

Εντοπισμός Πιθανών Διαφορών ανά Φύλο σε
Εισαγωγικό Προγραμματιστικό Μάθημα με χρήση
μεθόδων Ανάλυσης Δεδομένων σε αποτελέσματα
Εργασιών Φοιτητών, με σκοπό την προώθηση της
Ισότητας μεταξύ των Φύλων

Περίοδος: Επαναληπτική Εξεταστική 2022-2023 - Σεπτέμβριος 2023

Επιβλέπουσα Καθηγήτρια: Σατρατζέμη Μάγια

Συνεξετάστρια Καθηγήτρια: Κολωνιάρη Γεωργία

Συνεξεταστής Καθηγητής: Ευαγγελίδης Γεώργιος

Φοιτητής: Δαγκλής Ευάγγελος

Αριθμός Μητρώου: aid22012



Περιεχόμενα

- Σκοποί – Στόχοι Εργασίας – Ερευνητικά Ερωτήματα
- Δεδομένα – Τύπος και Δομή
- Στατιστικοί Έλεγχοι:
 - Κανονικότητας και Ισότητας Μέσων Όρων ή Ομοιογένειας
- Κανόνες Συσχέτισης
- Συσταδοποίηση:
 - Με Τυποποίηση και Ανάλυση Κύριων Συνιστωσών
 - Μόνο με Τυποποίηση
- Μοντέλα Μηχανικής Μάθησης
 - Αποτελέσματα 10-πλής Εκπαίδευσης/Επαλήθευσης και Προβλέψεων
 - Προβλέψεις μέχρι την 3^η και την 5^η εβδομάδα
- Επίλογος – Συγκρίσεις με Άλλες Έρευνες
- Παραπομπές

Σκοποί – Στόχοι της Εργασίας

- Κύριοι στόχοι της εργασίας συνοψίζονται από τα 3 **Ερευνητικά Ερωτήματα**:
 - 1) Τα λάθη και η ακαδημαϊκή απόδοση των φοιτητών στο αντικείμενο του προγραμματισμού διαφέρουν ανά φύλο και αν ναι, πώς;
 - Οπτικοποίηση και Στατιστικοί Έλεγχοι
 - 2) Μπορούν οι μέθοδοι της Συσταδοποίησης και των Κανόνων Συσχέτισης να συμβάλλουν στην κατανόηση των σχέσεων που δημιουργούνται μεταξύ των διαφορετικών οντοτήτων του μαθήματος;
 - 3) Είναι εφικτό να προβλεφθεί αν ένα άτομο θα περάσει ή όχι το μάθημα κάνοντας χρήση μόνο του αριθμού των λαθών μεταγλώττισης του κώδικα των εργασιών του και τον μέσο όρο του βαθμού των εργασιών του πριν τις εξετάσεις ή κατά προτίμηση και ακόμα νωρίτερα;
 - Δοκιμή διαφορετικών μοντέλων με δεδομένα μέχρι το τέλος του εξαμήνου και μέχρι συγκεκριμένες εβδομάδες νωρίς στο εξάμηνο.

Προηγούμενες Συναφείς Έρευνα

- Τα θέματα της εργασίας έχουν ήδη απασχολήσει τους ερευνητές στο παρελθόν και παραμένουν στο ενδιαφέρον τους.
- Στα πλαίσια εκπόνησης της διπλωματικής εργασίας πραγματοποιήθηκε βιβλιογραφική ανασκόπηση για την ήδη υπάρχουσα έρευνα επί των τομέων της εργασίας.
- Δύο σύντομες εκδοχές που περιγράφουν τις εν λόγω έρευνες και τα ευρήματά τους βρίσκονται στο [αντίστοιχο αρχείο](#).
- Κάποια από τα αποτελέσματα, τα πιο σχετικά με αυτά της τρέχουσας έρευνας, χρησιμοποιούνται για να συγκρίσεις και αντιπαραβολή στα συμπεράσματα.

Περιγραφή Εξεταζόμενου Μαθήματος

- Δεδομένα από ένα μόνο εισαγωγικό προγραμματιστικό μάθημα:

Μάθημα	Δομές Δεδομένων
Γλώσσα Προγραμματισμού	C
Υποχρεωτικές Εργασίες	Ναι
Επίπεδο Σπουδών	Προπτυχιακό
Εξάμηνο	2 ^ο
Τμήμα	Τμήμα Εφαρμοσμένης Πληροφορικής
Εκπαιδευτικό Ίδρυμα	Πανεπιστήμιο Μακεδονίας
Πόλη	Θεσσαλονίκη
Χώρα	Ελλάδα
Χρονιές εκπόνησης	2021, 2022
Συμμετέχοντες/-ουσες	1323 (924 το 2021 & 693 το 2022)

Δεδομένα – Τύπος και Δομή

- Κυρίως αρχεία που παράγονται από το πρόγραμμα αυτόματης διόρθωσης και βαθμολόγησης των εργασιών των φοιτητών “Diorthotis” (Karakasidis A. (2023)) με μορφή:

student_id,total_assignment_grade,current_exercise_name

=== COMPILATION ===

compilation_status_message

compilation_errors_notes_and_warnings

=== EXECUTION. TEST #XY ===

code_execution_result

=== EXECUTION RESULT: ===

code_execution_result_status_message

GRADE:current_exercise_grade

```
██████████10.0|"

```
*** EXERCISE a6f1.c ***

=== COMPILATION ===
COMPILED, BUT WITH ISSUES:
/usr/bin/ld: /tmp/ccYevD9f.o: in function `main':
a6f1.c:(.text+0x8e): warning: the `gets' function is dangerous and should not be used.

=== EXECUTION. TEST #1 ===
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
A E I O U Y

Give a string with uppercase characters: THISISATEST
Number of vowels: 4
Give a string with uppercase characters: N O T A T E S T
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Invalid character detected: -32
Number of vowels: 3
Give a string with uppercase characters: Another Test
Invalid character detected: n-110
Invalid character detected: o-111
Invalid character detected: t-116
Invalid character detected: h-104
Invalid character detected: e-101
Invalid character detected: r-114
Invalid character detected: -32
Invalid character detected: e-101
Invalid character detected: s-115
Invalid character detected: t-116
Number of vowels: 1

=== EXECUTION RESULT: ===
CORRECT
GRADE:10
```


```

Εξαγωγή Σφαλμάτων Μεταγλώττισης

- Μέσω ειδικού parser για τα αρχεία του «Διορθωτή», εξάγεται το πεδίο `compilation problems` και εκεί εφαρμόζονται μοτίβα Κανονικών Εκφράσεων που εντοπίζουν τα `compilation errors`, `notes` και `warnings` που εμπεριέχονται σε αυτό
 - όπως το: `r"error: unknown type name [^;]*; did you mean"`.
- Αν και, αρχικά, δημιουργήθηκαν μοτίβα για εξαγωγή 109 τέτοιων περιπτώσεων, που προέκυψαν από την μελέτη του περιεχομένου όλων των διαθέσιμων αρχείων, λόγω μη ικανοποιητικής συχνότητας εμφάνισής τους, τα λάθη που τελικά εξετάζονται περιορίστηκαν σε 46.
- *Κριτήριο αποδοχής*: το εξεταζόμενο σφάλμα εμφανίζεται με συχνότητα τουλάχιστον 1% της συχνότητας του πιο ευρέως εμφανιζόμενου σφάλματος στα αρχεία και των δύο ετών μαζί, δηλαδή 12 φορές.



Διάγραμμα Αριθμού Σφαλμάτων



• Τελευταίο πιο συχνό το 46.

• Το υπόλοιπο σχήμα.

#054 - All: 0007 M: 100.00%, F: 000.00%	error: excessive amount of items in character array initializer
#055 - All: 0007 M: 100.00%, F: 000.00%	error: array subscript (the index value inside the array[brackets]) is not an integer
#056 - All: 0006 M: 050.00%, F: 050.00%	warning: writing more bytes into a region of insufficient size, resulting in string overflow
#057 - All: 0006 M: 083.33%, F: 016.67%	error: too many arguments given to a function for it to work
#058 - All: 0006 M: 066.67%, F: 033.33%	error: invalid use of void expression, as a procedure that cannot return anything was tasked to return something
#059 - All: 0006 M: 050.00%, F: 050.00%	error: 'else' without a previous 'if'
#060 - All: 0006 M: 050.00%, F: 050.00%	warning: no semicolon at end of struct or union
#061 - All: 0006 M: 033.33%, F: 066.67%	note: to match the 'indicated symbol' (like '(')
#062 - All: 0006 M: 100.00%, F: 000.00%	error: an invalid suffix was used on a formatted integer constant (like referring to an integer as 's' instead of 'd')
#063 - All: 0006 M: 083.33%, F: 016.67%	warning: useless storage class specifier in empty declaration (probably due to a misplaced ;)
#064 - All: 0006 M: 050.00%, F: 050.00%	note: the 'indicated element' is defined in a '<library_name.h>'; did you forget to '#include <library_name.h>'?
#065 - All: 0005 M: 100.00%, F: 000.00%	error: invalid application of 'sizeof' to an incomplete type
#066 - All: 0004 M: 100.00%, F: 000.00%	warning: overflow in conversion from integer to character changes the value to something different
#067 - All: 0004 M: 050.00%, F: 050.00%	warning: unknown escape character (or sequence)
#068 - All: 0004 M: 075.00%, F: 025.00%	error: prototype declaration
#069 - All: 0004 M: 100.00%, F: 000.00%	error: macro names must be identifiers (came up because "#define <library_name.h>" was used instead of "#include <library_name.h>")
#070 - All: 0004 M: 075.00%, F: 025.00%	warning: comparison of different/distinct pointer types lacks a cast
#071 - All: 0003 M: 100.00%, F: 000.00%	error: argument doesn't match the procedure or function's prototype
#072 - All: 0003 M: 100.00%, F: 000.00%	error: the parameter has just a forward declaration, as it was not defined in a manner compatible with its calling in the function
#073 - All: 0003 M: 100.00%, F: 000.00%	warning: element redefined
#074 - All: 0003 M: 033.33%, F: 066.67%	warning: useless type name in empty declaration (a case of type_name ; where var_name is missing before ;)
#075 - All: 0003 M: 100.00%, F: 000.00%	error: subscripted value (the value inside the array[brackets]) is pointer to function
#076 - All: 0002 M: 100.00%, F: 000.00%	error: expected statement before an element or a token
#077 - All: 0002 M: 000.00%, F: 100.00%	warning: specified bound equals source length, resulting in string overflow
#078 - All: 0002 M: 000.00%, F: 100.00%	error: empty character constant
#079 - All: 0002 M: 100.00%, F: 000.00%	warning: one of the arguments, in a call to built-in function declared without prototype, is of the wrong type
#080 - All: 0002 M: 100.00%, F: 000.00%	warning: the used initializer-string for a char array is too long for the defined array size
#081 - All: 0002 M: 100.00%, F: 000.00%	error: the function is initialized with a pattern 'function_name = value' like as if it were a variable
#082 - All: 0002 M: 100.00%, F: 000.00%	error: the preprocessing directive of the form '#preprocessing_directive_name' used is invalid
#083 - All: 0002 M: 050.00%, F: 050.00%	warning: initialization of an element from an incompatible pointer type
#084 - All: 0002 M: 100.00%, F: 000.00%	error: variable redeclared as different kind of symbol (for example pointer and then non pointer)
#085 - All: 0002 M: 000.00%, F: 100.00%	error: type of formal parameter (the ones given as arguments in function calls) is incomplete
#086 - All: 0002 M: 050.00%, F: 050.00%	warning: file build-in function called on unallocated file object
#087 - All: 0002 M: 100.00%, F: 000.00%	error: switch quantity not an integer
#088 - All: 0002 M: 100.00%, F: 000.00%	error: case label not within a switch statement
#089 - All: 0002 M: 100.00%, F: 000.00%	error: break statement not within loop or switch
#090 - All: 0002 M: 100.00%, F: 000.00%	error: flexible array member not at the end of struct (something went wrong with an array definition)
#091 - All: 0002 M: 000.00%, F: 100.00%	error: the parsed extended character is not valid at the start of an identifier
#092 - All: 0002 M: 050.00%, F: 050.00%	fatal error: file_path_name.c: No such file or directory exists - Δεν υπάρχει τέτοιο αρχείο ή κατάλογος
#093 - All: 0001 M: 100.00%, F: 000.00%	error: element undeclared when not in a function, when outside of a function, but with similar enough name to a declared element
#094 - All: 0001 M: 100.00%, F: 000.00%	error: variably modified at file scope
#095 - All: 0001 M: 000.00%, F: 100.00%	error: expected specifier-qualifier-list before an element or a token
#096 - All: 0001 M: 100.00%, F: 000.00%	warning: comparison between pointer and zero character constant
#097 - All: 0001 M: 100.00%, F: 000.00%	error: lvalue required as unary '&' (or '*') operand
#098 - All: 0001 M: 000.00%, F: 100.00%	error: number of arguments doesn't match the procedure or function's prototype
#099 - All: 0001 M: 100.00%, F: 000.00%	note: did you mean to dereference the pointer?
#100 - All: 0001 M: 100.00%, F: 000.00%	warning: taking address of expression of type 'void' (as the examination of the expression stopped abruptly after a '&' with nothing afterwards)
#101 - All: 0001 M: 100.00%, F: 000.00%	warning: function declared with non-void being returned, but inside it void is returned
#102 - All: 0001 M: 000.00%, F: 100.00%	error: redefinition of an already defined parameter
#103 - All: 0001 M: 100.00%, F: 000.00%	error: a field was declared as a function
#104 - All: 0001 M: 100.00%, F: 000.00%	warning: 'sizeof' on an array function parameter will return the size of its pointer * type (here, 'sizeof' on array function parameter 'Word' will return size of 'char **')
#105 - All: 0001 M: 000.00%, F: 100.00%	error: the called object is not a function or function pointer
#106 - All: 0001 M: 100.00%, F: 000.00%	error: struct type value was used in a place where a scalar type value is required
#107 - All: 0001 M: 100.00%, F: 000.00%	error: conversion to a non-scalar type requested
#108 - All: 0001 M: 000.00%, F: 100.00%	warning: a trigraph was found and ignored, use -trigraphs to enable them
#109 - All: 0001 M: 000.00%, F: 100.00%	error: #include expects ""FILENAME"" or <FILENAME>

Διάγραμμα Αριθμού Σφαλμάτων

Δεδομένα – Τύπος και Δομή (2)

- Αρχεία λογιστικών φύλων, με ετήσια επιπρόσθετα δεδομένα φοιτητών:

Student ID	Name and Surname	Father's Name	Status	Entry Year	Enrol. Sem.	First Enroll.	Grade	Absences	Class	Enroll. Status	e-mail	Gender
()	()	()	Ενεργός	2020-21 X	2	NAI	7	0	Yci-Zfl	Κανονική	()@uom.edu.gr	M
()	()	()	Ενεργός	2014-15 X	14	OXI	---	0	Yci-Zfl	Κανονική	()@uom.edu.gr	F
()	()	()	Ενεργός	2020-21 X	2	NAI	7	0	Yci-Zfl	Κανονική	()@uom.edu.gr	M
()	()	()	Ενεργός	2019-20 X	4	OXI	---	0	Yci-Zfl	Κανονική	()@uom.edu.gr	M

- Κύριο ενδιαφέρον παρουσιάζει η Gender.
- Κάποιες ακόμα στήλες δεδομένων, επίσης χρησιμοποιούνται.
- Αυτά τα δεδομένα σε συνδυασμό με τα παράγωγα του parser των αρχείων του «Διορθωτή» χρησιμοποιούνται για την αναλυτική διαδικασία της εργασίας.
- Όπως και τα αποτελέσματα του «Διορθωτή», τα εν λόγω αρχεία παρουσιάζονται στην «πλήρη» τους μορφή, πριν από την διαδικασία της ανωνυμοποίησης που εφαρμόστηκε.

Περιεχόμενα Αρχείων 2021

- Το 2021 εκπονήθηκαν 12 εργασίες με το εξής περιεχόμενο:

Αριθμός Εργασίας	Ασκήσεις	Θέμα
1	a1f1, a5f1	Sets
2	a5f2, a16f2	Stack
3	a8f2, a17f2	Stack
4	a6f3, a12f3	Queue
5	a1f4, a30f4	List, Dynamic Stack, Queue
6	a2cf4, a2jf4, a2rf4	List, Dynamic Stack, Queue
7	a9f4, a10f4, a16f4	List, Dynamic Stack, Queue
8	a11f5, a29f5	Binary Search Tree
9	a25f5, a26f5	Binary Search Tree
10	a30f5	Binary Search Tree
11	a7f6	Hashing
12	a4f6	Hashing

Περιεχόμενα Αρχείων 2022

- Το 2022 εκπονήθηκαν 11 εργασίες με το εξής περιεχόμενο:

Αριθμός Εργασίας	Ασκήσεις	Θέμα
1	a2f1, a6f1	Sets
2	a7f2, a18f2	Stack
3	a1f3, a13f3	Queue
4	a23f4, a25f4	List, Dynamic Stack, Queue
5	a31f4, a32f4	List, Dynamic Stack, Queue
6	a2if4, a2gf4, a2qf4	List, Dynamic Stack, Queue
7	a11f4, a15f4	List, Dynamic Stack, Queue
8	a5f5, a31f5	Binary Search Tree
9	a8f5, a32f5	Binary Search Tree
10	a5f6	Hashing
11	a10f6	Hashing

Συγχώνευση σε δέκα εργασίες

- Καθώς ο αριθμός των εργασιών προς εκπόνηση είναι άνισος μεταξύ των δύο ετών, για την περάτωση ορισμένων πτυχών της εργασίας, οι ασκήσεις συγχωνεύτηκαν σε δέκα εργασίες ανά χρονιά.

Αριθμός Εργασίας	Ασκήσεις 2021	Ασκήσεις 2022
1	a1f1, a5f1	a2f1, a6f1
2	a5f2, a16f2, a8f2, a17f2	a7f2, a18f2
3	a6f3, a12f3	a1f3, a13f3
4	a1f4, a30f4	a23f4, a25f4, a31f4, a32f4
5	a2cf4, a2jf4, a2rf4	a2if4, a2gf4, a2qf4
6	a9f4, a10f4, a16f4	a11f4, a15f4
7	a11f5, a29f5	a5f5, a31f5
8	a25f5, a26f5, a30f5	a8f5, a32f5
9	a7f6	a5f6
10	a4f6	a10f6

Συγχώνευση σε δέκα εργασίες (2)

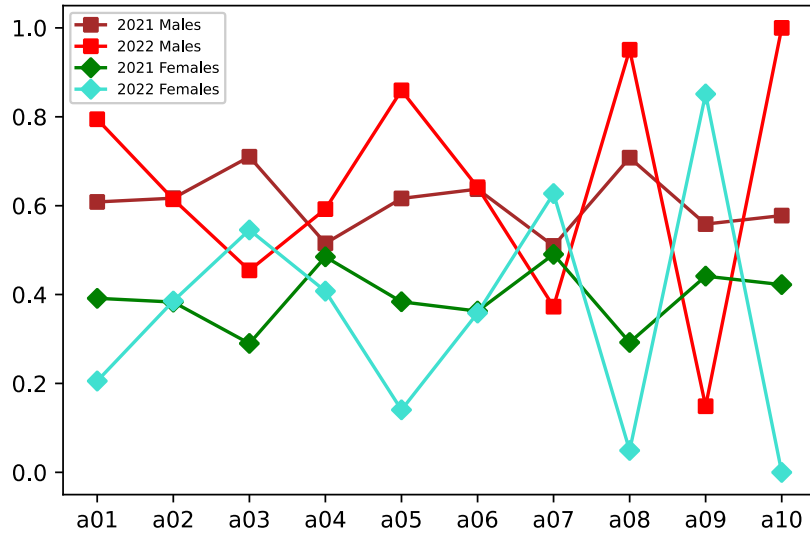
- Συγχώνευση σημαίνει ότι οι ασκήσεις των αναγραφόμενων εργασιών εκλήφθηκαν ότι ανήκουν σε μία μόνο νέα εργασία. Έγινε με γνώμονα την διατήρηση του διδακτικού σκοπού των αρχικών εργασιών.

Νέος Αριθμός Εργασίας	Αρχικοί Αριθμοί Εργασιών 2021	Αρχικοί Αριθμοί Εργασιών 2022	Θέμα Εργασιών
1	01	01	Sets
2	02, 03	02	Stack
3	04	03	Queue
4	05	04, 05	List, Dynamic Stack, Queue
5	06	06	List, Dynamic Stack, Queue
6	07	07	List, Dynamic Stack, Queue
7	08	08	Binary Search Tree
8	09, 10	09	Binary Search Tree
9	11	10	Hashing
10	12	11	Hashing

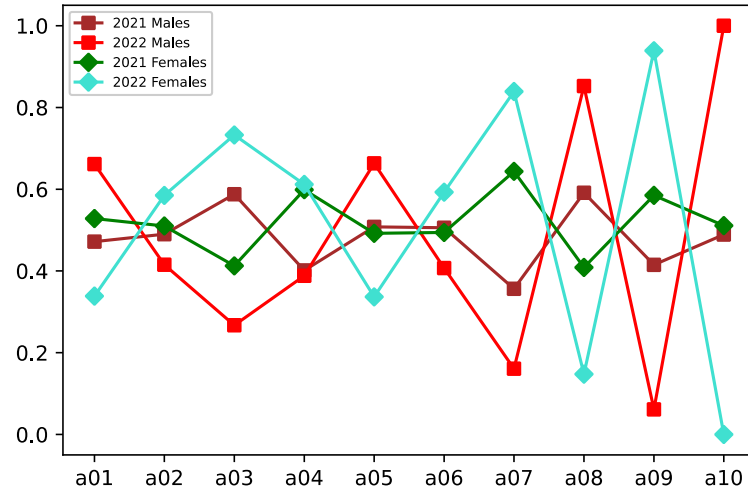
Οπτικοποιήσεις

- Σφάλματα ανά άτομο ανά φύλο χωρίς και με στάθμιση.

Line plot of errors per assignment from the ten assignments



Line plot of errors divided by exercise counts from the ten assignments

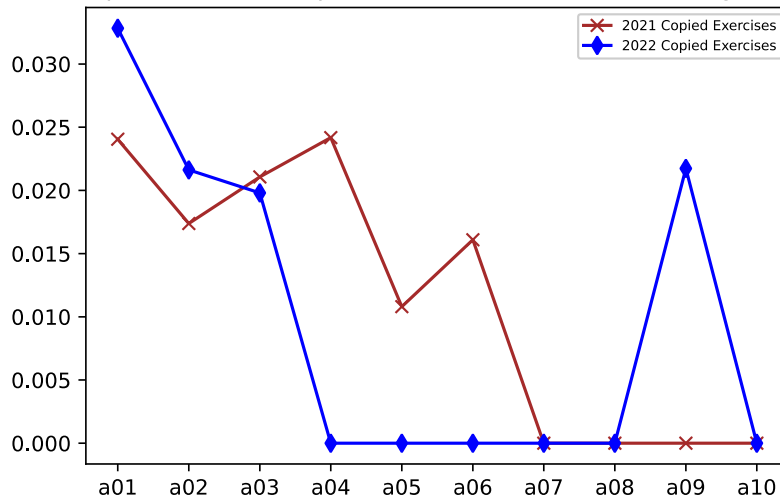


- Χωρίς την στάθμιση, οι άντρες φαίνονται να έχουν υψηλότερα ποσοστά, με στάθμιση βάσει του παραδομένου αριθμού ασκήσεων ανά φύλο για την κάθε εργασία, οι γυναίκες.

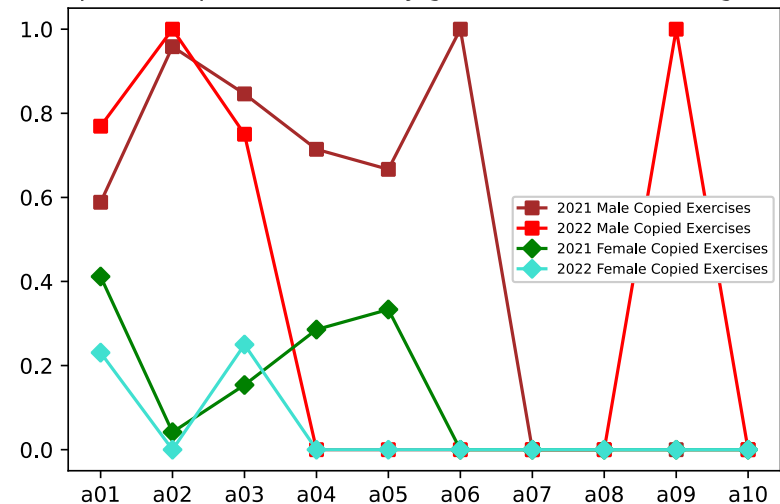
Οπτικοποιήσεις (2)

- Ποσοστό αντιγραμμένων ασκήσεων εργασιών απόλυτο και ανά φύλο.

Line plot of overall copied exercises from the ten assignments



Line plot of copied exercises by gender for the ten assignments

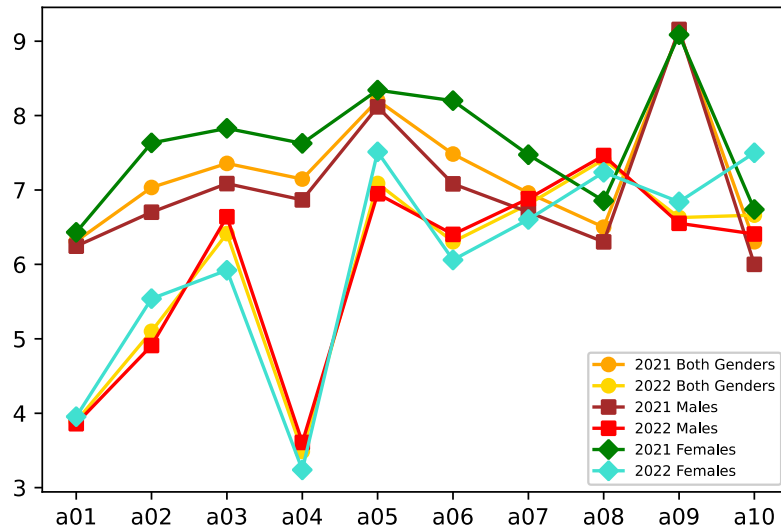


- Σχετικά χαμηλά ποσοστά (κάτω από 3.5%), συνολικά, το φαινόμενο παρατηρείται συχνότερα στο 2021, αλλά πιθανώς εντονότερα στο 2022.
- Παρατηρείται περισσότερο από τους άντρες παρά από τις γυναίκες.

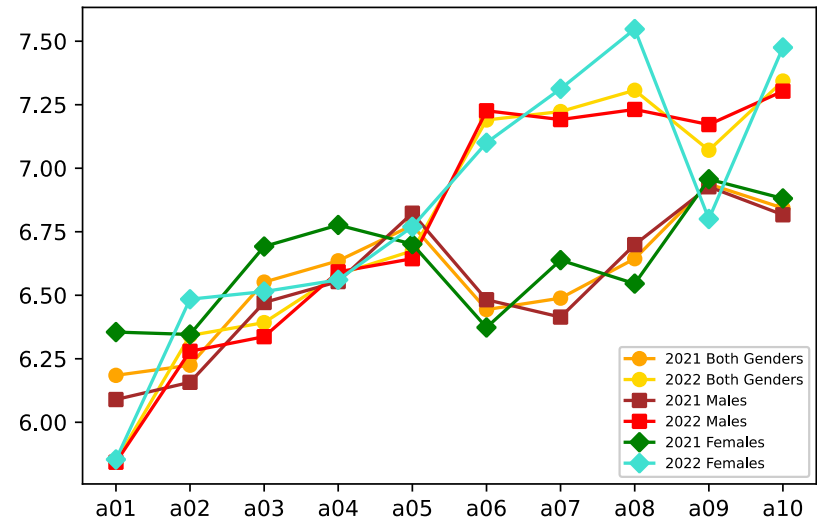
Οπτικοποιήσεις (3)

- Μέσοι όροι ανά φύλο από εργασίες και τελικούς βαθμούς.

Line plot of average assignment grades from the ten assignments



Line plot of average final course grades from the ten assignments



- Για εργασίες:
 - 2021 μεγαλύτεροι μ.ο. γυναικών για 9/10 εργασίες, το 2022 μόνο στις μισές, το 2022, αισθητά χαμηλότεροι μέσοι όροι μέχρι και την άσκηση 6.
- Για τελικούς βαθμούς:
 - Και στα δύο έτη, μεγαλύτεροι οι μ.ο. γυναικών σε 7/10 εργασίες, αισθητά καλύτεροι οι μ.ο. του 2022 από εργασία 6 και μετά (εκτός της 9).

Στατιστικοί Έλεγχοι Δεδομένων

- **Στόχος:** Διευκρίνηση αν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των δύο φύλων στον αριθμό λαθών τους, όπως επίσης στους βαθμούς των εργασιών και στον τελικό του μαθήματος.
- Πραγματοποιήθηκαν για δεδομένα ατόμων:
 - που παρέδωσαν καθεμιά από τις επιμέρους 12 ή 11 εργασίες της κάθε χρονιάς, σε επίπεδο εργασίας
 - των ετών 2021 και 2022 ξεχωριστά με ομαδοποίηση ανά έτος και
 - των ετών 2021 και 2022 μαζί
- και αφορούν:
 - τον αριθμό λαθών ανά φοιτητή,
 - τους βαθμούς των εργασιών τους και
 - τον τελικό βαθμό του μαθήματος.
- Επίπεδα σημαντικότητας που εξετάζονται: 90%, 95%, 99%.
- Έλεγχος Mann-Whitney για την Ομοιογένεια του Πληθυσμού.

Στατιστικοί Έλεγχοι Δεδομένων (2)

- Στα λάθη ανά άτομο:
 - Στατιστικά σημαντική διαφορά μόνο στην εργασία 11 του 2022 σε επίπεδα σημαντικότητας 95% και 90%.
- Στους βαθμούς εργασιών:
 - Σε επίπεδο σημαντικότητας 95%, διαφορά εμφανίζεται μόνο στην εργασία 7 του 2021, για 90% στην 7 και στις εργασίες 3, 4, 5 του 2021.
- Στους τελικούς βαθμούς του μαθήματος:
 - Δεν εντοπίζονται στατιστικά σημαντικές διαφορές μεταξύ των δύο φύλων σε καμιά περίπτωση και σε κανένα επίπεδο σημαντικότητας.
- Για επίπεδο σημαντικότητας 99%:
 - Δεν εντοπίζεται ποτέ καμία στατιστικά σημαντική διαφορά.

Κανόνες Συσχέτισης

- Ιδιότητες πειράματος:
 - **Αλγόριθμος:** Apriori | **Κάτω Όρια:** 0.1 Support, 0.01 Confidence
- Τα δεδομένα από τον parser και από τα αρχεία λογιστικού φύλλου μετατρέπονται σε δυαδική μορφή και προκύπτουν τα εξής:
 - *male/female*: το άτομο είναι ανδρικού ή γυναικείου φύλου,
 - *error_XY* ($XY \in [1, 46]$): το άτομο έκανε τουλάχιστον μια φορά το λάθος XY σε κάποια υποβληθείσα εργασία του,
 - *first_enrollment*: η εγγραφή του ατόμου στο μάθημα είναι η πρώτη του,
 - *took_course_in_its_semester*: το άτομο εγγράφηκε στο μάθημα στο δεύτερο εξάμηνο (δηλαδή στο εξάμηνο όπου διδάσκεται το μάθημα),
 - *submitted_assignment_XY* ($XY \in [1, 10]$): το άτομο παρέδωσε την εργασία XY
 - *assignment_XY_over_base* ($XY \in [1, 10]$): το άτομο πήρε πάνω από την βάση στην εργασία XY,
 - *passes_the_course/NOT_passes_the_course*: το άτομο περνάει ή δεν περνάει το μάθημα.

Κανόνες Συσχέτισης (2)

- Μερικοί από τους κανόνες όπου το άτομο περνάει το μάθημα είναι οι:

Κανόνας	Supp.	Conf.	Cos.	Lift	Jac.
{submitted_assignment_02}→{passes_the_course}	0.35	0.93	0.78	1.70	0.62
{submitted_assignment_03}→{passes_the_course}	0.34	0.93	0.76	1.70	0.59
{assignment_02_over_base}→{passes_the_course}	0.30	0.96	0.73	1.76	0.54
{took_course_in_its_semester}→{passes_the_course}	0.32	0.70	0.64	1.27	0.47
{took_course_in_its_semester, first_enrollment}→{passes_the_course}	0.32	0.70	0.64	1.27	0.47
{first_enrollment}→{passes_the_course}	0.33	0.69	0.64	1.26	0.47
{male}→{passes_the_course}	0.38	0.51	0.59	0.93	0.41
{error_43 (the 'gets' function is dangerous warning)}→{passes_the_course}	0.17	0.94	0.55	1.72	0.31
{male, took_course_in_its_semester}→{passes_the_course}	0.21	0.67	0.51	1.23	0.32
{male, first_enrollment, took_course_in_its_semester}→{passes_the_course}	0.21	0.67	0.51	1.23	0.32
{male, first_enrollment}→{passes_the_course}	0.21	0.66	0.51	1.21	0.32
{female}→{passes_the_course}	0.17	0.66	0.45	1.21	0.27
{took_course_in_its_semester, female}→{passes_the_course}	0.11	0.74	0.39	1.36	0.19
{took_course_in_its_semester, first_enrollment, female}→{passes_the_course}	0.11	0.74	0.39	1.36	0.19
{first_enrollment, female}→{passes_the_course}	0.11	0.74	0.39	1.35	0.19

Κανόνες Συσχέτισης (3)

- Οι κανόνες που παρατίθενται είναι ένα πολύ μικρό υποσύνολο από τους παραγμένους.
- Οι περισσότεροι κανόνες, ιδίως αυτοί με μεγαλύτερες τιμές Cosine Similarity είναι της μορφής:
 - “{submitted_assignment_XY} → {passes_the_course}”
 - “{assignment_XY_over_base} → {passes_the_course}”
 - “{submitted_assignment_XY, assignment_XY_over_base} → {passes_the_course}”
- Οι εργασίες με την σειρά που εμφανίζονται στους κανόνες, με φθίνουσα κατάταξη βάσει Cosine Similarity είναι οι 2, 3, 4, 1 και 5 με θέματα Stack, Queue, List – Dynamic Stack – Queue και Sets.
- Κανόνες που με ίδια ταξινόμηση περιλαμβάνουν “took_lesson_in_its_semester” ή/και “first_enrollment” εμφανίζονται πιο αργά, ενώ κανόνες που περιλαμβάνουν και φύλο ακόμα πιο αργά. Οι πάνω δύο παράγοντες συνδυασμένοι με ανδρικό φύλο εμφανίζονται πριν από τον συνδυασμό τους με γυναικείο φύλο. Όμως το γυναικείο φύλο από μόνο του έχει Lift > 1 εν αντιθέσει με το αντρικό.

Κανόνες Συσχέτισης (4)

- Ενδιαφέρον παρουσιάζει ότι από όλους τους κανόνες συσχέτισης που προέκυψαν, οι μοναδικοί που κατέληξαν στο αποτέλεσμα ότι το άτομο δεν περνάει το μάθημα και έχουν Lift > 1 είναι οι ακόλουθοι επτά:

Κανόνας	Supp.	Conf.	Cos.	Lift	Jac.
{NOT_took_course_in_its_semester}→{NOT_passes_the_course}	0.31	0.58	0.63	1.28	0.46
{male}→{NOT_passes_the_course}	0.37	0.49	0.63	1.09	0.44
{NOT_first_enrollment}→{NOT_passes_the_course}	0.30	0.58	0.62	1.28	0.45
{NOT_first_enrollment, NOT_took_course_in_its_semester}→ →{NOT_passes_the_course}	0.30	0.58	0.62	1.28	0.45
{male, NOT_took_course_in_its_semester}→{NOT_passes_the_course}	0.26	0.61	0.60	1.36	0.43
{male, NOT_first_enrollment}→{NOT_passes_the_course}	0.26	0.61	0.59	1.35	0.42
{male, NOT_took_course_in_its_semester, NOT_first_enrollment}→{NOT_passes_the_course}	0.26	0.61	0.59	1.35	0.42

- Φαίνεται ότι κάποιο άτομο δεν περνάει αν δεν πήρε το μάθημα στο 2^ο εξάμηνο φοίτησής του στο τμήμα, αν είναι αντρικού φύλου, αν δεν είναι η πρώτη φορά που παρακολουθεί το μάθημα ή πιο σίγουρα, αν συνδυάζει το αντρικού φύλο με τους εν λόγω παράγοντες.
- Δεν εντοπίζονται κανόνες τύπου error_XY → NOT_passes_the_course.

Συσταδοποίηση

- Έγιναν δύο δοκιμές για την συσταδοποίηση, μία με PCA και μία χωρίς.
- Στους πίνακες φαίνονται οι μεταβλητές κάθε δοκιμής και οι τιμές VIF.

- **Μέθοδοι Συσταδοποίησης:**

- Ιεραρχική, Φασματική και Μετατόπισης Μέσου

- **Μετρικές:**

- Μέσος Όρος Δεικτών Σιλουέτας (ή Σκόρ Σιλουέτας),
- Δείκτες Calinski-Harabasz και Davies-Bouldin,
- Αγνότητα κατά:
 - Φύλο, Χρονιά και Πέρασμα του μαθήματος.

Μόνο με Τυποποίηση Μεταβλητή VIF

error_count	1.189537
final_grade	4.717944
gender	1.412999
first_enrollment	2.196400
enrollment_semester	1.278370
has_cheated	3.523940
hw1_grade	3.714151
hw2_grade	6.907497
hw3_grade	5.477734
hw4_grade	6.273076
hw5_grade	4.745142
hw6_grade	5.936767
hw7_grade	6.384713
hw8_grade	6.110087
hw9_grade	4.606865
hw10_grade	3.093035

Με Τυποποίηση και PCA

Μεταβλητή VIF

first_enrollment	2.087287
enrollment_semester	1.210745
error_count	1.194092
has_cheated	3.957519
submitted_assignment_count	4.447209
final_grade	4.238331

Συσταδοποιήσεις με PCA

Ιεραρχική Συσταδοποίηση:

Clusters \ Metrics	2	3	4	5	6	7	8
Silhouette Score	0.505	0.588	0.664	0.659	0.577	0.557	0.533
Calinski-Harabaz	1460.086	2091.508	3601.752	3936.497	4290.55	4374.435	4434.682
Davies-Bouldin	0.854	0.647	0.476	0.49	0.633	0.565	0.676
Gender Purity	0.742	0.742	0.742	0.742	0.742	0.742	0.742
Year Purity	0.628	0.628	0.628	0.628	0.677	0.677	0.677
C. Pass Purity	0.832	0.832	0.832	0.85	0.873	0.873	0.873

Φασματική:

Clusters \ Metrics	2	3	4	5	6	7	8
Silhouette Score	0.547	0.612	0.695	0.537	0.544	0.524	0.52
Calinski-Harabaz	1866.141	2274.164	4554.084	1662.538	3545.597	3444.713	2739.474
Davies-Bouldin	0.735	0.515	0.431	0.991	0.648	0.715	0.719
Gender Purity	0.742	0.742	0.742	0.742	0.742	0.742	0.742
Year Purity	0.587	0.599	0.608	0.627	0.663	0.649	0.667
C. Pass Purity	0.811	0.835	0.843	0.839	0.853	0.853	0.866

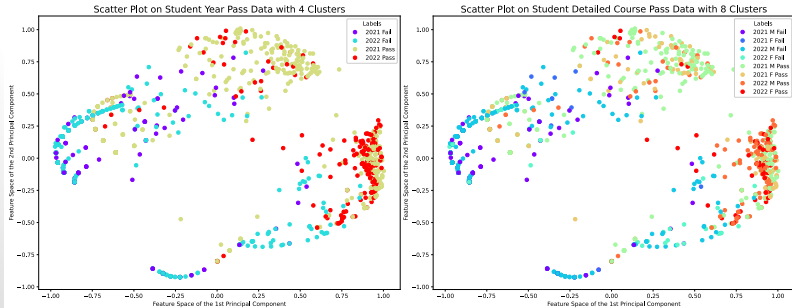
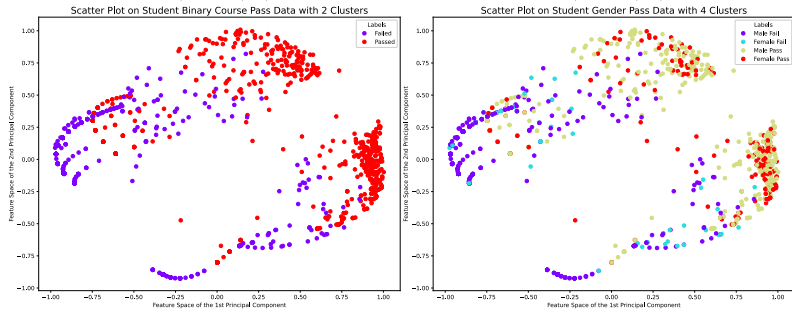
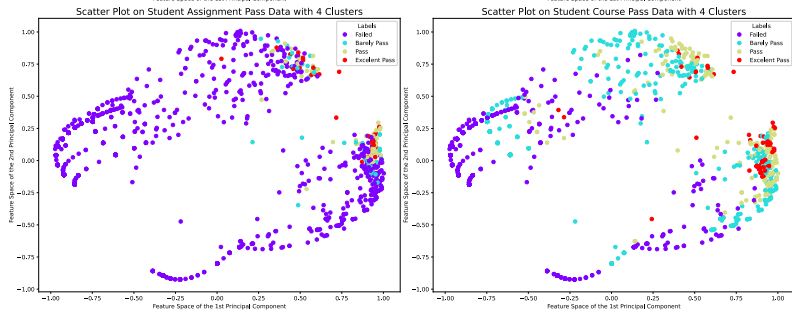
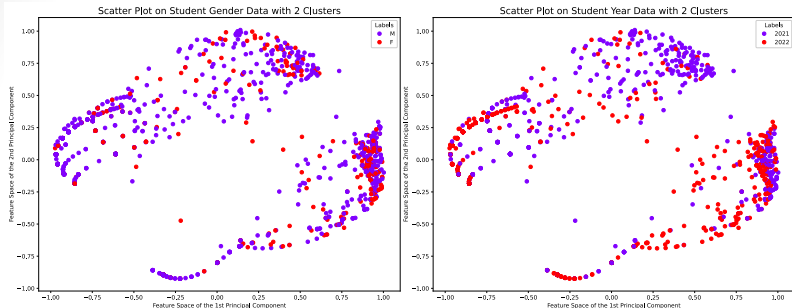
Μετατόπισης

Μέσου:

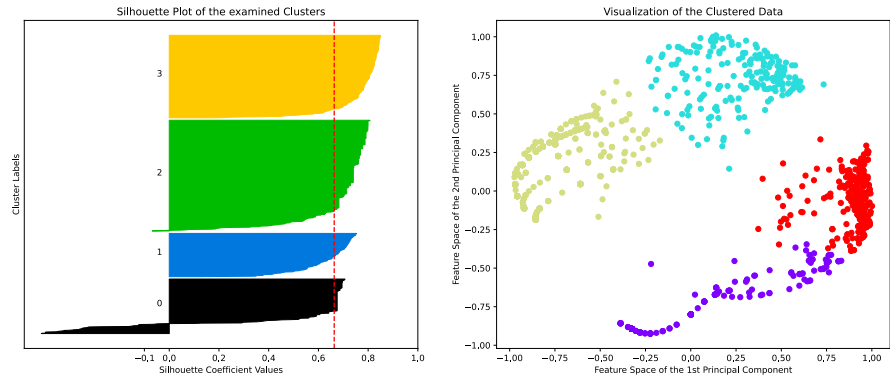
Clusters \ Metrics	4
Silhouette Score	0.695
Calinski-Harabaz	4568.141
Davies-Bouldin	0.434
Gender Purity	0.742
Year Purity	0.608
C. Pass Purity	0.841

Βέλτιστος αριθμός συστάδων, βάσει των περισσότερων μετρικών και όλων των μεθόδων: 4.

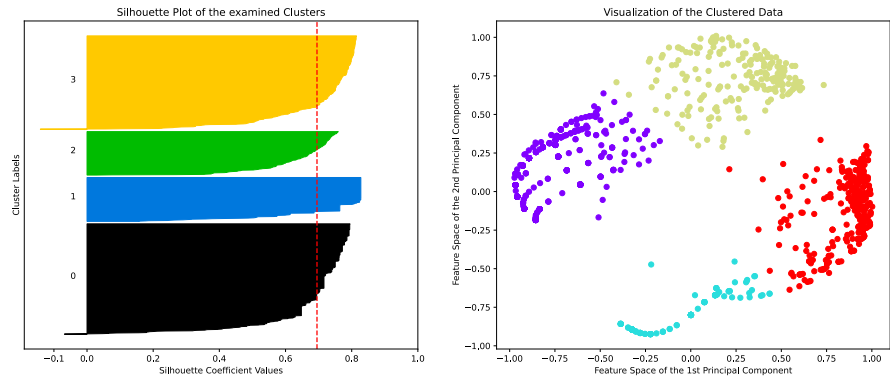
Συσταδοποιήσεις με PCA (2)



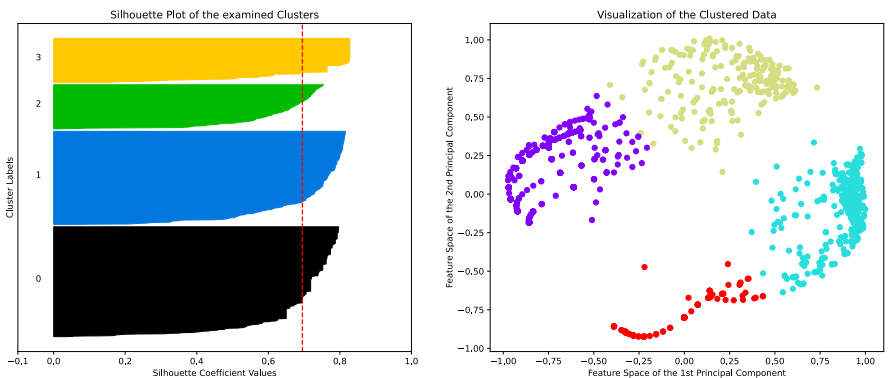
Silhouette Analysis for Hierarchical Clustering on Student Data after PCA with 4 Clusters



Silhouette Analysis for Spectral Clustering on Student Data after PCA with 4 Clusters



Silhouette Analysis for Mean Shift Clustering on Student Data after PCA with 4 Clusters



Συσταδοποιήσεις χωρίς PCA

Ιεραρχική Συσταδοποίηση:

Clusters \ Metrics	2	3	4	5	6	7	8
Silhouette Score	0.347	0.272	0.267	0.305	0.321	0.312	0.324
Calinski-Harabaz	699.535	528.552	466.832	439.129	435.042	410.49	387.232
Davies-Bouldin	1.281	1.575	1.549	1.537	1.418	1.339	1.342
Gender Purity	0.742	0.742	0.742	0.862	0.862	0.862	0.862
Year Purity	0.636	0.636	0.636	0.636	0.636	0.642	0.644
C. Pass Purity	0.763	0.763	0.763	0.763	0.783	0.783	0.783

Φασματική:

Clusters \ Metrics	2	3	4	5	6	7	8
Silhouette Score	0.362	0.29	0.313	0.316	0.327	0.314	0.329
Calinski-Harabaz	762.763	499.615	535.453	473.409	458.552	432.537	398.896
Davies-Bouldin	1.231	1.484	1.393	1.438	1.396	1.328	1.273
Gender Purity	0.742	0.742	0.742	0.849	0.893	0.878	0.878
Year Purity	0.616	0.589	0.624	0.624	0.616	0.621	0.665
C. Pass Purity	0.779	0.805	0.788	0.76	0.794	0.797	0.855

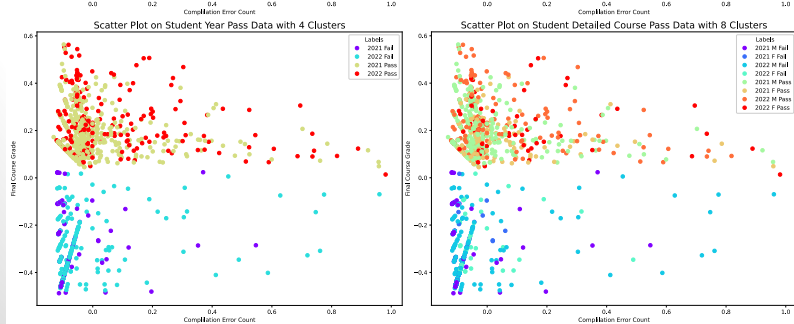
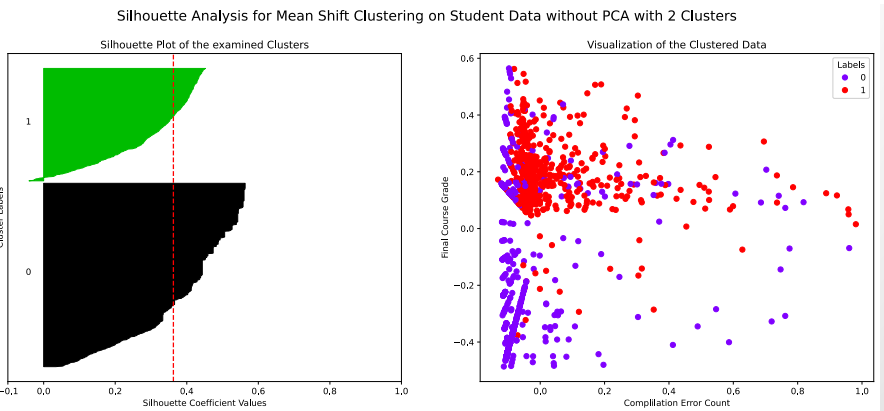
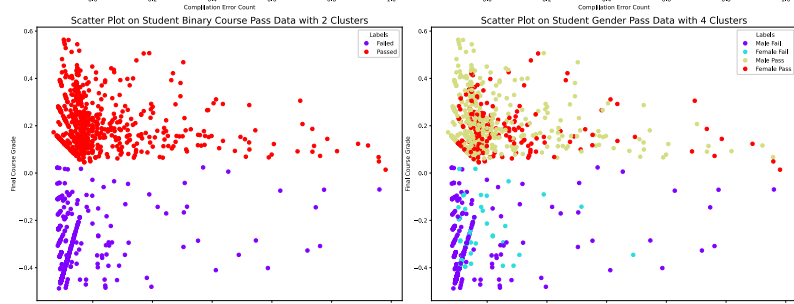
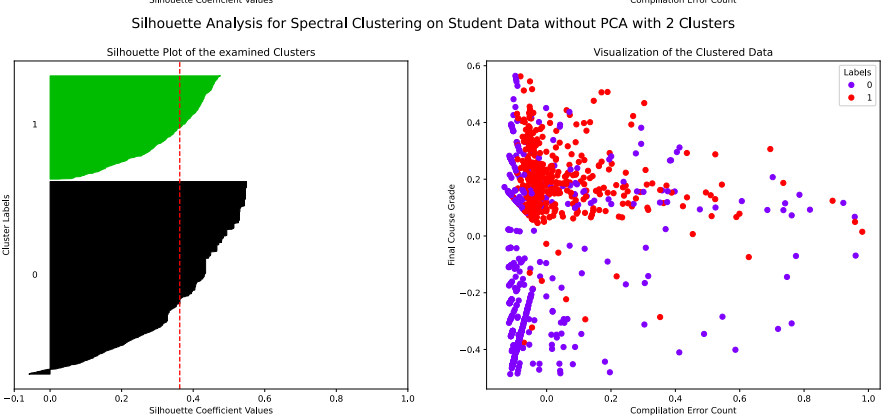
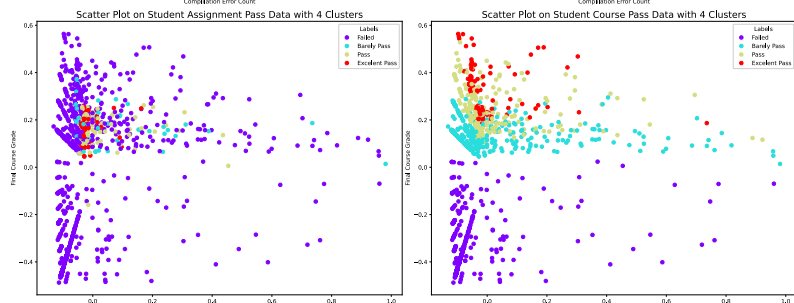
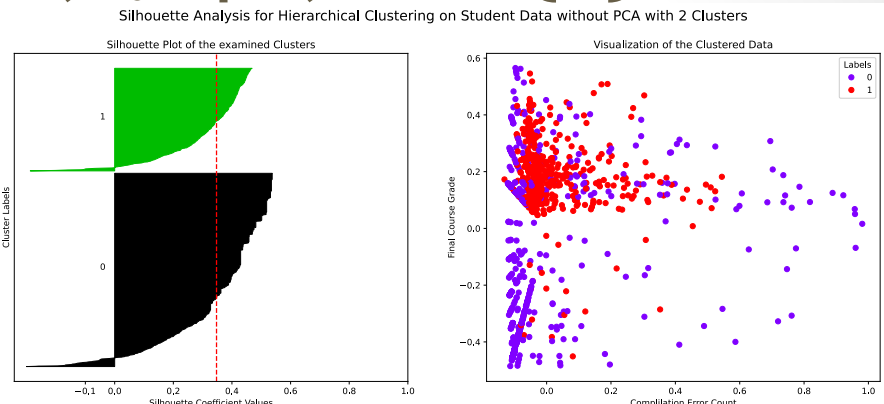
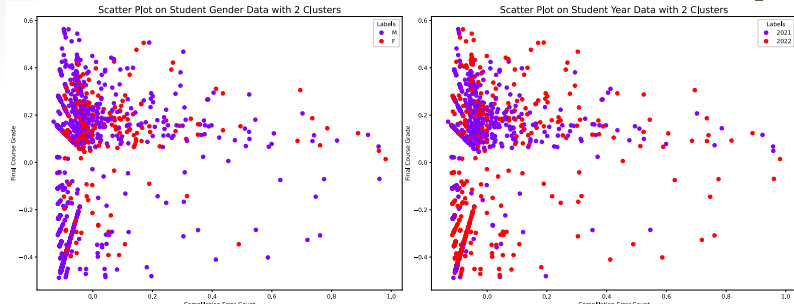
Μετατόπισης

Μέσου:

Clusters \ Metrics	2
Silhouette Score	0.363
Calinski-Harabaz	779.315
Davies-Bouldin	1.248
Gender Purity	0.742
Year Purity	0.621
C. Pass Purity	0.803

Βέλτιστος αριθμός συστάδων, βάσει των περισσότερων μετρικών και όλων των μεθόδων: 2.

Συσταδοποιήσεις χωρίς PCA (2)



Μοντέλα Μηχανικής Μάθησης

- Δημιουργήθηκαν 5 μοντέλα μηχανικής μάθησης:
 - Λογιστικής Παλινδρόμησης
 - Δένδρου Απόφασης
 - Μηχανής Διανύσματος Υποστήριξης
 - Νευρωνικού Δικτύου
 - Πολυεπίπεδος Αισθητήρας και Ακολουθιακό Νευρωνικό Δίκτυο.
- Ως στήλες δεδομένων, χρησιμοποιήθηκαν σε όλες τις περιπτώσεις:
 - ο αριθμός λαθών που έχει κάνει το εξεταζόμενο άτομο (λαμβάνοντας υπόψη τα 46 σφάλματα, όσες φορές εμφανίστηκε το καθένα),
 - ο μέσος όρος των εργασιών που παρέδωσε (με άνω όριο τις 10).
- Στήλη στόχος είναι το αν το άτομο πέρασε ή όχι το μάθημα
 - (τελικός βαθμός ≥ 5): 1 – περνάει | (τελικός βαθμός < 5): 0 – δεν περνάει
- Υλοποιείται K-πλή Διασταυρωμένη Επικύρωση (K-fold Cross Validation)
 - με $K=10$ για όλα τα μοντέλα.

Μοντέλα Μηχανικής Μάθησης (2)

- Για εκπαίδευση, επαλήθευση και δοκιμή:
 - χρησιμοποιήθηκαν όλα τα δεδομένα και των δύο ετών, άρα όλων των 1323 ατόμων, αλλά μετά από αφαίρεση των ανενεργών φοιτητών
 - δηλαδή όσοι έχουν: `error_count, submitted_assignment_count & final_grade = 0`
 - και στα παραδείγματα για κ-πλή εκπαίδευση και επαλήθευση (k-fold training and validation) και σε αυτά για δοκιμή (testing), ως νέα άγνωστα παραδείγματα, χρησιμοποιήθηκαν δεδομένα και από τα δύο έτη (με 3-fold διαχωρισμό βάσει έτους, εκ των οποίων τα δεδομένα των πρώτων 2 folds για εκπαίδευση/επαλήθευση και του 3^{ου} για δοκιμή)
 - με κριτήριο διαχωρισμού τους σε έτος το αν ένα id φοιτητή εμφανίζεται στο αρχείο δεδομένων του 2022 (οπότε θεωρείται ότι ανήκει στο 2022), ή όχι (άρα τότε ανήκει στο 2021), με διόρθωση όπου αν στο 2021 έχει υψηλότερο μ.ό. εργασιών από το 2022, καταχωρείται κατευθείαν στο 2021.
- Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι μετρικές:
 - Αποτελεσματικότητα, Ακρίβεια, Ανάκληση, Μετρική F1, Περιοχή Κάτω από την Καμπύλη ή A* και Μετρική Κάπα του Cohen. Χρησιμοποιούμενες εκδόσεις από Sci-Kit Learn.

Μοντέλα Μηχανικής Μάθησης (3)

feature	VIF
error_count	1.095056
average_assignment_grade	1.923464
final_grade	1.998182

- *Χαρακτηριστικά Υποδειγματοληψίας όλων των μοντέλων:*
 - **NearMiss**(version=3, n_neighbors=4, n_neighbors_ver3=4, n_jobs=-1)
- *Χαρακτηριστικά κατηγοριοποιητών Sci-Kit Learn:*
 - **LogisticRegression**(tol=1e-4, class_weight="balanced", max_iter=1000000, n_jobs=-1)
 - **DecisionTreeClassifier**(criterion="gini", class_weight="balanced", max_depth=6, random_state=1)
 - **LinearSVC**(max_iter=1000000, dual=False, tol=1e-24)
 - **MLPClassifier**(hidden_layer_sizes=(32,), solver="adam", random_state=1, max_iter=30000)

Μοντέλα Μηχανικής Μάθησης (4)

- *Χαρακτηριστικά Νευρωνικού Δικτύου Tensorflow:*
 - `model = Sequential()`
 - `Dense(640, kernel_initializer="normal", kernel_regularizer=regularizers.l2(0.001), bias_initializer="zeros", activation="relu", input_shape=(chosen_input_shape,))`
 - `Dense(1280, kernel_initializer="normal", kernel_regularizer=regularizers.l2(0.001), bias_initializer="zeros", activation="relu") x2`
 - `Dense(1, activation="sigmoid")`
 - `model.compile(loss=BinaryCrossentropy(from_logits=True), optimizer=SGD(learning_rate=16*1e-2, momentum=36*1e-2, nesterov=True), metrics=[Precision(name='precision'), Recall(name='recall'), AUC(name='auc'), BinaryAccuracy()])`
 - `model.fit(x_train, y_train, batch_size=24, epochs=320, validation_split=float(0.1), callbacks=[EarlyStopping(monitor='val_loss', patience=6)], shuffle=True)`

Μοντέλα Μηχανικής Μάθησης (5)

- Αποτελέσματα Εκπαίδευσης/Επαλήθευσης με δεδομένα μέχρι την 10^η εβδομάδα.

Model Metric	Logistic Regression	Decision Tree	Support Vector Machine	Sequential Neural Network	Multi-Layer Perceptron
Accuracy	0.689 (+/- 0.053)	0.701 (+/- 0.046)	0.674 (+/- 0.047)	0.671 (+/- 0.039)	0.684 (+/- 0.043)
Precision	0.925 (+/- 0.059)	0.869 (+/- 0.053)	0.948 (+/- 0.052)	0.928 (+/- 0.047)	0.868 (+/- 0.049)
Recall	0.638 (+/- 0.047)	0.712 (+/- 0.061)	0.599 (+/- 0.046)	0.61 (+/- 0.042)	0.684 (+/- 0.046)
F1 Metric	0.754 (+/- 0.043)	0.78 (+/- 0.037)	0.733 (+/- 0.042)	0.735 (+/- 0.035)	0.764 (+/- 0.033)
A* - AUC	0.739 (+/- 0.073)	0.691 (+/- 0.067)	0.749 (+/- 0.061)	0.732 (+/- 0.054)	0.683 (+/- 0.067)
Cohen's Kappa	0.365 (+/- 0.111)	0.322 (+/- 0.106)	0.365 (+/- 0.089)	0.346 (+/- 0.078)	0.3 (+/- 0.105)

- Όχι ιδιαίτερα υψηλές τιμές μετρικών – ιδιαιτέρως τιμές Kappa < 0.3 πάντα στην χειρότερη περίπτωση.
- Πιθανώς παρεπόμενο της Υποδειγματοληψίας.

Μοντέλα Μηχανικής Μάθησης (6)

- Αποτελέσματα Δοκιμής/Προβλέψεων με δεδομένα μέχρι την 10^η εβδομάδα.

Model \ Metric	Logistic Regression	Decision Tree	Support Vector Machine	Sequential Neural Network	Multi-Layer Perceptron
Accuracy	0.735 (+/- 0.004)	0.756 (+/- 0.011)	0.719 (+/- 0.009)	0.726 (+/- 0.018)	0.766 (+/- 0.007)
Precision	0.948 (+/- 0.004)	0.921 (+/- 0.011)	0.97 (+/- 0.005)	0.961 (+/- 0.01)	0.927 (+/- 0.008)
Recall	0.713 (+/- 0.005)	0.765 (+/- 0.02)	0.675 (+/- 0.009)	0.692 (+/- 0.03)	0.773 (+/- 0.012)
F1 Metric	0.814 (+/- 0.003)	0.836 (+/- 0.009)	0.796 (+/- 0.007)	0.804 (+/- 0.017)	0.843 (+/- 0.006)
A* - AUC	0.772 (+/- 0.006)	0.74 (+/- 0.018)	0.792 (+/- 0.01)	0.784 (+/- 0.009)	0.754 (+/- 0.014)
Cohen's Kappa	0.385 (+/- 0.008)	0.375 (+/- 0.021)	0.388 (+/- 0.015)	0.388 (+/- 0.017)	0.398 (+/- 0.017)

- Αρκετά ικανοποιητικές τιμές μετρικών, ιδιαιτέρως Kappa > 0.3 ακόμα και στην χειρότερη περίπτωση. Προβλέψεις εφικτές σε κάποιον βαθμό.
- Multi-Layer Perceptron το μοντέλο με την συνολικά καλύτερη προβλεπτική ικανότητα.

Μοντέλα Μηχανικής Μάθησης (7)

- Αποτελέσματα Δοκιμής/Προβλέψεων με δεδομένα μέχρι την 3^η εβδομάδα.

Model \ Metric	Logistic Regression	Support Vector Machine	Sequential Neural Network	Multi-Layer Perceptron
Accuracy	0.728 (+/- 0.003)	0.723 (+/- 0.004)	0.696 (+/- 0.022)	0.739 (+/- 0.005)
Precision	0.939 (+/- 0.006)	0.941 (+/- 0.007)	0.956 (+/- 0.016)	0.937 (+/- 0.01)
Recall	0.691 (+/- 0.002)	0.683 (+/- 0.003)	0.635 (+/- 0.039)	0.708 (+/- 0.008)
F1 Metric	0.796 (+/- 0.002)	0.791 (+/- 0.003)	0.762 (+/- 0.025)	0.807 (+/- 0.004)
A* - AUC	0.771 (+/- 0.007)	0.77 (+/- 0.009)	0.768 (+/- 0.011)	0.775 (+/- 0.011)
Cohen's Kappa	0.414 (+/- 0.009)	0.409 (+/- 0.012)	0.386 (+/- 0.021)	0.427 (+/- 0.014)

- Παράληψη του Decision Tree, καθώς καταφέρνει Kappa > 0.3 μόνο από την 7^η εβδομάδα και μετά.
- Εκτός από Kappa, αρκετά χαμηλότερες τιμές μετρικών, κάτι το αναμενόμενο. Συνεχίζουν να είναι σχετικά ικανοποιητικές.
- Multi-Layer Perceptron παραμένει η πιθανώς καλύτερη επιλογή.

Μοντέλα Μηχανικής Μάθησης (8)

- Αποτελέσματα Δοκιμής/Προβλέψεων με δεδομένα μέχρι την 5^η εβδομάδα.

Model	Logistic Regression	Support Vector Machine	Sequential Neural Network	Multi-Layer Perceptron
Accuracy	0.738 (+/- 0.004)	0.731 (+/- 0.004)	0.725 (+/- 0.016)	0.764 (+/- 0.005)
Precision	0.951 (+/- 0.003)	0.962 (+/- 0.002)	0.969 (+/- 0.009)	0.944 (+/- 0.007)
Recall	0.714 (+/- 0.007)	0.696 (+/- 0.005)	0.683 (+/- 0.025)	0.755 (+/- 0.007)
F1 Metric	0.816 (+/- 0.004)	0.808 (+/- 0.004)	0.801 (+/- 0.015)	0.839 (+/- 0.004)
A* - AUC	0.778 (+/- 0.004)	0.789 (+/- 0.004)	0.794 (+/- 0.009)	0.78 (+/- 0.012)
Cohen's Kappa	0.394 (+/- 0.006)	0.396 (+/- 0.007)	0.395 (+/- 0.016)	0.42 (+/- 0.014)

- Αποτελέσματα αρκετά κοντά σε αυτά της 10^{ης} εβδομάδας (ή και υψηλότερα, όπως τα Kappa), άρα ενδεχομένως η καλύτερη εβδομάδα για «πρώιμες» προβλέψεις.
- Multi-Layer Perceptron παραμένει και εδώ καλή συνολική επιλογή.
- Πιθανώς να είναι το προτιμότερο μοντέλο απ' όλα.

Επίλογος – Συγκρίσεις με Άλλες Έρευνες

- Αναφορικά με το 1^ο από τα 3 **Ερευνητικά Ερωτήματα**:
 - 1) Τα λάθη και η ακαδημαϊκή απόδοση των φοιτητών στο αντικείμενο του προγραμματισμού διαφέρουν ανά φύλο και αν ναι, πώς;
- Οι γυναίκες ενδεχομένως να έχουν λίγο μεγαλύτερη πιθανότητα σφάλματος, αλλά επίσης έχουν μεγαλύτερο μ.ο. εργασιών και τελικού βαθμού μαθήματος.
- Στατιστικά σημαντικές διαφορές μεταξύ των φύλων σπάνια συμβαίνει να αναγνωριστούν, αλλά όσες φορές συνέβη, ήταν υπέρ των γυναικών.
- Το εύρημα μοιάζει με αυτό της έρευνας των Pillay N. και Vikash R. (2005), όπου σχολιάζεται πώς το φύλο δεν επηρεάζει με στατιστικά σημαντικό τρόπο τον τελικό βαθμό στο μάθημα, αν και εδώ, την λίγο υψηλότερη απόδοση φαίνεται να είχαν οι άντρες.

Επίλογος – Συγκρίσεις (2)

- Αναφορικά με το 2^ο από τα 3 **Ερευνητικά Ερωτήματα**:
- 2) Μπορούν οι μέθοδοι της Συσταδοποίησης και των Κανόνων Συσχέτισης να συμβάλλουν στην κατανόηση των σχέσεων που δημιουργούνται μεταξύ των διαφορετικών οντοτήτων του μαθήματος;
- Με το 2^ο πείραμα συσταδοποίησης, φάνηκε ότι τα δεδομένα χωρίζονται σε δύο συστάδες, που μοιάζουν με τα ισχύοντα στοιχεία σχετικά με το αν το άτομο περνάει ή όχι, όπως επίσης με το φύλο.
- Από τους κανόνες συσχέτισης, προέκυψε ότι η ενασχόληση με εργασίες, ιδιαιτέρως τις 2, 3, 4, 1, 5 βοηθάει σημαντικότερα στο να περαστεί το μάθημα.
- Εύρημα παρόμοιο με έρευνες των (Höök, L. J., & Eckerdal, A. (2015, April)) και (Matetic M., Marija B. B., and Sabina S. (2015, June)) όπου περισσότερη ενασχόληση με το μάθημα σημαίνει ότι περνιέται ευκολότερα και (Damašeničius R. (2010)), όπου αν και αποτυχία στο αντικείμενο των λιστών οδηγεί σε αποτυχία στο μάθημα, εδώ ενασχόληση με εργασίες με λίστες βοηθάει στο να περάσει το μάθημα.

Επίλογος – Συγκρίσεις (3)

- Αναφορικά με το 3^ο από τα 3 **Ερευνητικά Ερωτήματα**:
- 3) Μπορεί να προβλεφθεί αν ένα άτομο θα περάσει ή όχι το μάθημα κάνοντας χρήση μόνο του αριθμού των λαθών μεταγλώττισης του κώδικα των εργασιών του και του μέσου όρου των εργασιών του πριν τις εξετάσεις ή κατά προτίμηση και ακόμα νωρίτερα;
- Γίνεται, σε σχετικά ικανοποιητικό βαθμό στο τέλος του εξαμήνου, πριν από τις εξετάσεις, καλούτσικα ήδη από την 3^η εβδομάδα από τις 10 και σχεδόν το ίδιο καλά με το τέλος του εξαμήνου στην 5^η.
- Οι μετρικές συνήθως είναι χαμηλότερες από αυτές άλλων ερευνών ((Mueen A., Zafar B. & Manzoor, U. (2016)), (Kostopoulos, G., Lipitakis, A. D., Kotsiantis, S., et al. (2017)), (Bucos M., & Drăgulescu, B. (2018)))
 - όμως και οι μεταβλητές που χρησιμοποιούνται συγκριτικά ελάχιστες.
- Όταν όμως οι προβλέψεις γίνονται στα μέσα του εξαμήνου, οι μετρικές τείνουν να είναι πιο συγκρίσιμες ή ενδεχομένως και καλύτερες σε σχέση με τις αντίστοιχες άλλων ερευνητικών εργασιών, όπως η (Kostopoulos, G., Lipitakis, A. D., Kotsiantis, S., et al. (2017)).

Τέλος Παρουσίασης

Ευχαριστώ για την διάθεση της προσοχής και
του χρόνου σας

Ερωτήσεις;

Παραπομπές

- Karakasidis, A. (2023). Diorthotis: A Parallel Batch Evaluator for Programming Assignments. In: 29th International European Conference on Parallel and Distributed Computing Euro-Par 2023: Parallel Processing Workshops.
- Pillay, N., & Jugoo, V. R. (2005). An investigation into student characteristics affecting novice programming performance. ACM Sigcse Bulletin, 37(4), 107-110.
- Höök, L. J., & Eckerdal, A. (2015, April). On the bimodality in an introductory programming course: An analysis of student performance factors. In 2015 International conference on learning and teaching in computing and engineering (pp. 79-86). IEEE.
- Matetic, M., Bakaric, M. B., & Sisovic, S. (2015, June). Association rule mining and visualization of introductory programming course activities. In Proceedings of the 16th International Conference on Computer Systems and Technologies (pp. 374-381).

Παραπομπές (2)

- Damaševičius, R. (2010). Analysis of academic results for informatics course improvement using association rule mining. *Information Systems Development: Towards a Service Provision Society*, 357-363.
- Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36.
- Kostopoulos, G., Lipitakis, A. D., Kotsiantis, S., & Gravvanis, G. (2017). Predicting student performance in distance higher education using active learning. In *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings* (pp. 75-86). Springer International Publishing.
- Bucos, M., & Drăgulescu, B. (2018). Predicting student success using data generated in traditional educational environments. *TEM Journal*, 7(3), 617.