0.938

```

Εικόνα 6.3: Random Forest - Αποτελέσματα 1 – Θετικών Σπουδών

Τα αποτελέσματα της παλινδρόμησης για την πρόβλεψη του βαθμού Φυσικής καταδεικνύουν εξαιρετικά υψηλή απόδοση του μοντέλου, με συντελεστή προσδιορισμού $R^2 \approx 0,77$, τιμή ιδιαίτερα σημαντική για εκπαιδευτικά δεδομένα. Παράλληλα, η πολύ μικρή διακύμανση που παρατηρήθηκε κατά τη διαδικασία cross-validation υποδηλώνει ισχυρή ικανότητα γενίκευσης και απουσία υπερπροσαρμογής. Τα ευρήματα αυτά επιβεβαιώνουν ότι το μοντέλο μπορεί να προβλέπει με πολύ μεγάλη ακρίβεια την επίδοση των μαθητών στη Φυσική, ενισχύοντας την αξιοπιστία των συμπερασμάτων της παρούσας ανάλυσης.

```

Confusion matrix [ [TN FP], [FN TP] ]:
[[1323  194]
 [ 205 1041]]

Classification report:
              precision    recall  f1-score   support

   fail         0.87         0.87         0.87         1517
   pass         0.84         0.84         0.84         1246

 accuracy              0.86         0.86         0.86         2763
 macro avg           0.85         0.85         0.85         2763
 weighted avg       0.86         0.86         0.86         2763

Feature importances (Classification):
ΜΑΘΗΜΑΤΙΚΑ Ο.Π.                0.428406
ΧΗΜΕΙΑ Ο.Π.                    0.420237
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.151357
dtype: float64

```

Εικόνα 6.4: Random Forest - Αποτελέσματα 2 – Θετικών Σπουδών

Η διερεύνηση της σημαντικότητας των χαρακτηριστικών στο μοντέλο παλινδρόμησης αναδεικνύει τη Χημεία Ομάδας Προσανατολισμού και τα Μαθηματικά Ομάδας Προσανατολισμού ως τους βασικούς προβλεπτικούς παράγοντες, οι οποίοι από κοινού εξηγούν σχεδόν το 93% της μεταβλητότητας της πρόβλεψης. Το εύρημα αυτό αντανακλά την ύπαρξη κοινού μαθηματικού και αναλυτικού υποβάθρου μεταξύ των τριών μαθημάτων, καθώς και τη μεταφορά δεξιοτήτων όπως η μοντελοποίηση φυσικών φαινομένων, η χρήση εξισώσεων και ο λογικός συλλογισμός. Αντίθετα, η Νεοελληνική Γλώσσα εμφανίζεται ως δευτερεύων δείκτης, λειτουργώντας κυρίως ως γενικός παράγοντας ακαδημαϊκής επάρκειας. Παιδαγωγικά, τα αποτελέσματα υποδηλώνουν ότι η επιτυχία στη Φυσική εξαρτάται πρωτίστως από την επάρκεια στα Μαθηματικά και τη Χημεία.

Αναφορικά με το μοντέλο ταξινόμησης για τη διάκριση επιτυχίας και αποτυχίας (όριο βαθμολογίας το 10), παρατηρείται πολύ καλή συνολική ακρίβεια ($Accuracy \approx 86\%$) και εξαιρετική διαχωριστική ικανότητα ($ROC AUC \approx 0,94$), γεγονός που καταδεικνύει υψηλή αξιοπιστία στην κατηγοριοποίηση των μαθητών. Η ανάλυση του πίνακα σύγκυσης δείχνει περιορισμένο αριθμό ψευδώς αρνητικών προβλέψεων, στοιχείο παιδαγωγικά σημαντικό, καθώς μειώνεται ο κίνδυνος άδικου χαρακτηρισμού μαθητών ως αποτυχόντων. Παράλληλα, το classification report καταγράφει ισορροπημένες τιμές precision και recall για τις δύο κατηγορίες, αποφεύγοντας τόσο την υπερβολική αυστηρότητα όσο και την αδικαιολόγητη «παραχώρηση» επιτυχιών.

Τέλος, στα χαρακτηριστικά της ταξινόμησης τα Μαθηματικά αποκτούν ελαφρώς μεγαλύτερη βαρύτητα σε σχέση με τη Χημεία, ενώ η Νεοελληνική Γλώσσα εμφανίζει αυξημένη σημασία στην κατηγοριοποίηση pass/fail, επιβεβαιώνοντας τον ρόλο της ως συμπληρωματικού δείκτη γενικής μαθησιακής ικανότητας.

Περίπτωση Ομάδα Σπουδών Υγείας

Όπως παρατηρήθηκε και στην Ομάδα Σπουδών Υγείας το μάθημα της Βιολογίας είχε τις χαμηλότερες επιδόσεις και γι' αυτό το λόγο επιλέχθηκε ως μάθημα στόχος (target) στην εφαρμογή του αλγορίθμου Random Forest.

Από το σύνολο των υποψηφίων **13.642**, οι **75 εγγραφές** αφαιρέθηκαν λόγω ελλιπών τιμών στον στόχο (Βιολογία Ο.Π.). Το τελικό δείγμα παραμένει **μεγάλο και στατιστικά επαρκές**, επιτρέποντας ασφαλή εκπαίδευση των μοντέλων.

```

Αφαιρέθηκαν 75 γραμμές με NaN στον στόχο για Regression.

=== REGRESSION (Predict: ΒΙΟΛΟΓΙΑ Ο.Π.) ===
MAE : 1.962
RMSE: 2.602
R^2 : 0.785
CV R^2 (5-fold): mean=0.780, std=0.009

Feature importances (Regression):
ΧΗΜΕΙΑ Ο.Π.                0.832686
ΦΥΣΙΚΗ Ο.Π.                0.093628
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.  0.073685
dtype: float64

Αφαιρέθηκαν 75 γραμμές με NaN στον στόχο για Classification.

=== CLASSIFICATION (Pass/Fail for: ΒΙΟΛΟΓΙΑ Ο.Π.) ===
Threshold: 10.0
Accuracy: 0.877
ROC AUC : 0.946

```

Εικόνα 6.5: Random Forest - Αποτελέσματα 1 - Σπουδών Υγείας

Τα αποτελέσματα της παλινδρόμησης για την πρόβλεψη του βαθμού Βιολογίας καταδεικνύουν εξαιρετικά υψηλή απόδοση του μοντέλου, με συντελεστή προσδιορισμού $R^2 \approx 0,79$, τιμή ιδιαίτερα σημαντική για εκπαιδευτικά δεδομένα. Η πολύ μικρή διακύμανση που παρατηρήθηκε κατά τη διαδικασία cross-validation επιβεβαιώνει την ισχυρή ικανότητα γενίκευσης και την απουσία φαινομένων υπερπροσαρμογής. Συνεπώς, το μοντέλο εμφανίζεται ιδιαίτερα αξιόπιστο ως προς την πρόβλεψη της επίδοσης των μαθητών στη Βιολογία, ενισχύοντας τη στατιστική εγκυρότητα της παρούσας ανάλυσης.

```

Confusion matrix [ [TN FP], [FN TP] ]:
[[ 740  140]
 [ 195 1639]]

Classification report:
              precision    recall  f1-score   support

   fail         0.79         0.84         0.82         880
   pass         0.92         0.89         0.91        1834

 accuracy              0.88         2714
 macro avg              0.86         0.87         0.86         2714
 weighted avg          0.88         0.88         0.88         2714

Feature importances (Classification):
ΧΗΜΕΙΑ Ο.Π.                0.417523
ΦΥΣΙΚΗ Ο.Π.                0.370612
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.  0.211865
dtype: float64

```

Εικόνα 6.6: Random Forest - Αποτελέσματα 2 - Σπουδών Υγείας

Η διερεύνηση της σχετικής σημασίας των χαρακτηριστικών στο μοντέλο παλινδρόμησης αναδεικνύει τη Χημεία Ομάδας Προσανατολισμού ως τον κυρίαρχο προβλεπτικό παράγοντα, με ποσοστό συνεισφοράς 0,83, ενώ η Φυσική και η Νεοελληνική Γλώσσα εμφανίζουν σαφώς μικρότερη επίδραση. Το εύρημα αυτό υποδηλώνει την ύπαρξη κοινής γνωστικής βάσης μεταξύ Χημείας και Βιολογίας, κυρίως σε επίπεδο μοριακής και βιοχημικής σκέψης, καθώς και στην κατανόηση μικροσκοπικών μηχανισμών. Η Φυσική λειτουργεί συμπληρωματικά, ενώ η Νεοελληνική Γλώσσα αποτελεί γενικό δείκτη ακαδημαϊκής επάρκειας. Παιδαγωγικά, τα αποτελέσματα υπογραμμίζουν ότι η επιτυχία στη Βιολογία εξαρτάται πρωτίστως από το επίπεδο χημικής κατανόησης των μαθητών.

Αναφορικά με το μοντέλο ταξινόμησης για τη διάκριση επιτυχίας και αποτυχίας (όριο βαθμολογίας το 10), παρατηρείται πολύ υψηλή συνολική ακρίβεια (Accuracy \approx 88%) και εξαιρετική διαχωριστική ικανότητα (ROC AUC \approx 0,95), γεγονός που καταδεικνύει μεγάλη αξιοπιστία στην κατηγοριοποίηση των μαθητών. Η ανάλυση του πίνακα σύγχυσης δείχνει ελαφρώς αυξημένο αριθμό ψευδώς αρνητικών προβλέψεων σε σχέση με τις ψευδώς θετικές, στοιχείο που υποδηλώνει ένα σχετικά αυστηρό μοντέλο. Παράλληλα, το classification report καταγράφει ιδιαίτερα υψηλό precision στην κατηγορία των επιτυχόντων, γεγονός που σημαίνει ότι σπάνια αποδίδονται λανθασμένα «χαρισμένες» επιτυχίες, ενώ η καλή ανάκληση δείχνει ότι εντοπίζεται η πλειονότητα των πραγματικά επιτυχόντων μαθητών. Στα χαρακτηριστικά της ταξινόμησης, η Φυσική αποκτά μεγαλύτερη βαρύτητα σε σχέση με το μοντέλο παλινδρόμησης, ενώ η Νεοελληνική Γλώσσα ενισχύει τον ρόλο της στην κατανόηση των θεμάτων, επιβεβαιώνοντας τη συνολική σημασία τόσο των θετικών επιστημών όσο και των γλωσσικών δεξιοτήτων στην πρόβλεψη της σχολικής επίδοσης.

Περίπτωση Ομάδα Σπουδών Οικονομίας & Πληροφορικής

Ο αλγόριθμος Random Forest εφαρμόστηκε αρχικά στο σύνολο των δεδομένων της Ομάδας Σπουδών Οικονομίας & Πληροφορικής που περιλαμβάνει **27.591** εγγραφές μαθητών και τέσσερις αριθμητικές μεταβλητές που αντιστοιχούν σε βαθμούς πανελλαδικώς εξεταζόμενων μαθημάτων και αφορούν τα: Μαθηματικά, Νεοελληνική Γλώσσα και Λογοτεχνία, Οικονομία και Πληροφορική. Ως μεταβλητή-στόχος επιλέχθηκε ο βαθμός της Πληροφορικής. Πριν την εκπαίδευση των μοντέλων αφαιρέθηκαν **465**

εγγραφές με ελλειπείς τιμές στη μεταβλητή-στόχο, ώστε να διασφαλιστεί η εγκυρότητα και η αξιοπιστία των αποτελεσμάτων.

Σύμφωνα με τον αλγόριθμο Random Forest Regression, ο οποίος βασίζεται στη δημιουργία πλήθους δέντρων απόφασης και στη συνάθροιση (ensemble) των προβλέψεών τους. Η απόδοση του μοντέλου αξιολογήθηκε με τις μετρικές Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) και τον συντελεστή προσδιορισμού R^2 .

```

=== REGRESSION (Predict Informatics grade) ===
MAE : 2.282
RMSE: 3.107
R^2 : 0.762
CV R^2 (5-fold): mean=0.763, std=0.010

Feature importances (Regression):
ΟΙΚΟΝΟΜΙΑ Ο.Π.                0.750885
ΜΑΘΗΜΑΤΙΚΑ Ο.Π.                0.165107
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.  0.084008
dtype: float64
Αφαιρέθηκαν 465 γραμμές με NaN στον αρχικό στόχο για Classification.

```

Εικόνα 6.7: Random Forest - Αποτελέσματα 1 -Σπουδών Οικονομίας & Πληροφορικής

Τα αποτελέσματα έδειξαν MAE ίσο με 2.28 και RMSE ίσο με 3.11, υποδηλώνοντας ότι η μέση απόκλιση μεταξύ προβλεπόμενων και πραγματικών τιμών κυμαίνεται σε αποδεκτά επίπεδα για κλίμακα βαθμολογίας 0–20. Ο συντελεστής R^2 ανήλθε σε 0.762, γεγονός που σημαίνει ότι το μοντέλο εξηγεί περίπου το 76.2% της συνολικής διακύμανσης του βαθμού Πληροφορικής.

Η γενικευσιμότητα του μοντέλου ελέγχθηκε μέσω πενταπλής διασταυρούμενης επικύρωσης (5-fold cross-validation), με μέση τιμή R^2 ίση με 0.763 και πολύ μικρή τυπική απόκλιση (0.010). Το αποτέλεσμα αυτό υποδεικνύει σταθερή απόδοση του μοντέλου και περιορισμένο κίνδυνο υπερπροσαρμογής, στοιχείο που συνάδει με τα πλεονεκτήματα των μεθόδων ensemble έναντι μεμονωμένων δέντρων απόφασης [20].

Η ανάλυση της σχετικής σημαντικότητας των χαρακτηριστικών κατέδειξε ότι ο βαθμός στην Οικονομία αποτελεί τον ισχυρότερο προβλεπτικό παράγοντα, συνεισφέροντας περίπου το 75% της συνολικής πληροφορίας του μοντέλου. Τα Μαθηματικά εμφανίζουν σημαντική αλλά σαφώς μικρότερη επίδραση (16.5%), ενώ η Νεοελληνική Γλώσσα και Λογοτεχνία παρουσιάζει περιορισμένη συνεισφορά (8.4%). Συμπερασματικά οι μαθητές που πάνε καλά στην Οικονομία και στα Μαθηματικά τείνουν να έχουν υψηλότερες επιδόσεις στην Πληροφορική.

```

=== CLASSIFICATION (Pass/Fail) ===
Threshold: 10.0
Accuracy: 0.860
ROC AUC : 0.937

Confusion matrix [ [TN FP], [FN TP] ]:
[[1851 393]
 [ 368 2814]]

Classification report:

```

	precision	recall	f1-score	support
fail	0.83	0.82	0.83	2244
pass	0.88	0.88	0.88	3182
accuracy			0.86	5426
macro avg	0.86	0.85	0.86	5426
weighted avg	0.86	0.86	0.86	5426

```

Feature importances (Classification):
ΟΙΚΟΝΟΜΙΑ Ο.Π. 0.418485
ΜΑΘΗΜΑΤΙΚΑ Ο.Π. 0.354989
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.226526

```

Εικόνα 6.8: Random Forest - Αποτελέσματα 2- Σπουδών Οικονομίας & Πληροφορικής

Πρόβλεψη Επιτυχίας ή Αποτυχίας (Classification)

Στο δεύτερο στάδιο της ανάλυσης, το πρόβλημα επαναδιατυπώθηκε ως πρόβλημα δυαδικής ταξινόμησης, όπου οι μαθητές κατηγοριοποιήθηκαν σε επιτυχία ή αποτυχία με βάση όριο βαθμολογίας ίσο με 10. Για τον σκοπό αυτό χρησιμοποιήθηκε αλγόριθμος Random Forest Classification, όπως υλοποιείται στη βιβλιοθήκη scikit-learn [25].

Η συνολική ακρίβεια του μοντέλου ανήλθε σε 86% (Accuracy), ενώ η τιμή της περιοχής κάτω από την καμπύλη ROC (ROC AUC) έφτασε το 0.937, ένδειξη εξαιρετικής ικανότητας διάκρισης μεταξύ των δύο κλάσεων. Τα αποτελέσματα αυτά υποδηλώνουν ότι το μοντέλο μπορεί να χρησιμοποιηθεί αποτελεσματικά για την πρόβλεψη της επιτυχίας ή αποτυχίας μαθητών σε πραγματικά εκπαιδευτικά σενάρια.

Η ανάλυση του πίνακα σύγχυσης (Confusion Matrix): 1851 σωστά εντοπισμένοι αποτυχημένοι, 393 false positives (προβλέφθηκαν pass αλλά απέτυχαν), 368 false negatives (προβλέφθηκαν fail αλλά πέρασαν), 2814 σωστά εντοπισμένοι επιτυχημένοι - έδειξε ισορροπημένη κατανομή σφαλμάτων μεταξύ ψευδώς θετικών και ψευδώς αρνητικών προβλέψεων, ενώ οι δείκτες precision 0.83%, recall 0.82% και F1-score 0.83% παρουσίασαν υψηλές και συγκρίσιμες τιμές για τις δύο κλάσεις. Το εύρημα αυτό υποδηλώνει ότι το μοντέλο δεν ευνοεί δυσανάλογα κάποια κατηγορία, γεγονός ιδιαίτερα σημαντικό σε εφαρμογές εκπαιδευτικής αξιολόγησης.

Η ανάλυση σημαντικότητας χαρακτηριστικών στο πρόβλημα της ταξινόμησης ανέδειξε εκ νέου την Οικονομία και τα Μαθηματικά ως τους κυριότερους παράγοντες πρόβλεψης, ενώ η Νεοελληνική Γλώσσα εμφάνισε αυξημένη συνεισφορά σε σύγκριση με το πρόβλημα της παλινδρόμησης.

Συνολική Αποτίμηση: Το Random Forest πέτυχε υψηλή απόδοση τόσο στην πρόβλεψη του ακριβούς βαθμού Πληροφορικής ($R^2 \approx 0.76$), όσο και στη διάκριση επιτυχίας/αποτυχίας (Accuracy = 86%, ROC AUC = 0.94). Η Οικονομία αναδείχθηκε ως ο σημαντικότερος προβλεπτικός παράγοντας, ενώ τα Μαθηματικά παίζουν κρίσιμο ρόλο στην επιτυχία. Τα αποτελέσματα είναι σταθερά και γενικεύσιμα, όπως επιβεβαιώνεται από το cross-validation.

Πίνακας 6.1: Συγκριτικός πίνακας Μαθημάτων

Μάθημα (Target)	Ομάδα	R ² (Regression)	Accuracy	ROC AUC	Recall (Pass)
Ιστορία Ο.Π.	Ανθρωπιστικών	0.718	0.846	0.927	0.82
Φυσική Ο.Π.	Θετικών	0.771	0.856	0.938	0.84
Βιολογία Ο.Π.	Υγείας	0.785	0.877	0.946	0.89
Πληροφορική Ο.Π.	Οικονομίας– Πληροφορικής	0.762	0.860	0.937	0.88

Από τον Πίνακα 6.1 διαπιστώνεται ότι η Βιολογία είναι το **πιο «προβλέψιμο» μάθημα** βάσει άλλων γνωστικών αντικειμένων με καλύτερη συνολική επίδοση. Η **Πληροφορική** πολύ κοντά στη Φυσική, με υψηλό Recall (0.88 & 0.89 αντίστοιχα), πράγμα που σημαίνει ότι ο αλγόριθμος Random Forest Regression εντοπίζει σωστά τους επιτυχόντες.

6.2. Gaussian Naive Bayes

Πρόβλεψη Επιτυχίας Υποψηφίων με τον Αλγόριθμο Gaussian Naive Bayes

Στόχος της παρούσας ανάλυσης είναι η πρόβλεψη της επιτυχίας των υποψηφίων των Πανελλαδικών Εξετάσεων, με βάση δημογραφικά και βαθμολογικά χαρακτηριστικά. Η επιτυχία ορίζεται ως δυαδική μεταβλητή με τιμές «ΠΕΡΑΣΕ» και «ΔΕΝ ΠΕΡΑΣΕ», γεγονός που καθιστά το πρόβλημα κατάλληλο για μεθόδους δυαδικής ταξινόμησης.

Η ανάλυση πραγματοποιήθηκε σε κάθε Ομάδα Προσανατολισμού ξεχωριστά με τη χρήση του αλγορίθμου **Gaussian Naive Bayes**, ο οποίος επιλέχθηκε λόγω της απλότητάς του, της υπολογιστικής αποδοτικότητάς του και της ικανότητάς του να χειρίζεται αριθμητικά χαρακτηριστικά.

Πριν την εκπαίδευση του μοντέλου εφαρμόστηκαν τα ακόλουθα βήματα προεπεξεργασίας:

- Μετατροπή των βαθμολογιών σε αριθμητική μορφή.
- Διαχείριση ελλিপών τιμών στα αριθμητικά χαρακτηριστικά με χρήση της μέσης τιμής.
- Κωδικοποίηση των κατηγορικών μεταβλητών (ΦΥΛΟ, Δ/ΝΣΗ Δ.Ε.) μέσω One-Hot Encoding.
- Διαχωρισμός του συνόλου δεδομένων σε σύνολο εκπαίδευσης (80%) και ελέγχου (20%), με **στρωματοποιημένη δειγματοληψία** ως προς τη μεταβλητή στόχου.

Περίπτωση Ομάδας Ανθρωπιστικών Σπουδών

Το σύνολο δεδομένων αφορά υποψηφίους της Ομάδας Ανθρωπιστικών Σπουδών και οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση είναι το φύλο, η γεωγραφική διεύθυνση δευτεροβάθμιας εκπαίδευσης, καθώς και οι βαθμολογίες των βασικών εξεταζόμενων μαθημάτων της ομάδας προσανατολισμού. Η μεταβλητή στόχος είναι η **ΕΠΙΤΥΧΩΝ**, με δυαδικές τιμές *ΠΕΡΑΣΕ* και *ΔΕΝ ΠΕΡΑΣΕ*.

Η τελική κατανομή της μεταβλητής στόχου έχει ως εξής: **14.155 υποψήφιοι** ανήκουν στην κατηγορία *ΠΕΡΑΣΕ*, ενώ **6.397 υποψήφιοι** στην κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*. Η κατανομή αυτή εμφανίζει

αριθμητική υπεροχή της κατηγορίας των επιτυχόντων, χωρίς ωστόσο η ανισοροπία να θεωρείται ακραία, επιτρέποντας την αξιόπιστη εκπαίδευση και αξιολόγηση ενός δυαδικού ταξινομητή.

Ο αλγόριθμος **Gaussian Naive Bayes** πέτυχε συνολική ακρίβεια (**Accuracy**) ίση με **87,6%**, ενώ η **Balanced Accuracy** ανήλθε στο **87,4%**. Η μικρή διαφορά μεταξύ των δύο δεικτών υποδηλώνει ότι το μοντέλο αποδίδει ισοροπημένα και στις δύο κατηγορίες και δεν επηρεάζεται σημαντικά από την ανισοροπία της κατανομής των δεδομένων. Οι τιμές αυτές καταδεικνύουν υψηλή συνολική απόδοση και ικανοποιητική ικανότητα γενίκευσης.

```

Κατανομή target - ΟΜΑΔΑ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ:
ΕΠΙΤΥΧΩΝ
ΠΕΡΑΣΕ          14155
ΔΕΝ ΠΕΡΑΣΕ       6397
Name: count, dtype: int64

Accuracy: 0.8759425930430552
Balanced Accuracy: 0.8737583064729777

Classification Report:
              precision    recall  f1-score   support

 ΔΕΝ ΠΕΡΑΣΕ      0.77      0.87      0.81      1280
  ΠΕΡΑΣΕ         0.94      0.88      0.91      2831

   accuracy              0.88      0.88      4111
  macro avg              0.85      0.87      0.86      4111
 weighted avg              0.88      0.88      0.88      4111

Confusion Matrix:
[[1111  169]
 [ 341 2490]]

Πρόβλεψη ΕΠΙΤΥΧΙΑΣ: ΠΕΡΑΣΕ
Πιθανότητες:
- ΔΕΝ ΠΕΡΑΣΕ: 0.000
- ΠΕΡΑΣΕ: 1.000
    
```

Εικόνα 6.9: Gaussian Naive Bayes - Αποτελέσματα Ανθρωπιστικών Σπουδών

Η ανάλυση της απόδοσης ανά κλάση δείχνει ότι για την κατηγορία «**ΔΕΝ ΠΕΡΑΣΕ**», το μοντέλο επιτυγχάνει **precision 77%**, **recall 87%** και **F1-score 81%**. Η υψηλή τιμή του recall υποδηλώνει ότι ο Gaussian Naive Bayes εντοπίζει αποτελεσματικά την πλειονότητα των υποψηφίων που δεν πέτυχαν, στοιχείο ιδιαίτερα σημαντικό σε εκπαιδευτικά σενάρια όπου ο έγκαιρος εντοπισμός πιθανής αποτυχίας μπορεί να αξιοποιηθεί για την παροχή υποστηρικτικών παρεμβάσεων. Η τιμή της precision δείχνει ότι, παρότι υπάρχουν ψευδώς θετικές προβλέψεις, το ποσοστό τους παραμένει σε αποδεκτά επίπεδα.

Για την κατηγορία «**ΠΕΡΑΣΕ**», τα αποτελέσματα είναι ιδιαίτερα ισχυρά, με **precision 94%**, **recall 88%** και **F1-score 91%**. Οι τιμές αυτές καταδεικνύουν ότι οι προβλέψεις επιτυχίας είναι σε μεγάλο βαθμό αξιόπιστες και ότι η πλειονότητα των πραγματικών επιτυχόντων αναγνωρίζεται σωστά από το μοντέλο, γεγονός που ενισχύει την πρακτική χρησιμότητά του.

Η ανάλυση του πίνακα σύγχυσης δείχνει ότι **2.490 υποψήφιοι** ταξινομήθηκαν σωστά ως *ΠΕΡΑΣΕ* και **1.111 υποψήφιοι** ως *ΔΕΝ ΠΕΡΑΣΕ*. Παράλληλα, **169 περιπτώσεις αποτυχίας** ταξινομήθηκαν λανθασμένα ως επιτυχία, ενώ **341 περιπτώσεις επιτυχίας** ταξινομήθηκαν ως αποτυχία. Η κατανομή των σφαλμάτων δείχνει ότι το μοντέλο τείνει ελαφρώς περισσότερο να χαρακτηρίζει επιτυχόντες ως μη επιτυχόντες, γεγονός που συνάδει με τη στρατηγική μεγιστοποίησης της ανάκλησης για την κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*.

Περίπτωση Ομάδας Θετικών Σπουδών

Το σύνολο δεδομένων αφορά υποψηφίους της Ομάδας Θετικών Σπουδών και οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση είναι το φύλο, η γεωγραφική διεύθυνση δευτεροβάθμιας εκπαίδευσης, καθώς και οι βαθμολογίες των βασικών εξεταζόμενων μαθημάτων της ομάδας προσανατολισμού. Η μεταβλητή στόχος είναι η **ΕΠΙΤΥΧΩΝ**, με δυαδικές τιμές *ΠΕΡΑΣΕ* και *ΔΕΝ ΠΕΡΑΣΕ*.

Το σύνολο δεδομένων αφορά υποψηφίους της **Ομάδας Θετικών Σπουδών** και περιλαμβάνει δημογραφικά και βαθμολογικά χαρακτηριστικά που σχετίζονται άμεσα με την επίδοσή τους στις Πανελλαδικές Εξετάσεις. Οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση περιλαμβάνουν το φύλο, τη γεωγραφική διεύθυνση δευτεροβάθμιας εκπαίδευσης, καθώς και τις βαθμολογίες των βασικών εξεταζόμενων μαθημάτων της ομάδας προσανατολισμού. Η μεταβλητή στόχος είναι η **ΕΠΙΤΥΧΩΝ**, με δυαδικές τιμές *ΠΕΡΑΣΕ* και *ΔΕΝ ΠΕΡΑΣΕ*.

```

Κατανομή target- ΟΜΑΔΑ ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ:
ΕΠΙΤΥΧΩΝ
ΠΕΡΑΣΕ          11323
ΔΕΝ ΠΕΡΑΣΕ      2509
Name: count, dtype: int64

Accuracy: 0.8760390314419949
Balanced Accuracy: 0.8676881876467639

Classification Report:

```

	precision	recall	f1-score	support
ΔΕΝ ΠΕΡΑΣΕ	0.61	0.85	0.71	502
ΠΕΡΑΣΕ	0.96	0.88	0.92	2265
accuracy			0.88	2767
macro avg	0.79	0.87	0.82	2767
weighted avg	0.90	0.88	0.88	2767

```

Confusion Matrix:
[[ 429   73]
 [ 270 1995]]

Πρόβλεψη ΕΠΙΤΥΧΙΑΣ: ΠΕΡΑΣΕ
Πιθανότητες:
- ΔΕΝ ΠΕΡΑΣΕ: 0.000
- ΠΕΡΑΣΕ: 1.000

```

Εικόνα 6.10: Gaussian Naive Bayes -Αποτελέσματα Θετικών Σπουδών

Η τελική κατανομή της μεταβλητής στόχου έχει ως εξής: **11.323 υποψήφιοι** ανήκουν στην κατηγορία *ΠΕΡΑΣΕ*, ενώ **2.509 υποψήφιοι** στην κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*. Η κατανομή αυτή παρουσιάζει σαφή αριθμητική υπεροχή της κατηγορίας των επιτυχόντων, χωρίς όμως η ανισορροπία να θεωρείται ακραία, επιτρέποντας την αξιόπιστη εκπαίδευση ενός δυαδικού ταξινομητή.

Ο αλγόριθμος **Gaussian Naive Bayes** πέτυχε συνολική ακρίβεια (**Accuracy**) ίση με **87,6%**, ενώ η **Balanced Accuracy** ανήλθε στο **86,8%**. Η ελαφρώς χαμηλότερη τιμή της **Balanced Accuracy** αντανακλά την ανισορροπία στην κατανομή των δεδομένων, ωστόσο παραμένει σε υψηλό επίπεδο, επιβεβαιώνοντας ότι το μοντέλο διατηρεί ικανοποιητική ικανότητα διάκρισης τόσο των επιτυχόντων όσο και των μη επιτυχόντων.

Η ανάλυση της απόδοσης ανά κλάση δείχνει ότι για την κατηγορία «**ΔΕΝ ΠΕΡΑΣΕ**», το μοντέλο επιτυγχάνει **recall 85%**, **precision 61%** και **F1-score 71%**. Η υψηλή τιμή του recall υποδηλώνει ότι ο Gaussian Naive Bayes εντοπίζει αποτελεσματικά την πλειονότητα των υποψηφίων που δεν πέτυχαν, στοιχείο ιδιαίτερα σημαντικό σε εκπαιδευτικά περιβάλλοντα όπου ο έγκαιρος εντοπισμός πιθανής αποτυχίας μπορεί να αξιοποιηθεί για παρεμβάσεις υποστήριξης. Η χαμηλότερη τιμή της precision υποδηλώνει ότι ένα ποσοστό των υποψηφίων που προβλέπονται ως μη επιτυχόντες ανήκει τελικά στην κατηγορία των επιτυχόντων, φαινόμενο αναμενόμενο σε περιβάλλοντα με αριθμητική υπεροχή της κατηγορίας *ΠΕΡΑΣΕ* και ενδεικτικό μιας συντηρητικής στάσης του μοντέλου ως προς την πρόβλεψη αποτυχίας.

Για την κατηγορία «**ΠΕΡΑΣΕ**», τα αποτελέσματα είναι ιδιαίτερα ισχυρά, με **precision 96%**, **recall 88%** και **F1-score 92%**. Οι τιμές αυτές καταδεικνύουν ότι οι προβλέψεις επιτυχίας είναι κατά κανόνα

αξιόπιστες και ότι η πλειονότητα των πραγματικών επιτυχόντων αναγνωρίζεται σωστά από το μοντέλο, γεγονός που ενισχύει τη χρησιμότητά του σε πρακτικά εκπαιδευτικά σενάρια.

Η ανάλυση του πίνακα σύγχυσης δείχνει ότι **1.995 υποψήφιοι** ταξινομήθηκαν σωστά ως *ΠΕΡΑΣΕ* και **429 υποψήφιοι** ως *ΔΕΝ ΠΕΡΑΣΕ*. Παράλληλα, **73 περιπτώσεις αποτυχίας** ταξινομήθηκαν λανθασμένα ως επιτυχία, ενώ **270 περιπτώσεις επιτυχίας** ταξινομήθηκαν ως αποτυχία. Η κατανομή των σφαλμάτων δείχνει ότι το μοντέλο τείνει ελαφρώς περισσότερο να χαρακτηρίζει επιτυχόντες ως μη επιτυχόντες, γεγονός που συνάδει με τη στρατηγική μεγιστοποίησης της ανάκλησης για την κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*.

Περίπτωση Ομάδας Σπουδών Υγείας

Κατανομή target- ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ:				
ΕΠΙΤΥΧΩΝ				
ΠΕΡΑΣΕ	10250			
ΔΕΝ ΠΕΡΑΣΕ	3688			
Name: count, dtype: int64				
Accuracy: 0.8597560975609756				
Balanced Accuracy: 0.85				
Classification Report:				
	precision	recall	f1-score	support
ΔΕΝ ΠΕΡΑΣΕ	0.70	0.83	0.76	738
ΠΕΡΑΣΕ	0.93	0.87	0.90	2050
accuracy			0.86	2788
macro avg	0.82	0.85	0.83	2788
weighted avg	0.87	0.86	0.86	2788
Confusion Matrix:				
[[612 126]				
[265 1785]]				
Πρόβλεψη ΕΠΙΤΥΧΙΑΣ: ΠΕΡΑΣΕ				
Πιθανότητες:				
- ΔΕΝ ΠΕΡΑΣΕ: 0.000				
- ΠΕΡΑΣΕ: 1.000				

Εικόνα 6.11: Gaussian Naive Bayes - Σπουδών Υγείας

Το σύνολο δεδομένων αφορά υποψηφίους της **Ομάδας Σπουδών Υγείας** και οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση είναι το φύλο, η γεωγραφική διεύθυνση δευτεροβάθμιας εκπαίδευσης, καθώς και οι βαθμολογίες των βασικών εξεταζόμενων μαθημάτων της ομάδας προσανατολισμού. Η μεταβλητή στόχος είναι η **ΕΠΙΤΥΧΩΝ**, με δυαδικές τιμές *ΠΕΡΑΣΕ* και *ΔΕΝ ΠΕΡΑΣΕ*.

Η τελική κατανομή της μεταβλητής στόχου έχει ως εξής: **10.250 υποψήφιοι** ανήκουν στην κατηγορία *ΠΕΡΑΣΕ*, ενώ **3.688 υποψήφιοι** στην κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*. Η κατανομή αυτή καταδεικνύει αριθμητική υπεροχή της κατηγορίας των επιτυχόντων, χωρίς ωστόσο η ανισορροπία να θεωρείται ακραία, επιτρέποντας την αξιόπιστη εκπαίδευση ενός δυαδικού ταξινομητή.

Ο αλγόριθμος **Gaussian Naive Bayes** πέτυχε συνολική ακρίβεια (**Accuracy**) ίση με **85,98%**, ενώ η **Balanced Accuracy** ανήλθε στο **85%**. Η υψηλή συνολική ακρίβεια υποδηλώνει ότι το μοντέλο ταξινομεί σωστά τη μεγάλη πλειονότητα των υποψηφίων. Η ελαφρώς χαμηλότερη τιμή της **Balanced Accuracy** αντανάκλα την ανισορροπία στην κατανομή των δεδομένων, ωστόσο παραμένει σε υψηλό επίπεδο, επιβεβαιώνοντας ότι το μοντέλο διατηρεί ικανοποιητική ικανότητα διάκρισης τόσο των επιτυχόντων όσο και των μη επιτυχόντων.

Η αξιολόγηση της απόδοσης πραγματοποιήθηκε και ανά κλάση, προκειμένου να αναλυθεί η συμπεριφορά του μοντέλου σε συνθήκες μη πλήρως ισορροπημένων δεδομένων. Για την κατηγορία «ΔΕΝ ΠΕΡΑΣΕ», ο Gaussian Naive Bayes παρουσίασε **recall 83%**, **precision 70%** και **F1-score 76%**. Η υψηλή τιμή του recall υποδηλώνει ότι το μοντέλο εντοπίζει αποτελεσματικά την πλειονότητα των υποψηφίων που δεν πέτυχαν, γεγονός ιδιαίτερα σημαντικό σε εκπαιδευτικά σενάρια όπου ο έγκαιρος εντοπισμός πιθανής αποτυχίας μπορεί να αξιοποιηθεί για στοχευμένες παρεμβάσεις. Η χαμηλότερη τιμή της precision υποδηλώνει την ύπαρξη ψευδώς θετικών προβλέψεων, φαινόμενο αναμενόμενο σε περιβάλλοντα με υπεροχή της κατηγορίας ΠΕΡΑΣΕ και ενδεικτικό μιας συντηρητικής στάσης του μοντέλου ως προς την πρόβλεψη αποτυχίας.

Για την κατηγορία «ΠΕΡΑΣΕ», τα αποτελέσματα είναι ιδιαίτερα ισχυρά, με **precision 93%**, **recall 87%** και **F1-score 90%**. Οι τιμές αυτές καταδεικνύουν ότι όταν το μοντέλο προβλέπει επιτυχία, η πρόβλεψη είναι κατά κανόνα ορθή και ότι η πλειονότητα των πραγματικών επιτυχόντων αναγνωρίζεται σωστά. Η συμπεριφορά αυτή καθιστά το μοντέλο αξιόπιστο για τη θετική κλάση, γεγονός που συνδέεται και με την αριθμητική της υπεροχή στο σύνολο δεδομένων.

Η ανάλυση του πίνακα σύγχυσης δείχνει ότι **1.785 υποψήφιοι** ταξινομήθηκαν σωστά ως ΠΕΡΑΣΕ και **612 υποψήφιοι** ως ΔΕΝ ΠΕΡΑΣΕ. Παράλληλα, **126 περιπτώσεις αποτυχίας** ταξινομήθηκαν λανθασμένα ως επιτυχία, ενώ **265 περιπτώσεις επιτυχίας** ταξινομήθηκαν ως αποτυχία. Η κατανομή αυτή δείχνει ότι το μοντέλο τείνει ελαφρώς περισσότερο να χαρακτηρίζει επιτυχόντες ως μη επιτυχόντες, επιλογή που συνάδει με τη στρατηγική μεγιστοποίησης της ανάκλησης για την κατηγορία ΔΕΝ ΠΕΡΑΣΕ.

Περίπτωση Ομάδας Σπουδών Οικονομίας & Πληροφορικής

```

Κατανομή target- ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ:
ΕΠΙΤΥΧΩΝ
ΠΕΡΑΣΕ          17995
ΔΕΝ ΠΕΡΑΣΕ      9954
Name: count, dtype: int64

Accuracy: 0.892128801431127
Balanced Accuracy: 0.8961424632574845

Classification Report:

```

	precision	recall	f1-score	support
ΔΕΝ ΠΕΡΑΣΕ	0.81	0.91	0.86	1991
ΠΕΡΑΣΕ	0.95	0.88	0.91	3599
accuracy			0.89	5590
macro avg	0.88	0.90	0.89	5590
weighted avg	0.90	0.89	0.89	5590

```

Confusion Matrix:
[[1812  179]
 [ 424 3175]]

Πρόβλεψη ΕΠΙΤΥΧΙΑΣ: ΠΕΡΑΣΕ
Πιθανότητες:
- ΔΕΝ ΠΕΡΑΣΕ: 0.000
- ΠΕΡΑΣΕ: 1.000

```

Εικόνα 6.12: Gaussian Naive Bayes -Σπουδών Οικονομίας & Πληροφορικής

Το σύνολο δεδομένων αφορά υποψηφίους της **Ομάδας Σπουδών Οικονομίας και Πληροφορικής** και περιλαμβάνει δημογραφικά και βαθμολογικά χαρακτηριστικά που σχετίζονται άμεσα με την επίδοσή τους στις Πανελλαδικές Εξετάσεις. Οι μεταβλητές που χρησιμοποιήθηκαν στην ανάλυση περιλαμβάνουν το φύλο, τη γεωγραφική διεύθυνση δευτεροβάθμιας εκπαίδευσης, καθώς και τις

βαθμολογίες των βασικών εξεταζόμενων μαθημάτων της συγκεκριμένης ομάδας προσανατολισμού. Η μεταβλητή στόχος είναι η **ΕΠΙΤΥΧΩΝ**, με δυαδικές τιμές *ΠΕΡΑΣΕ* και *ΔΕΝ ΠΕΡΑΣΕ*.

Η τελική κατανομή της μεταβλητής στόχου έχει ως εξής: **17.995 υποψήφιοι** ανήκουν στην κατηγορία *ΠΕΡΑΣΕ*, ενώ **9.954 υποψήφιοι** στην κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*. Η κατανομή αυτή εμφανίζει αριθμητική υπεροχή της κατηγορίας των επιτυχόντων, ωστόσο η παρουσία σημαντικού αριθμού αποτυχιών καθιστά το πρόβλημα κατάλληλο για την εκπαίδευση και αξιολόγηση ενός δυαδικού ταξινομητή χωρίς την ανάγκη ακραίων τεχνικών εξισορρόπησης.

Ο αλγόριθμος **Gaussian Naive Bayes** πέτυχε συνολική ακρίβεια (**Accuracy**) ίση με **89,2%**, ενώ η **Balanced Accuracy** ανήλθε στο **89,6%**. Η πολύ κοντινή τιμή των δύο δεικτών υποδηλώνει ότι το μοντέλο αποδίδει ισορροπημένα και στις δύο κατηγορίες, χωρίς να επηρεάζεται σημαντικά από την ανισορροπία της κατανομής των δεδομένων. Η υψηλή **Balanced Accuracy** επιβεβαιώνει την ικανότητα του μοντέλου να διακρίνει αποτελεσματικά τόσο τους επιτυχόντες όσο και τους μη επιτυχόντες.

Η ανάλυση της απόδοσης ανά κλάση δείχνει ότι για την κατηγορία «**ΔΕΝ ΠΕΡΑΣΕ**», το μοντέλο επιτυγχάνει **recall 91%**, **precision 81%** και **F1-score 86%**. Η ιδιαίτερα υψηλή τιμή του recall υποδηλώνει ότι ο Gaussian Naive Bayes εντοπίζει με μεγάλη αποτελεσματικότητα την πλειονότητα των υποψηφίων που δεν πέτυχαν, γεγονός κρίσιμο σε εκπαιδευτικά σενάρια όπου ο έγκαιρος εντοπισμός πιθανής αποτυχίας μπορεί να αξιοποιηθεί για στοχευμένες παρεμβάσεις. Παράλληλα, η σχετικά υψηλή precision δείχνει ότι το ποσοστό ψευδώς θετικών προβλέψεων παραμένει περιορισμένο.

Για την κατηγορία «**ΠΕΡΑΣΕ**», τα αποτελέσματα είναι επίσης ιδιαίτερα ισχυρά, με **precision 95%**, **recall 88%** και **F1-score 91%**. Οι τιμές αυτές καταδεικνύουν ότι οι προβλέψεις επιτυχίας είναι σε πολύ μεγάλο βαθμό αξιόπιστες και ότι η πλειονότητα των πραγματικών επιτυχόντων αναγνωρίζεται σωστά από το μοντέλο. Η ισορροπημένη απόδοση και στις δύο κλάσεις υποδηλώνει ότι το μοντέλο δεν παρουσιάζει έντονη μεροληψία υπέρ της πλειοψηφικής κατηγορίας.

Η ανάλυση του πίνακα σύγχυσης δείχνει ότι **3.175 υποψήφιοι** ταξινομήθηκαν σωστά ως *ΠΕΡΑΣΕ* και **1.812 υποψήφιοι** ως *ΔΕΝ ΠΕΡΑΣΕ*. Παράλληλα, **179 περιπτώσεις αποτυχίας** ταξινομήθηκαν λανθασμένα ως επιτυχία, ενώ **424 περιπτώσεις επιτυχίας** ταξινομήθηκαν ως αποτυχία. Η κατανομή των σφαλμάτων δείχνει μια ελαφρώς αυξημένη τάση του μοντέλου να χαρακτηρίζει επιτυχόντες ως μη επιτυχόντες, στοιχείο που συνδέεται με τη στρατηγική μεγιστοποίησης της ανάκλησης για την κατηγορία *ΔΕΝ ΠΕΡΑΣΕ*.

Η συγκριτική αξιολόγηση του αλγορίθμου Gaussian Naive Bayes στις τέσσερις Ομάδες Προσανατολισμού καταδεικνύει σταθερά καλή έως μέτρια προγνωστική απόδοση σε όλα τα σύνολα δεδομένων, με τιμές Accuracy που κυμαίνονται από 86% έως 89%. Η καλή και συγκρίσιμη **Balanced Accuracy** σε όλες τις ομάδες υποδηλώνει ότι το μοντέλο διατηρεί ικανοποιημένη ικανότητα διάκρισης μεταξύ επιτυχόντων και μη επιτυχόντων, παρά τις διαφοροποιήσεις στην κατανομή των κλάσεων. Ιδιαίτερα υψηλή απόδοση παρατηρείται στην Ομάδα Σπουδών Οικονομίας και Πληροφορικής, όπου τόσο η συνολική όσο και η ισορροπημένη ακρίβεια προσεγγίζουν το 90%, γεγονός που υποδηλώνει καλύτερη ευθυγράμμιση των χαρακτηριστικών με τις υποθέσεις του Gaussian Naive Bayes. Αντίθετα, στις Θετικές Σπουδές και στις Σπουδές Υγείας παρατηρείται χαμηλότερη precision για την κατηγορία «**ΔΕΝ ΠΕΡΑΣΕ**», στοιχείο που υποδηλώνει αυξημένο αριθμό ψευδώς θετικών προβλέψεων αποτυχίας, πιθανώς λόγω της αριθμητικής υπεροχής της κατηγορίας «**ΠΕΡΑΣΕ**».

Συνολικά, τα αποτελέσματα καταδεικνύουν ότι ο Gaussian Naive Bayes επιτυγχάνει σχετικά καλή και ισορροπημένη απόδοση στην πρόβλεψη της επιτυχίας των υποψηφίων όλων των Ομάδων Προσανατολισμού. Η απόδοση αυτή επιβεβαιώνει την καταλληλότητα του αλγορίθμου για προβλήματα

δυναμική ταξινόμηση σε εκπαιδευτικά δεδομένα, ιδιαίτερα όταν η κατανομή των κλάσεων είναι σχετικά ισορροπημένη και τα χαρακτηριστικά σχετίζονται άμεσα με την επίδοση των υποψηφίων.

6.3. Random Forest

Πρόβλεψη Επιτυχίας Υποψηφίων με τον Αλγόριθμο Random Forest

Για την πρόβλεψη της επιτυχίας των υποψηφίων της **Ομάδας Ανθρωπιστικών Σπουδών** εφαρμόστηκε αλγόριθμος Random Forest, με είσοδο αποκλειστικά τις βαθμολογίες των σχετικών μαθημάτων και στόχο τη δυαδική μεταβλητή «ΠΕΡΑΣΕ / ΔΕΝ ΠΕΡΑΣΕ».

Η αξιολόγηση πραγματοποιήθηκε με διαχωρισμό δεδομένων σε σύνολο εκπαίδευσης και ελέγχου (holdout), καθώς και με πενταπλή διασταυρούμενη επικύρωση (5-fold cross-validation).

=== ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΑΣ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ ===

Accuracy: 0.9715397713451714

Report: precision recall f1-score support

ΔΕΝ ΠΕΡΑΣΕ 0.96 0.95 0.95 1280

ΠΕΡΑΣΕ 0.98 0.98 0.98 2831

accuracy 0.97 4111

macro avg 0.97 0.97 0.97 4111

weighted avg 0.97 0.97 0.97 4111

=== Cross-Validation ===

Mean CV accuracy: 0.9710976603628861

Top 4 σημαντικά μαθήματα (Permutation Importance):

ΛΑΤΙΝΙΚΑ Ο.Π. 0.154318

ΙΣΤΟΡΙΑ Ο.Π. 0.071807

ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π. 0.039455

ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.039358

TOP 3 ΚΑΝΟΝΕΣ ΕΠΙΤΥΧΙΑΣ (fixed cutoffs, αξιολόγηση στο TEST set)

1. Αν $9.0 \leq \text{ΛΑΤΙΝΙΚΑ Ο.Π.} < 13.0$ ΚΑΙ $6.0 \leq \text{ΙΣΤΟΡΙΑ Ο.Π.} < 11.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} \geq 14.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=99, mean_model_p \approx 98.9%)

2. Αν $\text{ΙΣΤΟΡΙΑ Ο.Π.} \geq 11.0$ ΚΑΙ $\text{ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.} \geq 11.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 99.9%, n=1278, mean_model_p \approx 99.7%)

3. Αν ΛΑΤΙΝΙΚΑ Ο.Π. ≥ 13.0 ΚΑΙ ΙΣΤΟΡΙΑ Ο.Π. ≥ 11.0 ΚΑΙ ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π. ≥ 11.0 → πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 99.9%, n=1247, mean_model_p \approx 99.8%)

TOP 3 ΚΑΝΟΝΕΣ ΑΠΟΤΥΧΙΑΣ (fixed cutoffs, αξιολόγηση στο TEST set)

1. Αν ΛΑΤΙΝΙΚΑ Ο.Π. < 9.0 ΚΑΙ ΙΣΤΟΡΙΑ Ο.Π. < 6.0 ΚΑΙ ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. < 12.0 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 98.5%, n=720, mean_model_p \approx 2.7%)

2. Αν ΛΑΤΙΝΙΚΑ Ο.Π. < 9.0 ΚΑΙ ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. < 12.0 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 97.7%, n=767, mean_model_p \approx 3.8%)

3. Αν ΛΑΤΙΝΙΚΑ Ο.Π. < 9.0 ΚΑΙ ΙΣΤΟΡΙΑ Ο.Π. < 6.0 ΚΑΙ ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π. < 8.5 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 97.3%, n=890, mean_model_p \approx 3.2%)

Το μοντέλο Random Forest που εφαρμόστηκε στα δεδομένα της Ομάδας Ανθρωπιστικών Σπουδών παρουσίασε εξαιρετική προβλεπτική ικανότητα, επιτυγχάνοντας ακρίβεια 97.15% στο σύνολο ελέγχου, Από 4.111 υποψηφίους στο test set, το μοντέλο προβλέπει σωστά περίπου **3.994**. Ενώ η μέση ακρίβεια μέσω πενταπλής διασταυρούμενης επικύρωσης ανήλθε στο 97.11%. Η πολύ μικρή απόκλιση μεταξύ holdout και cross-validation υποδηλώνει καλή γενίκευση και απουσία υπερπροσαρμογής (overfitting). Επιπλέον, οι δείκτες precision και recall είναι υψηλοί και για τις δύο κατηγορίες («ΠΕΡΑΣΕ» και «ΔΕΝ ΠΕΡΑΣΕ»), γεγονός που δείχνει ότι το μοντέλο ταξινομεί με ισορροπημένο τρόπο τόσο τους επιτυχόντες όσο και τους μη επιτυχόντες, χωρίς εμφανή μεροληψία.

Η ανάλυση permutation importance ανέδειξε ως σημαντικότερο προβλεπτικό παράγοντα τα Λατινικά Ο.Π., με υπερδιπλάσια επίδραση σε σχέση με την Ιστορία Ο.Π. και σχεδόν τετραπλάσια σε σχέση με τα Αρχαία Ελληνικά και τη Νεοελληνική Γλώσσα. Η Ιστορία καταλαμβάνει τη δεύτερη θέση, ενώ τα Αρχαία και η Γλώσσα λειτουργούν συμπληρωματικά. Τα ευρήματα αυτά υποδεικνύουν ότι η επιτυχία στην Ομάδα Ανθρωπιστικών Σπουδών καθορίζεται από τα Λατινικά που διαδραματίζουν κεντρικό ρόλο και την Ιστορία να λειτουργεί ως βασικός παράγοντας αντιστάθμισης σε περιπτώσεις μέτριας επίδοσης.

Τέλος, οι ερμηνεύσιμοι κανόνες που εξήχθησαν από το μοντέλο επιβεβαιώνουν τη σημασία των παραπάνω μαθημάτων. Συνδυασμοί υψηλών βαθμολογιών σε Λατινικά, Ιστορία και Αρχαία οδηγούν σε ποσοστά επιτυχίας που προσεγγίζουν το 100%, ενώ ακόμη και μέτρια Λατινικά μπορούν να αντισταθμιστούν από υψηλή Ιστορία ή Γλώσσα. Αντίθετα, χαμηλές επιδόσεις στα Λατινικά, ιδιαίτερα όταν συνοδεύονται από χαμηλή Ιστορία ή Νεοελληνική Γλώσσα, σχετίζονται με πολύ υψηλή πιθανότητα αποτυχίας.

Για την πρόβλεψη της επιτυχίας των υποψηφίων της **Ομάδας Θετικών Σπουδών** προέκυψαν τα παρακάτω αποτελέσματα:

=== ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΑΣ ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ===

Accuracy: 0.9725334297072642

Report: precision recall f1-score support

ΔΕΝ ΠΕΡΑΣΕ	0.94	0.91	0.92	502
ΠΕΡΑΣΕ	0.98	0.99	0.98	2265
accuracy		0.97		2767
macro avg	0.96	0.95	0.95	2767
weighted avg	0.97	0.97	0.97	2767

=== Cross-Validation ===

Mean CV accuracy: 0.9715153363379632

Top 4 σημαντικά μαθήματα (Permutation Importance):

ΧΗΜΕΙΑ Ο.Π.	0.101193
ΦΥΣΙΚΗ Ο.Π.	0.087532
ΜΑΘΗΜΑΤΙΚΑ Ο.Π.	0.083845
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.	0.026455

dtype: float64

=====

TOP 3 ΚΑΝΟΝΕΣ ΕΠΙΤΥΧΙΑΣ (Θετικών, fixed cutoffs, αξιολόγηση στο TEST set)

=====

1. Αν $\Phi\Upsilon\text{ΣΙΚΗ Ο.Π.} \geq 15.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} \geq 14.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=463, mean_model_p \approx 99.7%)
2. Αν $10.0 \leq \text{ΜΑΘΗΜΑΤΙΚΑ Ο.Π.} < 16.0$ ΚΑΙ $10.0 \leq \Phi\Upsilon\text{ΣΙΚΗ Ο.Π.} < 15.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=452, mean_model_p \approx 99.8%)
3. Αν $10.0 \leq \text{ΜΑΘΗΜΑΤΙΚΑ Ο.Π.} < 16.0$ ΚΑΙ $\text{ΧΗΜΕΙΑ Ο.Π.} \geq 15.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=411, mean_model_p \approx 99.8%)

=====

TOP 3 ΚΑΝΟΝΕΣ ΑΠΟΤΥΧΙΑΣ (Θετικών, fixed cutoffs, αξιολόγηση στο TEST set)

=====

1. Αν $\text{ΜΑΘΗΜΑΤΙΚΑ Ο.Π.} < 10.0$ ΚΑΙ $\Phi\Upsilon\text{ΣΙΚΗ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 83.9%, n=286, mean_model_p \approx 17.4%)
2. Αν $\text{ΜΑΘΗΜΑΤΙΚΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 81.6%, n=294, mean_model_p \approx 19.2%)
3. Αν $\text{ΧΗΜΕΙΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 78.1%, n=310, mean_model_p \approx 22.6%)

Το μοντέλο Random Forest που εφαρμόστηκε στα δεδομένα της Ομάδας Θετικών Σπουδών παρουσίασε πολύ υψηλή προβλεπτική ικανότητα, επιτυγχάνοντας ακρίβεια 97.25% στο σύνολο ελέγχου, ενώ η μέση ακρίβεια μέσω πενταπλής διασταυρούμενης επικύρωσης ανήλθε στο 97.15%. Η σύγκλιση των δύο

μετρήσεων υποδηλώνει ισχυρή γενίκευση και απουσία υπερπροσαρμογής. Οι δείκτες precision και recall για την κατηγορία «ΠΕΡΑΣΕ» (0.98 και 0.99 αντίστοιχα) δείχνουν ότι το μοντέλο αναγνωρίζει σχεδόν όλους τους επιτυχόντες, ενώ για την κατηγορία «ΔΕΝ ΠΕΡΑΣΕ» (precision 0.94, recall 0.91) καταγράφεται επίσης ικανοποιητική απόδοση, παρά τη μικρότερη αναλογία αποτυχιών στο σύνολο δεδομένων.

Η ανάλυση permutation importance ανέδειξε ως σημαντικότερα μαθήματα τη Χημεία, τη Φυσική και τα Μαθηματικά, με τη Χημεία να παρουσιάζει τη μεγαλύτερη προβλεπτική συνεισφορά, ενώ η Νεοελληνική Γλώσσα εμφανίζει σαφώς μικρότερη επίδραση. Το εύρημα αυτό υποδεικνύει ότι, σε αντίθεση με την Ομάδα Ανθρωπιστικών Σπουδών, η επιτυχία στις Θετικές Σπουδές δεν εξαρτάται από ένα κυρίαρχο μάθημα, αλλά από έναν συνδυασμό βασικών θετικών επιστημών. Τα Μαθηματικά, η Φυσική και η Χημεία λειτουργούν συμπληρωματικά, συγκροτώντας έναν πολυπαραγοντικό μηχανισμό επίτευξης, ενώ η Γλώσσα παίζει επικουρικό ρόλο και επηρεάζει κυρίως τις οριακές περιπτώσεις.

Οι ερμηνεύσιμοι κανόνες που προέκυψαν επιβεβαιώνουν την παραπάνω εικόνα. Συνδυασμοί μέτρων ή υψηλών επιδόσεων σε δύο από τα βασικά θετικά μαθήματα (π.χ. Μαθηματικά–Φυσική ή Μαθηματικά–Χημεία) οδηγούν σχεδόν βέβαια σε επιτυχία, ακόμη και χωρίς άριστες βαθμολογίες, γεγονός που αναδεικνύει τη δυνατότητα αντιστάθμισης μεταξύ των μαθημάτων. Αντίθετα, χαμηλές επιδόσεις στα Μαθηματικά ή στη Χημεία, όταν συνοδεύονται από χαμηλή επίδοση στη Νεοελληνική Γλώσσα, συνδέονται με υψηλή πιθανότητα αποτυχίας. Συνολικά, τα αποτελέσματα δείχνουν ότι η επιτυχία στην Ομάδα Θετικών Σπουδών προϋποθέτει σταθερή επίδοση σε περισσότερα του ενός μαθήματα κορμού, υπογραμμίζοντας τον πολυδιάστατο χαρακτήρα των απαιτούμενων γνωστικών δεξιοτήτων.

Για την πρόβλεψη της επιτυχίας των υποψηφίων της **Ομάδας Σπουδών Υγείας** προέκυψαν τα παρακάτω αποτελέσματα:

=== ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΑΣ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ ===

Accuracy: 0.9601865136298422

Report: precision recall f1-score support

ΔΕΝ ΠΕΡΑΣΕ 0.93 0.92 0.92 738

ΠΕΡΑΣΕ 0.97 0.98 0.97 2050

accuracy 0.96 2788

macro avg 0.95 0.95 0.95 2788

weighted avg 0.96 0.96 0.96 2788

=== Cross-Validation ===

Mean CV accuracy: 0.9591042959755249

Top 4 σημαντικά μαθήματα (Permutation Importance):

ΒΙΟΛΟΓΙΑ Ο.Π. 0.108537

ΧΗΜΕΙΑ Ο.Π. 0.105022

ΦΥΣΙΚΗ Ο.Π. 0.062841

ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. 0.021162

=====

TOP 3 ΚΑΝΟΝΕΣ ΕΠΙΤΥΧΙΑΣ (Υγείας, fixed cutoffs, αξιολόγηση στο TEST set)

=====

1. Αν $10.0 \leq \text{ΒΙΟΛΟΓΙΑ Ο.Π.} < 15.0$ ΚΑΙ $\text{ΧΗΜΕΙΑ Ο.Π.} \geq 15.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=115, mean_model_p \approx 98.5%)
2. Αν $\text{ΒΙΟΛΟΓΙΑ Ο.Π.} \geq 15.0$ ΚΑΙ $\text{ΧΗΜΕΙΑ Ο.Π.} \geq 15.0$ ΚΑΙ $\text{ΦΥΣΙΚΗ Ο.Π.} < 10.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=112, mean_model_p \approx 97.2%)
3. Αν $10.0 \leq \text{ΒΙΟΛΟΓΙΑ Ο.Π.} < 15.0$ ΚΑΙ $10.0 \leq \text{ΦΥΣΙΚΗ Ο.Π.} < 15.0 \rightarrow$ πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate \approx 100.0%, n=99, mean_model_p \approx 97.5%)

=====

TOP 3 ΚΑΝΟΝΕΣ ΑΠΟΤΥΧΙΑΣ (Υγείας, fixed cutoffs, αξιολόγηση στο TEST set)

=====

1. Αν $\text{ΒΙΟΛΟΓΙΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΧΗΜΕΙΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 93.1%, n=331, mean_model_p \approx 8.4%)
2. Αν $\text{ΒΙΟΛΟΓΙΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 90.1%, n=354, mean_model_p \approx 11.5%)
3. Αν $\text{ΧΗΜΕΙΑ Ο.Π.} < 10.0$ ΚΑΙ $\text{ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.} < 12.0 \rightarrow$ υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate \approx 89.3%, n=354, mean_model_p \approx 12.1%)

Το μοντέλο Random Forest που εφαρμόστηκε στα δεδομένα της Ομάδας Σπουδών Υγείας και παρουσίασε υψηλή προβλεπτική ικανότητα, επιτυγχάνοντας ακρίβεια 96.02% στο σύνολο ελέγχου δηλαδή στους 2.788 υποψηφίους του συνόλου ελέγχου, περίπου 2.677 ταξινομήθηκαν σωστά, ενώ περίπου 111 περιπτώσεις παρουσίασαν λανθασμένη πρόβλεψη. Ενώ η μέση ακρίβεια μέσω πενταπλής διασταυρούμενης επικύρωσης ανήλθε στο 95.91%. Η μικρή απόκλιση μεταξύ των δύο μετρήσεων υποδηλώνει καλή γενίκευση και απουσία υπερπροσαρμογής. Οι δείκτες precision και recall για την κατηγορία «ΠΕΡΑΣΕ» (0.97 και 0.98 αντίστοιχα) δείχνουν ότι το μοντέλο αναγνωρίζει με πολύ μεγάλη ακρίβεια τους επιτυχόντες, ενώ για την κατηγορία «ΔΕΝ ΠΕΡΑΣΕ» (precision 0.93, recall 0.92) η απόδοση παραμένει ιδιαίτερα ικανοποιητική, παρά το μικρότερο πλήθος αποτυχιών στο δείγμα.

Η ανάλυση permutation importance ανέδειξε ως κυρίαρχους προβλεπτικούς παράγοντες τη Βιολογία και τη Χημεία, οι οποίες εμφανίζουν σχεδόν ισοδύναμη συμβολή, ακολουθούμενες από τη Φυσική, ενώ η Νεοελληνική Γλώσσα παρουσιάζει σαφώς μικρότερη επίδραση. Το εύρημα αυτό υποδηλώνει ότι στην Ομάδα Υγείας η επιτυχία βασίζεται πρωτίστως στην επάρκεια των βιοϊατρικών μαθημάτων, με τη Βιολογία και τη Χημεία να αποτελούν τον κεντρικό άξονα αξιολόγησης. Η Φυσική λειτουργεί υποστηρικτικά, ενώ η Γλώσσα επηρεάζει κυρίως τις οριακές περιπτώσεις και σχετίζεται περισσότερο με την αποτυχία παρά με την επιτυχία.

Οι ερμηνεύσιμοι κανόνες που εξήχθησαν επιβεβαιώνουν τον καθοριστικό ρόλο της Βιολογίας και της Χημείας. Συνδυασμοί μέτρων ή υψηλών επιδόσεων στα δύο αυτά μαθήματα οδηγούν σχεδόν βέβαια

σε επιτυχία, ακόμη και όταν η Φυσική βρίσκεται σε χαμηλό επίπεδο, γεγονός που αναδεικνύει ισχυρό μηχανισμό αντιστάθμισης. Αντίθετα, χαμηλές επιδόσεις στη Βιολογία ή τη Χημεία, όταν συνοδεύονται από χαμηλή επίδοση στη Νεοελληνική Γλώσσα, συνδέονται με πολύ υψηλή πιθανότητα αποτυχίας (άνω του 89%). Συνολικά, τα αποτελέσματα δείχνουν ότι η επιτυχία στην Ομάδα Υγείας προϋποθέτει επαρκή γνωστική βάση στα μαθήματα κορμού Βιολογίας και Χημείας, ενώ η Γλώσσα λειτουργεί ως κρίσιμος παράγοντας κινδύνου σε περιπτώσεις ακαδημαϊκής αδυναμίας.

Για την πρόβλεψη της επιτυχίας των υποψηφίων της **Ομάδας Σπουδών Οικονομίας & Πληροφορικής** προέκυψαν τα εξής αποτελέσματα:

=== ΑΞΙΟΛΟΓΗΣΗ ΟΜΑΔΑΣ ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ ===

Accuracy: 0.9778175313059034

Report:	precision	recall	f1-score	support
ΔΕΝ ΠΕΡΑΣΕ	0.97	0.97	0.97	1991
ΠΕΡΑΣΕ	0.98	0.98	0.98	3599
accuracy		0.98	5590	
macro avg	0.98	0.98	0.98	5590
weighted avg	0.98	0.98	0.98	5590

=== Cross-Validation ===

Mean CV accuracy: 0.9773874826318372

Top 6 σημαντικά μαθήματα (Permutation Importance):

ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.	0.118104
ΟΙΚΟΝΟΜΙΑ Ο.Π.	0.111127
ΜΑΘΗΜΑΤΙΚΑ Ο.Π.	0.044007
ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.	0.026869

=====

TOP 3 ΚΑΝΟΝΕΣ ΕΠΙΤΥΧΙΑΣ (fixed cutoffs, αξιολόγηση στο TEST set)

=====

1. Αν $ΜΑΘΗΜΑΤΙΚΑ Ο.Π. < 10.0$ ΚΑΙ $ΟΙΚΟΝΟΜΙΑ Ο.Π. \geq 15.0$ ΚΑΙ $ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. \geq 14.0$ → πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate≈100.0%, n=316, mean_model_p≈99.8%)

2. Αν $ΜΑΘΗΜΑΤΙΚΑ Ο.Π. < 10.0$ ΚΑΙ $ΟΙΚΟΝΟΜΙΑ Ο.Π. \geq 15.0$ ΚΑΙ $10.0 \leq ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π. < 15.0$ → πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate≈100.0%, n=271, mean_model_p≈99.6%)

3. Αν $10.0 \leq ΜΑΘΗΜΑΤΙΚΑ Ο.Π. < 16.0$ ΚΑΙ $10.0 \leq ΟΙΚΟΝΟΜΙΑ Ο.Π. < 15.0$ → πολύ υψηλή πιθανότητα ΕΠΙΤΥΧΙΑΣ (pass_rate≈100.0%, n=181, mean_model_p≈98.1%)

=====

TOP 3 ΚΑΝΟΝΕΣ ΑΠΟΤΥΧΙΑΣ (fixed cutoffs, αξιολόγηση στο TEST set)

-
1. Αν ΜΑΘΗΜΑΤΙΚΑ Ο.Π. < 10.0 ΚΑΙ ΟΙΚΟΝΟΜΙΑ Ο.Π. < 10.0 ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π. < 10.0 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate≈92.9%, n=1808, mean_model_p≈7.9%)
 2. Αν ΟΙΚΟΝΟΜΙΑ Ο.Π. < 10.0 ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π. < 10.0 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate≈92.4%, n=1817, mean_model_p≈8.3%)
 3. Αν ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π. < 10.0 ΚΑΙ ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π. < 12.0 → υψηλή πιθανότητα ΑΠΟΤΥΧΙΑΣ (fail_rate≈91.5%, n=1564, mean_model_p≈9.4%)

Το μοντέλο Random Forest που εφαρμόστηκε στην ομάδα Σπουδών Οικονομίας & Πληροφορικής και παρουσίασε εξαιρετικά υψηλή προβλεπτική απόδοση, με ακρίβεια 97.78% στο σύνολο ελέγχου και μέση ακρίβεια 97.74% μέσω πενταπλής διασταυρούμενης επικύρωσης. Η σχεδόν ταυτόσημη επίδοση των δύο μετρήσεων υποδηλώνει ισχυρή γενίκευση και σταθερότητα του μοντέλου. Σε απόλυτους αριθμούς, από τους 5.590 υποψηφίους του test set, περίπου 5.465 ταξινομήθηκαν σωστά, ενώ μόλις ~125 περιπτώσεις παρουσίασαν λανθασμένη πρόβλεψη. Οι δείκτες precision και recall είναι εξαιρετικά υψηλοί και για τις δύο κατηγορίες, γεγονός που δείχνει ότι το μοντέλο διακρίνει αποτελεσματικά τόσο τους επιτυχόντες όσο και τους αποτυχόντες.

Η ανάλυση permutation importance ανέδειξε ως κυρίαρχους παράγοντες την Πληροφορική και την Οικονομία, με σαφώς μεγαλύτερη συνεισφορά από τα Μαθηματικά και τη Νεοελληνική Γλώσσα. Το εύρημα αυτό υποδηλώνει ότι στην συγκεκριμένη ομάδα σπουδών η επιτυχία καθορίζεται πρωτίστως από τα μαθήματα προσανατολισμού (Πληροφορική–Οικονομία), ενώ τα Μαθηματικά λειτουργούν περισσότερο υποστηρικτικά και η Γλώσσα έχει περιορισμένο αλλά μη αμελητέο ρόλο, κυρίως σε οριακές περιπτώσεις αποτυχίας. Η ισχυρή βαρύτητα των δύο κεντρικών μαθημάτων εξηγεί και τη μεγάλη καθαρότητα των κανόνων που εξήχθησαν.

Οι ερμηνεύσιμοι κανόνες επιβεβαιώνουν ότι υψηλές ή έστω μέτριες επιδόσεις σε Πληροφορική και Οικονομία αρκούν συχνά για να εξασφαλίσουν επιτυχία, ακόμη και όταν τα Μαθηματικά βρίσκονται σε χαμηλό επίπεδο, γεγονός που καταδεικνύει ισχυρό μηχανισμό αντιστάθμισης. Αντίθετα, ταυτόχρονα χαμηλοί βαθμοί σε Πληροφορική και Οικονομία οδηγούν σε αποτυχία με πιθανότητα άνω του 92%, ενώ η χαμηλή επίδοση στη Γλώσσα επιτείνει περαιτέρω τον κίνδυνο. Συνολικά, τα αποτελέσματα δείχνουν ότι στην ομάδα αυτή η επιτυχία εξαρτάται σχεδόν αποκλειστικά από τις επιδόσεις στα μαθήματα ειδικότητας, με τα Μαθηματικά και τη Γλώσσα να παίζουν δευτερεύοντα, διορθωτικό ρόλο.

6.4. Συγκριτική ανάλυση των αλγορίθμων Gaussian Naive Bayes και Random Forest

Ο **Gaussian Naive Bayes** παρουσίασε σταθερά «καλή αλλά μέτρια» επίδοση στις τέσσερις ομάδες, με Accuracy περίπου 86%–89% (και αντίστοιχα υψηλές Balanced Accuracy), γεγονός που δείχνει ότι ως απλός πιθανοκρατικός ταξινομητής μπορεί να αποδώσει ικανοποιητικά σε εκπαιδευτικά δεδομένα και να παρέχει ισορροπημένη διάκριση μεταξύ «ΠΕΡΑΣΕ» και «ΔΕΝ ΠΕΡΑΣΕ». Ωστόσο, σε ομάδες με εντονότερη υπεροχή της κλάσης «ΠΕΡΑΣΕ» (π.χ. Θετικές, Υγείας) παρατηρήθηκε μειωμένη precision για την κλάση «ΔΕΝ ΠΕΡΑΣΕ», άρα περισσότερες περιπτώσεις που προβλέπονται ως αποτυχία αλλά τελικά ανήκουν στους επιτυχόντες. Παρ' όλα αυτά, το μοντέλο διατήρησε υψηλό recall για τους μη επιτυχόντες, κάτι που είναι χρήσιμο όταν προτεραιότητα είναι ο εντοπισμός υποψηφίων με αυξημένο κίνδυνο αποτυχίας.

Αντίθετα, ο **Random Forest** εμφάνισε σαφώς ανώτερη απόδοση σε όλες τις ομάδες, με Accuracy περίπου 96%–98% και πολύ κοντινά αποτελέσματα μεταξύ holdout και 5-fold cross-validation, ένδειξη ισχυρής γενίκευσης. Η υπεροχή αυτή είναι αναμενόμενη, καθώς ο Random Forest μπορεί να μοντελοποιεί μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ βαθμών, κάτι που ο Naive Bayes (με την υπόθεση ανεξαρτησίας χαρακτηριστικών) δεν αποτυπώνει επαρκώς. Επιπλέον, ο Random Forest προσέφερε αυξημένη ερμηνευσιμότητα σε επίπεδο εφαρμογής μέσω permutation importance και κανόνων (fixed cutoffs), αναδεικνύοντας διαφορετικούς «μηχανισμούς επιτυχίας» ανά ομάδα (π.χ. κυριαρχία Λατινικών στις Ανθρωπιστικές, τριπλέτα Χημεία–Φυσική–Μαθηματικά στις Θετικές, άξονας Βιολογία–Χημεία στην Υγείας, και ισχυρή εξάρτηση από Πληροφορική–Οικονομία στην Οικονομίας & Πληροφορικής). Συνολικά, τα ευρήματα δείχνουν ότι ο Naive Bayes μπορεί να λειτουργήσει ως αξιόπιστη λύση, αλλά ο Random Forest είναι καταλληλότερος όταν ζητείται μέγιστη ακρίβεια και παραγωγή πρακτικά αξιοποιήσιμων κανόνων για καθοδήγηση υποψηφίων.

Κεφάλαιο 7^ο: Εξαγωγή Κανόνων συσχέτισης

7.1. Κανόνες συσχέτισης σχετικά με τις Προτιμήσεις των Υποψηφίων

Για την εφαρμογή του αλγορίθμου **Apriori** πραγματοποιήθηκε συγχώνευση των αρχείων **ΦΥΛΟ**, **ΠΡΟΤΙΜΗΣΕΙΣ** και **ΑΠΟΤΕΛΕΣΜΑΤΑ** που αφορούν τα δεδομένα των Πανελλαδικών Εξετάσεων του έτους 2024. Η ενοποίηση έγινε με κοινό αναγνωριστικό τον *Κωδικό Υποψηφίου* και οδήγησε στη δημιουργία του αρχείου **MERGE_FYLO_PROT_APOT_2024.csv**, το οποίο περιλαμβάνει συνολικά **1.048.576** εγγραφές όπως βλέπουμε και στην Εικόνα 7.1.

Το παραγόμενο σύνολο δεδομένων συγκεντρώνει δημογραφικά χαρακτηριστικά, πληροφορίες εκπαιδευτικών προτιμήσεων και τελικά αποτελέσματα των υποψηφίων, παρέχοντας μια ολοκληρωμένη αναπαράσταση του πληθυσμού μελέτης. Η συγκεκριμένη διαδικασία ολοκλήρωσης δεδομένων κρίθηκε απαραίτητη, καθώς επέτρεψε τη μετατροπή των αρχικών πινάκων σε ενιαία συναλλακτική μορφή, κατάλληλη για την εξαγωγή κανόνων συσχέτισης μέσω του Apriori.

Στη συνέχεια πραγματοποιήθηκε έλεγχος πληρότητας, απομάκρυνση μη σχετικών πεδίων και τυποποίηση των κατηγορικών μεταβλητών, ώστε κάθε εγγραφή να αντιστοιχεί σε σύνολο χαρακτηριστικών έτοιμο προς ανάλυση. Η διαδικασία αυτή εξασφάλισε τη συνέπεια των δεδομένων και τη συμβατότητά τους με τις απαιτήσεις του αλγορίθμου.

ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟ ΦΥΛΟ	Σειρά Επιλογή	ΟΝΟΜΑ ΣΧΟΛΗΣ	Δ/ΝΣΗ Δ.Ε.	ΕΙΔΟΣ ΛΥΚΕΙΟΥ	ΗΜΕΡΗΣΙΟ/ΕΣΠΕΡΙΝΟ	ΟΜΑΔΑ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΥ - ΕΙΔ ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ	ΕΠΙΤΥΧΩΝ
00004bbc3e	Θήλυ		Δ/ΝΣΗ Δ.Ε. ΧΑΝΙΩΝ	ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ ΚΑΙ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚ ΔΕΝ ΠΕΡΑΣΕ	
0000c99e2f	Άρρεν						
0000f3f6b0	Άρρεν						
0001f97e02	Θήλυ	1 ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ Ι Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	2 ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	3 ΝΑΥΤΙΛΙΑΚΩΝ ΣΠΟΥΔΩΝ (Ι Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	4 ΔΙΕΘΝΩΝ ΚΑΙ ΕΥΡΩΠΑΪΚΩΙ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	5 ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	6 ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	7 ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ Ι Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	8 ΠΛΗΡΟΦΟΡΙΚΗΣ (ΑΘΗΝΑ) Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	9 ΜΑΡΚΕΤΙΝΓΚ ΚΑΙ ΕΠΙΚΟΙΝΩ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
0001f97e02	Θήλυ	10 ΔΙΕΘΝΩΝ ΚΑΙ ΕΥΡΩΠΑΪΚΩΙ Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. ΑΤΤΙΚΗΣ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤ ΠΕΡΑΣΕ	
00022a64cb	Άρρεν	1 ΤΟΥΡΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ (Γ Δ/ΝΣΗ Δ.Ε. ΔΩΔΕΚΑΝΗΣΟΥ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΤΟΥΡΙΣΤΙΚΩΝ ΣΠΟ' ΠΕΡΑΣΕ	
00022a64cb	Άρρεν	2 ΔΙΟΙΚΗΣΗΣ ΤΟΥΡΙΣΜΟΥ (Α Δ/ΝΣΗ Δ.Ε. ΔΩΔΕΚΑΝΗΣΟΥ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΤΟΥΡΙΣΤΙΚΩΝ ΣΠΟ' ΠΕΡΑΣΕ	
00022a64cb	Άρρεν	3 ΔΙΟΙΚΗΣΗΣ ΤΟΥΡΙΣΜΟΥ (Π Δ/ΝΣΗ Δ.Ε. ΔΩΔΕΚΑΝΗΣΟΥ		ΓΕΝΙΚΟ	ΗΜΕΡΗΣΙΟ	ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΠΛΗΡ ΤΟΥΡΙΣΤΙΚΩΝ ΣΠΟ' ΠΕΡΑΣΕ	

Εικόνα 7.1: Στιγμιότυπο αρχείου ΦΥΛΟ_ΠΡΟΤΙΜΗΣΕΙΣ_ΑΠΟΤΕΛΕΣΜΑΤΑ

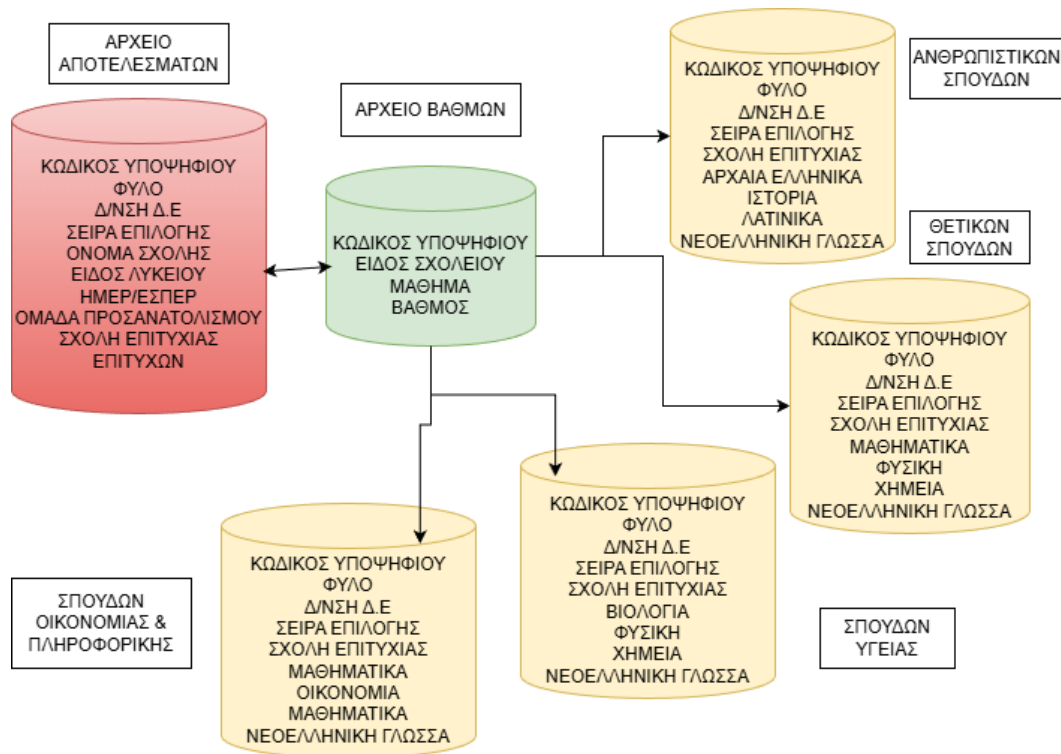
Στο αρχείο αυτό πραγματοποιήθηκαν επιπλέον διεργασίες, κατά τις οποίες διατηρήθηκαν αποκλειστικά οι υποψήφιοι των **Γενικών Λυκείων (ΓΕΛ) που Πέτυχαν την εισαγωγή τους σε κάποια σχολή**. Στη συνέχεια, παρέμεινε μόνο η Σειρά Επιλογής στην οποία το Όνομα Σχολής ταυτιζόταν με τη Σχολή Επιτυχίας. Το αρχείο αυτό ονομάστηκε **SEIRA_EPITYXIAS_2024.csv**.

Ακολούθως, το αρχείο αυτό συγχωνεύτηκε με το αρχείο **ΒΑΘΜΟΙ**, και με τη χρήση κώδικα Python τοποθετήθηκαν οι βαθμοί των τεσσάρων βασικών μαθημάτων σε ξεχωριστές στήλες. Τέλος, εφαρμόζοντας φίλτρο ως προς την Ομάδα Προσανατολισμού, δημιουργήθηκε ξεχωριστό αρχείο για κάθε ομάδα.

Για την **Ομάδα Ανθρωπιστικών Σπουδών** προέκυψε ένα αρχείο με 13.150 γραμμές, με τις εξής στήλες (Βλέπε Εικόνα 7.2 και Σχήμα 7.1):

ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ	ΦΥΛΟ	Σειρά Επιλογής	Δ/ΝΣΗ Δ.Ε.	ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ	ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ	ΙΣΤΟΡΙΑ Ο.Π.	ΛΑΤΙΝΙΚΑ Ο.Π.	ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩ
000cfd391	Θήλυ		1	Δ/ΝΣΗ Δ.Ε. ΜΕΣΣΗΝ ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗ	12.1	18.2	15.8	15
000dc94db2	Θήλυ		2	Δ/ΝΣΗ Δ.Ε. ΦΛΩΡΙΝ/ΑΣΤΥΦΥΛΑΚΩΝ (MC	12.9	14.3	18	12.6
000dc94db2	Θήλυ		2	Δ/ΝΣΗ Δ.Ε. ΦΛΩΡΙΝ/ΑΣΤΥΦΥΛΑΚΩΝ (MC	12.9	14.3	18	12.6
0010b971d6	Άρρεν		2	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝ ΝΟΜΙΚΗΣ (ΘΕΣΣΑΛ	16.1	18.6	19.9	16.2
0015c1d425	Θήλυ		7	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝ ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (Α	8.9	5	9.7	14.4
00164ff291	Θήλυ		3	Δ/ΝΣΗ Δ.Ε. ΔΥΤ. ΘΕ ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗ	16.7	18.2	19.7	15.8
00191288c6	Άρρεν		9	Δ/ΝΣΗ Δ.Ε. ΙΩΑΝΝΙΝ ΦΙΛΟΛΟΓΙΑΣ (ΙΩΑΝΝ	10.6	6.7	15.6	13.5
001ec674c2	Θήλυ		1	Δ/ΝΣΗ Δ.Ε. ΛΑΚΩΝ/ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤ	10.5	14.5	12	11.2
001f46b321	Θήλυ		5	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝ ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΕΚ	16.5	10.9	19.3	15.8
0020cc605f	Άρρεν		5	Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. Θ ΙΣΤΟΡΙΑΣ ΚΑΙ ΑΡΧΑ	7.8	11.9	8.9	12.6
002651ffc3	Θήλυ		7	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝ ΚΟΙΝΩΝΙΚΗΣ ΘΕΟΛ	12.3	3.9	13.8	12.9
00268825fc	Άρρεν		1	Δ/ΝΣΗ Δ.Ε. ΛΑΣΙΘΙΟ ΦΙΛΟΛΟΓΙΑΣ (ΡΕΘΥ	10.8	5.7	11.8	14
0028cce21c	Άρρεν		2	Δ/ΝΣΗ Δ.Ε. ΗΜΑΘΙΑ ΙΣΤΟΡΙΑΣ ΚΑΙ ΕΘΝ	8.7	4.6	9.9	14.1
002cb94931	Άρρεν		3	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝ ΕΠΙΚΟΙΝΩΝΙΑΣ ΚΑΙ Ι	14.8	17.3	17.2	13
002e20ccf6	Θήλυ		2	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝ ΑΓΩΓΗΣ ΚΑΙ ΦΡΟΝΤ	9.9	15.2	16	13.1

Εικόνα 7.2: Στιγμιότυπο αρχείου ΦΥΛΟ_ΣΧΟΛΗ_ΕΠΙΤΥΧΙΑΣ_ΒΑΘΜΟΙ



Σχήμα 7.1: Δημιουργία αρχείων κατάλληλων για τον εξαγωγή Κανόνων Συσχέτισης

Κατόπιν στην επόμενη φάση στο αρχείο προτιμήσεων όπως δόθηκε από το Υπουργείο Παιδείας, του οποίου ενδεικτικό απόσπασμα βλέπουμε παρακάτω (Εικόνα 7.3):

ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ	Σειρά Επιλ	ΟΝΟΜΑ ΣΧΟΛΗΣ
c2ab679ed7	1	ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ)
c2ab679ed7	2	ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΚΕΡΚΥΡΑ)
5c7ec7688f	1	ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ, ΑΡΧΕΙΟΝΟΜΙΑΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΗΣΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ)
5c7ec7688f	2	ΕΠΙΣΤΗΜΩΝ ΠΡΟΣΧΟΛΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΕΚΠΑΙΔΕΥΣΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ)
cf15b76fee	1	ΦΙΛΟΣΟΦΙΑΣ (ΑΘΗΝΑ)
cf15b76fee	2	ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤΕΡΟΒΑΘΜΙΑΣ ΕΚΠΑΙΔΕΥΣΗΣ (ΑΘΗΝΑ)
cf15b76fee	3	ΦΙΛΟΛΟΓΙΑΣ (ΠΑΤΡΑ)
cf15b76fee	4	ΘΕΩΡΙΑΣ ΚΑΙ ΙΣΤΟΡΙΑΣ ΤΗΣ ΤΕΧΝΗΣ (ΑΘΗΝΑ)
0229b83cf4	1	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ) - ΕΚΠΑ
0229b83cf4	2	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ) - ΠΑΝΤΕΙΟ
0229b83cf4	3	ΨΥΧΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ)
0229b83cf4	4	ΨΥΧΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΟ)
0229b83cf4	5	ΨΥΧΟΛΟΓΙΑΣ (ΦΛΩΡΙΝΑ)
0229b83cf4	6	ΨΥΧΟΛΟΓΙΑΣ (ΔΙΔΥΜΟΤΕΙΧΟ)

Εικόνα 7.3: Στιγμιότυπο αρχείου ΠΡΟΤΙΜΗΣΕΙΣ

Επιλέχθηκαν οι τέσσερις πρώτες προτιμήσεις (ΠΡΟΤ1, ΠΡΟΤ2, ΠΡΟΤ3, ΠΡΟΤ4) και τοποθετήθηκαν σε κάθε στήλη ξεχωριστά, μία γραμμή για κάθε Κωδικό Υποψηφίου σύμφωνα με την Εικόνα 7.4:

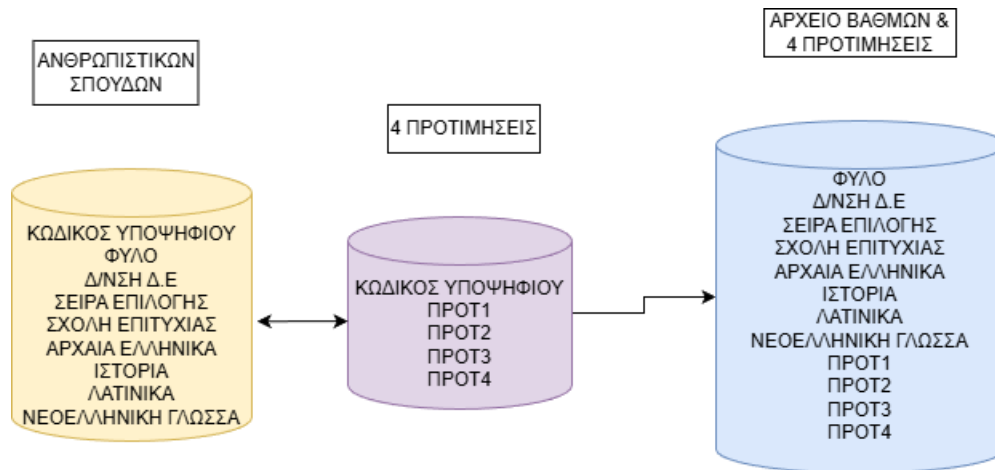
ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ	ΠΡΟΤ1	ΠΡΟΤ2	ΠΡΟΤ3	ΠΡΟΤ4
000cfd391	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ
000dc94db2	ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΑΣΤΩΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ
0010b971d6	ΝΟΜΙΚΗΣ (ΑΘΗΝΑ)	ΝΟΜΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΝΟΜΙΚΗΣ (ΚΟΜΟΤΗΝΗ)	ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
0015c1d425	ΚΟΙΝΩΝΙΚΗΣ ΕΡΓΑΣΙΑΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ
00164ff291	ΨΥΧΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΙΩΑΝΝΙΝΑ)
00191288c6	ΑΓΓΛΙΚΗΣ ΓΛΩΣΣΑΣ	ΑΓΓΛΙΚΗΣ ΓΛΩΣΣΑΣ	ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΑΣΤΩΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)
001ec674c2	ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤΕΡΟΒΑΘΜΙΑΣ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΑΡΧΑΙΟΛΟΓΙΑΣ	ΙΣΤΟΡΙΑΣ, ΑΡΧΑΙΟΛΟΓΙΑΣ	ΙΣΤΟΡΙΑΣ, ΑΡΧΑΙΟΛΟΓΙΑΣ
001f46b321	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΨΥΧΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΗ)	ΚΟΙΝΩΝΙΚΗΣ ΕΡΓΑΣΙΑΣ

Εικόνα 7.4: Στιγμιότυπο αρχείου ΠΡΟΤΙΜΗΣΕΙΣ 1,2,3,4

Κατόπιν συγχωνεύτηκαν τα δύο αρχεία: ΦΥΛΟ_ΣΧΟΛΗ_ΕΠΙΤΥΧΙΑΣ_ΒΑΘΜΟΙ και ΠΡΟΤΙΜΗΣΕΙΣ 1,2,3,4 σε ένα αρχείο MERGE_VATHMOI_PROT1234_ANTHROP_2024.csv όπου αφαιρέθηκε ο “ΚΩΔΙΚΟΣ ΥΠΟΨΗΦΙΟΥ”. Παρακάτω παρουσιάζεται ενδεικτικό τμήμα του αρχείου (Εικόνα 7.5) και η δημιουργία των αρχείων στο Σχήμα 7.2:

ΦΥΛΟ	Σειρά Επιλογής	Δ/ΝΣΗ Δ.Ε.	ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ	ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ	ΙΣΤΟΡΙΑ Ο.Π.	ΛΑΤΙΝΙΚΑ Ο.Π.	ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ	ΠΡΟΤ1	ΠΡΟΤ2	ΠΡΟΤ3	ΠΡΟΤ4
Θήλυ	1	Δ/ΝΣΗ Δ.Ε. ΜΕΣΣΟΝΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	12.1	18.2	15.8	15	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ
Θήλυ	2	Δ/ΝΣΗ Δ.Ε. ΦΛΩΡΙΝΑΣ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)	12.9	14.3	18	12.6	ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΑΣΤΩΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ
Θήλυ	2	Δ/ΝΣΗ Δ.Ε. ΦΛΩΡΙΝΑΣ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)	12.9	14.3	18	12.6	ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΑΣΤΩΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ
Άρρεν	2	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝΑΣ	ΝΟΜΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	16.1	18.6	19.9	16.2	ΝΟΜΙΚΗΣ (ΑΘΗΝΑ)	ΝΟΜΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΝΟΜΙΚΗΣ (ΚΟΜΟΤΗΝΗ)	ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
Θήλυ	7	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝΑΣ	ΚΟΙΝΩΝΟΛΟΓΙΑΣ (ΦΙΛΟΛΟΓΙΑΣ)	8.9	5	9.7	14.4	ΚΟΙΝΩΝΙΚΗΣ ΕΡΓΑΣΙΑΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ
Θήλυ	3	Δ/ΝΣΗ Δ.Ε. ΔΥΤ. ΘΕΣΣΑΛΟΝΙΚΗΣ	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	16.7	18.2	19.7	15.8	ΨΥΧΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΙΩΑΝΝΙΝΑ)
Άρρεν	9	Δ/ΝΣΗ Δ.Ε. ΙΩΑΝΝΙΝΩΝ	ΦΙΛΟΛΟΓΙΑΣ (ΙΩΑΝΝΙΝΑ)	10.6	6.7	15.6	13.5	ΑΓΓΛΙΚΗΣ ΓΛΩΣΣΑΣ	ΑΓΓΛΙΚΗΣ ΓΛΩΣΣΑΣ	ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΑΣΤΩΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟΝΟΤΕΚΝΟΝ)
Θήλυ	1	Δ/ΝΣΗ Δ.Ε. ΛΑΚΩΝΙΑΣ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤΕΡΟΒΑΘΜΙΑΣ	10.5	14.5	12	11.2	ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤΕΡΟΒΑΘΜΙΑΣ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΑΡΧΑΙΟΛΟΓΙΑΣ	ΙΣΤΟΡΙΑΣ, ΑΡΧΑΙΟΛΟΓΙΑΣ	ΙΣΤΟΡΙΑΣ, ΑΡΧΑΙΟΛΟΓΙΑΣ
Θήλυ	5	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝΑΣ	ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΕΡΕΥΝΑΣ	16.5	10.9	19.3	15.8	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝΑ)	ΨΥΧΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ)	ΨΥΧΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΗ)	ΚΟΙΝΩΝΙΚΗΣ ΕΡΓΑΣΙΑΣ
Άρρεν	5	Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. Θ ΑΘΗΝΑΣ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΑΡΧΑΙΟΛΟΓΙΑΣ	7.8	11.9	8.9	12.6	ΦΙΛΟΣΟΦΙΑΣ ΚΑΙ ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ	ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ	ΕΠΙΣΤΗΜΩΝ ΠΡΟΣΧΗΜΑΤΟΣ	ΕΠΙΣΤΗΜΩΝ ΠΡΟΣΧΗΜΑΤΟΣ
Θήλυ	7	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝΑΣ	ΚΟΙΝΩΝΙΚΗΣ ΘΕΟΛΟΓΙΑΣ	12.3	3.9	13.8	12.9	ΚΟΙΝΩΝΙΚΗΣ ΕΡΓΑΣΙΑΣ	ΚΟΙΝΩΝΙΚΗΣ ΑΝΘΡΩΠΟΛΟΓΙΑΣ	ΚΟΙΝΩΝΙΚΗΣ ΠΟΛΙΤΙΚΗΣ	ΚΟΙΝΩΝΟΛΟΓΙΑΣ (ΦΙΛΟΛΟΓΙΑΣ)
Άρρεν	1	Δ/ΝΣΗ Δ.Ε. ΛΑΣΙΘΙΟΥ	ΦΙΛΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΗ)	10.8	5.7	11.8	14	ΦΙΛΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΗ)	ΦΙΛΟΛΟΓΙΑΣ (ΠΑΤΡΑΣ)	ΦΙΛΟΛΟΓΙΑΣ (ΙΩΑΝΝΙΝΑ)	ΦΙΛΟΛΟΓΙΑΣ (ΚΑΛΑΜΑΤΑ)
Άρρεν	2	Δ/ΝΣΗ Δ.Ε. ΗΜΑΘΙΑΣ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΕΘΝΟΓΡΑΦΙΑΣ	8.7	4.6	9.9	14.1	ΚΟΙΝΩΝΙΚΗΣ ΠΟΛΙΤΙΚΗΣ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΕΘΝΟΓΡΑΦΙΑΣ	ΚΟΙΝΩΝΟΛΟΓΙΑΣ (ΦΙΛΟΛΟΓΙΑΣ)	ΚΟΙΝΩΝΟΛΟΓΙΑΣ (ΦΙΛΟΛΟΓΙΑΣ)

Εικόνα 7.5: Στιγμιότυπο αρχείου MERGE_VATHMOI_PROT1234_ANTHROP_2024.csv



Σχήμα 7.2: Δημιουργία αρχείων κατάλληλων για τον εξαγωγή Κανόνων Συσχέτισης

Στο τελικό αυτό αρχείο εφαρμόστηκε ο αλγόριθμος **Apriori**, ένας κλασικός αλγόριθμος εξόρυξης δεδομένων για την ανακάλυψη κανόνων συσχέτισης, με την αξιολόγηση των παραγόμενων κανόνων να βασίζεται στα μέτρα ενδιαφέροντος **support**, **confidence** και **lift** ως εξής:

$$\text{Confidence} \geq 0.6, \text{ support} \geq 0.005 \ \& \ \text{lift} \geq 1.2$$

Στη συνέχεια παρουσιάζονται οι σημαντικότεροι κανόνες συσχέτισης που προέκυψαν από την εφαρμογή του αλγορίθμου, οι οποίοι εμφανίζουν υψηλές τιμές στα μέτρα ενδιαφέροντος **confidence**, **support** και **lift**, και είναι ταξινομημένοι κατά φθίνουσα σειρά με βάση το κριτήριο του **lift**.

ΟΜΑΔΑ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ

✓ Basket dimensions: (13150, 560)

✓ Υπολογίστηκαν frequent itemsets

✓ Παρήχθησαν κανόνες: 3954

Οι 15 καλύτεροι κανόνες συσχέτισης (ΠΡΟΤ1 / ΠΡΟΤ2 → ΠΡΟΤ2 / ΠΡΟΤ3 / ΠΡΟΤ4)

1. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΓΑΛΛΙΚΗΣ ΓΛΩΣΣΑΣ ΚΑΙ ΦΙΛΟΛΟΓΙΑΣ (ΑΘΗΝΑ), τότε με πιθανότητα 55.9% επιλέγει επίσης: ΠΡΟΤ2_ΓΑΛΛΙΚΗΣ ΓΛΩΣΣΑΣ ΚΑΙ ΦΙΛΟΛΟΓΙΑΣ (ΘΕΣΣΑΛΟΝΙΚΗ).

(support=0.144%, lift=282.64)

2. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΑΝΩΤΕΡΗ ΣΧΟΛΗ ΤΟΥΡΙΣΤΙΚΗΣ ΕΚΠΑΙΔΕΥΣΗΣ ΚΡΗΤΗΣ (ΑΣΤΕΚ), τότε με πιθανότητα 85.7% επιλέγει επίσης: ΠΡΟΤ2_ΑΝΩΤΕΡΗ ΣΧΟΛΗ ΤΟΥΡΙΣΤΙΚΗΣ ΕΚΠΑΙΔΕΥΣΗΣ ΡΟΔΟΥ (ΑΣΤΕΡ).

(support=0.274%, lift=274.91)

3. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΑΘΗΝΑ), τότε με πιθανότητα 38.5% επιλέγει επίσης: ΠΡΟΤ3_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.114%, lift=210.74)
4. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΑΘΗΝΑ), τότε με πιθανότητα 57.1% επιλέγει επίσης: ΠΡΟΤ3_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΚΕΡΚΥΡΑ). (support=0.122%, lift=183.28)
5. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΨΥΧΟΛΟΓΩΝ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗ, τότε με πιθανότητα 69.1% επιλέγει επίσης: ΠΡΟΤ2_ΝΟΜΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗ. (support=0.357%, lift=178.22)
6. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 53.1% επιλέγει επίσης: ΠΡΟΤ3_ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ (ΚΕΡΚΥΡΑ). (support=0.129%, lift=170.39)
7. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΦΙΛΟΛΟΓΙΑΣ (ΚΑΛΑΜΑΤΑ), τότε με πιθανότητα 57.7% επιλέγει επίσης: ΠΡΟΤ2_ΙΣΤΟΡΙΑΣ, ΑΡΧΑΙΟΛΟΓΙΑΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗΣ ΠΟΛΙΤΙΣΜΙΚΩΝ ΑΓΑΘΩΝ (ΚΑΛΑΜΑΤΑ). (support=0.114%, lift=168.59)
8. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΘΕΑΤΡΙΚΩΝ ΣΠΟΥΔΩΝ (ΑΘΗΝΑ), τότε με πιθανότητα 29.6% επιλέγει επίσης: ΠΡΟΤ2_ΘΕΑΤΡΙΚΩΝ ΣΠΟΥΔΩΝ (ΝΑΥΠΛΙΟ). (support=0.122%, lift=155.85)
9. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΠΙΚΟΙΝΩΝΙΑΣ ΚΑΙ ΨΗΦΙΑΚΩΝ ΜΕΣΩΝ (ΚΑΣΤΟΡΙΑ), τότε με πιθανότητα 41.0% επιλέγει επίσης: ΠΡΟΤ2_ΨΗΦΙΑΚΩΝ ΜΕΣΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΣ (ΑΡΓΟΣΤΟΛΙ). (support=0.122%, lift=154.14)
10. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΝΟΜΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗ, τότε με πιθανότητα 48.8% επιλέγει επίσης: ΠΡΟΤ2_ΨΥΧΟΛΟΓΩΝ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗ. (support=0.297%, lift=152.63)
11. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΣΧΟΛΗ ΔΟΚΙΜΩΝ ΣΗΜΑΙΟΦΟΡΩΝ ΛΙΜΕΝΙΚΟΥ ΣΩΜΑΤΟΣ - ΕΛΛΗΝΙΚΗΣ ΑΚΤΟΦΥΛΑΚΗΣ (Σ.Δ.Σ.Λ.Σ. - ΕΛ.ΑΚΤ.) (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ),

τότε με πιθανότητα 53.2% επιλέγει επίσης: ΠΡΟΤ2_ΣΧΟΛΗ ΔΟΚΙΜΩΝ ΛΙΜΕΝΟΦΥΛΑΚΩΝ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ). (support=0.190%, lift=148.82)

12. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΟ),

τότε με πιθανότητα 35.3% επιλέγει επίσης: ΠΡΟΤ2_ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (ΜΥΤΙΛΗΝΗ). (support=0.137%, lift=132.61)

13. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΑΞΙΩΜΑΤΙΚΩΝ ΠΥΡΟΣΒΕΣΤΙΚΗΣ ΑΚΑΔΗΜΙΑΣ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ),

τότε με πιθανότητα 62.5% επιλέγει επίσης: ΠΡΟΤ2_ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΩΝ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ). (support=0.380%, lift=124.53)

14. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (ΡΕΘΥΜΝΟ),

τότε με πιθανότητα 26.7% επιλέγει επίσης: ΠΡΟΤ3_ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (ΜΥΤΙΛΗΝΗ). (support=0.122%, lift=116.89)

15. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΑΙΔΑΓΩΓΙΚΟ ΝΗΠΙΑΓΩΓΩΝ (ΙΩΑΝΝΙΝΑ), τότε με πιθανότητα 29.8% επιλέγει επίσης: ΠΡΟΤ2_ΑΓΩΓΗΣ ΚΑΙ ΦΡΟΝΤΙΔΑΣ ΣΤΗΝ ΠΡΩΙΜΗ ΠΑΙΔΙΚΗ ΗΛΙΚΙΑ (ΙΩΑΝΝΙΝΑ).

(support=0.106%, lift=97.93)

ΟΜΑΔΑ ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ

✓ Basket dimensions: (10541, 936)

✓ Υπολογίστηκαν frequent itemsets

✓ Παρήχθησαν κανόνες: 4230

Οι 12 καλύτεροι κανόνες συσχέτισης (ΠΡΟΤ1 / ΠΡΟΤ2 → ΠΡΟΤ2 / ΠΡΟΤ3 / ΠΡΟΤ4)

1. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ),

τότε με πιθανότητα 75.0% επιλέγει επίσης: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ).

(support=0.114%, lift=564.70)

2. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ),

τότε με πιθανότητα 60.0% επιλέγει επίσης: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.142%, lift=332.87)

3. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΣΤΑΤΙΣΤΙΚΗΣ (ΑΘΗΝΑ),

τότε με πιθανότητα 40.6% επιλέγει επίσης: ΠΡΟΤ2_ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ (ΠΕΙΡΑΙΑΣ). (support=0.123%, lift=225.38)

4. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΦΥΣΙΚΗΣ (ΛΑΜΙΑ),

τότε με πιθανότητα 55.0% επιλέγει επίσης: ΠΡΟΤ2_ΦΥΣΙΚΗΣ (ΚΑΒΑΛΑ) - ΠΡΩΗΝ ΤΜΗΜΑ ΤΟΥ ΔΙ.ΠΑ.Ε.. (support=0.104%, lift=222.98)

5. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΥΕΛΠΙΔΩΝ (ΣΣΕ) - ΣΩΜΑΤΑ,

τότε με πιθανότητα 78.6% επιλέγει επίσης: ΠΡΟΤ2_ΕΥΕΛΠΙΔΩΝ (ΣΣΕ) - ΟΠΛΑ.
(support=0.209%, lift=212.36)

6. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΑΘΗΝΑ), τότε με πιθανότητα 52.8% επιλέγει επίσης: ΠΡΟΤ2_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.180%, lift=206.05)

7. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΙΚΑΡΩΝ (ΣΙ) ΜΕΤΕΩΡΟΛΟΓΩΝ,

τότε με πιθανότητα 52.4% επιλέγει επίσης: ΠΡΟΤ3_ΙΚΑΡΩΝ (ΣΙ) ΕΛΕΓΚΤΕΣ ΑΕΡΑΜΥΝΑΣ.
(support=0.104%, lift=184.05)

8. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΣΧΟΛΗ ΠΛΟΙΑΡΧΩΝ,

τότε με πιθανότητα 51.8% επιλέγει επίσης: ΠΡΟΤ2_ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ.
(support=0.275%, lift=181.96)

9. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΑΘΗΝΑ),

τότε με πιθανότητα 65.6% επιλέγει επίσης: ΠΡΟΤ3_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΠΕΙΡΑΙΑΣ).
(support=0.199%, lift=168.72)

10. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΦΥΣΙΚΗΣ (ΛΑΜΙΑ),

τότε με πιθανότητα 40.7% επιλέγει επίσης: ΠΡΟΤ3_ΦΥΣΙΚΗΣ (ΚΑΒΑΛΑ) - ΠΡΩΗΝ ΤΜΗΜΑ ΤΟΥ ΔΙ.ΠΑ.Ε.. (support=0.104%, lift=165.17)

11. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΙΚΑΡΩΝ (ΣΙ) ΜΗΧΑΝΙΚΟΙ (ΣΜΑ), τότε με πιθανότητα 32.0% επιλέγει επίσης: ΠΡΟΤ2_ΙΚΑΡΩΝ (ΣΙ) ΜΕΤΕΩΡΟΛΟΓΩΝ. (support=0.152%, lift=160.62)

12. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΑΞΙΩΜΑΤΙΚΩΝ ΠΥΡΟΣΒΕΣΤΙΚΗΣ ΑΚΑΔΗΜΙΑΣ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ), τότε με πιθανότητα 54.0% επιλέγει επίσης: ΠΡΟΤ2_ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΩΝ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ). (support=0.323%, lift=153.75)

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΥΓΕΙΑΣ

✓ Basket dimensions: (9214, 565)

✓ Υπολογίστηκαν frequent itemsets

✓ Παρήχθησαν κανόνες: 5142

Οι 10 καλύτεροι κανόνες συσχέτισης (ΠΡΟΤ1 / ΠΡΟΤ2 → ΠΡΟΤ2 / ΠΡΟΤ3 / ΠΡΟΤ4)

1. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 44.0% επιλέγει επίσης: ΠΡΟΤ4_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΣΕΡΡΕΣ). (support=0.119%, lift=225.23)

2. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΚΤΗΝΙΑΤΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 36.1% επιλέγει επίσης: ΠΡΟΤ3_ΚΤΗΝΙΑΤΡΙΚΗΣ (ΚΑΡΔΙΤΣΑ). (support=0.141%, lift=195.72)

3. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΖΑΚΥΝΘΟΣ), τότε με πιθανότητα 50.0% επιλέγει επίσης: ΠΡΟΤ3_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΜΥΤΙΛΗΝΗ). (support=0.109%, lift=170.63)

4. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΑΘΗΝΑ), τότε με πιθανότητα 31.2% επιλέγει επίσης: ΠΡΟΤ3_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΣΕΡΡΕΣ). (support=0.109%, lift=143.97)

5. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΜΥΤΙΛΗΝΗ), τότε με πιθανότητα 50.0% επιλέγει επίσης: ΠΡΟΤ3_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΖΑΚΥΝΘΟΣ). (support=0.119%, lift=124.51)

6. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΟΔΟΝΤΙΑΤΡΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗΣ,
τότε με πιθανότητα 46.0% επιλέγει επίσης: ΠΡΟΤ3_ΦΑΡΜΑΚΕΥΤΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗΣ.
(support=0.250%, lift=111.54)

7. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΛΟΓΟΘΕΡΑΠΕΙΑΣ (ΠΑΤΡΑ),
τότε με πιθανότητα 50.0% επιλέγει επίσης: ΠΡΟΤ2_ΛΟΓΟΘΕΡΑΠΕΙΑΣ (ΙΩΑΝΝΙΝΑ).
(support=0.217%, lift=98.02)

8. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΜΥΤΙΛΗΝΗ),
τότε με πιθανότητα 45.5% επιλέγει επίσης: ΠΡΟΤ2_ΠΕΡΙΒΑΛΛΟΝΤΟΣ (ΛΑΡΙΣΑ).
(support=0.109%, lift=87.25)

9. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΛΟΓΟΘΕΡΑΠΕΙΑΣ (ΙΩΑΝΝΙΝΑ),
τότε με πιθανότητα 45.8% επιλέγει επίσης: ΠΡΟΤ2_ΛΟΓΟΘΕΡΑΠΕΙΑΣ (ΠΑΤΡΑ).
(support=0.119%, lift=75.41)

10. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΙΑΤΡΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗΣ,
τότε με πιθανότητα 40.9% επιλέγει επίσης: ΠΡΟΤ2_ΟΔΟΝΤΙΑΤΡΙΚΟ (ΣΣΑΣ) ΘΕΣ/ΝΙΚΗΣ.
(support=0.488%, lift=75.39)

ΟΜΑΔΑ ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ

- ✓ Basket dimensions: (16341, 715)
- ✓ Υπολογίστηκαν frequent itemsets
- ✓ Παρήχθησαν κανόνες: 4014

Οι 15 καλύτεροι κανόνες συσχέτισης (ΠΡΟΤ1 / ΠΡΟΤ2 → ΠΡΟΤ2 / ΠΡΟΤ3 / ΠΡΟΤ4)

1. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ),
τότε με πιθανότητα 41.2% επιλέγει επίσης: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ3_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.630%, lift=64.12)

2. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ3_ΠΛΗΡΟΦΟΡΙΚΗΣ

(ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 98.1% επιλέγει επίσης: ΠΡΟΤ1_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.630%, lift=64.12)

3. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ3_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 98.1% επιλέγει επίσης: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.630%, lift=48.43)

4. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ (ΑΘΗΝΑ), ΠΡΟΤ3_ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ (ΑΘΗΝΑ), τότε με πιθανότητα 94.3% επιλέγει επίσης: ΠΡΟΤ2_ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ (ΠΕΙΡΑΙΑΣ). (support=0.606%, lift=47.85)

5. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΑΞΙΩΜΑΤΙΚΩΝ ΠΥΡΟΣΒΕΣΤΙΚΗΣ ΑΚΑΔΗΜΙΑΣ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ), τότε με πιθανότητα 50.8% επιλέγει επίσης: ΠΡΟΤ2_ΣΧΟΛΗ ΠΥΡΟΣΒΕΣΤΩΝ (ΜΟΝΟ ΓΙΑ ΠΟΛΙΤΕΣ). (support=0.936%, lift=45.64)

6. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ (ΑΘΗΝΑ), ΠΡΟΤ3_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΠΕΙΡΑΙΑΣ),
τότε με πιθανότητα 53.9% επιλέγει επίσης: ΠΡΟΤ2_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΑΘΗΝΑ), ΠΡΟΤ4_ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ (ΑΘΗΝΑ). (support=0.979%, lift=43.80)

7. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 32.4% επιλέγει επίσης: ΠΡΟΤ4_ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ3_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=0.722%, lift=42.72)

8. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ4_ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 81.9% επιλέγει επίσης: ΠΡΟΤ3_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ), ΠΡΟΤ1_ΠΛΗΡΟΦΟΡΙΚΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ).
(support=0.722%, lift=42.65)

9. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ ΕΚΠΑΙΔΕΥΣΗΣ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 37.0% επιλέγει επίσης: ΠΡΟΤ2_ΠΑΙΔΑΓΩΓΙΚΟ ΔΗΜΟΤΙΚΗΣ ΕΚΠΑΙΔΕΥΣΗΣ (ΒΟΛΟΣ). (support=0.594%, lift=39.80)

10. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΠΟΛΟΓΙΣΤΩΝ (ΘΕΣΣΑΛΟΝΙΚΗ), τότε με πιθανότητα 80.4% επιλέγει επίσης: ΠΡΟΤ2_ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (ΘΕΣΣΑΛΟΝΙΚΗ). (support=1.230%, lift=39.69)

11. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΤΟΥΡΙΣΜΟΥ (ΚΕΡΚΥΡΑ), τότε με πιθανότητα 50.7% επιλέγει επίσης: ΠΡΟΤ3_ΟΙΚΟΝΟΜΙΚΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΤΟΥΡΙΣΜΟΥ (ΧΙΟΣ). (support=0.208%, lift=97.56)

12. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ2_ΙΚΑΡΩΝ (ΣΙ) ΕΡΕΥΝΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ, τότε με πιθανότητα 39.7% επιλέγει επίσης: ΠΡΟΤ4_ΙΚΑΡΩΝ (ΣΙ) ΕΦΟΔΙΑΣΤΩΝ. (support=0.165%, lift=95.42)

13. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΚΑΙ ΤΟΥΡΙΣΜΟΥ (ΗΡΑΚΛΕΙΟ), τότε με πιθανότητα 33.8% επιλέγει επίσης: ΠΡΟΤ2_ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ (ΗΡΑΚΛΕΙΟ). (support=0.141%, lift=93.68)

14. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ (ΚΟΖΑΝΗ), τότε με πιθανότητα 42.5% επιλέγει επίσης: ΠΡΟΤ2_ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ (ΚΑΒΑΛΑ) - ΠΡΩΗΝ ΤΜΗΜΑ ΤΟΥ ΔΙ.ΠΑ.Ε.. (support=0.104%, lift=86.81)

15. ► Εάν ένας μαθητής επιλέξει: ΠΡΟΤ1_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΣΕΡΡΕΣ), τότε με πιθανότητα 78.1% επιλέγει επίσης: ΠΡΟΤ2_ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗΣ ΑΓΩΓΗΣ ΚΑΙ ΑΘΛΗΤΙΣΜΟΥ (ΚΟΜΟΤΗΝΗ). (support=0.153%, lift=79.79)

Συμπεράσματα

Στην Ομάδα Ανθρωπιστικών Σπουδών, οι ισχυρότεροι κανόνες αφορούν ομοειδή τμήματα σε διαφορετικές πόλεις (π.χ. Γαλλικής Φιλολογίας Αθήνα/Θεσσαλονίκη, Μουσικών Σπουδών Αθήνα/Θεσσαλονίκη/Κέρκυρα, Θεατρικών Σπουδών Αθήνα/Ναύπλιο), γεγονός που υποδηλώνει ότι οι υποψήφιοι διατηρούν σταθερό το γνωστικό πεδίο και «μετακινούν» τη γεωγραφία ως εναλλακτική. Παράλληλα, εμφανίζονται χαρακτηριστικοί κανόνες εντός ειδικών κατηγοριών (π.χ. ΣΣΑΣ: Ψυχολόγων/Νομικό), όπου οι επιλογές φαίνεται να συγκροτούν «πακέτα» λόγω κοινών θεσμικών/επαγγελματικών χαρακτηριστικών. Στις Θετικές Σπουδές και στην Ομάδα Οικονομίας & Πληροφορικής, οι κανόνες συγκεντρώνονται έντονα γύρω από συγγενείς ειδικότητες Πληροφορικής/Εφαρμοσμένης Πληροφορικής και Οικονομικών σε διαφορετικά ιδρύματα ή κατευθύνσεις.

Στην Ομάδα Σπουδών Υγείας, οι κανόνες αναδεικνύουν επίσης ομάδες επιλογών που συνδέονται είτε με γεωγραφική εναλλαγή (π.χ. Περιβάλλοντος Μυτιλήνη/Ζάκυνθος ή σύνδεση με Λάρισα), είτε με συγγενείς σχολές/κλάδους (π.χ. Κτηνιατρική Θεσσαλονίκη/Καρδίτσα, Λογοθεραπεία Πάτρα/Ιωάννινα), είτε με ειδικές στρατιωτικές/παραγωγικές σχολές (ΣΣΑΣ: Ιατρικό/Οδοντιατρικό). Συνολικά, οι κανόνες συσχέτισης προσφέρουν μια «χαρτογράφηση» του τρόπου με τον οποίο οι υποψήφιοι οργανώνουν τις επιλογές τους, αποτυπώνοντας ότι οι μετακινήσεις στις προτιμήσεις γίνονται κυρίως εντός του ίδιου ή πολύ συγγενούς γνωστικού πεδίου, με μεταβολές που αφορούν κυρίως την πόλη/ίδρυμα ή την εξειδίκευση. Αυτό το εύρημα είναι ιδιαίτερα χρήσιμο για την ερμηνεία της συμπεριφοράς επιλογών και μπορεί να αξιοποιηθεί τόσο στην εκπαιδευτική συμβουλευτική όσο και στη μελέτη των προτιμήσεων εισαγωγής ανά πεδίο.

7.2. Κανόνες συσχέτισης που αφορούν το Φύλο και τη Σειρά Επιτυχίας των Υποψηφίων ως προς την προτίμησή τους

Αρχικά, για την Ομάδα Ανθρωπιστικών Σπουδών χρησιμοποιήθηκε το αρχείο **MERGE_VATHMOI_PROT1234_ANTHROP_2024.csv**, στο οποίο διατηρήθηκαν οι στήλες: «ΦΥΛΟ», «Δ/ΝΣΗ Δ.Ε.», «ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ», «ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.», «ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.», «ΙΣΤΟΡΙΑ Ο.Π.», «ΛΑΤΙΝΙΚΑ Ο.Π.» και «Σειρά Επιλογής».

Όπου η στήλη «Σειρά Επιλογής» υπέστη διαδικασία διακριτοποίησης στις κατηγορίες «1η επιλογή», «2η-3η», «4η-7η», «8η-12η» και «λοιπές επιλογές», σύμφωνα με τη θέση κατάταξης της Σχολής Επιτυχίας στο μηχανογραφικό δελτίο κάθε υποψηφίου. Κατόπιν της ανωτέρω επεξεργασίας δημιουργήθηκε το αρχείο **SEIRA_EPILOGHS_EPITYCHIA_ANTHROP_2024.csv**, από το οποίο παρουσιάζεται στη συνέχεια ενδεικτικό τμήμα (Εικόνα 7.6).

ΦΥΛΟ	Δ/ΝΣΗ Δ.Ε.	ΣΧΟΛΗ ΕΠΙΤΥΧΙΑΣ	ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ	ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩ	ΙΣΤΟΡΙΑ Ο.Π.	ΛΑΤΙΝΙΚΑ Ο.Π.	ΣΕΙΡΑ ΕΠΙΤΥΧΙΑΣ
Θήλυ	Δ/ΝΣΗ Δ.Ε. ΜΕΣΣΗΝ	ΕΠΙΣΤΗΜΗΣ ΦΥΣΙΚΗ	12.1	15	18.2	15.8	1η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. ΦΑΩΡΙΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟ	12.9	12.6	14.3	18	2η-3η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. ΦΑΩΡΙΝ	ΑΣΤΥΦΥΛΑΚΩΝ (ΜΟ	12.9	12.6	14.3	18	2η-3η επιλογή
Άρρεν	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝ	ΝΟΜΙΚΗΣ (ΘΕΣΣΑΛ	16.1	16.2	18.6	19.9	2η-3η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝ	ΚΟΙΝΩΝΙΟΛΟΓΙΑΣ (Α	8.9	14.4	5	9.7	4η-7η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. ΔΥΤ. ΘΕ	ΨΥΧΟΛΟΓΙΑΣ (ΑΘΗΝ	16.7	15.8	18.2	19.7	2η-3η επιλογή
Άρρεν	Δ/ΝΣΗ Δ.Ε. ΙΩΑΝΝΙΝ	ΦΙΛΟΛΟΓΙΑΣ (ΙΩΑΝΝ	10.6	13.5	6.7	15.6	8η-12η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. ΛΑΚΩΝΙ	ΠΑΙΔΑΓΩΓΙΚΟ ΔΕΥΤ	10.5	11.2	14.5	12	1η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. Δ ΑΘΗΝ	ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΕΚ	16.5	15.8	10.9	19.3	4η-7η επιλογή
Άρρεν	Δ/ΝΣΗ Δ.Ε. ΑΝΑΤ. Θ	ΙΣΤΟΡΙΑΣ ΚΑΙ ΑΡΧΑ	7.8	12.6	11.9	8.9	4η-7η επιλογή
Θήλυ	Δ/ΝΣΗ Δ.Ε. Β ΑΘΗΝ	ΚΟΙΝΩΝΙΚΗΣ ΘΕΟΛ	12.3	12.9	3.9	13.8	4η-7η επιλογή
Άρρεν	Δ/ΝΣΗ Δ.Ε. ΛΑΣΙΘΙΟ	ΦΙΛΟΛΟΓΙΑΣ (ΡΕΘΥ	10.8	14	5.7	11.8	1η επιλογή

Εικόνα 7.6. Στιγμιότυπο αρχείου ΒΑΘΜΟΙ_ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ

Πάνω στο οποίο εφαρμόστηκε ο αλγόριθμος Apriori με την εξής διακριτοποίηση στους βαθμούς:

`bins = [0, 10, 13, 15, 18, 20]`

`labels = ["ΜΕΤΡΙΑ", "ΣΧΕΤΙΚΑ ΚΑΛΗ", "ΚΑΛΗ", "ΠΟΛΥ ΚΑΛΗ", "ΑΡΙΣΤΗ"]`

Και με βάση τα εξής μέτρα ενδιαφέροντος:

min_support=0.01, confidence =0.2, lift > 1.0

Με αποτέλεσμα να προκύψουν οι παρακάτω 15 καλύτεροι Κανόνες Συσχέτισης σχετικά με το ΦΥΛΟ, το ΜΟ των Μαθημάτων και τη Σειρά Επιτυχίας:

1. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΑΡΙΣΤΗ, ΦΥΛΟ=Άρρεν
3. τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=99.4%, support=1.25%, lift=2.70

2. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΠΟΛΥ ΚΑΛΗ, ΦΥΛΟ=Άρρεν
4. τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=44.2%, support=2.75%, lift=1.20

3. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΠΟΛΥ ΚΑΛΗ, ΦΥΛΟ=Άρρεν
5. τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=2η-3η επιλογή με confidence=36.0%, support=2.24%, lift=1.24

4. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΚΑΛΗ, ΦΥΛΟ=Άρρεν τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=42.4%, support=1.75%, lift=1.15

5. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΚΑΛΗ, ΦΥΛΟ=Άρρεν τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=2η-3η επιλογή με confidence=29.5%, support=1.22%, lift=1.02

6. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΣΧΕΤΙΚΑ ΚΑΛΗ, ΦΥΛΟ=Άρρεν τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=2η-3η επιλογή με confidence=31.9%, support=1.96%, lift=1.10

7. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΣΧΕΤΙΚΑ ΚΑΛΗ, ΦΥΛΟ=Άρρεν τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=4η-7η επιλογή με confidence=24.7%, support=1.52%, lift=1.12
6. -----
8. Εάν: ΦΥΛΟ=Θήλυ, ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΑΡΙΣΤΗ, ΣΧΟΛΗ_ΕΠΙΤΥΧΙΑΣ =ΝΟΜΙΚΗΣ (ΑΘΗΝΑ), τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=100.0%, support=1.83%, lift=2.72

9. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΑΡΙΣΤΗ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=99.0%, support=4.68%, lift=2.69

10. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΠΟΛΥ ΚΑΛΗ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=1η επιλογή με confidence=44.2%, support=10.84%, lift=1.20

11. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΠΟΛΥ ΚΑΛΗ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=2η-3η επιλογή με confidence=34.5%, support=8.47%, lift=1.19

12. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΚΑΛΗ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ= 2η-3η επιλογή με confidence=33.3%, support=5.32%, lift=1.15

13. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΣΧΕΤΙΚΑ ΚΑΛΗ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=2η-3η επιλογή με confidence=29.0%, support=7.19%, lift=1.00

14. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΜΕΤΡΙΑ, ΦΥΛΟ=Θήλυ

τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=4η-7η επιλογή με confidence=24.7%, support=2.33%, lift=1.12

15. Εάν: ΜΟ_ΚΑΤΗΓΟΡΙΑ=ΜΕΤΡΙΑ, ΦΥΛΟ=Θήλυ τότε: ΣΕΙΡΑ_ΕΠΙΤΥΧΙΑΣ=υπόλοιπες επιλογές με confidence=20.5%, support=1.93%, lift=5.44

Κεντρικό εύρημα αποτελεί η εξαιρετικά ισχυρή συσχέτιση της κατηγορίας «Άριστη» στον μέσο όρο βαθμολογίας με την επίτευξη εισαγωγής στην πρώτη επιλογή, ανεξαρτήτως φύλου, με τιμές εμπιστοσύνης που προσεγγίζουν ή αγγίζουν το 100%. Το αποτέλεσμα αυτό καταδεικνύει ότι η πολύ υψηλή ακαδημαϊκή επίδοση αποτελεί καθοριστικό παράγοντα επιτυχίας στην κορυφή των προτιμήσεων. Παράλληλα, στους υποψηφίους με «Πολύ Καλή» ή «Καλή» επίδοση παρατηρείται σταδιακή μετατόπιση της σειράς επιτυχίας προς τη 2η-3η επιλογή, γεγονός που αποτυπώνει με σαφήνεια την κλιμακωτή σχέση μεταξύ μέσου όρου βαθμολογίας και ανταγωνιστικότητας των σχολών επιλογής.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι διαφοροποιήσεις ως προς το φύλο, καθώς για τις μαθήτριες εντοπίζονται ισχυρότεροι και συχνότεροι κανόνες που συνδέουν υψηλή επίδοση με επιτυχία στην 1η επιλογή, αλλά και σαφέστερη συσχέτιση χαμηλότερων κατηγοριών μέσου όρου με επιτυχία σε μεταγενέστερες επιλογές. Οι κανόνες αυτοί υποδηλώνουν μεγαλύτερη διαβάθμιση στη σειρά επιτυχίας σε συνάρτηση με τη συνολική επίδοση, ενώ αναδεικνύουν και τη συνύπαρξη ακαδημαϊκών και κοινωνικών παραμέτρων στη διαδικασία επιλογής και εισαγωγής.

7.3. Κανόνες Συσχέτισης σχετικές με την Επίδοση των Μαθημάτων

Οι κανόνες συσχέτισης που εξάγονται επικεντρώνονται κυρίως στις σχέσεις μεταξύ των επιπέδων επίδοσης στα υπό μελέτη μαθήματα, τα οποία κατηγοριοποιήθηκαν σε χαμηλή (<10), μέτρια (10-15) και υψηλή (≥15) βαθμολογία. Η εξαγωγή των κανόνων πραγματοποιήθηκε με ελάχιστη υποστήριξη (min_support) ίση με 0.05, ελάχιστη εμπιστοσύνη (confidence) ≥ 0.60 και δείκτη ανύψωσης (lift) ≥ 1.05.

Ομάδα Σπουδών Οικονομίας & Πληροφορικής

✓ Φορτώθηκε το αρχείο

Σύνολο γραμμών: 16341

✓ Στήλες βαθμών: ['ΜΑΘΗΜΑΤΙΚΑ Ο.Π.', 'ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.', 'ΟΙΚΟΝΟΜΙΑ Ο.Π.', 'ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.']

✓ Basket dimensions: (16341, 12)

✓ Frequent itemsets: 70

✓ Παρήχθησαν κανόνες: 230

✓ Τελικοί κανόνες: 10

ΟΙ 10 ΚΑΛΥΤΕΡΟΙ ΚΑΝΟΝΕΣ ΕΠΙΔΟΣΕΩΝ (ΚΛΙΜΑΚΑ 0-20)

Διακριτοποίηση βαθμών:

Χαμηλή: <10 | Μέτρια: 10-15 | Υψηλή: ≥15

1. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 98.9% έχει επίσης: ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.=Υψηλή. (support=8.6%, lift=1.70)
2. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 97.6% έχει επίσης: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Υψηλή. (support=8.5%, lift=1.70)

3. ► Εάν ένας υποψήφιος έχει: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 95.4% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Χαμηλή. (support=11.4%, lift=1.62)
4. ► Εάν ένας υποψήφιος έχει: ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.=Χαμηλή, τότε με πιθανότητα 93.7% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Χαμηλή. (support=12.3%, lift=1.59)
5. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 84.8% έχει επίσης: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Υψηλή. (support=26.8%, lift=1.48)
6. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 84.7% έχει επίσης: ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.=Υψηλή. (support=26.8%, lift=1.46)
7. ► Εάν ένας υποψήφιος έχει: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 85.0% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Χαμηλή. (support=25.3%, lift=1.45)
8. ► Εάν ένας υποψήφιος έχει: ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.=Μέτρια, τότε με πιθανότητα 84.7% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Χαμηλή. (support=24.4%, lift=1.44)
9. ► Εάν ένας υποψήφιος έχει: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Υψηλή, τότε με πιθανότητα 77.9% έχει επίσης: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Υψηλή. (support=18.0%, lift=1.36)
10. ► Εάν ένας υποψήφιος έχει: ΠΛΗΡΟΦΟΡΙΚΗ Ο.Π.=Υψηλή, τότε με πιθανότητα 77.7% έχει επίσης: ΟΙΚΟΝΟΜΙΑ Ο.Π.=Υψηλή. (support=45.2%, lift=1.35)

Ομάδα Ανθρωπιστικών Σπουδών

- ✓ Φορτώθηκε το αρχείο
Σύνολο γραμμών: 13150
- ✓ Στήλες βαθμών: ['ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.', 'ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.', 'ΙΣΤΟΡΙΑ Ο.Π.', 'ΛΑΤΙΝΙΚΑ Ο.Π.']
- ✓ Basket dimensions: (13150, 12)
- ✓ Frequent itemsets: 78
- ✓ Παρήχθησαν κανόνες: 238
- ✓ Τελικοί κανόνες: 10

ΟΙ 10 ΚΑΛΥΤΕΡΟΙ ΚΑΝΟΝΕΣ ΕΠΙΔΟΣΕΩΝ (ΚΛΙΜΑΚΑ 0-20)

Διακριτοποίηση βαθμών:
Χαμηλή: <10 | Μέτρια: 10-15 | Υψηλή: ≥15

1. ► Εάν ένας υποψήφιος έχει: ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 78.1% έχει επίσης: ΙΣΤΟΡΙΑ Ο.Π.=Υψηλή. (support=18.4%, lift=2.35)
2. ► Εάν ένας υποψήφιος έχει: ΛΑΤΙΝΙΚΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 88.6% έχει επίσης: ΙΣΤΟΡΙΑ Ο.Π.=Χαμηλή. (support=8.3%, lift=2.05)

3. ► Εάν ένας υποψήφιος έχει: ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 79.7% έχει επίσης: ΙΣΤΟΡΙΑ Ο.Π.=Χαμηλή. (support=14.8%, lift=1.84)
4. ► Εάν ένας υποψήφιος έχει: ΛΑΤΙΝΙΚΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 74.8% έχει επίσης: ΙΣΤΟΡΙΑ Ο.Π.=Χαμηλή. (support=20.8%, lift=1.73)
5. ► Εάν ένας υποψήφιος έχει: ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 81.3% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Υψηλή. (support=19.2%, lift=1.73)
6. ► Εάν ένας υποψήφιος έχει: ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 98.7% έχει επίσης: ΛΑΤΙΝΙΚΑ Ο.Π.=Υψηλή. (support=23.3%, lift=1.59)
7. ► Εάν ένας υποψήφιος έχει: ΑΡΧΑΙΑ ΕΛΛΗΝΙΚΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 80.6% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Μέτρια. (support=15.0%, lift=1.56)
8. ► Εάν ένας υποψήφιος έχει: ΙΣΤΟΡΙΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 95.0% έχει επίσης: ΛΑΤΙΝΙΚΑ Ο.Π.=Υψηλή. (support=31.6%, lift=1.53)
9. ► Εάν ένας υποψήφιος έχει: ΙΣΤΟΡΙΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 71.3% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Υψηλή. (support=23.7%, lift=1.51)
10. ► Εάν ένας υποψήφιος έχει: ΛΑΤΙΝΙΚΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 73.5% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Μέτρια. (support=20.4%, lift=1.42)

Ομάδα Θετικών Σπουδών

- ✓ Φορτώθηκε το αρχείο
Σύνολο γραμμών: 10541
- ✓ Στήλες βαθμών: ['ΜΑΘΗΜΑΤΙΚΑ Ο.Π.', 'ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.', 'ΦΥΣΙΚΗ Ο.Π.', 'ΧΗΜΕΙΑ Ο.Π.']
- ✓ Basket dimensions: (10541, 12)
- ✓ Frequent itemsets: 86
- ✓ Παρήχθησαν κανόνες: 298
- ✓ Τελικοί κανόνες: 10

ΟΙ 10 ΚΑΛΥΤΕΡΟΙ ΚΑΝΟΝΕΣ ΕΠΙΔΟΣΕΩΝ (ΚΛΙΜΑΚΑ 0-20)

Διακριτοποίηση βαθμών:

Χαμηλή: <10 | Μέτρια: 10-15 | Υψηλή: ≥15

1. ► Εάν ένας υποψήφιος έχει: ΦΥΣΙΚΗ Ο.Π.=Υψηλή, τότε με πιθανότητα 90.2% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Υψηλή. (support=22.3%, lift=2.30)
2. ► Εάν ένας υποψήφιος έχει: ΦΥΣΙΚΗ Ο.Π.=Υψηλή, τότε με πιθανότητα 94.0% έχει επίσης: ΧΗΜΕΙΑ Ο.Π.=Υψηλή. (support=23.3%, lift=2.07)
3. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Χαμηλή,

- τότε με πιθανότητα 91.7% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=14.2%, lift=2.00)
4. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Χαμηλή,
τότε με πιθανότητα 91.7% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=15.8%, lift=2.00)
 5. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Υψηλή,
τότε με πιθανότητα 84.1% έχει επίσης: ΧΗΜΕΙΑ Ο.Π.=Υψηλή.
(support=32.9%, lift=1.85)
 6. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Υψηλή,
τότε με πιθανότητα 72.5% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Υψηλή.
(support=32.9%, lift=1.85)
 7. ► Εάν ένας υποψήφιος έχει: ΦΥΣΙΚΗ Ο.Π.=Υψηλή,
τότε με πιθανότητα 64.3% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ
Γ.Π.=Υψηλή.
(support=15.9%, lift=1.49)
 8. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Μέτρια,
τότε με πιθανότητα 66.4% έχει επίσης: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Μέτρια.
(support=24.8%, lift=1.46)
 9. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Μέτρια,
τότε με πιθανότητα 66.5% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=24.8%, lift=1.45)
 10. ► Εάν ένας υποψήφιος έχει: ΜΑΘΗΜΑΤΙΚΑ Ο.Π.=Μέτρια,
τότε με πιθανότητα 62.7% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=28.5%, lift=1.37)

Ομάδα Σπουδών Υγείας

- ✓ Φορτώθηκε το αρχείο
- Σύνολο γραμμών: 9214
- ✓ Στήλες βαθμών: ['ΒΙΟΛΟΓΙΑ Ο.Π.', 'ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.', 'ΦΥΣΙΚΗ Ο.Π.', 'ΧΗΜΕΙΑ Ο.Π.']
- ✓ Basket dimensions: (9214, 12)
- ✓ Frequent itemsets: 76
- ✓ Παρήχθησαν κανόνες: 230
- ✓ Τελικοί κανόνες: 10

ΟΙ 10 ΚΑΛΥΤΕΡΟΙ ΚΑΝΟΝΕΣ ΕΠΙΔΟΣΕΩΝ (ΚΛΙΜΑΚΑ 0-20)

Διακριτοποίηση βαθμών:

Χαμηλή: <10 | Μέτρια: 10-15 | Υψηλή: ≥15

1. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Χαμηλή,
τότε με πιθανότητα 97.9% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=12.9%, lift=1.93)
2. ► Εάν ένας υποψήφιος έχει: ΦΥΣΙΚΗ Ο.Π.=Υψηλή,
τότε με πιθανότητα 98.0% έχει επίσης: ΧΗΜΕΙΑ Ο.Π.=Υψηλή.
(support=22.5%, lift=1.90)
3. ► Εάν ένας υποψήφιος έχει: ΒΙΟΛΟΓΙΑ Ο.Π.=Χαμηλή,
τότε με πιθανότητα 95.6% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή.
(support=10.4%, lift=1.89)

4. ► Εάν ένας υποψήφιος έχει: ΦΥΣΙΚΗ Ο.Π.=Υψηλή, τότε με πιθανότητα 97.6% έχει επίσης: ΒΙΟΛΟΓΙΑ Ο.Π.=Υψηλή. (support=22.4%, lift=1.69)
5. ► Εάν ένας υποψήφιος έχει: ΒΙΟΛΟΓΙΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 82.2% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή. (support=25.8%, lift=1.62)
6. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Μέτρια, τότε με πιθανότητα 81.6% έχει επίσης: ΦΥΣΙΚΗ Ο.Π.=Χαμηλή. (support=28.8%, lift=1.61)
7. ► Εάν ένας υποψήφιος έχει: ΒΙΟΛΟΓΙΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 63.0% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Μέτρια. (support=6.8%, lift=1.52)
8. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 87.6% έχει επίσης: ΒΙΟΛΟΓΙΑ Ο.Π.=Υψηλή. (support=45.1%, lift=1.52)
9. ► Εάν ένας υποψήφιος έχει: ΒΙΟΛΟΓΙΑ Ο.Π.=Υψηλή, τότε με πιθανότητα 78.1% έχει επίσης: ΧΗΜΕΙΑ Ο.Π.=Υψηλή. (support=45.1%, lift=1.52)
10. ► Εάν ένας υποψήφιος έχει: ΧΗΜΕΙΑ Ο.Π.=Χαμηλή, τότε με πιθανότητα 62.6% έχει επίσης: ΝΕΟΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΛΟΓΟΤΕΧΝΙΑ Γ.Π.=Μέτρια. (support=8.2%, lift=1.51)

Οι κανόνες συσχέτισης που εξήχθησαν για τις τέσσερις Ομάδες Προσανατολισμού αποκαλύπτουν ισχυρά και συστηματικά πρότυπα συνύπαρξης επιδόσεων μεταξύ των μαθημάτων, επιβεβαιώνοντας την ύπαρξη κοινών γνωστικών και δεξιολογικών βάσεων εντός κάθε επιστημονικού πεδίου. Οι υψηλές τιμές εμπιστοσύνης (confidence), σε συνδυασμό με δείκτες ανύψωσης (lift) σημαντικά μεγαλύτερους της μονάδας, καταδεικνύουν ότι οι συσχετίσεις αυτές δεν είναι τυχαίες αλλά αντανακλούν ουσιαστικές εκπαιδευτικές σχέσεις.

Στην Ομάδα Οικονομίας και Πληροφορικής, τα Μαθηματικά Ομάδας Προσανατολισμού αναδεικνύονται ως ο κεντρικός άξονας επίδοσης, με εξαιρετικά ισχυρές συσχετίσεις τόσο με την Πληροφορική όσο και με την Οικονομία, σε όλα τα επίπεδα βαθμολογίας. Αντίστοιχα, στην Ομάδα Ανθρωπιστικών Σπουδών, τα Αρχαία Ελληνικά και τα Λατινικά εμφανίζονται ως καθοριστικοί παράγοντες για την επίδοση στην Ιστορία και στη Νεοελληνική Γλώσσα, γεγονός που υποδηλώνει έντονη διαθεματική σύνδεση μέσω γλωσσικών, ερμηνευτικών και μνημονικών δεξιοτήτων. Οι υψηλές τιμές lift σε αρκετούς κανόνες ενισχύουν την άποψη ότι η φιλολογική επάρκεια λειτουργεί αθροιστικά και όχι αποσπασματικά.

Αντίστοιχα ισχυρά μοτίβα παρατηρούνται στις Ομάδες Θετικών Σπουδών και Σπουδών Υγείας, όπου οι συσχετίσεις μεταξύ Μαθηματικών, Φυσικής, Χημείας και Βιολογίας είναι ιδιαίτερα έντονες. Η υψηλή επίδοση στη Φυσική ή στη Χημεία συνοδεύεται με πολύ μεγάλη πιθανότητα από υψηλή επίδοση στα υπόλοιπα μαθήματα, ενώ η χαμηλή επίδοση σε ένα από αυτά συχνά συνδέεται με γενικευμένη μαθησιακή δυσκολία στο επιστημονικό πεδίο. Τα ευρήματα αυτά επιβεβαιώνουν ότι οι θετικές επιστήμες απαιτούν κοινό αναλυτικό και εννοιολογικό υπόβαθρο, ενώ παράλληλα αναδεικνύουν τον ρόλο της Νεοελληνικής Γλώσσας ως υποστηρικτικού παράγοντα κατανόησης και επεξεργασίας των εξεταστικών θεμάτων. Συνολικά, οι κανόνες συσχέτισης προσφέρουν πολύτιμη εικόνα της δομής της

μαθητικής επίδοσης και μπορούν να αξιοποιηθούν τόσο για ερμηνευτικούς όσο και για παιδαγωγικούς σκοπούς.

Κεφάλαιο 8^ο: Συμπεράσματα και Μελλοντική Έρευνα

8.1. Συμπεράσματα

Από τη στατιστική ανάλυση και την εφαρμογή τεχνικών εξόρυξης δεδομένων στα δεδομένα των Πανελλαδικών Εξετάσεων των ετών 2023 και 2024 προέκυψαν σημαντικά και πολυεπίπεδα συμπεράσματα σχετικά με τις επιδόσεις, τις επιλογές και τη συμπεριφορά των υποψηφίων.

Αρχικά, διαπιστώθηκε ότι το πλήθος των υποψηφίων παρουσίασε μικρή αύξηση το 2024 σε σύγκριση με το 2023, με τους μαθητές των Γενικών Λυκείων να αποτελούν τη συντριπτική πλειονότητα. Η κατανομή των υποψηφίων ανά Ομάδα Προσανατολισμού χαρακτηρίζεται από σχετική σταθερότητα, με την Ομάδα Σπουδών Οικονομίας και Πληροφορικής να διατηρεί τη μεγαλύτερη συμμετοχή και για τα δύο έτη. Παράλληλα, παρατηρήθηκε ενίσχυση του ενδιαφέροντος για τις Ανθρωπιστικές Σπουδές και ελαφρά μείωση του ποσοστού υποψηφίων στις Σπουδές Υγείας, γεγονός που ενδέχεται να συνδέεται με τον αυξημένο ανταγωνισμό και τις υψηλές βαθμολογικές απαιτήσεις των συγκεκριμένων σχολών.

Η ανάλυση των επιτυχόντων και μη επιτυχόντων ανέδειξε σημαντικές διαφοροποιήσεις μεταξύ των Ομάδων Προσανατολισμού αλλά και μεταξύ των φύλων, τόσο ως προς τα ποσοστά επιτυχίας όσο και ως προς τις βαθμολογικές επιδόσεις. Η εφαρμογή της Ελάχιστης Βάσης Εισαγωγής φαίνεται να έχει επηρεάσει ουσιαστικά τη δυνατότητα πρόσβασης σε συγκεκριμένα Τμήματα, περιορίζοντας τον αριθμό των υποψηφίων που μπορούν να δηλώσουν σχολές με υψηλές απαιτήσεις, χωρίς ωστόσο να επηρεάζει τη μεταξύ τους κατάταξη.

Η συσταδοποίηση των δεδομένων μέσω των αλγορίθμων **k-means** και **DBSCAN** κατέδειξε την ύπαρξη διακριτών ομάδων υποψηφίων με κοινά χαρακτηριστικά ως προς τις επιδόσεις και τις εκπαιδευτικές επιλογές. Οι ομάδες αυτές αποτυπώνουν διαφορετικά προφίλ υποψηφίων, επιβεβαιώνοντας ότι οι μαθητές δεν αποτελούν ένα ομοιογενές σύνολο, αλλά παρουσιάζουν σημαντικές εσωτερικές διαφοροποιήσεις.

Η κατηγοριοποίηση με τη χρήση του αλγορίθμου **Random Forest Regression** επέτρεψε την αποτελεσματική μοντελοποίηση της σχέσης μεταξύ βαθμολογικών επιδόσεων και τελικού αποτελέσματος εισαγωγής, αναδεικνύοντας τη σημασία συγκεκριμένων μαθημάτων και παραγόντων στη διαμόρφωση των μορίων. Ο **Gaussian Naive Bayes** παρουσίασε σταθερά «καλή αλλά μέτρια» επίδοση στην πρόβλεψη της επιτυχίας στις τέσσερις ομάδες, με Accuracy περίπου 86%–89%. Αντίθετα, ο **Random Forest** εμφάνισε σαφώς ανώτερη απόδοση σε όλες τις ομάδες, με Accuracy περίπου 96%–98% και πολύ κοντινά αποτελέσματα μεταξύ holdout και 5-fold cross-validation, ένδειξη ισχυρής γενίκευσης. Και τέλος, οι κανόνες συσχέτισης που προέκυψαν από την εφαρμογή του αλγορίθμου **Apriori** ανέδειξαν ισχυρές σχέσεις μεταξύ των προτιμήσεων των υποψηφίων στο μηχανογραφικό δελτίο και των τελικών αποτελεσμάτων εισαγωγής, ιδιαίτερα όταν η ανάλυση πραγματοποιήθηκε ανά Ομάδα Προσανατολισμού.

Συνολικά, η εργασία καταδεικνύει ότι η αξιοποίηση τεχνικών εξόρυξης δεδομένων στον τομέα της εκπαίδευσης μπορεί να προσφέρει ουσιαστική γνώση σχετικά με τις εκπαιδευτικές τάσεις και τη συμπεριφορά των υποψηφίων. Τα αποτελέσματα μπορούν να αξιοποιηθούν τόσο από την εκπαιδευτική κοινότητα όσο και από τους φορείς χάραξης εκπαιδευτικής πολιτικής, συμβάλλοντας στη λήψη πιο τεκμηριωμένων αποφάσεων και στη βελτίωση του συστήματος εισαγωγής στην τριτοβάθμια εκπαίδευση.

8.2. Μελλοντική Έρευνα

Η παρούσα διπλωματική εργασία ανέδειξε τη δυναμική που παρουσιάζει η ανάλυση Πανελλαδικών δεδομένων των Γενικών Λυκείων μέσω συνδυασμού στατιστικών μεθόδων και τεχνικών εξόρυξης γνώσης. Παρά τα σημαντικά ευρήματα που προέκυψαν, υπάρχουν αρκετές δυνατότητες περαιτέρω εμβάθυνσης και επέκτασης της έρευνας.

Μία προφανής κατεύθυνση μελλοντικής εργασίας αφορά την ανάλυση Πανελλαδικών δεδομένων των Επαγγελματικών Λυκείων. Επίσης η ενσωμάτωση δεδομένων περισσότερων ετών, ώστε να καταστεί δυνατή η πληρέστερη διαχρονική μελέτη των τάσεων στις επιδόσεις και στις εκπαιδευτικές επιλογές των υποψηφίων. Η επέκταση αυτή θα επέτρεπε την πιο αξιόπιστη αποτίμηση των επιπτώσεων θεσμικών αλλαγών, όπως η Ελάχιστη Βάση Εισαγωγής και οι συντελεστές βαρύτητας, σε βάθος χρόνου.

Μια ακόμη ενδιαφέρουσα προοπτική αποτελεί ο εμπλουτισμός των υπάρχοντων συνόλων δεδομένων με πρόσθετα κοινωνικοοικονομικά χαρακτηριστικά (όπου αυτό είναι εφικτό και σύμφωνο με το ισχύον πλαίσιο προστασίας προσωπικών δεδομένων), όπως το μορφωτικό επίπεδο των γονέων ή η γεωγραφική περιοχή κατοικίας. Η ενσωμάτωση τέτοιων μεταβλητών θα μπορούσε να συμβάλει στην ανάδειξη παραγόντων εκπαιδευτικών ανισοτήτων και στη διαμόρφωση πιο στοχευμένων πολιτικών παρέμβασης.

Τέλος, μία πρακτική εφαρμογή των αποτελεσμάτων, μια web εφαρμογή, θα μπορούσε να είναι η ανάπτυξη ενός διαδραστικού πληροφοριακού συστήματος ή πίνακα ελέγχου (dashboard), το οποίο θα παρέχει οπτικοποιημένες αναλύσεις και βασικές προβλέψεις σε μαθητές, γονείς και εκπαιδευτικούς. Ένα τέτοιο εργαλείο θα μπορούσε να λειτουργήσει υποστηρικτικά στη διαδικασία λήψης αποφάσεων, προσφέροντας τεκμηριωμένη πληροφόρηση σχετικά με τις πιθανότητες επιτυχίας και τις εκπαιδευτικές επιλογές.

Συνολικά, η παρούσα εργασία μπορεί να αποτελέσει αφετηρία για περαιτέρω έρευνα στον τομέα της εκπαιδευτικής εξόρυξης δεδομένων, με στόχο την ουσιαστικότερη κατανόηση του συστήματος εισαγωγής στην τριτοβάθμια εκπαίδευση και τη συμβολή στη διαμόρφωση πιο δίκαιων και αποτελεσματικών εκπαιδευτικών πολιτικών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Omilo, “Πως να μπεις σε ένα ελληνικό πανεπιστήμιο;,” Omilo. Accessed: Feb. 10, 2026. [Online]. Available: <https://omilo.com/el/greek-university/>
- [2] “Ελάχιστη Βάση Εισαγωγής,” *Βικιπαίδεια*. Aug. 03, 2025. Accessed: Feb. 10, 2026. [Online]. Available: https://el.wikipedia.org/w/index.php?title=%CE%95%CE%BB%CE%AC%CF%87%CE%B9%CF%83%CF%84%CE%B7_%CE%92%CE%AC%CF%83%CE%B7_%CE%95%CE%B9%CF%83%CE%B1%CE%B3%CF%89%CE%B3%CE%AE%CF%82&oldid=11219626#cite_note-6
- [3] “ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΘΡΗΣΚΕΥΜΑΤΩΝ & ΑΘΛΗΤΙΣΜΟΥ - 14-04-22 Για πρώτη φορά συντελεστές μαθημάτων Πανελλαδικών από τα ίδια τα ΑΕΙ.” Accessed: Feb. 06, 2026. [Online]. Available: <https://www.minedu.gov.gr/news/51862-14-04-22-gia-proti-fora-syntelestes-mathimatou-panelladikon-apo-ta-idia-ta-aei>
- [4] S. Ougiaroglou and G. Paschalis, “Association Rules Mining from the Educational Data of ESOG Web-Based Application,” in *Artificial Intelligence Applications and Innovations*, vol. 382, L. Pliadis, I. Maglogiannis, H. Papadopoulos, K. Karatzas, and S. Sioutas, Eds., in IFIP Advances in Information and Communication Technology, vol. 382. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 105–114. doi: 10.1007/978-3-642-33412-2_11.
- [5] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI Mag.*, vol. 17, no. 3, pp. 37–37, 1996.
- [6] J. Han, M. Kamber, and J. Pei, “Data mining: Concepts and,” *Tech. Walth. Morgan Kaufmann Publ.*, 2012, Accessed: Feb. 07, 2026. [Online]. Available: <http://homes.di.unimi.it/ceselli/IM/2012-13/slides/02-KnowYourData.pdf>
- [7] V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal, “A Survey of Algorithms for Dense Subgraph Discovery,” in *Managing and Mining Graph Data*, vol. 40, C. C. Aggarwal and H. Wang, Eds., in Advances in Database Systems, vol. 40. , Boston, MA: Springer US, 2010, pp. 303–336. doi: 10.1007/978-1-4419-6045-0_10.
- [8] E. Frank and M. A. Hall, *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, 2011.
- [9] V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal, “A Survey of Algorithms for Dense Subgraph Discovery,” in *Managing and Mining Graph Data*, vol. 40, C. C. Aggarwal and H. Wang, Eds., in Advances in Database Systems, vol. 40. , Boston, MA: Springer US, 2010, pp. 303–336. doi: 10.1007/978-1-4419-6045-0_10.
- [10] B. Fry, *Visualizing data: Exploring and explaining data with the processing environment*. O’Reilly Media, Inc., 2007. Accessed: Feb. 09, 2026. [Online]. Available: [https://books.google.com/books?hl=el&lr=&id=6jsVAiULQBgC&oi=fnd&pg=PR3&dq=Visualizing+Data+Fry,+B.++\(2008\).+Visualizing+Data.+O%2%80%99Reilly.&ots=2BuR19jbMV&sig=VfSCY-UxgOaAdUCCL1bp_8sPJro](https://books.google.com/books?hl=el&lr=&id=6jsVAiULQBgC&oi=fnd&pg=PR3&dq=Visualizing+Data+Fry,+B.++(2008).+Visualizing+Data.+O%2%80%99Reilly.&ots=2BuR19jbMV&sig=VfSCY-UxgOaAdUCCL1bp_8sPJro)
- [11] “Methodologies for Data Quality Measurement and Improvement,” in *Data Quality*, in Data-Centric Systems and Applications. , Springer Berlin Heidelberg, 2006, pp. 161–200. doi: 10.1007/3-540-33173-5_7.
- [12] D. Pyle, *Data preparation for data mining*. morgan kaufmann, 1999. Accessed: Feb. 09, 2026. [Online]. Available: [https://books.google.com/books?hl=el&lr=&id=hhdVr9F-JfAC&oi=fnd&pg=PA1&dq=Data+Preparation+for+Data+Mining+Pyle,+D.++\(1999\).+Data+Preparation+for+Data+Mining.+Morgan+Kaufmann.&ots=6igOdQOw8w&sig=1sBvVugLcnKHK_Fop0Nu_AGo4Cw](https://books.google.com/books?hl=el&lr=&id=hhdVr9F-JfAC&oi=fnd&pg=PA1&dq=Data+Preparation+for+Data+Mining+Pyle,+D.++(1999).+Data+Preparation+for+Data+Mining.+Morgan+Kaufmann.&ots=6igOdQOw8w&sig=1sBvVugLcnKHK_Fop0Nu_AGo4Cw)
- [13] C. C. Aggarwal, *Data Mining: The Textbook*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [14] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. in Springer Series in Statistics. New York, NY: Springer New York, 2001. doi: 10.1007/978-0-387-21606-5.
- [15] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006. Accessed: Feb. 06, 2026. [Online]. Available: <https://link.springer.com/9780387310732>
- [16] R. L. Thorndike, “Who belongs in the family?,” *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.
- [17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *kdd*, 1996, pp. 226–231. Accessed: Feb. 06, 2026.

[Online]. Available: https://cdn.aaai.org/KDD/1996/KDD96-037.pdf?source=post_page-----

- [18] R. Agarwal and R. Srikant, “Fast algorithms for mining association rules,” in *Proc. of the 20th VLDB Conference*, Santiago, Chile, 1994, p. 499. Accessed: Feb. 07, 2026. [Online]. Available: [https://www.columbia.edu/~rd2537/docu/apriori\(abstract\).pdf](https://www.columbia.edu/~rd2537/docu/apriori(abstract).pdf)
- [19] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [20] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning.” Springer series in statistics New-York, 2009. Accessed: Feb. 07, 2026. [Online]. Available: <https://www.academia.edu/download/31156736/10.1.1.158.8831.pdf>
- [21] G. James, D. Witten, T. Hastie, and R. Tibshirani, “Linear Model Selection and Regularization,” in *An Introduction to Statistical Learning*, in Springer Texts in Statistics. , New York, NY: Springer US, 2021, pp. 225–288. doi: 10.1007/978-1-0716-1418-1_6.
- [22] P. Domingos and M. Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss,” *Mach. Learn.*, vol. 29, no. 2–3, pp. 103–130, Nov. 1997, doi: 10.1023/A:1007413511361.
- [23] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, “PREDICTING STUDENTS’ PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES,” *Appl. Artif. Intell.*, vol. 18, no. 5, pp. 411–426, May 2004, doi: 10.1080/08839510490442058.
- [24] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, “Benefits, Adoption Barriers and Myths of Open Data and Open Government,” *Inf. Syst. Manag.*, vol. 29, no. 4, pp. 258–268, Sep. 2012, doi: 10.1080/10580530.2012.716740.
- [25] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

ΠΑΡΑΡΤΗΜΑ: Κώδικες σε Python

Κώδικας σε Python1

```
# Ποσοστά Υποψηφίων ανά Είδος Λυκείου
#=====
import matplotlib.pyplot as plt

# =====
# ΔΕΔΟΜΕΝΑ
# =====
labels = ["ΓΕΛ", "ΕΠΑΛ"]

values_2023 = [77649, 20617]
values_2024 = [80108, 21264]

# =====
# ΣΥΝΑΡΤΗΣΗ ΠΙΤΑΣ
# =====
def pie_school_type(values, year):
    plt.figure(figsize=(7, 6))
    plt.pie(
        values,
        labels=labels,
        autopct=lambda p: f"{p:.2f}%".replace(".", ","),
        startangle=90
    )
    plt.title(
        f"Ποσοστά Υποψηφίων Ανά Είδος Λυκείου {year}",
        fontsize=14,
        weight="bold"
    )
    plt.axis("equal") # τέλειος κύκλος
    plt.show()

# =====
# ΓΡΑΦΗΜΑΤΑ
# =====
pie_school_type(values_2023, 2023)
pie_school_type(values_2024, 2024)
```

Κώδικας σε Python 2

```
# Επιτυχόντες και Αποτυχόντες ανά Έτος
# και ανά Ομάδα Προσανατολισμού
#=====
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# =====
# 1) ΔΕΔΟΜΕΝΑ (ΣΥΝΟΛΑ ανά ομάδα)
# =====

groups_2023 = [
    "ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ",
    "ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ & ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ)",
    "ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ & ΥΓΕΙΑΣ (Ε.Π. ΥΓΕΙΑΣ)",
    "ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ",
    "ΜΟΥΣΙΚΩΝ ΣΠΟΥΔΩΝ"
]

success_2023 = np.array([13598, 10669, 9952, 17534, 240])
fail_2023     = np.array([ 6676,  2943, 4612, 11296, 129])

groups_2024 = [
    "ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΣΠΟΥΔΩΝ",
    "ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ & ΥΓΕΙΑΣ (Ε.Π. ΘΕΤΙΚΩΝ)",
    "ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ & ΥΓΕΙΑΣ (Ε.Π. ΥΓΕΙΑΣ)",
    "ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ"
]

success_2024 = np.array([14218, 11363, 10293, 18073])
fail_2024     = np.array([ 7561,  2728, 4041, 11831])

# =====
# 2) ΣΥΝΑΡΤΗΣΗ: Πίνακας + Γράφημα ποσοστών
# =====
def percentages_table_and_chart(groups, success, fail, year):
    total = success + fail
    success_pct = success / total * 100
    fail_pct = fail / total * 100

    # --- ΠΙΝΑΚΑΣ ---
    df = pd.DataFrame({
        "ΕΠΙΤΥΧΟΝΤΕΣ": success,
        "ΜΗ ΕΠΙΤΥΧΟΝΤΕΣ": fail,
        "ΣΥΝΟΛΟ": total,
        "% ΕΠΙΤΥΧΟΝΤΕΣ": success_pct,
```

```

        "% ΜΗ ΕΠΙΤΥΧΟΝΤΕΣ": fail_pct
    }, index=groups)

    # Μορφοποίηση για πιο ωραία εμφάνιση
    df_display = df.copy()
    df_display["% ΕΠΙΤΥΧΟΝΤΕΣ"] = df_display["%
ΕΠΙΤΥΧΟΝΤΕΣ"].map(lambda x: f"{x:.2f}%")
    df_display["% ΜΗ ΕΠΙΤΥΧΟΝΤΕΣ"] = df_display["% ΜΗ
ΕΠΙΤΥΧΟΝΤΕΣ"].map(lambda x: f"{x:.2f}%")

    print(f"\nΠοσοστά Επιτυχόντων & Μη Επιτυχόντων ανά Ομάδα - {year}")
    print(df_display[["% ΕΠΙΤΥΧΟΝΤΕΣ", "% ΜΗ ΕΠΙΤΥΧΟΝΤΕΣ"]])

    # --- ΓΡΑΦΗΜΑ 100% STACKED ---
    x = np.arange(len(groups))

    plt.figure(figsize=(12, 6))
    plt.bar(x, success_pct, label="Επιτυχόντες")
    plt.bar(x, fail_pct, bottom=success_pct, label="Μη Επιτυχόντες")

    # Ετικέτες ποσοστών πάνω στις μπάρες
    for i in range(len(groups)):
        plt.text(i, success_pct[i] / 2, f"{success_pct[i]:.1f}%",
ha="center", va="center")
        plt.text(i, success_pct[i] + fail_pct[i] / 2,
f"{fail_pct[i]:.1f}%", ha="center", va="center")

    plt.xticks(x, groups, rotation=25, ha="right")
    plt.ylabel("Ποσοστό (%)")
    plt.title(f"Ποσοστά Επιτυχόντων & Μη Επιτυχόντων ανά Ομάδα
Προσανατολισμού - {year}")
    plt.ylim(0, 100)
    plt.legend()
    plt.tight_layout()
    plt.show()

    return df # αν θες να το χρησιμοποιήσεις μετά

# =====
# 3) ΕΚΤΕΛΕΣΗ ΓΙΑ 2023 & 2024
# =====
df_2023 = percentages_table_and_chart(groups_2023, success_2023,
fail_2023, 2023)
df_2024 = percentages_table_and_chart(groups_2024, success_2024,
fail_2024, 2024)

```

Κώδικας σε Python 3

```
# =====
# 1 Ανάλυση Πανελλαδικών με Clustering K-Means
# Με χρωματική κωδικοποίηση επιδόσεων σε clusters
# Γράφημα σε Parallel Coordinates
# =====

# -----
# 2 Εισαγωγή βιβλιοθηκών
# -----
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import DBSCAN, KMeans
import matplotlib.patches as mpatches

!pip install -q kneed
from kneed import KneeLocator

from google.colab import drive

# -----
# 3 Σύνδεση με Google Drive
# -----
drive.mount('/content/drive')

# -----
# 4 Φόρτωση αρχείου
# -----
path = "/content/drive/MyDrive/Colab Notebooks/ΑΠΟΤ-
2024/CLUSTERING/VATHMOI_OMADA_THETIKON_WIDE.csv"

df = pd.read_csv(path, sep=';', encoding='windows-1253')
df = df.replace(',', '.', regex=True)

for col in df.columns:
    try:
        df[col] = pd.to_numeric(df[col])
    except:
        pass

# -----
# 5 Επιλογή αριθμητικών στηλών & δημιουργία μέσου όρου
# -----
data = df.select_dtypes(include=['int64', 'float64']).copy()
data = data.fillna(data.mean(numeric_only=True))
```

```

# μέσος όρος ανά μαθητή (για ταξινόμηση clusters)
df["mean_score"] = data.mean(axis=1)

# -----
# 5 Κανονικοποίηση
# -----
scaler = StandardScaler()
X_scaled = scaler.fit_transform(data)

# -----
# 8 K-Means: εύρεση βέλτιστου K
# -----
inertia = []
K_range = range(2, 9)

for k in K_range:
    model = KMeans(n_clusters=k, random_state=42)
    model.fit(X_scaled)
    inertia.append(model.inertia_)

kneedle_k = KneeLocator(K_range, inertia, curve='convex',
direction='decreasing')
best_k = kneedle_k.knee if kneedle_k.knee is not None else 4

kmeans = KMeans(n_clusters=best_k, random_state=42)
df["KMeans_raw"] = kmeans.fit_predict(X_scaled)

# -----
# 9 Επαναταξινόμηση clusters βάσει επίδοσης
# -----
cluster_means = df.groupby("KMeans_raw")["mean_score"].mean()
sorted_clusters = cluster_means.sort_values().index.tolist()

# mapping: 0 = χαμηλές → K-1 = υψηλές
mapping = {raw_id: new_id for new_id, raw_id in
enumerate(sorted_clusters)}

df["KMeans_cluster"] = df["KMeans_raw"].map(mapping)

# -----
# 10 Χρώματα & ετικέτες
# -----
custom_colors = {
    0: "#F9A825", # yellow (χαμηλές)
    1: "#0D47A1", # dark blue (μέτριες)
    2: "#388E3C", # green (καλές)
    3: "#D32F2F" # red (άριστοι)
}

```

```

legend_labels = {
    0: "0 - Χαμηλές Επιδόσεις",
    1: "1 - Μέτριες Επιδόσεις",
    2: "2 - Πολύ Καλές Επιδόσεις",
    3: "3 - Άριστοι"
}

# -----
# [10] Sample 50 τελειόφοιτοι για καθαρό γράφημα
# -----
sample_df = df.sample(n=min(200, len(df)), random_state=42).copy()

# -----
# [10] Custom Parallel Coordinates ΧΩΡΙΣ κανονικοποίηση
# -----
plt.figure(figsize=(18, 6))

cols = data.columns.tolist()

for idx, row in sample_df.iterrows():
    c = custom_colors[row['KMeans_cluster']]
    plt.plot(range(len(cols)), row[cols], color=c, alpha=0.65)

plt.xticks(range(len(cols)), cols, rotation=45)
plt.ylabel("Πραγματικοί Βαθμοί (0-20)")
plt.title("Parallel Coordinates χωρίς κανονικοποίηση - K-Means - ΟΜΑΔΑΣ
ΘΕΤΙΚΩΝ ΣΠΟΥΔΩΝ")

# Legend
legend_patches = [
    mpatches.Patch(color=custom_colors[c], label=legend_labels[c])
    for c in sorted(sample_df["KMeans_cluster"].unique())
]

plt.legend(handles=legend_patches, bbox_to_anchor=(1.02, 1), loc="upper
left")
plt.grid(True)
plt.show()

```

Κώδικας σε Python 4

```
# =====
# Εξόρυξη Κανόνων Συσχέτισης από δηλώσεις προτιμήσεων υποψηφίων
# ΣΠΟΥΔΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ
# =====

import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
from google.colab import drive
import warnings

warnings.filterwarnings("ignore")

# -----
# 1 Σύνδεση με Google Drive
# -----
drive.mount('/content/drive')

# -----
# 2 Ανάγνωση αρχείου δεδομένων
# -----
df = pd.read_csv(
    "/content/drive/MyDrive/Colab Notebooks/ΑΠΟΤ-2024/ASSOCIATION
RULES/MERGE_VATHMOI_PROT1234_PLHROF_2024.csv",
    sep=";",
    encoding="utf-8",
    dtype=str
)

print("✓ Φορτώθηκε το αρχείο")
print("Στήλες:", df.columns.tolist())

# -----
# 3 Επιλογή στηλών προτιμήσεων (ΠΡΟΤ1-ΠΡΟΤ4)
# -----
pref_cols = [c for c in df.columns if "ΠΡΟΤ" in c or "ΠΡΟΤ" in c]
prefs = df[pref_cols].replace("nan", pd.NA)

print("✓ Στήλες προτιμήσεων:", pref_cols)

# -----
# 4 Μετατροπή σε one-hot encoded basket
# -----
basket = pd.get_dummies(prefs, prefix=prefs.columns, dummy_na=False)

print("✓ Basket dimensions:", basket.shape)

# -----
```

```

# 5) Apriori - Εύρεση συχνών συνόλων
# -----
frequent_itemsets = apriori(
    basket,
    min_support=0.001,      # 0.1% του δείγματος
    use_colnames=True
)

print("✓ Υπολογίστηκαν frequent itemsets")

# -----
# 6) Κανόνες συσχέτισης
# -----
rules = association_rules(
    frequent_itemsets,
    metric="lift",
    min_threshold=2.0
)

print("✓ Παρήχθησαν κανόνες:", len(rules))

# -----
# 7) Τελικό φιλτράρισμα κανόνων
# -----
final_rules = rules[
    # MONO 1 → 1
    (rules["antecedents"].apply(lambda x: len(x) == 1)) &
    (rules["consequents"].apply(lambda x: len(x) == 1)) &

    # Antecedent: ΠΡΟΤ1 ή ΠΡΟΤ2
    (rules["antecedents"].astype(str)
     .str.contains("ΠΡΟΤ1|ΠΡΟΤ1|ΠΡΟΤ2|ΠΡΟΤ2")) &

    # Consequent: ΠΡΟΤ2 ή ΠΡΟΤ3 ή ΠΡΟΤ4
    (rules["consequents"].astype(str)
     .str.contains("ΠΡΟΤ2|ΠΡΟΤ2|ΠΡΟΤ3|ΠΡΟΤ3|ΠΡΟΤ4|ΠΡΟΤ4"))
]

# Ταξινόμηση με βάση lift → confidence
final_rules = final_rules.sort_values(
    ["lift", "confidence"],
    ascending=False
).head(20)

print("✓ Τελικοί κανόνες:", len(final_rules))

# -----
# 8) Περιγραφή κανόνων

```

```

# -----
def explain_rule(rule):
    ant = list(rule["antecedents"])[0]
    con = list(rule["consequents"])[0]
    conf = rule["confidence"]
    sup = rule["support"]
    lift = rule["lift"]

    return (
        f"► Εάν ένας μαθητής επιλέξει: {ant},\n"
        f"   τότε με πιθανότητα {conf:.1%} επιλέγει επίσης: {con}.\n"
        f"   (support={sup:.3%}, lift={lift:.2f})\n"
    )

# -----
# 9 Εκτύπωση αποτελεσμάτων
# -----
print("\n OI 20 ΚΑΛΥΤΕΡΟΙ ΚΑΝΟΝΕΣ "
      "(ΠΡΟΤ1 / ΠΡΟΤ2 → ΠΡΟΤ2 / ΠΡΟΤ3 / ΠΡΟΤ4)\n")

for i, (_, rule) in enumerate(final_rules.iterrows(), 1):
    print(f"{i}. {explain_rule(rule)}")

```