



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ

«ΠΛΗΡΟΦΟΡΙΚΗ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΑΣΦΑΛΕΙΑ, ΔΙΑΧΕΙΡΙΣΗ
ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΡΟΣΟΜΟΙΩΣΗ»

**«Εξόρυξη δεδομένων- Δημιουργία μοντέλου πρόβλεψης
επιτυχίας ή μη στο μάθημα των Μαθηματικών»**

Παρασκευοπούλου Ευσταθία

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων καθηγητής:

Κωνσταντίνος Κολομβάτσος

Λαμία, έτος 2020



UNIVERSITY OF THESSALY

SCHOOL OF SCIENCE

INFORMATICS AND COMPUTATIONAL BIOMEDICINE

**“Data mining – Creating a model for predicting success or not
in the subject of Mathematics”**

Paraskevopoulou Efstathia

Master thesis

Supervisor

Konstantinos Kolomvatsos

Lamia , 2020

«Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης»

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο «Εξόρυξη δεδομένων- Δημιουργία μοντέλου πρόβλεψης επιτυχίας ή μη στο μάθημα των Μαθηματικών» αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Η ΔΗΛΟΥΣΑ

Παρασκευοπούλου Ευσταθία

Ημερομηνία

Υπογραφή

**«Εξόρυξη δεδομένων- Δημιουργία μοντέλου πρόβλεψης επιτυχίας
ή μη στο μάθημα των Μαθηματικών»**

Παρασκευοπούλου Ευσταθία

Τριμελής Επιτροπή:

Κωνσταντίνος Κολομβάτσος, Επικ. Καθηγητής (επιβλέπων)

Αθανάσιος Λουκόπουλος, Επικ. Καθηγητής

Νικόλαος Τζιρίτας, Επικ. Καθηγητής

ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία μελετάμε τον τρόπο με τον οποίο μπορούν να εφαρμοστούν τεχνικές Εξόρυξης Γνώσης-Data Mining στην εκπαίδευση. Ο επιστημονικός τομέας ο οποίος ερευνά και αναπτύσσει τεχνικές προκειμένου να ανακαλύψει γνώση από δεδομένα τα οποία προέρχονται από την εκπαίδευση ονομάζεται Εξόρυξη Γνώσης από Εκπαιδευτικά δεδομένα-Educational Data Mining (EDM). Ο στόχος της εξόρυξης δεδομένων είναι να παραχθεί ένα μοντέλο που μπορεί να χρησιμοποιηθεί για την πραγματοποίηση ταξινόμησης, πρόβλεψης, εκτίμησης ή άλλων παρόμοιων εργασιών.

Στην εργασία αυτή εκτός από τη θεωρητική μελέτη των αλγορίθμων και των τεχνικών που διέπουν την εξόρυξη γνώσης από δεδομένα γενικά, γίνεται μια λεπτομερέστερη μελέτη και παρουσίαση της Κατηγοριοποίησης-Classification και συγκεκριμένα θα μελετήσουμε τα Νευρωνικά Δίκτυα διότι αυτά θα χρησιμοποιήσουμε στη φάση της υλοποίησης του μοντέλου. Τα νευρωνικά δίκτυα είναι μια μεθοδολογία για ανάπτυξη εφαρμογών τεχνητής νοημοσύνης, τα οποία αποτελούνται από ένα σύνολο απλών, διασυνδεδεμένων και προσαρμοστικών μονάδων, οι οποίες συνιστούν ένα πολύπλοκο υπολογιστικό μοντέλο. Τα Τεχνητά Νευρωνικά Δίκτυα-Artificial Neural Network χρησιμοποιούνται όλο και περισσότερο καθώς με τη βοήθειά τους έχουμε τη δυνατότητα να αντιμετωπίσουμε πολύπλοκα προβλήματα με εφαρμογή σε πολλούς και διαφορετικούς τομείς της επιστήμης και της τεχνολογίας.

Στην συνέχεια η εργασία θα επικεντρωθεί στην έρευνα που πραγματοποιήθηκε ώστε να συλλεχθούν δεδομένα, τα οποία θα χρησιμοποιηθούν για τη δημιουργία ενός μοντέλου πρόβλεψης επιτυχίας ή μη στις πανελλαδικές εξετάσεις, από τις επιδόσεις των μαθητών στο μάθημα των μαθηματικών. Οι μέθοδοι πρόβλεψης συμβάλλουν στη σωστή και έγκαιρη λήψη αποφάσεων. Στόχος των προβλέψεων είναι να είναι όσο το δυνατόν πιο εύστοχες και να ελαχιστοποιηθεί η απόκλιση τους από τις πραγματικές μελλοντικές τιμές που προσπαθούν να περιγράψουν.

Έπειτα αναλύουμε και αποδεικνύουμε πως με την εφαρμογή ενός γενετικού αλγορίθμου N γενεών, βρίσκουμε τις καλύτερες παραμέτρους και έχουμε το νευρωνικό που με συνεχείς μεταλλάξεις και διασταυρώσεις του πληθυσμού θα πετύχει την καλύτερη επίδοση. Η ιδέα του συνδυασμού των νευρωνικών δικτύων και ενός γενετικού αλγορίθμου προέκυψε από τις δυνατότητες τους να βρίσκουν λύση εκεί που άλλες παραδοσιακές μέθοδοι αποτυγχάνουν. Από την μία πλευρά, τα Τεχνητά Νευρωνικά Δίκτυα- ΤΝΔ μιμούνται τις λειτουργίες του ανθρώπινου εγκεφάλου και από την άλλη, οι γενετικοί αλγόριθμοι μιμούνται την διαδικασία της εξέλιξης στην φύση.

Λέξεις κλειδιά: Εξόρυξη δεδομένων, Κατηγοριοποίηση, Νευρωνικά Δίκτυα, Γενετικός Αλγόριθμος

ABSTRACT

In this master thesis, we study how various Knowledge of Data Mining techniques, and how data mining can be used for educational purposes. In this point we have to report Educational Data Mining-EDM which is a scientific field that researches and develops techniques, in order to discover knowledge from data, coming from education. The purpose of data mining is to produce a model that it could be used to perform classification, prediction, estimation or other similar tasks.

In this work, apart from the theoretical study of a set of algorithms and techniques governing the extraction of knowledge from data, we will present a more detailed study and presentation of Classification. Specifically, we will study Neural Networks because we will use them in the implementation phase of the model. Neural Networks assist in performing artificial intelligence, consisting of a set of simple, interconnected and adaptive units (i.e., neurons), which constitute a complex computational model. The Artificial Neural Networks are increasingly adopted by the research community as, with their assistance, we have the ability to deal complex problems which they apply to many different fields of science and technology.

Furthermore, this work focuses on the research carried out to collect data, which will be used to create a model for predicting success or failure in National exams in the subject of mathematics. The aim of the proposed forecasting is to be as accurate as possible and to minimize the deviation of these predictions, from the actual future values which they are trying to describe.

Then, we analyze and prove how the application of a genetic algorithm, involving N generations, can give us the neural network, which achieve the best performance using continuous mutations and crosses of the population. The idea of combining neural networks and genetic algorithms emerged by their innate ability finding solutions where traditional methods fail. On one hand, ANNs imitate the human brain procedures and on the other hand, genetic algorithms imitate the physical evolution process.

Keywords: data mining, classification, neural networks, genetic algorithm

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 2-1 Η εξόρυξη γνώσης ως συμβολή πολλών αρχών.....	13
Εικόνα 2-2 Η διαδικασία της Ανεύρεσης Γνώσης στις Βάσεις Δεδομένων.....	15
Εικόνα 2-3 Κύριοι τομείς που σχετίζονται με την εξόρυξη εκπαιδευτικών δεδομένων.....	17
Εικόνα 2-4 Μηχανική Μάθηση.....	19
Εικόνα 2-5 Βασικές τεχνικές Data Mining.....	24
Εικόνα 2-6 Σχηματική Απεικόνιση Κατηγοριοποίησης.....	25
Εικόνα 2-7 Κατηγορίες Μηχανικής Μάθησης.....	27
Εικόνα 2-8 Παράδειγμα σύγκρισης Ταξινόμησης- Παλινδρόμησης-Ομαδοποίησης.....	28
Εικόνα 2-9 Σύγκριση φυσικού νευρώνα με στοιχειώδη τεχνητό νευρώνα.....	30
Εικόνα 2-10 Χαρακτηριστικά ΤΝΔ.....	31
Εικόνα 2-11 ΤΝΔ πρόσθιας τροφοδότησης.....	31
Εικόνα 2-12 Νευρωνικό δίκτυο τροφοδοσίας δύο επιπέδων.....	32
Εικόνα 3-1 Δείγμα αποτελεσμάτων EXCEL.....	36
Εικόνα 4-1 Οπτικοποίηση των attributes.....	55
Εικόνα 4-2 Τελική Μορφή Αρχείου Εισόδου CSV.....	56
Εικόνα 4-3 Βιολογικό πρότυπο και γενετικοί αλγόριθμοι.....	58
Εικόνα 4-4 Τελική εκτέλεση προγράμματος.....	68

ΠΙΝΑΚΑΣ ΓΡΑΦΗΜΑΤΩΝ

Γράφημα 3-1 Φύλο Συμμετεχόντων	38
Γράφημα 3-2 Ηλικία Συμμετεχόντων.....	38
Γράφημα 3-3 Κατοικία Συμμετεχόντων	39
Γράφημα 3-4 Επίπεδο εκπαίδευσης Συμμετεχόντων	39
Γράφημα 3-5 Εργασία Συμμετεχόντων	40
Γράφημα 3-6 Οικονομική κατάσταση Συμμετεχόντων.....	40
Γράφημα 3-7 Μορφωτικό επίπεδο γονέων Συμμετεχόντων	41
Γράφημα 3-8 Είδος εκπαίδευσης	42
Γράφημα 3-9 Συμμετοχή στις Πανελλαδικές Εξετάσεις.....	42
Γράφημα 3-10 Ημερήσιες Ώρες Μελέτης την χρονία των Πανελλαδικών	43
Γράφημα 3-11 Δημόσια εκπαίδευση ή Φροντιστήριο/Ιδιαίτερα μαθήματα	43
Γράφημα 3-12 Σύγκριση άγχους κατά τη διάρκεια της χρονιάς και τις ημέρες των πανελλαδικών.....	44
Γράφημα 3-13 Μόρια Συμμετεχόντων στις πανελλαδικές εξετάσεις	45
Γράφημα 3-14 Επιτυχία του Συμμετέχοντα σε 1η Σχολή Προτίμησης ή όχι.....	45
Γράφημα 3-15 Κατεύθυνση Υποψηφίων	46
Γράφημα 3-16 Μέσος όρος βαθμολογίας Θετικών μαθημάτων στις 3 τάξεις του Λυκείου .	47
Γράφημα 3-17 Συγκριτικά Διαγράμματα προφορικών και γραπτών Λυκείου.....	47
Γράφημα 3-18 Σύγκριση γραπτών πανελλαδικών και ενδοσχολικών Γ' Λυκείου	48
Γράφημα 3-19 Ημερήσιες ώρες μελέτης την χρονιά των Πανελλαδικών για τα Μαθηματικά	48
Γράφημα 3-20 Συγκριτικό γράφημα εξοικείωσης ατόμων με τα μαθηματικά γενικότερα και με τα μαθηματικά της Γ' Λυκείου	49

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	5
ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ.....	7
ΠΙΝΑΚΑΣ ΓΡΑΦΗΜΑΤΩΝ	8
1 ΕΙΣΑΓΩΓΗ.....	11
1.1 Στόχοι Εργασίας.....	11
1.2 Θεματική Παρουσίαση Κεφαλαίων	11
2 ΕΞΟΥΞΗ ΔΕΔΟΜΕΝΩΝ	13
2.1 Εισαγωγή στην εξόρυξη δεδομένων.....	13
2.2 Η Εξόρυξη ως στάδιο ανεύρεσης γνώσης.....	15
2.3 Εξόρυξη Δεδομένων στην εκπαίδευση	17
2.4 Μηχανική Μάθηση	19
2.4.1 Μάθηση με επίβλεψη	21
2.4.2 Μη Επιβλεπόμενη Μάθηση	21
2.4.3 Ενισχυτική Μάθηση	22
2.5 Βασικά στάδια της Εξόρυξης Δεδομένων	23
2.6 Τεχνικές Εξόρυξης Δεδομένων	24
2.6.1 Κατηγοριοποίηση.....	25
2.6.2 Παλινδρόμηση	26
2.6.3 Ομαδοποίηση	26
2.6.4 Σύγκριση Κατηγοριοποίηση-Παλινδρόμηση-Ομαδοποίηση.....	28
2.7 Τεχνητά Νευρωνικά Δίκτυα	29
2.7.1 Δομή και Λειτουργία των Τεχνητών νευρωνικών δικτύων	29
3 ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΕΡΕΥΝΑΣ.....	33
3.1 Εισαγωγή.....	33
3.2 Ορισμός Δείγματος- Στόχοι της έρευνας	33
3.3 Σχεδιασμός και υλοποίηση έρευνας ερωτηματολογίου	33
3.4 Αξιοπιστία και εγκυρότητα ερωτηματολογίου	34
3.5 Ερωτηματολόγιο Έρευνας.....	34
3.6 Μέθοδος επεξεργασίας και ανάλυσης δεδομένων	36
3.7 Παρουσίαση αποτελεσμάτων της έρευνας	37
3.7.1 Δημογραφικά στοιχεία συμμετεχόντων	38
3.7.2 Ερωτήσεις Σχετικά με τις Πανελλαδικές	42

3.7.3	Ερωτήσεις σχετικές με τα Μαθηματικά.....	46
3.7.4	Γενικές Ερωτήσεις.....	49
4	ΜΟΝΤΕΛΟ.....	53
4.1	Το σύνολο δεδομένων.....	53
4.2	Εφαρμογή Νευρωνικών Δικτύων και ανάγκη Γενετικών Αλγορίθμων.....	56
4.2.1	Γενετικοί Αλγόριθμοι.....	57
4.3	Αλγόριθμος.....	59
4.4	Βασικά σημεία Αλγορίθμου.....	62
4.4.1	Η συνάρτηση του νευρωνικού δικτύου.....	62
4.4.2	Διάβασμα αρχείου και καταχώρηση σε λίστες.....	63
4.4.3	Δημιουργία αρχικού συνόλου χρωμοσωμάτων.....	63
4.4.4	Γενετικός Αλγόριθμος.....	66
4.5	Εκτέλεση του προγράμματος.....	68
5	ΕΠΙΛΟΓΟΣ.....	69
5.1	Συμπεράσματα.....	69
5.2	Μελλοντικές Επεκτάσεις και Βελτιώσεις.....	69
6	Βιβλιογραφία.....	70
	ΠΑΡΑΡΤΗΜΑ.....	73

1 ΕΙΣΑΓΩΓΗ

1.1 Στόχοι Εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η έκδοση πληροφοριών και αποτελεσμάτων με χρήση μεθόδων εξόρυξης δεδομένων για τη δημιουργία μοντέλου πρόβλεψης επιτυχίας ή μη στο μάθημα των μαθηματικών στις Εθνικές εξετάσεις. Ο γενικός στόχος της διαδικασίας εξόρυξης δεδομένων είναι η εξαγωγή πληροφοριών από ένα σύνολο δεδομένων και ταυτόχρονα η μετατροπή σε μια δομή για περαιτέρω χρήση.

Κύριο ζητούμενο είναι η ανάπτυξη μοντέλου πρόβλεψης επιτυχίας η μη ενός μαθητή στις πανελλαδικές εξετάσεις σε σχέση με τις επιδόσεις του στο μάθημα των μαθηματικών, το άγχος και την μόνιμη κατοικία. Για το λόγο αυτό δημιουργήθηκε ένα ερωτηματολόγιο και πραγματοποιήθηκε έρευνα ώστε να συλλεχθούν και να αξιολογηθούν δεδομένα, που θα χρησιμοποιηθούν ως το σύνολο δεδομένων πάνω στο οποίο θα εφαρμοστεί το μοντέλο μας.

Για τη μελέτη της αποδοτικότητας του μοντέλου θα πρέπει να γίνει ο κατάλληλος σχεδιασμός για την προσέγγιση της διαδικασίας ελέγχου ώστε να βρεθεί ο πιο κατάλληλος αλγόριθμος από τους διαθέσιμες μεθόδους εξόρυξης δεδομένων που θα ανταποκρίνεται με όσο το δυνατόν μεγαλύτερη ακρίβεια στα αποτελέσματα.

1.2 Θεματική Παρουσίαση Κεφαλαίων

Σε αυτή την ενότητα παρουσιάζεται μία σύντομη περιγραφή της διάρθρωσης της διπλωματικής εργασίας η οποία αποτελείται από τα παρακάτω κεφάλαια:

Στο κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο της εξόρυξης δεδομένων και γνώσης. Αναλύονται και ορίζονται οι όροι Εξόρυξη Δεδομένων - Data Mining (DM) και Ανακάλυψη Γνώσης από Βάσεις Δεδομένων, οι οποίοι έχουν αντικατασταθεί από έναν σχετικά καινούργιο όρο την Επιστήμη των Δεδομένων - Data Science. Μελετάμε πως γίνεται η εξόρυξη γνώσης από βάσεις δεδομένων και ποια είναι η βασική μεθοδολογία εφαρμογής αυτών των τεχνικών. Αναλύουμε τον επιστημονικό κλάδο της εξόρυξης δεδομένων από εκπαιδευτικά δεδομένα-Educational Data Mining (EDM) και κάνουμε μία γρήγορη ανάλυση στη Μηχανική Μάθηση και στις επιμέρους κατηγορίες που χωρίζεται. Στο τέλος του δεύτερου κεφαλαίου αναφέρουμε τους βασικούς αλγόριθμους της εξόρυξης γνώσης και αναλύουμε την μέθοδο της κατηγοριοποίησης-Classification που θα χρησιμοποιήσουμε στην εργασία μας. Οι τεχνικές της κατηγοριοποίησης είναι περισσότερο διαδεδομένες από τις τεχνικές μηχανικής μάθησης για χρήση στην εκπαίδευση. Έπειτα γίνεται μια εισαγωγή στις βασικές έννοιες των Τεχνητών Νευρωνικών Δίκτυων που χρησιμοποιήθηκαν για τη δημιουργία του μοντέλου, καθώς είναι μια διαδεδομένη μεθοδολογία που χρησιμοποιείται για κατηγοριοποίηση και για τη δημιουργία προγνωστικών μοντέλων.

Στο κεφάλαιο 3 περιγράφουμε την έρευνα που πραγματοποιήσαμε στα πλαίσια της εν λόγω διπλωματικής εργασίας και αναλύουμε τα αποτελέσματα που προέκυψαν. Από τα

δεδομένα που επιλέχθηκαν έπειτα από την προεπεξεργασία τους, πήραμε τα τελικά δεδομένα που χρησιμοποιήθηκαν ως το τελικό σύνολο δεδομένων του μοντέλου μας.

Στη συνέχεια στο κεφάλαιο 4 γίνεται μια αναφορά στο γενετικό αλγόριθμο και παράλληλα εξηγούμε τους λόγους που μας οδήγησαν στο συνδυασμό των γενετικών αλγορίθμων και των νευρωνικών δικτύων, τονίζοντας τη σημασία του. Έπειτα αναλύουμε τον αλγόριθμο που κατασκευάστηκε για τη δημιουργία του μοντέλου μας.

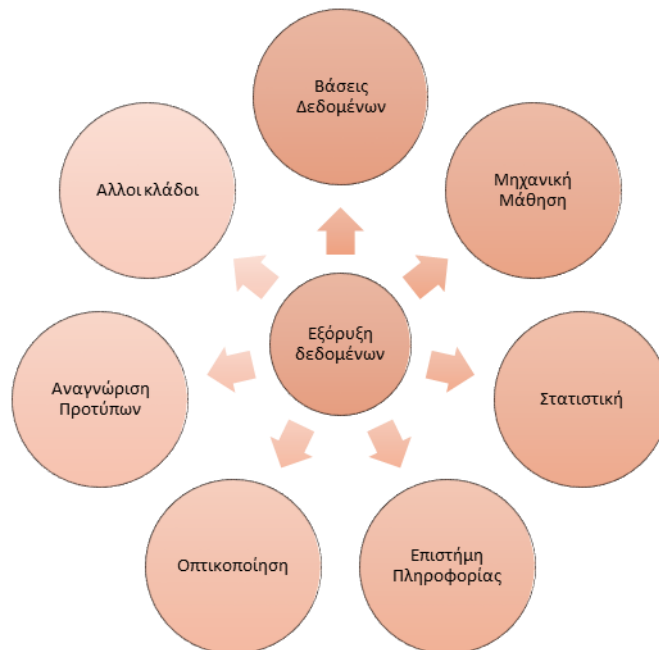
Στο Κεφάλαιο 5 παρουσιάζονται τα γενικά συμπεράσματα της διπλωματικής εργασίας και προτείνονται συγκεκριμένα πεδία για μελλοντικές επεκτάσεις. Τέλος παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για τη συγγραφή της παρούσας διπλωματικής εργασίας και τα παραρτήματα.

2 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

2.1 Εισαγωγή στην εξόρυξη δεδομένων

Στην σύγχρονη εποχή οι περισσότεροι οργανισμοί αλλά και εταιρείες διατηρούν μεγάλες βάσεις δεδομένων στις οποίες καταγράφουν διάφορες πληροφορίες, πολλές φορές εξαιρετικής σημασίας, όπως για παράδειγμα στοιχεία των εργαζομένων τους, κινήσεις πελατών κ.α.. Το ζητούμενο ήταν να βρεθεί ένας τρόπος ώστε τα δεδομένα να χρησιμοποιηθούν με τέτοιο τρόπο ώστε να είναι χρήσιμα για τους κατόχους τους. Τη λύση την έδωσε η εξόρυξη δεδομένων (αγγλ. Data Mining), η οποία αφορά όλες τις τεχνικές και διαδικασίες, εύρεσης νέας και πιθανόν χρήσιμης γνώσης από δεδομένα, που στην πράξη είναι μεγάλες Βάσεις Δεδομένων.

Η εξόρυξη γνώσης από δεδομένα ορίζεται ως η διαδικασία εξαγωγής σημαντικών, μη προφανών και πιθανώς χρήσιμων πληροφοριών ή προτύπων από βάσεις δεδομένων. Η διαδικασία της Εξόρυξης δεδομένων η οποία αποτελεί και τη βασική δραστηριότητα στην Εξόρυξη Γνώσης (Knowledge Discovery) στις βάσεις δεδομένων, ασχολείται με την εύρεση προτύπων στα δεδομένα. Επιπλέον η εξόρυξη δεδομένων επιτυγχάνεται με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων. Οι μέθοδοι της προέρχονται από διάφορους κλάδους συμπεριλαμβάνοντας την επιστήμη των υπολογιστών, τα μαθηματικά, την αποθήκευση δεδομένων και την κατανομημένη επεξεργασία.



Εικόνα 2-1 Η εξόρυξη γνώσης ως συμβολή πολλών αρχών

Υπάρχουν πολλοί ορισμοί για την εξόρυξη δεδομένων. Η Εξόρυξη Δεδομένων συχνά ορίζεται ως "η ανακάλυψη γνώσης που είναι κρυμμένες σε μία Βάση Δεδομένων". Ένας άλλος ορισμός που μπορούμε να αποδώσουμε για την Εξόρυξη Γνώσης είναι "η εξερευνητική ανάλυση δεδομένων που οδηγεί σε ανακάλυψη πληροφορίας από συμπερασματική μάθηση". Στη συνέχεια παρουσιάζονται δύο πιο «αυστηροί» ορισμοί της εξόρυξης δεδομένων.

Ορισμός 1 : «Εξόρυξη γνώσης είναι η ανάλυση, συνήθως τεράστιων, παρατηρούμενων συνόλων δεδομένων, έτσι ώστε να βρεθούν μη παρατηρηθείσες σχέσεις και να συνοψιστούν τα δεδομένα με καινοφανείς τρόπους οι οποίοι να είναι κατανοητοί και χρήσιμοι στον κάτοχο των δεδομένων.» (David Hand, Heikki Mannila, Padhraic Smyth, 2001)

Ορισμός 2: «Εξόρυξη γνώσης είναι η διαδικασία εξαγωγής υπονοούμενης γνώσης, άγνωστης αλλά ενδεχομένως χρήσιμης γνώσης, υπό την μορφή συσχετίσεων, προτύπων και τάσεων, μέσω διαδικασίας εξέτασης, ανάλυσης και επεξεργασίας βάσεων δεδομένων, συνδυάζοντας και χρησιμοποιώντας τεχνικές από την μηχανική μάθηση, την αναγνώριση προτύπων, την στατιστική, τις βάσεις δεδομένων και την οπτικοποίηση». (Larose, 2006)

Είναι μια νέα τεχνολογία που βοηθάει τις επιχειρήσεις να επικεντρωθούν στη σημαντική πληροφορία που βρίσκεται μέσα στις αποθήκες δεδομένων τους (data warehouses). Οι τεχνικές της, είναι σε θέση να ψάξουν γρήγορα και λεπτομερειακά βάσεις δεδομένων για την αναζήτηση κρυμμένων προτύπων (patterns), ανακαλύπτοντας πληροφορίες που οι ειδικοί μπορεί να χάσουν ή να παραβλέψουν. Ο πραγματικός στόχος της εξόρυξης δεδομένων είναι η αυτόματη ή ημιαυτόματη ανάλυση μεγάλων ποσοτήτων δεδομένων για την εξαγωγή κάποιου ενδιαφέροντος προτύπου που ήταν άγνωστο μέχρι εκείνη τη στιγμή, έτσι ώστε τα πρότυπα που θα προκύψουν να έχουν κατανοητή δομή ώστε να βοηθήσουν τον άνθρωπο να βγάλει σωστά συμπεράσματα και να πάρει στο μέλλον κατάλληλες αποφάσεις.

Η εξόρυξη δεδομένων αποτελείται από διάφορα βήματα για την ανακάλυψη κανόνων, σχέσεων και προτύπων στα δεδομένα που επεξεργάζεται. Το πρώτο βήμα είναι η περιγραφή των δεδομένων, η συλλογή περιληπτικών πληροφοριών και στατιστικών στοιχείων, η δημιουργία γραφημάτων και η εύρεση πιθανών σχέσεων ανάμεσα στις μεταβλητές. Στη συνέχεια κάποιος μπορεί να δημιουργήσει ένα μοντέλο πρόβλεψης από τις σχέσεις που προέβλεψε και να το ελέγξει για το κατά πόσο ισχύει σε θεωρητικό επίπεδο. Το τελευταίο βήμα είναι να αναλυθεί εμπειρικά πλέον το μοντέλο, ώστε να βρεθεί κατά πόσο ισχύει στον πραγματικό κόσμο και με βάση αυτό να γίνει προγραμματισμός για το μέλλον.

Με λίγα λόγια οι εργασίες που πραγματοποιούνται κατά την Εξόρυξη δεδομένων χωρίζονται σε εργασίες για περιγραφή και πρόβλεψη. Η πρόβλεψη προϋποθέτει τη χρησιμοποίηση διαφόρων γνωστών μεταβλητών για την εκτίμηση μελλοντικών άγνωστων τιμών και η περιγραφή αφορά τη δημιουργία κατανοητών για τον άνθρωπο μοντέλων που θα περιγράψουν τα δεδομένα .

2.2 Η Εξόρυξη ως στάδιο ανεύρεσης γνώσης

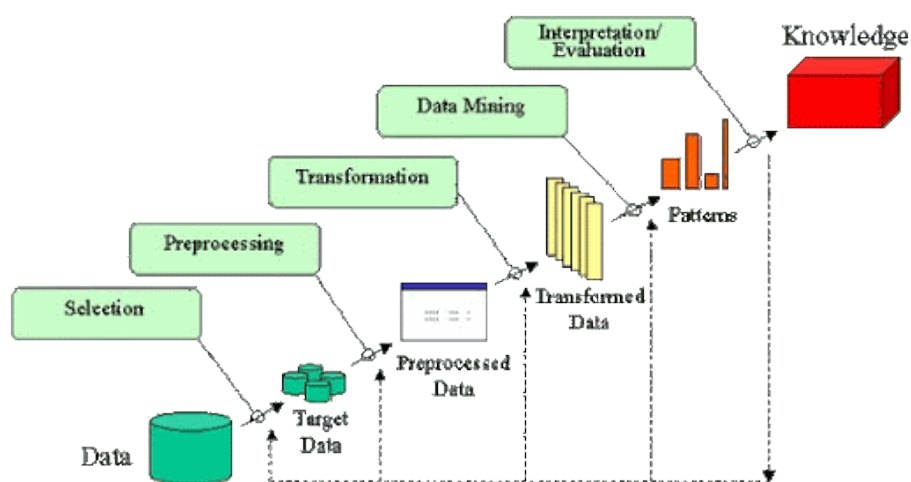
Στη διεθνή βιβλιογραφία υπάρχει μια γενικότερη σύγχυση ανάμεσα στους όρους «Εξόρυξη Γνώσης» (Data mining) και «Ανεύρεση γνώσης στις βάσεις δεδομένων» (Knowledge discovery in databases, KDD). Σε πολλές περιπτώσεις αξίζει να σημειωθεί ότι οι δύο αυτοί όροι ταυτίζονται, ενώ στην πραγματικότητα η εξόρυξη δεδομένων αποτελεί ένα μόνο επιμέρους βήμα (ή πιο απλά τμήμα) της ανεύρεσης γνώσης, συγκροτώντας το πυρήνα αυτής.

Η ανεύρεση γνώσης είναι μια επαναληπτική διαδικασία που αποτελείται από μια σειρά βημάτων, τα οποία οδηγούν από τη συλλογή των δεδομένων στην ανακάλυψη και εξαγωγή χρήσιμης πληροφορίας από αυτά. Συνεπώς οι όροι ανακάλυψη γνώσης σε βάσεις δεδομένων και εξόρυξη γνώσης από δεδομένα συχνά χρησιμοποιούνται εναλλακτικά για την ίδια έννοια. Τα τελευταία χρόνια ο όρος KDD έχει χρησιμοποιηθεί για να εκφράσει μια διαδικασία που αποτελείται από πολλά βήματα, ένα από τα οποία είναι η εξόρυξη γνώσης από δεδομένα.

Ορισμός 3: «Η Ανακάλυψη Γνώσης από Βάσεις Δεδομένων ορίζεται ως μία ντετερμινιστική διαδικασία αναγνώρισης καινοτόμων, πρωτότυπων, έγκυρων και ενδεχομένως χρήσιμων και κατανοητών προτύπων από τα δεδομένα» (Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, 1996)

Για να κατανοηθεί πλήρως η έννοια της εξόρυξης δεδομένων, παρατίθενται το παρακάτω σχήμα που περιγράφει την διαδικασία ανεύρεσης γνώσης μέσα από τα βήματα:

- data cleaning – Καθαρισμός Δεδομένων
- data integration – Ενσωμάτωση Δεδομένων
- data selection – Επιλογή Δεδομένων
- data transformation – Τροποποίηση Δεδομένων
- data mining – Εξόρυξη Δεδομένων
- pattern evaluation / interpretation – Αξιολόγηση προτύπων
- knowledge representation – Αναπαράσταση γνώσης



Εικόνα 2-2 Η διαδικασία της Ανεύρεσης Γνώσης στις Βάσεις Δεδομένων (KDD) (Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, 1996)

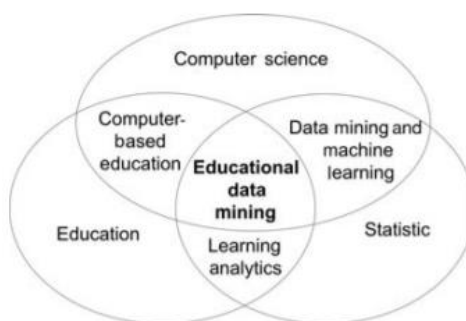
Πιο αναλυτικά στον καθαρισμό δεδομένων γίνεται η εξάλειψη του θορύβου και των ασυνεπών δεδομένων, ενώ στην ενσωμάτωση των δεδομένων πολλαπλές πηγές δεδομένων μπορούν να συνδυαστούν. Στην συνέχεια γίνεται η επιλογή των δεδομένων κατά την οποία τα δεδομένα που σχετίζονται με την εργασία ανάλυσης ανακτώνται από τη βάση δεδομένων ώστε να μετασχηματιστούν. Έπειτα στο μετασχηματισμό δεδομένων τα δεδομένα μετατρέπονται σε έντυπα κατάλληλα για εξόρυξη εκτελώντας διαδικασίες περίληψης ώστε να γίνει η εξόρυξη δεδομένων, η οποία είναι η βασική διαδικασία όπου εφαρμόζονται οι μέθοδοι η οποία έχει ως σκοπό να εξαχθούν πρότυπα δεδομένων. Τα πρώτα 4 βήματα που προαναφέρθηκαν είναι μορφές προ-επεξεργασίας δεδομένων, διαδικασία απαραίτητη για την εξόρυξη. Για να προσδιοριστούν τα πρότυπα που μας ενδιαφέρουν γίνεται η αξιολόγηση μοτίβων ώστε τα πρότυπα να αντιπροσωπεύουν τη γνώση με βάση ορισμένα μέτρα ενδιαφέροντος. Τελικό στάδιο της διαδικασίας είναι η παρουσίαση της γνώσης όπου οι τεχνικές απεικόνισης και εκπροσώπησης της γνώσης χρησιμοποιούνται για να παρουσιάσουν την εξόρυξη γνώσης στον χρήστη.

Πολλές φορές κάποια από τα παραπάνω βήματα μπορούν να συνδυαστούν μεταξύ τους για το καλύτερο δυνατό αποτέλεσμα ή και να χρειαστεί κάποια βήματα να επαναληφθούν. Συμπεραίνουμε λοιπόν ότι η εξόρυξη δεδομένων είναι μια διαδραστική και επαναληπτική διαδικασία με πολλαπλά βήματα και αποφάσεις που αποτελεί την διαδικασία κλειδί για την ανεύρεση γνώσης.

2.3 Εξόρυξη Δεδομένων στην Εκπαίδευση

Τα τελευταία χρόνια, ερευνητές από διάφορες επιστημονικές περιοχές ξεκίνησαν να εξετάζουν και να ερευνούν το πώς η Εξόρυξη Γνώσης από συλλογές δεδομένων μπορεί να επηρεάσει τη βελτίωση της Εκπαίδευσης μέσω ερευνών σ' αυτό το πεδίο. Η εξόρυξη γνώσης για εκπαιδευτικούς σκοπούς (ή Εκπαιδευτική Εξόρυξη Δεδομένων, ή εξόρυξη εκπαιδευτικών δεδομένων)- Educational Data Mining (EDM) είναι ένας αναδυόμενος διεπιστημονικός τομέας έρευνας, στον οποίο έχουν αναπτυχθεί μέθοδοι και τεχνικές για την διερεύνηση δεδομένων που προέρχονται από διάφορα εκπαιδευτικά συστήματα πληροφοριών. Συνεπώς, η EDM ασχολείται με την ανάπτυξη μεθόδων για την εξερεύνηση των μοναδικών τύπων δεδομένων που προέρχονται από εκπαιδευτικά περιβάλλοντα (Cristobal Romero, Sebastian Ventura, 2013).

Οι τρεις κυριότερες επιστημονικές περιοχές που σχετίζονται με την εξόρυξη εκπαιδευτικών δεδομένων¹ είναι η Επιστήμη των Υπολογιστών, η Εκπαίδευση και η Στατιστική. Όπως παρατηρούμε και στο επόμενο σχήμα η τομή των περιοχών αυτών σχηματίζει άλλες υποπεριοχές της EDM όπως την περιοχή της Μαθησιακής Ανάλυσης (Learning Analytics), των Υπολογιστικών Περιβαλλόντων, της Εξόρυξης Δεδομένων και της Μηχανικής Μάθησης (Cristobal Romero, Sebastian Ventura, 2013).



Εικόνα 2-3 Κύριοι τομείς που σχετίζονται με την εξόρυξη εκπαιδευτικών δεδομένων (Cristobal Romero, Sebastian Ventura, 2013)

Η αποτύπωση ενός ορισμού για την Εκπαιδευτική Εξόρυξη Δεδομένων αποδείχτηκε μία δύσκολη διαδικασία. Στη συνέχεια παρουσιάζονται δύο ορισμοί για την EDM όπου έχουν ως κοινή βάση την έμφαση που δίνεται στην αναδυόμενη γνώση από εκπαιδευτικά δεδομένα με σκοπό τη βελτίωση των εκπαιδευτικών συστημάτων.

Ορισμός 4: «Η εξόρυξη γνώσης από εκπαιδευτικά δεδομένα- EDM ορίζεται ως το αναδυόμενο επιστημονικό πεδίο που πραγματεύεται την ανάπτυξη μεθόδων που εξερευνά τη μοναδικότητα των τύπων δεδομένων που προέρχονται από εκπαιδευτικές δομές και κάνουν χρήση των μεθόδων αυτών με σκοπό την καλύτερη κατανόηση των μαθητών και των συνθηκών κάτω από τις οποίες επιτυγχάνεται η γνώση» (Kenneth Koedinger, Ryan Baker, Kyle Cunningham, Alida Skogsholm, Brett Leber, John Stamper, 2010)

¹ Με τον όρο «εκπαιδευτικά δεδομένα» εννοούμε τα δεδομένα που προέρχονται από βάσεις δεδομένων που διατηρούν εκπαιδευτικά ιδρύματα με τη χρήση κάποιων συστημάτων καταγραφής αναφορών και πληροφοριών που αφορούν τους εκάστοτε εκπαιδευομένους.

Ορισμός 5: «Η εξόρυξη γνώσης από εκπαιδευτικά δεδομένα είναι μια επιστήμη γνώσης και παράλληλα μία πλούσια περιοχή εφαρμογής για εξόρυξη γνώσης, δεδομένου της αυξανόμενης διαθεσιμότητας εκπαιδευτικών δεδομένων. Θεωρείται ότι ενισχύει την καθοδηγούμενη από τα δεδομένα λήψη αποφάσεων με σκοπό τη βελτίωση της τρέχουσας εκπαιδευτικής πρακτικής και του εκπαιδευτικού υλικού» (Toon Calders, Mykola Pechenizkiy, 2012)

Με πιο απλά λόγια η εξόρυξη γνώσης για εκπαιδευτικούς σκοπούς είναι η διαδικασία με την οποία μετατρέπουμε τα δεδομένα τα οποία έχουμε συλλέξει σε χρήσιμες πληροφορίες τις οποίες μπορούμε να χρησιμοποιήσουμε για να πάρουμε καλύτερες αποφάσεις αλλά και για να απαντήσουμε σε διάφορα ερευνητικά ερωτήματα. Η χρήση τεχνικών μηχανικής μάθησης για εκπαιδευτικούς σκοπούς είναι ένας πολλά υποσχόμενος τομέας ανάπτυξη μεθόδων εξερεύνησης δεδομένων από υπολογιστικούς εκπαιδευτικούς χώρους και ανακάλυψη σημαντικών προτύπων. Τα αποθηκευμένα δεδομένα όπως τα δημογραφικά και ακαδημαϊκά δεδομένα μαθητών, πληροφορίες σχετικά με τις βαθμολογίες των μαθητών είναι χρήσιμα για αλγόριθμους μηχανικής μάθησης (Kotsiantis, 2011) .

Η EDM μετατρέπει ανεπεξέργαστα δεδομένα που προέρχονται από εκπαιδευτικά συστήματα σε χρήσιμες πληροφορίες που θα μπορούσαν ενδεχομένως να έχουν μεγάλο αντίκτυπο στην εκπαιδευτική έρευνα και πρακτική. Αρκετοί ερευνητές εφάρμοσαν μεθόδους εξόρυξης δεδομένων όπως ομαδοποίηση, ταξινόμηση, κανόνες συσχέτισης και εξόρυξη κειμένων στο εκπαιδευτικό πλαίσιο (Ashish Dutt, Maizatul Akmar Ismail & Tutut Herawan, 2016) . Μια έρευνα που πραγματοποιήθηκε το 2007, παρείχε μια περιεκτική πηγή δημοσιεύσεων μεταξύ 1995 και 2005 σχετικά με το EDM των Romero & Ventura (Cristobal Romero, Sebastian Ventura, 2013). Αυτή η έρευνα παρουσιάζει και αναθεωρεί βασικά ορόσημα της τότε τρέχουσας κατάστασης στον τομέα της EDM, μαζί με συγκεκριμένες εφαρμογές και εργαλεία. Συνεπώς, καλύπτει την εφαρμογή της εξόρυξης δεδομένων από τα παραδοσιακά εκπαιδευτικά ιδρύματα στη διαδικτυακή διαχείριση της μάθησης και σε έξυπνα προσαρμοσμένα εκπαιδευτικά συστήματα.

Σε μια άλλη έρευνα από την Peña-Ayala , αναλύθηκαν περίπου 240 δείγματα έργων EDM που δημοσιεύθηκαν μεταξύ του 2010 και το 2013 (Pena-Ayala, 2013) . Στόχοι της συγκεκριμένης έρευνας ήταν να διατηρήσει και να ενισχύσει τα χρονικά των πρόσφατων εξελίξεων της EDM, καθώς επίσης να οργανώσει και να αναλύσει τα δεδομένα που προκύπτουν από μια προσέγγιση της κλασικής εξόρυξης δεδομένων. Ένα από τα βασικά ευρήματα αυτή της έρευνας ήταν ότι το μεγαλύτερο μέρος των ερευνητικών έργων EDM επικεντρώθηκε σε τρία είδη εκπαιδευτικών συστημάτων, στα εκπαιδευτικά καθήκοντα, σε μεθόδους και αλγόριθμους. Η εφαρμογή τεχνικών της εξόρυξης δεδομένων για μελέτη on-line μαθημάτων προτάθηκε από τους Zaiane & Luo, οι οποίοι πρότειναν μια μη παραμετρική τεχνική ομαδοποίησης δεδομένων για δραστηριότητα ιστού εκτός σύνδεσης των μαθητών (Osmar Zaiane, Jun Luo, 2001) .

Οι μέθοδοι στη χρήση τεχνικών μηχανικής μάθησης για εκπαιδευτικές προτάσεις περιλαμβάνουν αλγόριθμους ταξινόμησης και παλινδρόμησης, κανόνες συσχέτισης, ανάλυση διαδοχικών προτύπων, καθώς και ομαδοποίηση και εξόρυξη ιστού. Το πεδίο της EDM έχει επεκταθεί σημαντικά τα τελευταία 15 χρόνια, καθώς από το 2005 και μετά έχουν ακολουθήσει διεθνή συνέδρια, βιβλιογραφία, επιστημονικά περιοδικά, με τη δημιουργία επιστημονικών κοινοτήτων να έχει καθιερώσει την επιστημονική υπόσταση της εν λόγω περιοχής , εκ των οποίων τα σημαντικότερα αναλύθηκαν παραπάνω.

2.4 Μηχανική Μάθηση

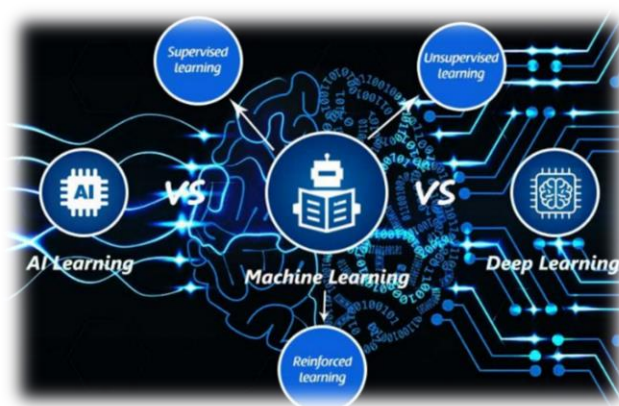
Η Μηχανική Μάθηση - Machine Learning είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη και αποτελεί έναν από τους σημαντικότερους τομείς της. Είναι δηλαδή η δημιουργία ή βελτίωση ενός γνωστικού μοντέλου από ένα εκπαιδευτικό σύνολο δεδομένων. Στην συνέχεια παρουσιάζονται δύο ορισμοί αρκετά σημαντική για την κατανόηση της Μηχανικής Μάθησης.

Ορισμός 6: Η Μηχανική Μάθηση είναι το Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί. (Simon, 2013)

Ορισμός 7: Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E . (Mitchell, 1997)

Η Μηχανική μάθηση αποτελεί έναν από τους σημαντικότερους τομείς έρευνας της τεχνητής νοημοσύνης² και είναι το φαινόμενο κατά το οποίο ένα σύστημα βελτιώνει την απόδοσή του κατά την εκτέλεση μιας συγκεκριμένης εργασίας, χωρίς να υπάρχει ανάγκη να προγραμματιστεί εκ νέου και έχει ως σκοπό τη δημιουργία συστημάτων-μηχανών ικανών να μαθαίνουν και να βελτιώνουν την απόδοσή τους σε κάποιους τομείς μέσω της αξιοποίησης προηγούμενης γνώσης και εμπειρίας (Δημητριάδης, 2015) .

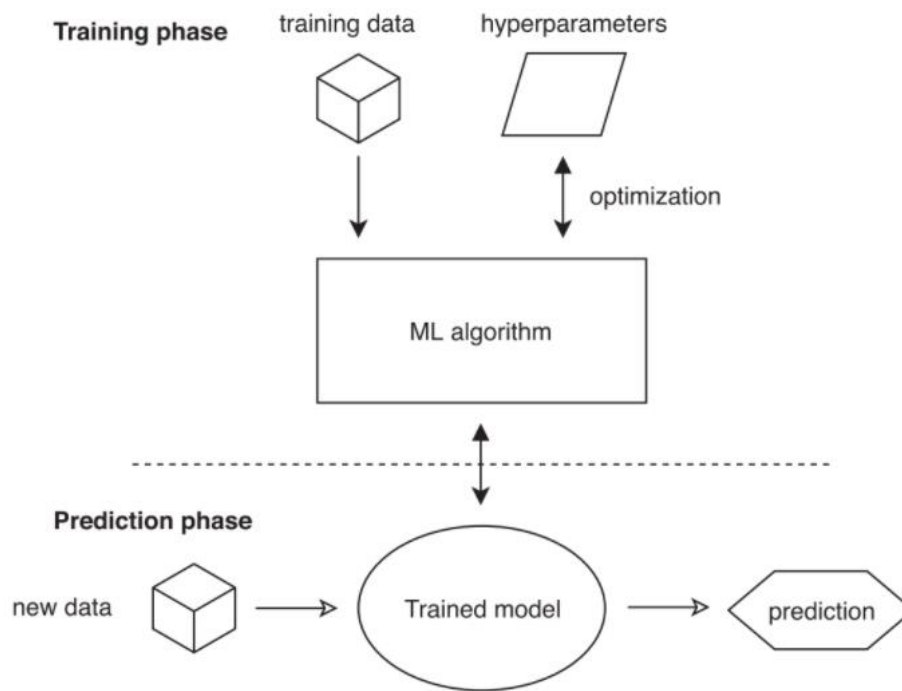
Με πιο απλά λόγια, Μηχανική Μάθηση είναι η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων και οι τρόποι μάθησης στον τομέα της είναι ανάλογοι με τους τρόπους που μαθαίνει ο άνθρωπος. Έτσι, διακρίνουμε τα είδη Μηχανικής Μάθησης σε μάθηση με επίβλεψη, μη επιβλεπόμενη μάθηση και ενισχυτική μάθηση τα οποία αναλύονται στη συνέχεια.



Εικόνα 2-4 Μηχανική Μάθηση

² Ο όρος τεχνητή νοημοσύνη αναφέρεται στον κλάδο της πληροφορικής ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα. (Wikipedia.org)

Στην επόμενη εικόνα παρουσιάζεται η γενική επισκόπηση της μηχανικής μάθησης. Κατά τη διάρκεια της εκπαίδευσης ένα μοντέλο Μηχανικής Μάθησης –ML βελτιστοποιείται χρησιμοποιώντας εκπαιδευτικά δεδομένα και συντονίζοντας υπερ-παραμέτρους. Στη συνέχεια, το εκπαιδευμένο μοντέλο αναπτύσσεται για να παρέχει προβλέψεις για νέα δεδομένα που τροφοδοτούνται στο σύστημα (Joost Verbraeken, Matthijs Wolting, Jonathan Katzy , Jeroen Kloppenborg, Tim Verbelen, Jan Rellermeyer, 2020) .



Εικόνα 2-5 Γενική επισκόπηση Μηχανικής Μάθησης

2.4.1 Μάθηση με επίβλεψη

Η μάθηση με επίβλεψη ή Επιβλεπόμενη Μάθηση (Supervised Learning) όπως λέγεται είναι μια κατηγορία Μηχανικής Μάθησης όπου το σύστημα πρέπει να «μάθει» με επαγωγική μέθοδο μια συνάρτηση που ονομάζεται "συνάρτηση στόχος" (target function) και που αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής, που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου, βάσει των τιμών ενός συνόλου μεταβλητών, που ονομάζονται ανεξάρτητες μεταβλητές ή μεταβλητές εισόδου ή χαρακτηριστικά (Κοτζαπαναγιώτου, 2019) .

Στη Μάθηση με Επίβλεψη εντάσσονται δύο κατηγορίες προβλημάτων (Learning tasks) εξόρυξης γνώσης: Τα προβλήματα ταξινόμησης ή κατηγοριοποίησης (classification) και τα προβλήματα παρεμβολής ή παλινδρόμησης (regression). Η ταξινόμηση όπου θα ασχοληθούμε στην παρούσα εργασία αφορά στη δημιουργία μοντέλων πρόβλεψης διακριτών τάξεων (κλάσεων ή κατηγοριών), όπως για παράδειγμα τα στοιχεία αιτήσεων χορήγησης τραπεζικών δανείων ή η ομάδα αίματος ενώ η παλινδρόμηση αφορά στη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών, όπως για παράδειγμα η πρόβλεψη ισοτιμίας νομισμάτων ή τιμής μετοχής.

2.4.2 Μη Επιβλεπόμενη Μάθηση

Η Μη Επιβλεπόμενη Μάθηση (ή Μη Επιτηρούμενη Μάθηση)- Unsupervised Learning είναι μια κατηγορία Μηχανικής Μάθησης στην οποία ο αλγόριθμος, ο οποίος δεν του παρέχεται κάποια προηγούμενη γνώση, έχει μόνο τα δεδομένα εισόδου και προσπαθεί να ανακαλύψει πρότυπα σε αυτά βασιζόμενο στις ιδιότητές τους, να βρει δηλαδή την δομή των δεδομένων εισόδου. Κάθε ένα από τα πρότυπα που προκύπτουν περιγράφει ένα μέρος από τα δεδομένα. Συνεπώς σε αυτήν την κατηγορία Μηχανικής Μάθησης το Σύστημα έχει ως στόχο να ανακαλύψει "συσχετίσεις" και "ομάδες" από τη Βάση Γνώσης που του παρέχεται, βασιζόμενο αποκλειστικά και μόνο στις ιδιότητές τους. Σαν αποτέλεσμα προκύπτουν τα πρότυπα (περιγραφές), κάθε ένα από τα οποία περιγράφει ένα μέρος από τα δεδομένα.

Στη μη Επιβλεπόμενη Μάθηση εντάσσονται δύο κατηγορίες προβλημάτων εξόρυξης γνώσης: οι Κανόνες συσχέτισης (association rules) και η Συσταδιοποίηση (clustering). Οι κανόνες συσχέτισης σχετίζονται με την ανακάλυψη και διατύπωση σχέσεων που υπάρχουν στα δεδομένα, ενώ η συσταδιοποίηση οργάνωση μιας συλλογής από αντικείμενα ή στοιχεία (objects) σε συστάδες (clusters) με βάση κάποιο μέτρο ομοιότητας τα οποία αναλύονται στην συνέχεια.

2.4.3 Ενισχυτική Μάθηση

Η Ενισχυτική Μάθηση (Reinforcement Learning) είναι η τελευταία κατηγορία Μηχανικής Μάθησης στην οποία ο αλγόριθμος μαθαίνει αλληλοεπιδρώντας άμεσα με το περιβάλλον. Ένα πρόγραμμα αλληλοεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος χωρίς να υπάρχει κάποιος καθοδηγητής ο οποίος να αναφέρει ρητά αν έχει επιτευχθεί ο στόχος ή όχι .

Αυτό το είδος μάθησης είναι εμπνευσμένο από την επιβράβευση και την τιμωρία που χρησιμοποιούμε στη μάθηση των έμβιων όντων. Το σύστημα προσπαθεί να ανακαλύψει, χωρίς την καθοδήγηση εξωτερικού επιβλέποντα, τις ενέργειες εκείνες που θα του αποφέρουν το μεγαλύτερο κέρδος.

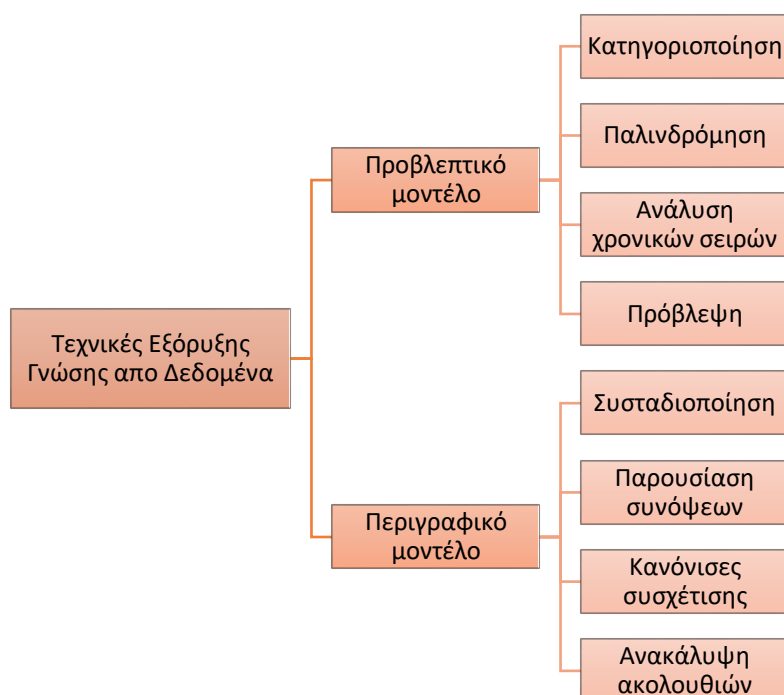
2.5 Βασικά στάδια της Εξόρυξης Δεδομένων

Υπάρχει ένα πλήθος από αλγόριθμους της διαδικασίας της εξόρυξης δεδομένων οι οποίοι έχουν ως στόχο την εξέλιξη του μεγέθους του συνόλου δεδομένων που εισάγονται σε αυτούς. Οι περισσότεροι αλγόριθμοι μπορούν να τοποθετηθούν στα όρια ενός απλού πλαισίου, το οποίο αποτελείται από τα βασικά στάδια της διαδικασίας της εξόρυξης δεδομένων δηλαδή την περιγραφή του μοντέλου, την αξιολόγηση και τον αλγόριθμο αναζήτησης.

- *Περιγραφή μοντέλου:* σε αυτό το στάδιο γίνεται προσπάθεια να γίνει κατανοητή και να δηλωθεί η λειτουργία του μοντέλου και να καθοριστούν ο στόχοι που θέτουμε όπως για παράδειγμα η ταξινόμηση, η συσταδιοποίηση και η παλινδρόμηση που αναλύονται στην συνέχεια. Επιπλέον αξίζει να σημειωθεί ότι μας απασχολεί και η παραστατική μορφή του μοντέλου, δηλαδή η απεικόνιση του ώστε να μπορούμε να το ερμηνεύσουμε σωστά. Τα πιο ευρέως γνωστά μοντέλα αποτελούν τα νευρωνικά δίκτυα, τα δένδρα και οι κανόνες απόφασης, τα δίκτυα Bayes, τα συγγενικά μοντέλα, κλπ.
- *Αξιολόγηση μοντέλου:* σε συνέχεια της δημιουργίας του μοντέλου εξετάζουμε αν το μοντέλο μας ταιριάζει με τις συνθήκες της ανεύρεσης γνώσης από βάσεις δεδομένων. Κρίνουμε αν είναι έγκυρα τα πρότυπα και πόση ακρίβεια και πόσο χρήσιμο είναι το μοντέλο.
- *Αλγόριθμος αναζήτησης:* σε αυτό το στάδιο συγκρίνουμε το μοντέλο και τις παραμέτρους έχοντας ως δεδομένα το σύνολο των δεδομένων, την οικογένεια μοντέλων και το κριτήριο αξιολόγησης. Οι πιο βασικοί τύποι αλγορίθμων αναζήτησης είναι αυτοί που αναζητούν παραμέτρους βελτιστοποίησης ενός κριτηρίου (αλγόριθμοι αναζήτησης παραμέτρων) και αυτοί που αναζητούν μοντέλα αντιπροσώπευσης των δεδομένων (αλγόριθμοι αναζήτησης μοντέλου) (Σταυλιώτης, 2008).

2.6 Τεχνικές Εξόρυξης Δεδομένων

Οι βασικοί στόχοι της εξόρυξης γνώσης είναι η εφαρμογή τεχνικών πρόβλεψης (prediction) και περιγραφής (description) σε μεγάλες βάσεις δεδομένων. Η πρόβλεψη αφορά νέα στιγμιότυπα του προβλήματος των οποίων προβλέπει την τιμή κάποιου συγκεκριμένου γνωρίσματός του, ενώ η περιγραφή στοχεύει στην εύκολη ανεύρεση ερμηνεύσιμων προτύπων από τα δεδομένα του προβλήματος. Έτσι μπορούμε να πούμε ότι υπάρχουν δύο είδη τεχνικών εξόρυξης δεδομένων, το προβλεπτικό μοντέλο (predictive model) και το περιγραφικό μοντέλο (descriptive model). Στην επόμενη εικόνα παρουσιάζονται οι βασικές εργασίες για το κάθε μοντέλο, προβλεπτικό και περιγραφικό.



Εικόνα 2-6 Βασικές τεχνικές Data Mining

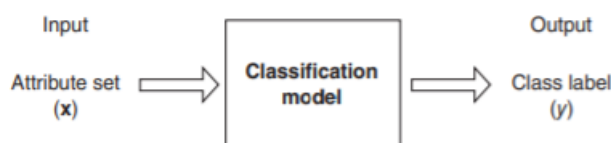
Οι κυριότερες μέθοδοι εξόρυξης δεδομένων είναι οι εξής: κατηγοριοποίηση, παλινδρόμηση και η ομαδοποίηση. Στην παρούσα εργασία θα ασχοληθούμε κυρίως με την κατηγοριοποίηση η οποία εκπαιδεύει μία συνάρτηση και κατηγοριοποιεί κάποια δεδομένα σε μια από διάφορες κλάσεις που δημιουργούνται .

Στις δύο επόμενες ενότητες κάνουμε μία σύντομη ανάλυση της κατηγοριοποίησης και της παλινδρόμησης όπου και στις δύο περιπτώσεις, η είσοδος είναι ένα σύνολο δεδομένων και παράγεται ένα μοντέλο το οποίο μπορεί στο μέλλον να χρησιμοποιηθεί για να προβλέψει τις τιμές της κατηγορίας για τα νέα δεδομένα και τέλος στην τελευταία ενότητα την τμηματοποίηση. Σε αυτό το σημείο πρέπει να σημειωθεί ότι ο όρος «μοντελοποίηση πρόβλεψης» (predictive modeling) αναφέρεται τόσο στην κατηγοριοποίηση όσο και στην παλινδρόμηση.

2.6.1 Κατηγοριοποίηση

Η Κατηγοριοποίηση ή Ταξινόμηση, όπως ήδη αναφέραμε αποτελεί βασική διαδικασία Μηχανικής Μάθησης με επίβλεψη, και αποτελεί μία από τις βασικές εργασίες εξόρυξης δεδομένων (Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2018). Η κατηγοριοποίηση συχνά συγχέεται με το γενικό όρο της πρόβλεψης όμως βασίζεται στην εξέταση των χαρακτηριστικών ενός νέου αντικειμένου (μη κατηγοριοποιημένο) το οποίο με βάση τα χαρακτηριστικά αυτά αντιστοιχίζεται σε ένα προκαθορισμένο σύνολο κλάσεων. Τα αντικείμενα που πρόκειται να κατηγοριοποιηθούν αναπαριστάνονται γενικά από τις εγγραφές της Βάσης Δεδομένων και η διαδικασία της κατηγοριοποίησης αποτελείται από την ανάθεση κάθε εγγραφής σε κάποιες από τις προκαθορισμένες κατηγορίες. Η εργασία της κατηγοριοποίησης χαρακτηρίζεται από έναν καλά καθορισμένο ορισμό των κατηγοριών και το σύνολο που χρησιμοποιείται για την εκπαίδευση του μοντέλου αποτελείται από προκατηγοριοποιημένα παραδείγματα.

Το επόμενο σχήμα απεικονίζει την γενική ιδέα πίσω από την κατηγοριοποίηση. Τα δεδομένα για κατηγοριοποίηση αποτελούνται από μια συλλογή περιστατικών (περιπτώσεων). Κάθε περίπτωση έχει χαρακτηριστεί από ένα ζεύγος (x, y) , όπου x είναι το σύνολο των χαρακτηριστικών τιμών που περιγράφουν το περιστατικό και y είναι η ετικέτα της κλάσης του περιστατικού. Το σύνολο των χαρακτηριστικών x μπορεί να περιέχει χαρακτηριστικά οποιασδήποτε κλάσης, ενώ το y πρέπει να είναι κατηγορηματικό.



Εικόνα 2-7 Σχηματική Απεικόνιση Κατηγοριοποίησης (Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2018)

Η βασική εργασία είναι να δημιουργηθεί ένα μοντέλο το οποίο θα μπορούσε να εφαρμοστεί για να κατηγοριοποιεί δεδομένα που δεν έχουν ακόμη κατηγοριοποιηθεί δεν έχουν δηλαδή ανατεθεί σε κάποια από τις κατηγορίες. Στις περισσότερες περιπτώσεις, υπάρχει ένας περιορισμένος αριθμός κατηγοριών και εμείς θα πρέπει να αναθέσουμε κάθε εγγραφή στην κατάλληλη κατηγορία. Για το σκοπό αυτό χρησιμοποιούνται συνήθως κάποιες τεχνικές, τις οποίες μπορούμε να κατατάξουμε σε δύο κατηγορίες. Η πρώτη χρησιμοποιεί Δέντρα Αποφάσεων (Decision Trees) και η δεύτερη Νευρωνικά Δίκτυα (Neural Networks). Και οι δύο στηρίζονται στην ιδέα της εκπαίδευσης (training) με τη βοήθεια ενός υποσυνόλου δεδομένων που ονομάζεται σύνολο εκπαίδευσης (training set). Το υποσύνολο αυτό επιλέγεται σαν αντιπροσωπευτικό δείγμα του συνολικού όγκου δεδομένων. Με την εφαρμογή της διαδικασίας εκπαίδευσης καθορίζονται κάποια πρότυπα για τις κατηγορίες δεδομένων. Έτσι, όταν προκύψει ένα νέο στοιχείο τότε μπορεί εύκολα να κατηγοριοποιηθεί.

Με πιο απλά λόγια η μέθοδος της κατηγοριοποίησης βασίζεται στην εξέταση των χαρακτηριστικών ενός νέου αντικειμένου το οποίο με βάση τα χαρακτηριστικά αυτά αντιστοιχίζεται σε ένα προκαθορισμένο σύνολο κλάσεων που όπως προαναφέρθηκε χρησιμοποιεί μεθόδους μάθησης με επίβλεψη. Οι τεχνικές της κατηγοριοποίησης χρησιμοποιούν κατά κανόνα ένα σύνολο εκπαίδευσης, όπου όλα τα αντικείμενα είναι ήδη συνδεδεμένα με γνωστές κλάσεις. Ο αλγόριθμος ταξινόμησης «μαθαίνει» από αυτό το

σύνολο, χρησιμοποιώντας την μάθηση αυτή για την κατασκευή ενός μοντέλου. Το μοντέλο αυτό στην συνέχεια ταξινομεί νέα αντικείμενα στις κατάλληλες κλάσεις. Συμπερασματικά μπορούμε να πούμε ότι η κατηγοριοποίηση μαθαίνει σε μία λειτουργία να χαρτογραφεί (ταξινομεί) ένα στοιχείο δεδομένων σε μία από τις διάφορες προκαθορισμένες κατηγορίες.

Από τις παραπάνω Τεχνικές Μηχανικής Μάθησης με Επίβλεψη, σε αυτήν τη διπλωματική εργασία θα ασχοληθούμε με τα Τεχνικά Νευρωνικά Δίκτυα, τα οποία αναλύονται στο κεφάλαιο 2.7

2.6.2 Παλινδρόμηση

Η παλινδρόμηση (regression) αποτελεί την πιο παλιά και συγχρόνως την πιο ευρέως γνωστή στατιστική μέθοδο, η οποία υλοποιείται εντός των πλαισίων της εξόρυξης δεδομένων και όπως και η κατηγοριοποίηση σχετίζεται με την πρόβλεψη τιμής ενός πεδίου με βάση τις τιμές άλλων πεδίων. Η διαφορά της κατηγοριοποίησης με την παλινδρόμηση είναι πως στην πρώτη η κατηγορία³ είναι διακριτή ενώ στην δεύτερη η εργασία είναι συνεχής (Ντάλλα, 2009). Με πιο απλά λόγια η διαφορά της παλινδρόμησης από την κατηγοριοποίηση είναι πως η παλινδρόμηση αντιστοιχεί τα αντικείμενα ενός συνόλου δεδομένων σε μια πραγματική μεταβλητή πρόβλεψης.

Όπως και στην κατηγοριοποίηση έχοντας ως δεδομένο ένα σύνολο δεδομένων, μόνο ένα τμήμα του χρησιμοποιείται για το μοντέλο πρόβλεψης. Αυτό το τμήμα των δεδομένων αποτελεί το σύνολο εκπαίδευσης. Το σύνολο των δεδομένων που απομένει χρησιμοποιείται για την αξιολόγηση της ικανότητας πρόβλεψης του μοντέλου και αποκαλείται σύνολο δοκιμής (testing set). Το σύνολο δοκιμής χρησιμοποιείται για να υπολογιστεί η απόδοση του μοντέλου πάνω σε νέα και άγνωστα δεδομένα, καθώς επίσης και για να υπολογιστεί η εγκυρότητα των προτύπων σε νέα δεδομένα.

Πιο συγκεκριμένα η παλινδρόμηση χρησιμοποιεί μια βάση από αριθμητικά δεδομένα, βάσει των οποίων δημιουργεί μια μαθηματική σχέση που ταιριάζει στα δεδομένα αυτά. Στην συνέχεια η μαθηματική αυτή σχέση χρησιμοποιείται ώστε να προβλεφθεί η μελλοντική συμπεριφορά, με εφαρμογή κάθε φορά σε αυτή νέων αριθμητικών δεδομένων. Έτσι, στηριζόμενοι στα παραπάνω, μπορούμε πλέον να πούμε με μαθηματικές εκφράσεις ότι η παλινδρόμηση αντιστοιχεί ένα δεδομένο x σε μία μεταβλητή πρόβλεψης y . Η συνάρτηση στην οποία γίνεται εκμάθηση είναι της μορφής $y = f(x)$ και έτσι τελικά η παλινδρόμηση είναι η μέθοδος ή η διαδικασία εκτίμησης της εξαρτημένης μεταβλητής από την ανεξάρτητη.

2.6.3 Ομαδοποίηση

Η ομαδοποίηση (clustering) ή αλλιώς συσταδοποίηση ή τμηματοποίηση-κατάτμηση είναι μια περιγραφική μέθοδος και χρησιμοποιείται για να προσδιορίσει τους φυσικούς σχηματισμούς ομάδων από τα δεδομένα με βάση ένα σύνολο κοινών ιδιοτήτων. Όπως προαναφέρθηκε είναι μία βασική διαδικασία μη επιβλεπόμενης μάθησης στον τομέα της Μηχανικής Μάθησης και αυτό οφείλεται στο γεγονός ότι στη συσταδοποίηση ο αριθμός και

³ το πεδίο στόχος αποκαλείται κατηγορία (class) ή εξαρτημένη μεταβλητή στη γλώσσα της στατιστικής.

η σύσταση των ομάδων δεν είναι προκαθορισμένα, καθώς επίσης δεν υπάρχει κάποιο παράδειγμα που να δείχνει ποιες από τις επιθυμητές σχέσεις θα ήταν έγκυρες.

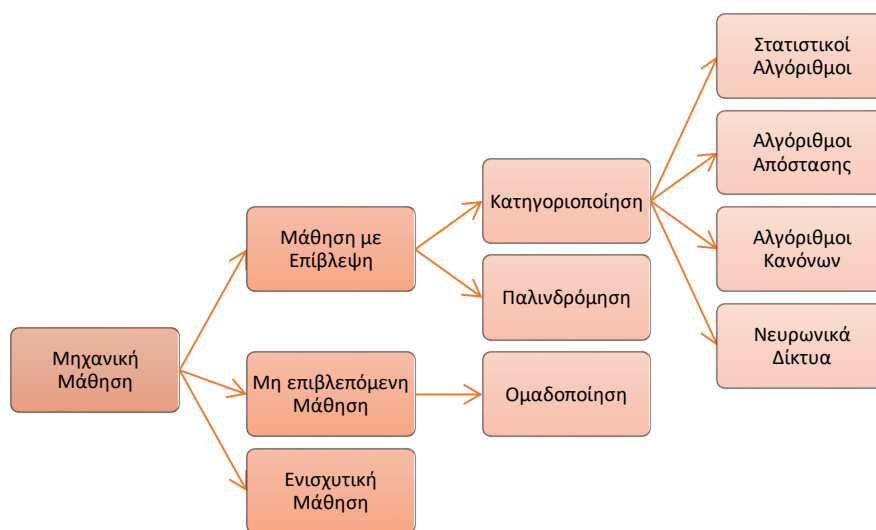
Στόχος της συσταδοποίησης είναι η δημιουργία συστάδων⁴ (clusters), δηλαδή ομάδων (τμημάτων), οι οποίες θα περιέχουν παρόμοια ή παρεμφερή δείγματα δοθέντος του συνόλου δεδομένων. Ένα cluster δηλαδή είναι μια συλλογή αντικειμένων που είναι πανομοιότυπα και διαφέρουν από τα υπόλοιπα αντικείμενα των υπολοίπων τμημάτων (Ντάλλα, 2009) .

Στην ουσία αναζητάμε ένα πεπερασμένο σύνολο συστάδων, για να περιγράψουν τα δεδομένα, καθώς σκοπός της ομαδοποίησης είναι να επιτευχθεί υψηλή ομοιότητα μεταξύ των αντικειμένων εντός των τμημάτων (συστάδων) και χαμηλή ομοιότητα μεταξύ των αντικειμένων που ανήκουν σε διαφορετικά τμήματα. Οι κατηγορίες μπορεί να είναι αμοιβαία απολυόμενες και εξαντλητικές ή να έχουν μία πιο σύνθετη αναπαράσταση (Βασίλειος Βερούκιος, Βασίλειος Καγκλής, Ηλίας Σταυρόπουλος, 2015) .

Οι αλγόριθμοι συσταδοποίησης βασίζονται στο σύνολό τους σε δύο αρχές:

- τη μεγιστοποίηση της ομοιότητας μεταξύ των αντικειμένων της ίδιας συστάδας (intracluster similarity)
- τη ταυτόχρονη ελαχιστοποίηση της ομοιότητας των αντικειμένων διαφορετικών συστάδων (inter-class similarity)

Στο επόμενο σχήμα αναγράφονται οι κατηγορίες της μηχανικής μάθησης και οι υποκατηγορίες που αυτές χωρίζονται.



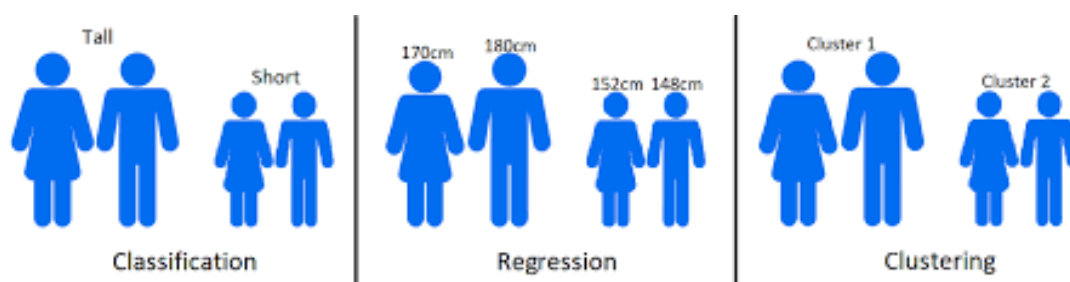
Εικόνα 2-8 Κατηγορίες Μηχανικής Μάθησης

⁴ Η συστάδα ορίζεται να είναι μια συλλογή από αντικείμενα που φέρουν κάποια κοινά χαρακτηριστικά

2.6.4 Σύγκριση Κατηγοριοποίηση-Παλινδρόμηση-Ομαδοποίηση

Όπως προαναφέρθηκε στην κατηγοριοποίηση, το αποτέλεσμα που θέλουμε να προβλέψουμε είναι η κλάση των δειγμάτων, η οποία κλάση μπορεί να πάρει διακριτές τιμές από ένα πεπερασμένο σύνολο. Αντίθετα, κατά την πρόβλεψη με χρήση τεχνικών όπως η παλινδρόμηση, η μεταβλητή-στόχος μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός (Βασίλειος Βερούκιος, Βασίλειος Καγκλής, Ηλίας Σταυρόπουλος, 2015) .

Για να διευκρινίσουμε τις διαφορές ανάμεσα στην κατηγοριοποίηση, την παλινδρόμηση και την ομαδοποίηση θα εξετάσουμε ένα σύνολο δεδομένων με βάση τα χαρακτηριστικά μιας ομάδας ατόμων. Η κατηγοριοποίηση περιλαμβάνει την προσπάθεια επισήμανσης κάθε ατόμου ως «κοντού» ή «ψηλού», η παλινδρόμηση προσπαθεί να προβλέψει πόσο ψηλό είναι κάθε άτομο και στην ομαδοποίηση όπου δεν υπάρχουν ετικέτες ύψους, ο αλγόριθμος θα συγκεντρώσει σε ομάδες τα άτομα λαμβάνοντας υπόψιν τα παρόμοια τους χαρακτηριστικά (π.χ. παρόμοια βάρη ή μεγέθη παπουτσιών που μπορεί να υποδηλώνουν ότι κάποιος είναι ψηλός ή κοντός) (Lawrence, 2017).



Εικόνα 2-9 Παράδειγμα σύγκρισης Ταξινόμησης- Παλινδρόμησης-Ομαδοποίησης (Lawrence, 2017)

2.7 Τεχνητά Νευρωνικά Δίκτυα

Το αντικείμενο των Τεχνητών Νευρωνικών Δικτύων- ΤΝΔ ή απλά Νευρωνικά Δίκτυα, γνωρίζει ραγδαία ανάπτυξη τα τελευταία 35 χρόνια και αποτελεί πλέον ένα ευρύ και αυτόνομο επιστημονικό πεδίο, που σχετίζεται με το γενικότερο πλαίσιο της τεχνητής νοημοσύνης και των ευφύων συστημάτων (Διαμαντάρας, 2007) . Τα Τεχνητά Νευρωνικά Δίκτυα – ΤΝΔ αποτελούν μια μορφή τεχνητής νοημοσύνης και αποτελούνται από ένα σύνολο απλών, διασυνδεδεμένων και προσαρμοστικών μονάδων, οι οποίες συνιστούν ένα πολύπλοκο υπολογιστικό μοντέλο (Θεοδόση-Κόκκινου, 2013). Με πιο απλά λόγια τα ΤΝΔ αποτελούν ένα μαθηματικό μοντέλο προσομοίωσης της λειτουργίας του ανθρώπινου εγκεφάλου.

Ορισμός 8: "Ένα Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ) είναι ένας παράλληλος καταμετρημένος επεξεργαστής που αποτελείται από τους νευρώνες ,οι οποίοι είναι απλές ανεξάρτητες υπολογιστικές μονάδες επεξεργασίας και είναι διασυνδεδεμένοι μεταξύ τους σε στρώματα. Διαθέτει μία φυσική κλίση στο να αποθηκεύει εμπειρική γνώση , την οποία και χρησιμοποιεί. Τα δύο σημεία όπου το ΤΝΔ μοιάζει με τον ανθρώπινο εγκέφαλο είναι ότι το δίκτυο μέσω μίας διαδικασίας εκπαίδευσης λαμβάνει γνώση από το περιβάλλον και πως η γνώση αποθηκεύεται στα συνοπτικά βάρη που υπάρχουν στις διασυνδέσεις μεταξύ των νευρώνων (I. Aleksander and H. Morton, 1990) .

Ο κλάδος της τεχνητής νοημοσύνης ασχολείται με την κατανόηση της λειτουργίας του εγκέφαλου, έτσι τα ΤΝΔ έχοντας ως στόχο την προσέγγιση της λειτουργίας του, και μέσω της μίμησης έχουν ως απώτερο σκοπό να κατασκευάσουν αποδοτικότερους αλγόριθμους για την επίλυση προβλημάτων διαφόρων ειδών. Υπάρχουν αρκετά μοντέλα ΤΝΔ⁵ τα οποία έχουν ως κοινό χαρακτηριστικό ότι όλα κατηγοριοποιούνται με βάση τον τρόπο εκπαίδευσής τους, ενώ ο αναγνώστης εισάγεται στο γενικότερο πρόβλημα της μάθησης και της αυτοπροσαρμογής ενός υπολογιστικού συστήματος (Διαμαντάρας, 2007). Τα ΤΝΔ για να χρησιμοποιηθούν, πρέπει πρώτα να εκπαιδευτούν για να μάθουν. Κατά τη διαδικασία της μάθησης οι συντελεστές βάρους των νευρωνικών δικτύων αποκτούν κατάλληλες τιμές, ώστε το ΤΝΔ να εκτελεί τους κατάλληλους υπολογισμούς και να λύνει το εκάστοτε πρόβλημα.

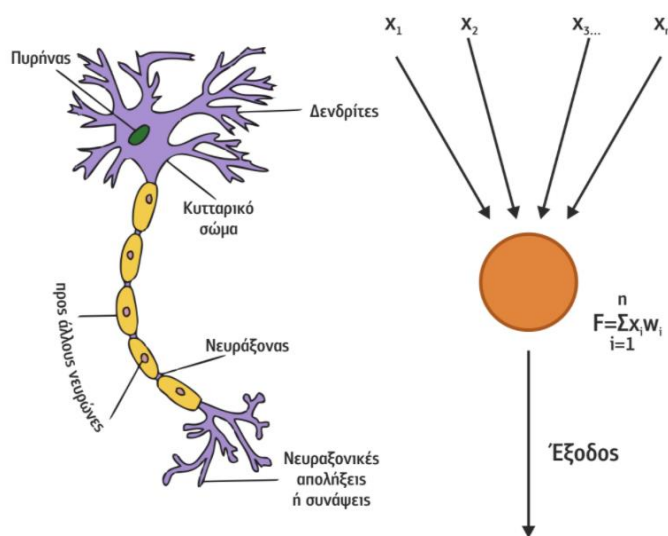
2.7.1 Δομή και Λειτουργία των Τεχνητών νευρωνικών δικτύων

Η δομή του ΤΝΔ μιμείται όσο το δυνατόν καλύτερα την δομή του βιολογικού νευρωνικού δικτύου με στόχο να εμφανίζει παρόμοιες ιδιότητες, καθώς αποτελεί ένα απλοποιημένο μοντέλο του. Επομένως όπως και το δίκτυο νευρώνων του εγκεφάλου, έτσι και ένα ΤΝΔ αποτελείται από ένα σύνολο αλληλοεπιδρώντων τεχνητών νευρώνων που συνδέονται μεταξύ τους με τις συνάψεις (synapses) ή νευροαξονικές απολήξεις όπως ονομάζονται.

Κάθε ζεύγος νευρώνων αλληλοεπιδρά σε διαφορετικό βαθμό με το περιβάλλον, γεγονός που αντικατοπτρίζεται στα συνοπτικά βάρη (synaptic weights). Πιο συγκεκριμένα, καθώς το

⁵ Τα σημαντικότερα μοντέλα ΤΝΔ είναι: ο στοιχειώδης *Perceptron* του ενός νευρώνα, τα δίκτυα *Perceptron* πολλών στρωμάτων, ο αλγόριθμος εκπαίδευσης *Back Propagation*, τα δίκτυα *Radial Basis Function (RBF)*, τα αυτο-οργανούμενα δίκτυα, τα γραμμικά και μη-γραμμικά *Χεμπιανά* μοντέλα μάθησης, οι μηχανές διανυσμάτων υποστήριξης (*SVM*), τα δυναμικά μοντέλα και άλλα.

ΤΝΔ έρχεται σε αλληλεπίδραση με το περιβάλλον και μαθαίνει, οι δεσμοί αναπροσαρμόζονται αφού τα συνοπτικά βάρη μεταβάλλονται συνέχεια, ενδυναμώνοντας είτε αποδυναμώνοντας τον εκάστοτε δεσμό. Συνεπώς η εμπειρία που αποκτά το ΤΝΔ από το περιβάλλον κωδικοποιείται στα συνοπτικά βάρη προσδίδοντας την ικανότητα στο δίκτυο να εξελίσσεται και να προσαρμόζεται (Κολοσιώνης, 2018). Στο επόμενο σχήμα παρουσιάζεται μια σύγκριση ενός φυσικού νευρώνα με ένα στοιχειώδη τεχνητό νευρώνα Perceptron , δηλαδή ενός ΤΝΔ που αποτελείται από έναν μόνο νευρώνα . Κάθε τεχνητός νευρώνας αποτελείται από πολλές εισόδους x_i και μία μόνο έξοδο y . Κάθε είσοδος x_i «ζυγίζεται» με ένα βάρος w_i και τα αποτελέσματα αθροίζονται μέσω της συνάρτησης αθροίσματος (summation function) F (Γεωργούλη, 2015) .



Εικόνα 2-10 Σύγκριση φυσικού νευρώνα με στοιχειώδη τεχνητό νευρώνα (Γεωργούλη, 2015)

Όπως προαναφέρθηκε τα ΤΝΔ είναι συστήματα επεξεργασίας δεδομένων, τα οποία αποτελούνται από ένα πλήθος τεχνητών νευρώνων. Η λειτουργία των τεχνητών νευρώνων αυτών είναι παρόμοια με την λειτουργία των νευρώνων του ανθρώπινου εγκεφάλου, καθώς και στις δύο περιπτώσεις οι νευρώνες εργάζονται παράλληλα, με στόχο την επίλυση προβλημάτων. Τα συνήθη ΤΝΔ χρησιμοποιούν πολύ απλοποιημένα μοντέλα νευρώνων τέτοια ώστε να διατηρούν μόνο τα πολύ αδρά χαρακτηριστικά των λεπτομερών μοντέλων που χρησιμοποιούνται στη νευρολογία.

Υπάρχουν πολλοί διαφορετικοί τύποι ΝΔ, με διαφορετικές αρχιτεκτονικές και εφαρμογές. Τα νευρικά δίκτυα επεξεργάζονται γρήγορα νέα δεδομένα επειδή είναι συμπαγή μοντέλα. Ωστόσο, σε αντίθεση απαιτούν συνήθως ένα μεγάλο ποσό υπολογισμού που πρέπει να εκπαιδευτεί. Επιπλέον προσαρμόζονται εύκολα και είναι ικανά για προβλήματα παλινδρόμησης και ταξινόμησης (Mohammad Saeid Mahdavejad, Mohammadreza Rezvan , Mohammadamin Barekatin , Peyman Adibi , Payam Barnaghi , Amit P. Sheth, 2018) .

Επιπροσθέτως, τα νευρωνικά δίκτυα έχουν εξαιρετική ικανότητα στο να λαμβάνουν σύνθετη και υψηλού επιπέδου γνώση λόγω της μη γραμμικής τους αναπαράστασης, καθώς ένα ΤΝΔ αποτελείται από διασυνδεδεμένους μη γραμμικούς νευρώνες. Η εκπαίδευσή τους πραγματοποιείται βάσει ορισμένων παραδειγμάτων, τα οποία επιλέγονται προσεκτικά για

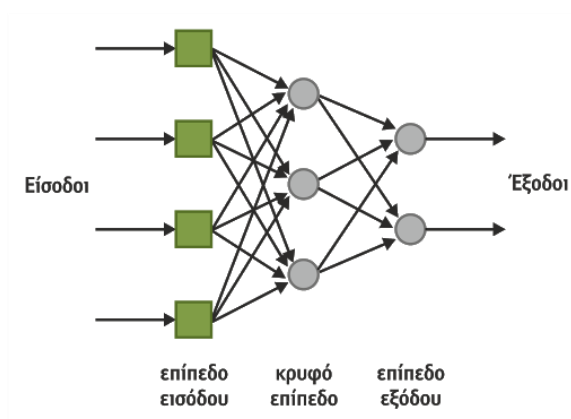
αποφυγή μη σωστής λειτουργίας του δικτύου ή σπατάλης χρήσιμου χρόνου, σύμφωνα με τα οποία μαθαίνουν το περιβάλλον τους. Έτσι λοιπόν, είναι πολύ σημαντική η συνεισφορά τους για παράδειγμα στην ιατρική, τις χρηματιστηριακές προβλέψεις, την εκπαίδευση, τη ρομποτική κ.α. Εξίσου σημαντική είναι και η συμβολή τους στην αναγνώριση προτύπων και ταξινόμηση δεδομένων μέσα από μια διαδικασία εκπαίδευσης.

Υπάρχουν έξι βασικά χαρακτηριστικά των ΤΝΔ τα οποία είναι σημαντικά για την τεχνολογία και παρουσιάζονται με την βοήθεια του παρακάτω σχήματος (Vidushi Sharma , Sachin Rai , Anurag Dev, 2012):



Εικόνα 2-11 Χαρακτηριστικά ΤΝΔ (Vidushi Sharma , Sachin Rai , Anurag Dev, 2012)

Τα ΤΝΔ χαρακτηρίζονται, επιπλέον, με βάση τον τρόπο με τον οποίο είναι συνδεδεμένοι οι κόμβοι τους δηλαδή αν είναι πρόσθιας τροφοδότησης (feed forward) ή οπίσθιας τροφοδότησης (feed backward ή recurrent). Στην πλειοψηφία των εφαρμογών χρησιμοποιούνται δίκτυα πρόσθιας τροφοδότησης ενός κρυφού επιπέδου με πλήρως συνδεδεμένους κόμβους. Ένα παράδειγμα ΤΝΔ πρόσθιας τροφοδότησης παρουσιάζεται στην επόμενη εικόνα:



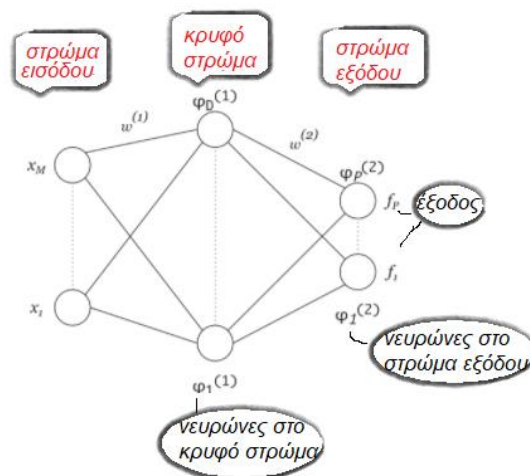
Εικόνα 2-12 ΤΝΔ πρόσθιας τροφοδότησης (Γεωργούλη, 2015)

Το Feed Forward Neural Networks (FFNN) είναι γνωστό ως Multilayer Perceptrons (MLP) και είναι ο πιο κοινός τύπος νευρωνικών δικτύων στις πρακτικές εφαρμογές. Η μορφή της κατηγοριοποίησης ή της παλινδρόμησης είναι:

$$f(\mathbf{x}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) = \Phi^{(2)}(\Phi^{(1)}(\mathbf{x}^T \mathbf{w}^{(1)})^T \mathbf{w}^{(2)}) \quad [\text{σχέση 1}]$$

όπου $\mathbf{w}^{(1)} = (w_1^{(1)}, \dots, w_M^{(1)})^T$, $\Phi^{(1)} = (\phi_1^{(1)}, \dots, \phi_D^{(1)})^T$, $\mathbf{w}^{(2)} = (w_1^{(2)}, \dots, w_D^{(2)})^T$ και $\Phi^{(2)} = (\phi_1^{(2)}, \dots, \phi_p^{(2)})^T$

Στην συνέχεια απεικονίζεται ένα νευρωνικό δίκτυο τροφοδοσίας δύο επιπέδων. Κάθε νευρώνας εξόδου είναι συνδεδεμένος σε κάθε νευρώνα εισόδου, είναι δηλαδή ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο. Τα στοιχεία του διανύσματος εισόδου \mathbf{x} είναι μονάδες (νευρώνες) στο επίπεδο είσοδο, το $\phi_i^{(1)}$ είναι οι μονάδες στο κρυφό επίπεδο και το $\phi_i^{(2)}$ είναι οι μονάδες στο επίπεδο εξόδου, το οποίο εξάγει την f . Οι δραστηριότητες των μονάδων σε κάθε ένα επίπεδο καθορίζονται από μια μη γραμμική συνάρτηση των δραστηριοτήτων του προηγούμενο επιπέδου. Στην μηχανική μάθηση $\phi(\cdot)$ ονομάζεται η συνάρτηση ενεργοποίησης. Η συνάρτηση ενεργοποίησης επιλέγεται στο τελευταίο επίπεδο σε σχέση με την εργασία επεξεργασίας των δεδομένων (Mohammad Saeid Mahdavejad, Mohammadreza Rezvan, Mohammadamin Barekatin, Peyman Adibi, Payam Barnaghi, Amit P. Sheth, 2018).



Εικόνα 2-13 Νευρωνικό δίκτυο τροφοδοσίας δύο επιπέδων

3 ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΕΡΕΥΝΑΣ

3.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται ο τρόπος με τον οποίο σχεδιάστηκε και πραγματοποιήθηκε η έρευνα της παρούσας διπλωματικής εργασίας. Πιο συγκεκριμένα θα παρουσιαστούν οι στόχοι και τα ερευνητικά ερωτήματα της παρούσας έρευνας, το δείγμα, τα εργαλεία συλλογής δεδομένων και τα αποτελέσματα.

3.2 Ορισμός Δείγματος- Στόχοι της έρευνας

Σκοπός της παρούσας έρευνας είναι η συλλογή δεδομένων για τη δημιουργία μοντέλου πρόβλεψης επιτυχίας ή μη στις πανελλαδικές εξετάσεις στο μάθημα των μαθηματικών, λαμβάνοντας υπόψιν τις γραπτές και προφορικές βαθμολογίες του στις τρεις τάξεις του Λυκείου, την μόνιμη κατοικία του, την οικονομική του κατάσταση και το άγχος, που εμφανώς επηρεάζει τα αποτελέσματα των υποψηφίων.

Η επιλογή του δείγματος πρέπει να προσδιορίζει το σωστό υποσύνολο του πληθυσμού γι' αυτό το λόγω στην έρευνα μας έλαβαν μέρος άτομα ηλικίας 18 χρονών και άνω. Τα αποτελέσματα γίνονται πιο αξιόπιστα καθώς το μέγεθος του δείγματος μεγαλώνει γι' αυτό το λόγω η έρευνα πραγματοποιήθηκε σε διάφορα μέρη της Ελλάδος, χωρίς να υπάρχει γεωγραφικός περιορισμός. Είναι απαραίτητο να βρεθεί ένα αντιπροσωπευτικό δείγμα έτσι ώστε να γίνει εφικτή η γενίκευση των συμπερασμάτων της έρευνας.

Η έρευνα πραγματοποιήθηκε κατά τον μήνα Ιούνιο του έτους 2020 και έλαβαν μέρος 340 άτομα, 18 ετών και άνω όπως περιγράφεται στην συνέχεια.

3.3 Σχεδιασμός και υλοποίηση έρευνας ερωτηματολογίου

Για τη δημιουργία του ερωτηματολογίου έρευνας μελετήθηκαν παρόμοιες εργασίες και ερωτηματολόγια προηγούμενων ερευνών και εργασιών. Επιλέχθηκε η δημιουργία ενός νέου ερωτηματολογίου που δεν βασίζεται σε κάποιο υπάρχον. Η δημιουργία ενός νέου ερωτηματολογίου αν και σχετικά χρονοβόρα και ριψοκίνδυνη διαδικασία, καθώς ενέχει πάντα την πιθανότητα αμφισβητήσιμων αποτελεσμάτων, χρησιμοποιείται δεδομένου ότι είναι η μόνη που μπορεί να καλύψει όλα τα δεδομένα που μας ενδιαφέρουν.

Τα ερωτηματολόγια συνήθως αποτελούνται από «κλειστές» και «ανοικτές» ερωτήσεις. Οι ανοικτές ερωτήσεις δεν έχουν προκαθορισμένες απαντήσεις, αλλά επιτρέπουν στους συμμετέχοντες να απαντήσουν οτιδήποτε επιθυμούν. Οι ανοικτές ερωτήσεις απαιτούν

χρόνο ,έτσι στο ερωτηματολόγιο που χρησιμοποιήσαμε υπάρχουν μόνο δύο τέτοιες ερωτήσεις. Στις κλειστές ερωτήσεις οι ερευνητές προκαθορίζουν τις απαντήσεις, οπότε οι συμμετέχοντες καλούνται να επιλέξουν μεταξύ των απαντήσεων που έχουν προκαθοριστεί. Γενικά προτιμάται οι ερωτήσεις να είναι κλειστού τύπου, έτσι και εμείς επιλέξαμε οι πλειονότητα των ερωτήσεων να είναι κλειστού τύπου. Αυτό που απαιτεί προσοχή στις κλειστές ερωτήσεις είναι οι κατηγορίες των απαντήσεων να καλύπτουν ολόκληρο το φάσμα των πιθανών απαντήσεων των συμμετεχόντων.

3.4 Αξιοπιστία και εγκυρότητα ερωτηματολογίου

Η αξιοπιστία (credibility) και η εγκυρότητα (validity) είναι δύο χαρακτηριστικά τα οποία πρέπει απαραίτητως να διαθέτει μία έρευνα έτσι ώστε μετά το πέρας της να γίνει αποδεκτή και να έχει ισχύ. **Το ερωτηματολόγιο που χρησιμοποιήθηκε ήταν ανώνυμο** ώστε να διασφαλισθεί η αξιοπιστία στις απαντήσεις που δόθηκαν από τους συμμετέχοντες. Παράλληλα, κατά το σχεδιασμό του ερωτηματολογίου, έγινε προσπάθεια να αποφευχθούν σύνθετες ερωτήσεις.

3.5 Ερωτηματολόγιο Έρευνας

Για τις ανάγκες της έρευνας δημιουργήθηκε ένα τελικό ερωτηματολόγιο με 32 ερωτήματα, το οποίο στάλθηκε με τη μορφή υπερσυνδέσμου στους ερωτώμενους μέσω της εφαρμογής του GoogleDrive. Το ερωτηματολόγιο, όπως παρουσιάζεται στο παράρτημα της διπλωματικής εργασίας και όπως αναφέρθηκε παραπάνω απαρτιζόταν από ερωτήσεις σε πλειονότητα κλειστού αλλά και ανοικτού τύπου. Κρίθηκε αναγκαίο να είναι σχετικά σύντομο, να είναι δηλαδή μικρό σε έκταση και να χρησιμοποιηθεί απλή γλώσσα, ώστε να είναι κατανοητό και εύκολα διαχειρίσιμο από τους ερωτώμενους. Το εν λόγω ερωτηματολόγιο αποτελείται από τέσσερις ενότητες:

- Στην 1^η ενότητα περιλαμβάνονται 7 ερωτήσεις με δημογραφικές και προσωπικές πληροφορίες του ερωτώμενου, όπως το φύλο, η ηλικία, η κατοικία, το επίπεδο εκπαίδευσης , η εργασία, οικονομική κατάσταση, κ.α. Αυτές οι πληροφορίες μας επιτρέπουν να εντοπίσουμε αν τα αποτελέσματα της μελέτης ποικίλουν σε διαφορετικές ομάδες ανθρώπων.
- Στην 2^η ενότητα περιλαμβάνονται 9 ερωτήσεις σχετικά με τις Πανελλαδικές εξετάσεις όπως ο χρόνος μελέτης και το άγχος κατά τη διάρκεια της χρονιάς αλλά και την ημέρα των πανελλαδικώς εξεταζόμενων μαθημάτων, αν οι ερωτώμενοι βασίστηκαν στην δημόσια εκπαίδευση ή όχι, τα μόρια που έγραψε ο ερωτώμενος στις πανελλαδικές εξετάσεις, αν πέτυχε την πρώτη σχολή της προτίμησης του και άλλα.
- Στην 3^η ενότητα περιλαμβάνονται 13 ερωτήσεις και ζητούνται πληροφορίες σχετικά με τα Μαθηματικά, όπως η βαθμολογία (τριμήνων και γραπτών) και στις τρεις τάξεις του Λυκείου, η βαθμολογία στα μαθηματικά στις πανελλαδικές, μέσος χρόνος μελέτης και η εξοικείωση του ερωτώμενου με τα μαθηματικά.

- Στην 4^η και τελευταία ενότητα υπάρχουν 3 Γενικές ερωτήσεις σχετικά με τα μαθηματικά, οι δύο ανοικτού τύπου όπως για παράδειγμα οι εφαρμογές στην καθημερινή ζωή, κλπ.

3.6 Μέθοδος επεξεργασίας και ανάλυσης δεδομένων

Το ερωτηματολόγιο απαντήθηκε από 340 ενήλικες. Μετά την συλλογή των δεδομένων άρχισε το στάδιο της επεξεργασίας τους. Με την ολοκλήρωση των απαντήσεων το πρώτο βήμα ήταν η καταχώρηση όλων των απαντήσεων σε ένα αρχείο excel, ένα δείγμα του οποίου παραθέτουμε παρακάτω:

1. Φύλο	2. Ηλικία	3. Κατοικία	4. Επίπεδο Εκπαίδευσης	5. Εργασία	6. Οικονομική Κατάσταση	7. Μορφωτικό επίπεδο γονέων	8. Είδος Εκπαίδευσης	9. Πόσες φορές έχετε δώ	10. Πόσες ώρες
Γυναίκα	25-30	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		4 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 4-6
Γυναίκα	>40	Αγροτική Περιοχή	Απόφοιτος/η Λυκείου	Δημόσιος Τομέας		3 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		2 6-8
Γυναίκα	30-35	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Δημόσιος Τομέας		3 Μεταπτυχιακό	Δημόσιο Σχολείο		1 >8
Γυναίκα	20-25	Αγροτική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Άνεργος/Άνεργη		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 2-4
Γυναίκα	25-30	Ημιαστική Περιοχή	Κάτοχος Μεταπτυχιακού	Δημόσιος Τομέας		4 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		1 4-6
Γυναίκα	18-20	Αγροτική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Φοιτητής/Φοιτήτρια		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 2-4
Γυναίκα	20-25	Ημιαστική Περιοχή	Απόφοιτος/η Λυκείου	Φοιτητής/Φοιτήτρια		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 4-6
Γυναίκα	>40	Αγροτική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		3 Άγροτες	Δημόσιο Σχολείο		1 6-8
Ανδρας	>40	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 4-6
Γυναίκα	25-30	Ημιαστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Δημόσιος Τομέας		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 2-4
Γυναίκα	>40	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		3 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		1 2-4
Ανδρας	20-25	Αγροτική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Άνεργος/Άνεργη		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 2-4
Γυναίκα	30-35	Ημιαστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 4-6
Γυναίκα	30-35	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 4-6
Ανδρας	>40	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Ιδιωτικός Τομέας		3 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		1 6-8
Γυναίκα	35-40	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 4-6
Γυναίκα	25-30	Ημιαστική Περιοχή	Κάτοχος Μεταπτυχιακού	Ιδιωτικός Τομέας		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 4-6
Γυναίκα	25-30	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Ιδιωτικός Τομέας		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 4-6
Γυναίκα	20-25	Αστική Περιοχή	Απόφοιτος/η Λυκείου	Φοιτητής/Φοιτήτρια		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 2-4
Ανδρας	25-30	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		5 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 0-2
Γυναίκα	18-20	Αστική Περιοχή	Απόφοιτος/η Λυκείου	Φοιτητής/Φοιτήτρια		3 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		1 2-4
Γυναίκα	20-25	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		4 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 2-4
Ανδρας	25-30	Αγροτική Περιοχή	Απόφοιτος/η Λυκείου	Ιδιωτικός Τομέας		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 0-2
Γυναίκα	20-25	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Ιδιωτικός Τομέας		3 Πανεπιστημιακής Εκπαίδευσης	Ιδιωτικό Σχολείο		1 4-6
Ανδρας	>40	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Δημόσιος Τομέας		3 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		2 6-8
Γυναίκα	30-35	Ημιαστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		4 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		1 6-8
Γυναίκα	30-35	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Ιδιωτικός Τομέας		3 Απόφοιτοι Γυμνασίου	Δημόσιο Σχολείο		2 2-4
Γυναίκα	>40	Αστική Περιοχή	Κάτοχος Μεταπτυχιακού	Δημόσιος Τομέας		3 Πανεπιστημιακής Εκπαίδευσης	Δημόσιο Σχολείο		1 4-6
Ανδρας	>40	Αστική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		4 Δημοτικού	Δημόσιο Σχολείο		2 2-4
Γυναίκα	35-40	Αγροτική Περιοχή	Απόφοιτος/η ΑΕΙ/ΤΕΙ	Αυταπασχόληση		2 Απόφοιτοι Λυκείου	Δημόσιο Σχολείο		1 2-4

Εικόνα 3-1 Δείγμα αποτελεσμάτων EXCEL

Όπως παρατηρούμε κάθε στήλη έχει έναν τίτλο που αντιστοιχεί στην ερώτηση που κλήθηκε να απαντήσει ο ερωτώμενος. Κάθε γραμμή αντιστοιχεί σε μία σειρά απαντήσεων ενός συγκεκριμένου ατόμου.

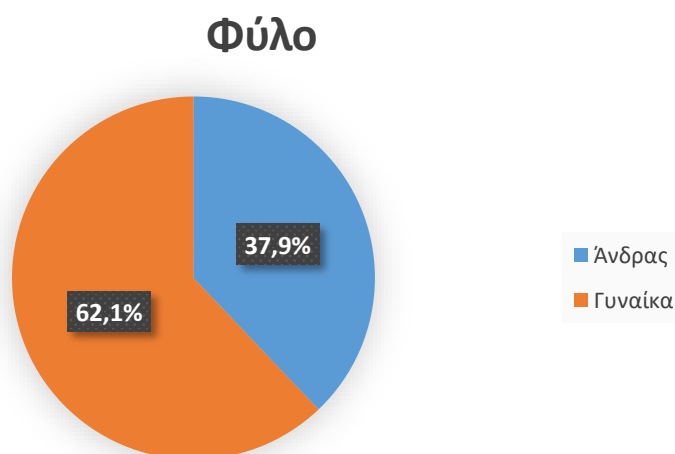
Στην προηγούμενη εικόνα εμφανίζονται οι απαντήσεις σχετικά με τα δημογραφικά χαρακτηριστικά των ατόμων που συμμετείχαν στην έρευνα. Για την δημιουργία του μοντέλου μας επιλέξαμε συγκεκριμένες ερωτήσεις και κάθε κατηγορία ορίστηκε σαν αριθμός από το 0 έως N ανάλογα με τις N κατηγορίες που υπάρχουν σε κάθε ερώτηση, όπως αναλύονται στο κεφάλαιο 4.1 .

3.7 Παρουσίαση αποτελεσμάτων της έρευνας

Όπως προαναφέρθηκε συλλέξαμε τις απόψεις και τα στοιχεία 340 ατόμων. Στο κεφάλαιο αυτό γίνεται παράθεση των στατιστικών αποτελεσμάτων της έρευνας και παρουσιάζονται τα αποτελέσματα όσο αφορά τα γενικά στοιχεία των ατόμων του δείγματος που συμμετείχαν καθώς επίσης γραφήματα και πίνακες που αναλύουν τις απαντήσεις τους.

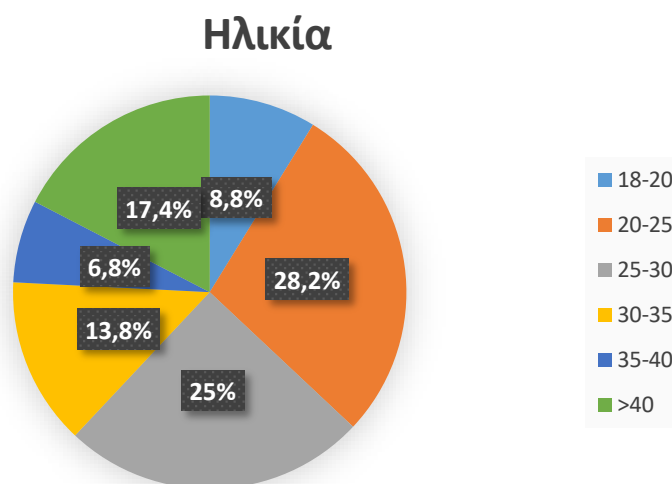
3.7.1 Δημογραφικά στοιχεία συμμετεχόντων

Σύμφωνα με έρευνα του PISA 2015 (Σοφianoπούλου, Εμβαλωτής, Πίτσια & Καρακολίδης, 2017) οι γυναίκες και οι άνδρες δεν διαφέρουν στατιστικά σημαντικά στην επίδοσή τους στα Μαθηματικά. Από τα δημογραφικά στοιχεία παρατηρούμε ότι το μεγαλύτερο ποσοστό που συμμετείχε ήταν γυναίκες, πιο αναλυτικά 129 άνδρες δηλαδή το 37,9% του δείγματος και 211 γυναίκες, δηλαδή το 62,1% και παρουσιάζονται σχηματικά παρακάτω:



Γράφημα 3-1 Φύλο Συμμετεχόντων

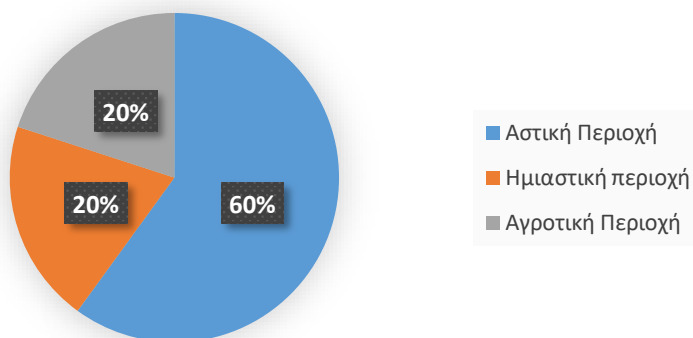
Ως προς την ηλικία το μεγαλύτερο ποσοστό που συμμετείχε ήταν νέοι 20 έως 25 ετών (28,2%) και το μικρότερο ποσοστό άτομα ηλικίας 35 έως 40 (6,8%). Αναλυτικότερα συμμετείχαν 30 άτομα ηλικίας 18 έως 20 ετών, 96 άτομα ηλικίας 20 έως 25 ετών, 85 άτομα ηλικίας 25-30 ετών 47 άτομα ηλικίας 30 έως 35 ετών, 23 άτομα ηλικίας 35 έως 40 και τέλος 59 άτομα ηλικίας 40 ετών και άνω.



Γράφημα 3-2 Ηλικία Συμμετεχόντων

Σύμφωνα με έρευνα (Σοφιανοπούλου, Εμβαλωτής, Πίτσια & Καρακολίδης, 2017) μαθητές που φοιτούν σε σχολεία των αστικών περιοχών είναι αυτοί που κατά μέσο όρο έχουν τις υψηλότερες επιδόσεις στο σχολείο, ενώ αντίστοιχα οι μαθητές των αγροτικών περιοχών έχουν τις χαμηλότερες επιδόσεις. Στην συγκεκριμένη έρευνα, σχετικά με την κατοικία το μεγαλύτερο μέρος του δείγματος είναι κάτοικοι Αστικής περιοχής, 204 άτομα δηλαδή το 60% του δείγματος και 68 άτομα ημιαστικής και 68 άτομα αγροτικής περιοχής. Παρουσιάζεται σχηματικά παρακάτω:

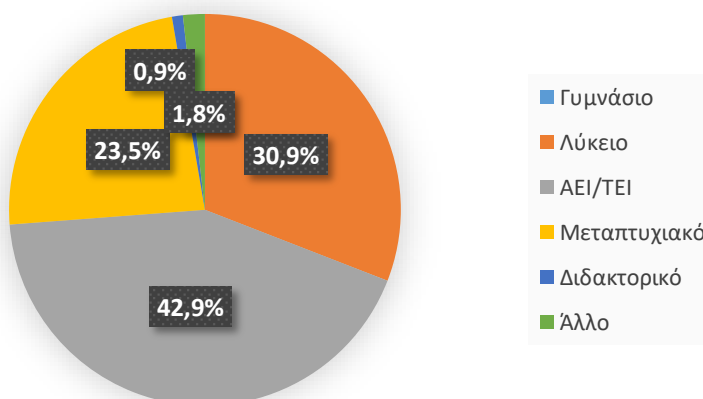
Κατοικία



Γράφημα 3-3 Κατοικία Συμμετεχόντων

Σχετικά με το επίπεδο εκπαίδευσης των ερωτώμενων τα περισσότερα άτομα ήταν απόφοιτοι Πανεπιστημιακής και Τεχνολογικής εκπαίδευσης (ΑΕΙ/ΤΕΙ) ,καθώς συμμετείχαν 146 άτομα (42,9%). Πιο αναλυτικά δεν έλαβε μέρος κανένα άτομο που το ανώτατο επίπεδο εκπαίδευσης να είναι Γυμνάσιο, 105 απόφοιτοι Λυκείου (30,9%), όπως προαναφέρθηκε 156 απόφοιτοι ΑΕΙ/ΤΕΙ, 80 άτομα που είναι κάτοχοι Μεταπτυχιακού Διπλώματος δηλαδή το 23,5% του δείγματος, 3 άτομα κάτοχοι Διδακτορικού Διπλώματος και τα υπόλοιπα άτομα που περιέχονται στην κατηγορία «άλλο» είναι απόφοιτοι ΙΕΚ, ΕΠΑΣ, κ.α. και καταλαμβάνουν το 1,8% του δείγματος.

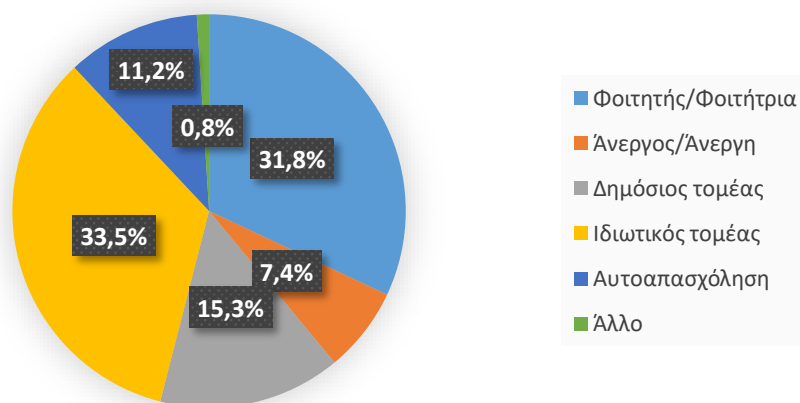
Επίπεδο εκπαίδευσης



Γράφημα 3-4 Επίπεδο εκπαίδευσης Συμμετεχόντων

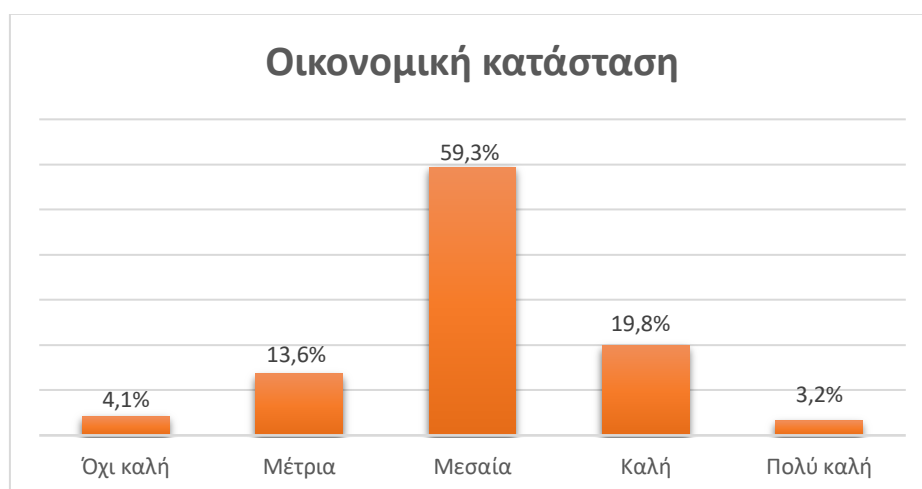
Σχετικά με τον εργασιακό τομέα των ατόμων που έλαβαν μέρος στην έρευνα, το μεγαλύτερο ποσοστό ήταν εργαζόμενοι Ιδιωτικού Τομέα. Αναλυτικότερα συμμετείχαν 108 φοιτητές και φοιτήτριες (31,8%), 25 άνεργοι (7,4%), 52 υπάλληλοι Δημόσιου τομέα (15,3%), 114 άτομα που εργάζονται στον Ιδιωτικό τομέα (33,5%), 38 άτομα που αυτοαπασχολούνται δηλαδή το 11,2% του δείγματος και το 0,8% δεν ανήκει σε καμία από τις προαναφερθέντες κατηγορίες, είναι συνταξιούχοι, ελεύθεροι επαγγελματίες, κ.α.

Εργασία



Γράφημα 3-5 Εργασία Συμμετεχόντων

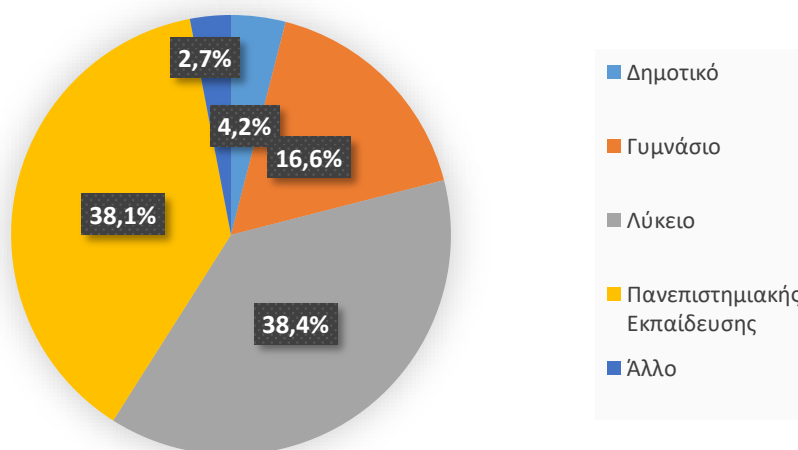
Στην συνέχεια παρουσιάζεται σχηματικά και κατά προσέγγιση η οικονομική κατάσταση των ατόμων που έλαβαν μέρος στην έρευνα και παρατηρούμε ότι 201 άτομα δηλαδή το μεγαλύτερο ποσοστό του δείγματος ,για ακρίβεια το 59,3%, είναι μεσαίας οικονομικής κατάστασης.ς. Το 4,1%, δηλαδή 14 άτομα δεν έχουν καθόλου καλή οικονομική κατάσταση, ενώ 11 άτομα δηλαδή το 3,2% του δείγματος έχουν άριστη κατάσταση.



Γράφημα 3-6 Οικονομική κατάσταση Συμμετεχόντων

Η τελευταία ερώτηση που κλήθηκαν να απαντήσουν οι συμμετέχοντες είναι σχετικά με το μορφωτικό επίπεδο των γονέων τους. Οι απαντήσεις που έδωσαν αφορούν έστω έναν από τους δύο γονείς. Πιο αναλυτικά 14 άτομα απάντησαν πως το ανώτατο επίπεδο εκπαίδευσης των γονέων του είναι Δημοτικό, 55 άτομα Γυμνάσιο, 127 άτομα Λύκειο, 126 άτομα Πανεπιστημιακής εκπαίδευσης και τα υπόλοιπα άτομα επέλεξαν την επιλογή «άλλο», δηλαδή είναι κάτοχοι μεταπτυχιακού, κ.α. Τα στατιστικά αποτελέσματα παρατίθενται παρακάτω:

Μορφωτικό επίπεδο γονέων

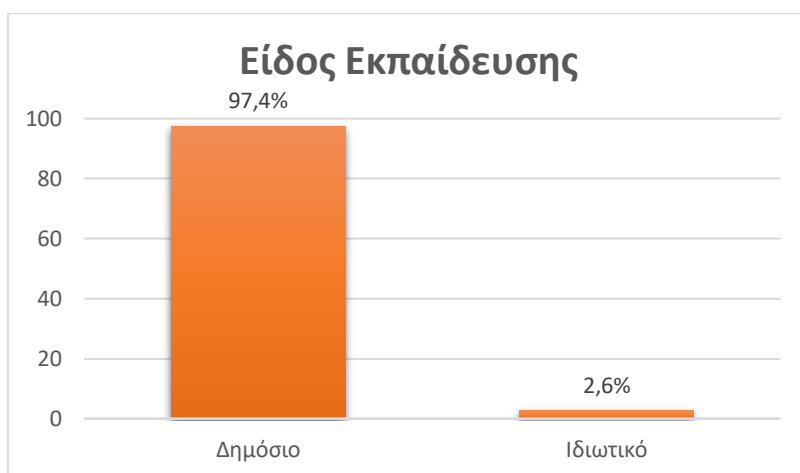


Γράφημα 3-7 Μορφωτικό επίπεδο γονέων Συμμετεχόντων

3.7.2 Ερωτήσεις Σχετικά με τις Πανελλαδικές

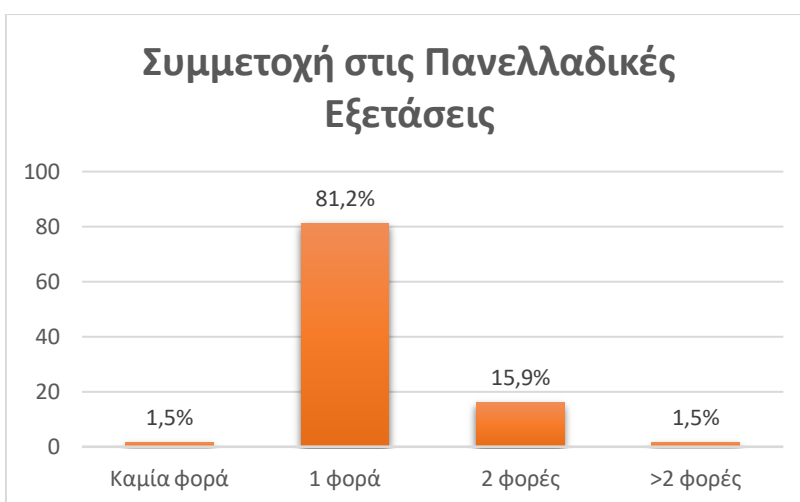
Το Δεύτερο μέρος της έρευνας περιλάμβανε ερωτήσεις σχετικά με τις Πανελλαδικές εξετάσεις. Οι πανελλαδικές εξετάσεις θεωρούνται το πλέον σημαντικό γεγονός της μαθητικής ζωής, έτσι γίνεται η περισσότερο αγχογόνα κατάσταση που η πλειοψηφία των μαθητών έχει περάσει ως τότε.

Η πρώτη ερώτηση αφορούσε το είδος της εκπαίδευσης των συμμετεχόντων, αν είχαν φοιτήσει δηλαδή σε δημόσιο ή ιδιωτικό σχολείο. Το μεγαλύτερο μέρος του δείγματος για την ακρίβεια 331 άτομα από τα 340 (97,4%) ήταν απόφοιτοι Δημόσιας εκπαίδευσης, ενώ τα υπόλοιπα 9 άτομα, δηλαδή το 2,6% είχαν πάει σε Ιδιωτικό σχολείο.



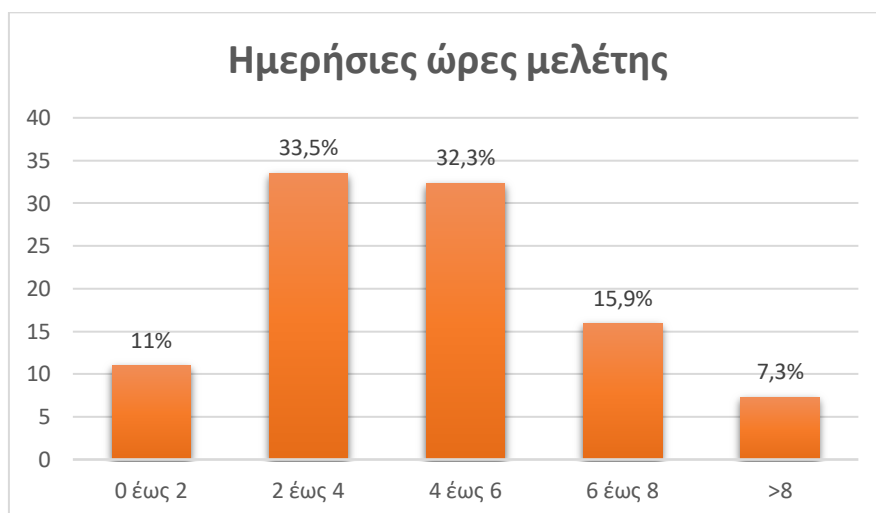
Γράφημα 3-8 Είδος εκπαίδευσης

Η επόμενη ερώτηση αφορά το πλήθος των φορών που το κάθε άτομο συμμετείχε στις πανελλαδικές εξετάσεις. Το μεγαλύτερο μέρος, δηλαδή 276 άτομα έχουν εξεταστεί μία φορά, 54 άτομα δύο φορές, 5 άτομα περισσότερες από δύο φορές ενώ καμία φορά επίσης 5 άτομα. Τα στατιστικά παρουσιάζονται στο επόμενη γράφημα.



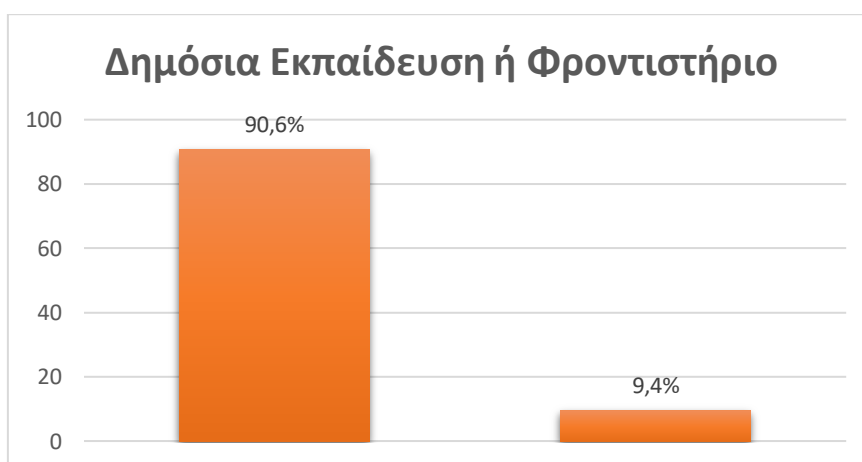
Γράφημα 3-9 Συμμετοχή στις Πανελλαδικές Εξετάσεις

Στην συνέχεια ακολούθησε ερώτηση σχετικά με τις ώρες ημερήσιας μελέτης κατά τη διάρκεια της χρονιάς που ο κάθε συμμετέχοντας έλαβε μέρος στις πανελλαδικές εξετάσεις. Παρατηρήθηκε πως η πλειονότητα των συμμετεχόντων 110 άτομα, δηλαδή το 33,5% διάβαζε από 2 έως 4 ώρες καθημερινά, ακολούθησαν το 32,3% του δείγματος όπου διάβαζε από 4 έως 6 ώρες ημερησίως, στην συνέχεια από 6 έως 8 ώρες το 15,9%, από 0 έως 2 ώρες το 11% και η μειονότητα των ατόμων διάβαζε περισσότερο από 8 ώρες δηλαδή το 7,3%.



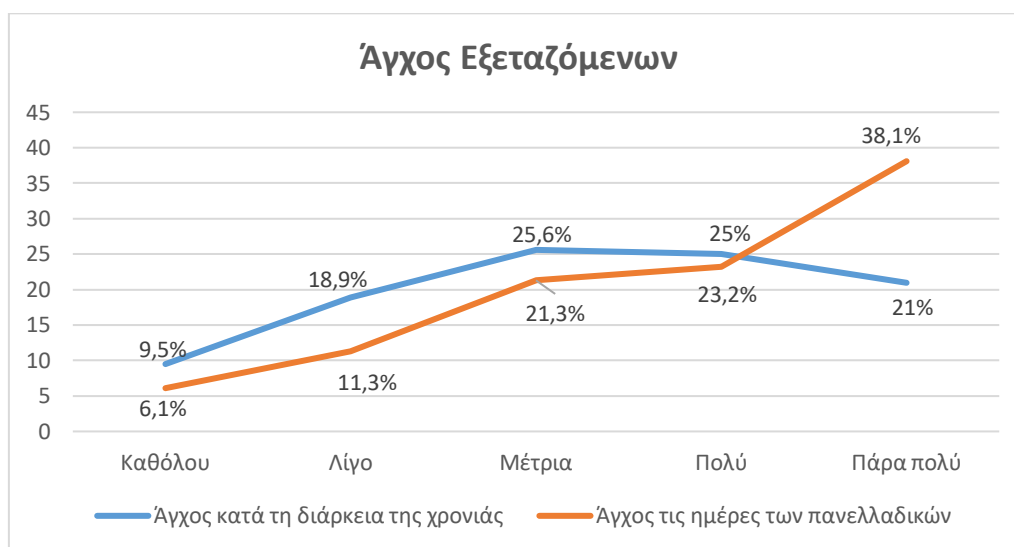
Γράφημα 3-10 Ημερήσιες Ώρες Μελέτης την χρονία των Πανελλαδικών

Η επόμενη ερώτηση αφορούσε την επιλογή των τότε μαθητών να βασιστούν στην δημόσια εκπαίδευση ή αν έκαναν φροντιστήριο και ιδιαίτερα. Από τα 340 άτομα που ερωτήθηκαν τα 308 έκανα φροντιστήριο (90,6%) ενώ μόνο 32 άτομα βασίστηκαν στην δημόσια εκπαίδευση (9,4%).



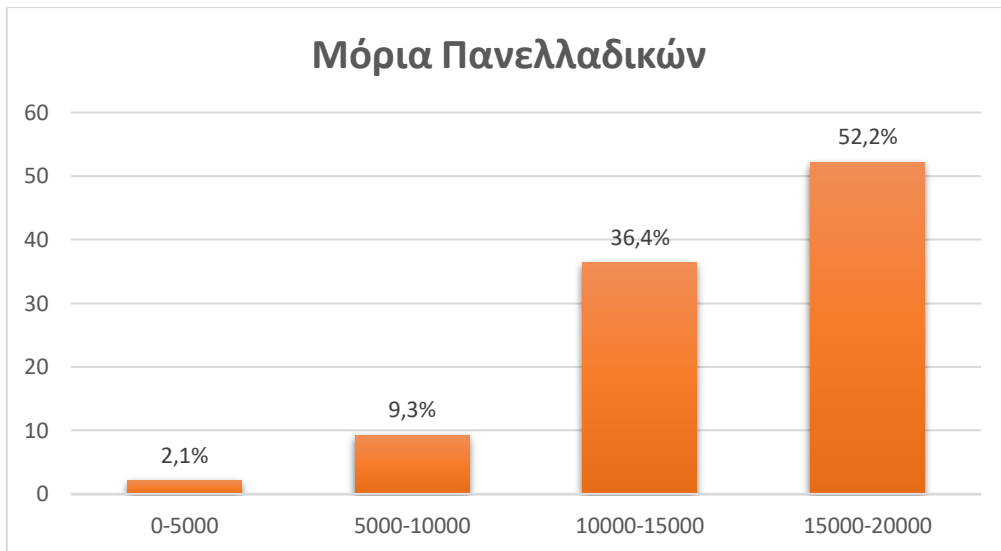
Γράφημα 3-11 Δημόσια εκπαίδευση ή Φροντιστήριο/Ιδιαίτερα μαθήματα

Ακολούθησαν δύο ερωτήσεις σχετικές με το άγχος. Το άγχος που βιώνει ο μαθητής για τα μαθηματικά σχετίζεται με τα αρνητικά αισθήματα που μπορεί να νιώθει όταν προσπαθεί να διαχειριστεί μία άσκηση ή ένα πρόβλημα, τόσο στην καθημερινότητα του όσο και στην σχολική του ζωή (Στεργιανού, 2018). Η μία από τις επόμενες ερωτήσεις αφορούσε το άγχος κατά την διάρκεια της χρονιάς που ο κάθε υποψήφιος εξετάστηκε και η άλλη το άγχος τις ημέρες των πανελλαδικών εξετάσεων. Όπως φαίνεται και στο παρακάτω διάγραμμα κατά την διάρκεια της χρονιάς το μεγαλύτερο ποσοστό είχε άγχος σε μέτριο επίπεδο, ενώ τις ημέρες των πανελλαδικών η πλειονότητα των υποψηφίων είχε πάρα πολύ άγχος.



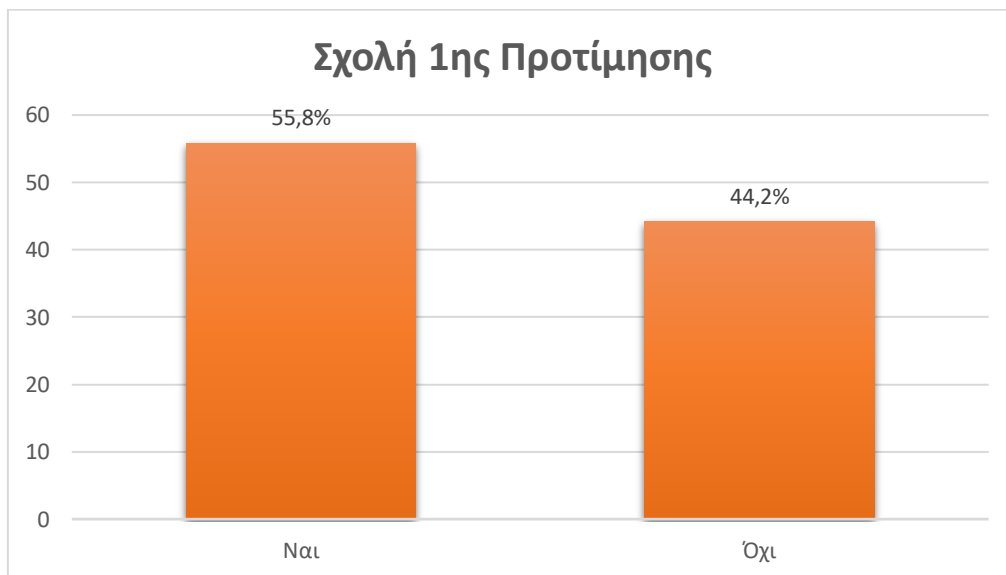
Γράφημα 3-12 Σύγκριση άγχους κατά τη διάρκεια της χρονιάς και τις ημέρες των πανελλαδικών

Στην συνέχεια παρουσιάζονται τα μόρια που έγραψε ο κάθε υποψήφιος στις πανελλαδικές εξετάσεις. Παρατηρούμε ότι η πλειονότητα του δείγματος έγραψε από 15 έως 20 χιλιάδες μόρια και η μειονότητα από 0 έως 5 χιλιάδες.



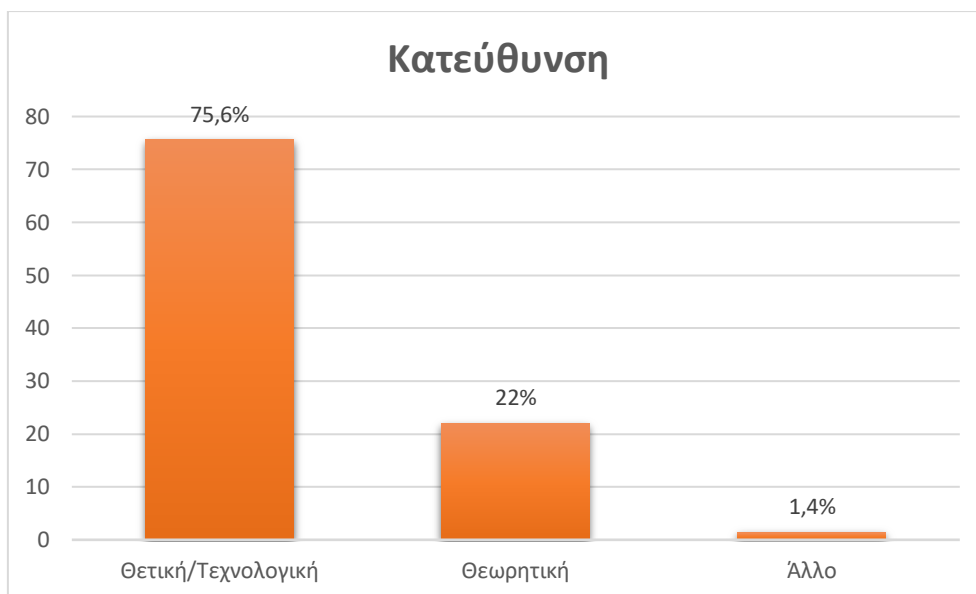
Γράφημα 3-13 Μόρια Συμμετεχόντων στις πανελλαδικές εξετάσεις

Στην ερώτηση αν ο υποψήφιος πέτυχε την σχολή της πρώτης προτίμησης του το 55,8% απάντησε θετικά ενώ το 44,2% απάντησε αρνητικά.



Γράφημα 3-14 Επιτυχία του Συμμετέχοντα σε 1η Σχολή Προτίμησης ή όχι

Τελευταία ερώτηση αυτής της ενότητας ήταν η κατεύθυνση που είχε επιλέξει ο κάθε υποψήφιος. Το μεγαλύτερο ποσοστό των συμμετεχόντων είχαν επιλέξει Θετική/Τεχνολογική κατεύθυνση (75,6%), το 22% Θεωρητική ενώ το υπόλοιπο 2,4% επέλεξε την επιλογή «άλλο» είχαν εξεταστεί με Δέσμες, ήταν απόφοιτη ΕΠΑΛ,κ.α.



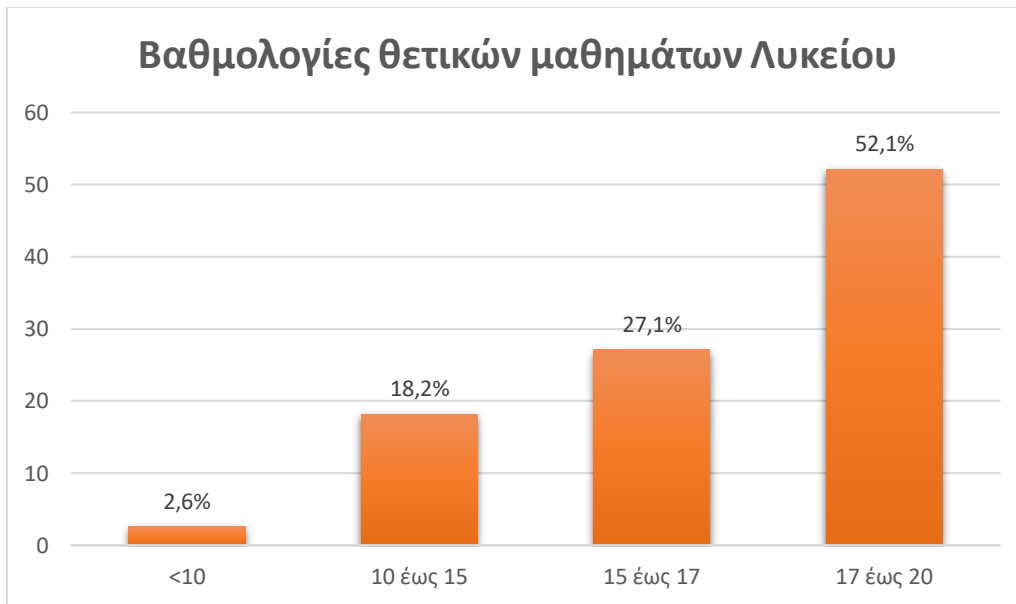
Γράφημα 3-15 Κατεύθυνση Υποψηφίων

Σε αυτό το σημείο αξίζει να σημειωθεί πως για να είναι όσο γίνεται πιο ακριβές το μοντέλο μας και δεδομένου ότι πλέον πανελλαδικώς εξετάζονται στα μαθηματικά μόνο όσα άτομα έχουν επιλέξει θετική (ή τεχνολογική) κατεύθυνση, δεν θα ληφθούν υπόψιν τα δεδομένα των ατόμων που είχαν επιλέξει θεωρητική κατεύθυνση.

3.7.3 Ερωτήσεις σχετικές με τα Μαθηματικά

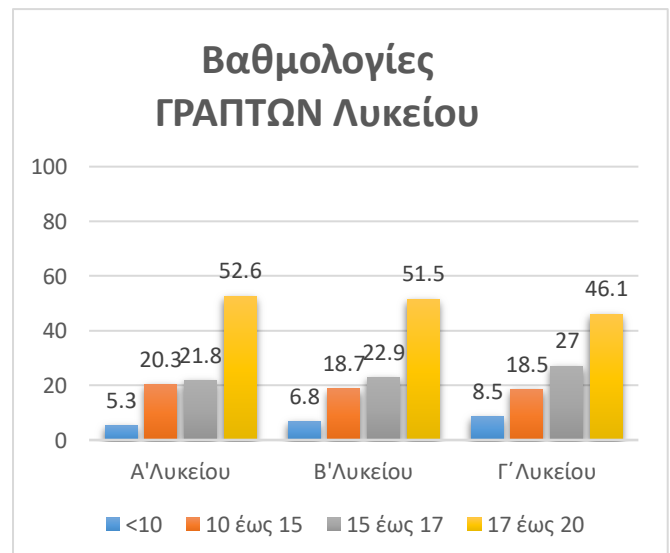
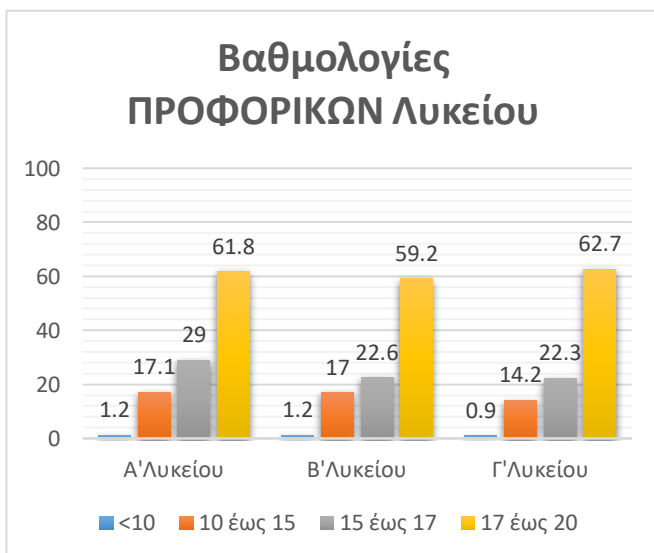
Το να μιλάει κανείς για την εκπαίδευση των μαθηματικών στην Ελλάδα μοιάζει να είναι ένα ευαίσθητο θέμα. Οι ανησυχίες γύρω από τη διδασκαλία και τις χαμηλές επιδόσεις στο μάθημα των μαθηματικών είναι παρόμοιες με εκείνες άλλων χωρών και φαίνεται να ευθυγραμμίζονται με τις διεθνείς τάσεις. Οι μαθηματικές επιδόσεις και γενικότερα η ικανότητα στα μαθηματικά έχουν γίνει αποτελεσματικοί μηχανισμοί που διέπουν την ανθρώπινη συμπεριφορά. Έτσι οι χαμηλές επιδόσεις των υποψηφίων στις πανελλαδικές εξετάσεις, ειδικά στο μάθημα των μαθηματικών, κάνει ζωτικής σημασίας την αύξηση των επιδόσεων της ελληνικής νεολαίας η οποία θα μπορούσε να είναι σημαντική για πολλούς λόγους (Valero, 2017) .

Η πρώτη ερώτηση αυτής της ενότητας αφορούσε το μέσο όρο που είχε ο κάθε συμμετέχοντας στα θετικά μαθήματα πλην των μαθηματικών, δηλαδή στην φυσική , την χημεία, τη βιολογία κατά τη διάρκεια και των τριών τάξεων του Λυκείου. Το μεγαλύτερο ποσοστό του δείγματος είχε βαθμολογία από 17 έως 20, ενώ το μικρότερο κάτω από την βάση.



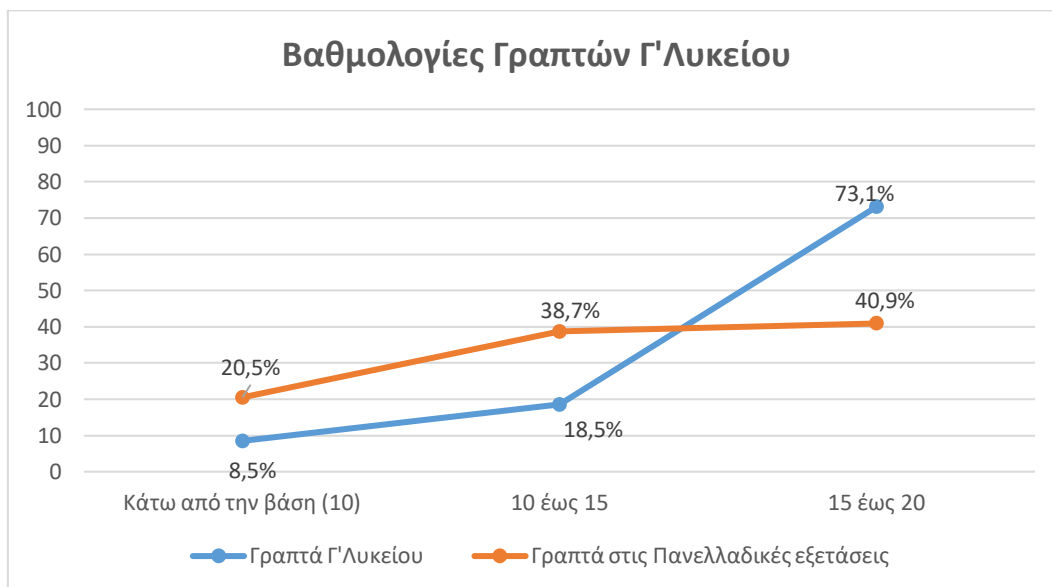
Γράφημα 3-16 Μέσος όρος βαθμολογίας θετικών μαθημάτων στις 3 τάξεις του Λυκείου

Στην συνέχεια παρουσιάζονται δύο διαγράμματα με τα ποσοστά των βαθμολογιών των συμμετεχόντων στο μάθημα των μαθηματικών. Οι ερωτώμενοι κλήθηκαν να απαντήσουν 6 ερωτήσεις σχετικές με την βαθμολογίες τους ξεχωριστά για κάθε τάξη του Λυκείου. Το αριστερό διάγραμμα περιλαμβάνει τις βαθμολογίες των τετραμήνων, δηλαδή τους προφορικούς βαθμούς όπως λέγονται και το δεξί διάγραμμα περιλαμβάνει τις βαθμολογίες των γραπτών (διαγωνίσματα, τεστ, εξετάσεις). Συγκριτικά βλέπουμε πως οι βαθμολογίες των γραπτών είναι πιο χαμηλές σε όλες τις τάξεις σε σχέση με του βαθμούς των τετραμήνων, κυρίως στην Γ' Λυκείου.



Γράφημα 3-17 Συγκριτικά Διαγράμματα προφορικών και γραπτών Λυκείου

Στην συνέχεια παρουσιάζεται ένα συγκριτικό διάγραμμα με τις βαθμολογίες των γραπτών στα μαθηματικά στην Γ' Λυκείου (ενδοσχολικά) και τη βαθμολογία στα μαθηματικά στις πανελλαδικές εξετάσεις. Παρατηρείται πως το ποσοστό των γραπτών κάτω από τη βάση είναι εμφανώς μεγαλύτερο στις πανελλαδικές εξετάσεις (πορτοκαλί γραμμή) και αντιστοίχως το ποσοστό των γραπτών που είναι από 15 και άνω στις πανελλαδικές εξετάσεις είναι σχεδόν το μισό σε σχέση με το ποσοστό των γραπτών ενδοσχολικά.



Γράφημα 3-18 Σύγκριση γραπτών πανελλαδικών και ενδοσχολικών Γ' Λυκείου

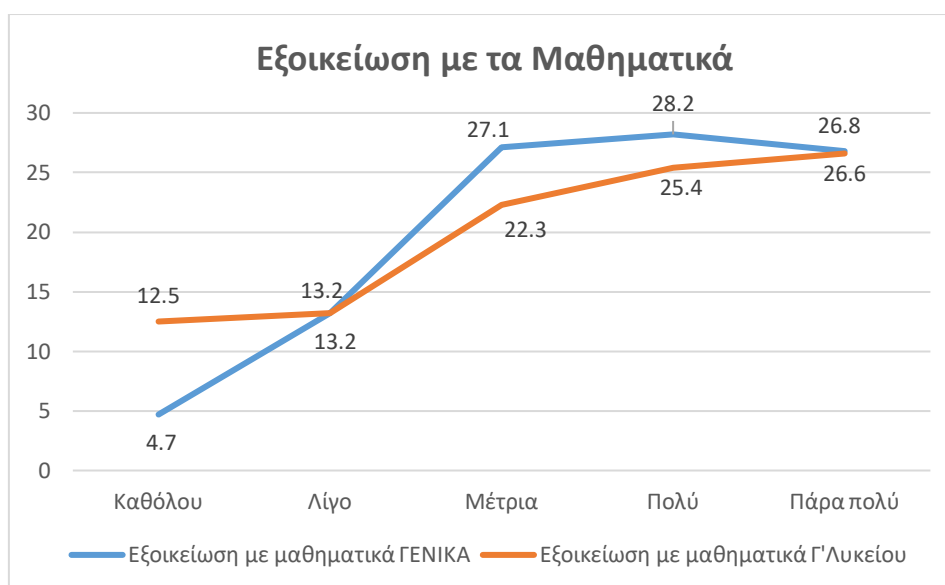
Ακολουθεί ένα διάγραμμα με τις ημερήσιες ώρες μελέτης για τα μαθηματικά. Παρατηρούμε ότι το μεγαλύτερο ποσοστό του πληθυσμού (41,4%) διάβαζε μαθηματικά από 0 έως 2 ώρες, ακολούθησε το 39,6% να μελετά από 2 έως 4 ώρες και συνεχίζεται η φθίνουσα πορεία καταλήγοντας πως κανένα άτομο δεν αφιέρωσε περισσότερες από 8 ώρες καθημερινά για να μελετήσει μαθηματικά. Σε σύγκριση με το γράφημα 3.10 παρατηρούμε πως σε σχέση με τις συνολικές ώρες μελέτης ένα μικρό ποσοστό από αυτές διατίθενται για τα μαθηματικά.



Γράφημα 3-19 Ημερήσιες ώρες μελέτης την χρονιά των Πανελλαδικών για τα Μαθηματικά

Η επόμενη ερώτηση αφορούσε το πόσο ικανοποιημένο είναι το κάθε άτομο από την επίδοση του στα μαθηματικά στις πανελλαδικές εξετάσεις. Περίπου το ίδιο ποσοστό ατόμων απάντησε θετικά και αρνητικά, καθώς το 46% των ατόμων είναι ικανοποιημένο από την βαθμολογία του, ενώ το 54% των ατόμων που συμμετείχαν στην έρευνα δεν είναι ικανοποιημένο από τη βαθμολογία του στο συγκεκριμένο μάθημα.

Οι επόμενες δύο ερωτήσεις είχαν να κάνουν με την εξοικείωση των συμμετεχόντων με τα μαθηματικά, η μία γενικά και η επόμενη συγκεκριμένα με τα μαθηματικά της Γ' Λυκείου. Παρατηρούμε ότι το ποσοστό των ατόμων που δεν είναι καθόλου εξοικειωμένα με τα μαθηματικά της Γ' Λυκείου είναι πολύ μεγαλύτερο συγκριτικά με το ποσοστό των ατόμων που δεν είναι εξοικειωμένα γενικότερα (σχεδόν τριπλάσιο). Τελικά περίπου το ίδιο ποσοστό των ατόμων νιώθουν πολύ μεγάλη εξοικείωση και με τους δύο τομείς.



Γράφημα 3-20 Συγκριτικό γράφημα εξοικείωσης ατόμων με τα μαθηματικά γενικότερα και με τα μαθηματικά της Γ' Λυκείου

Τελευταία ερώτηση της ενότητας ήταν αν χρειάστηκαν τα μαθηματικά στις προπτυχιακές σπουδές του κάθε ατόμου. Το 80,3% απάντησε θετικά, ενώ μόνο το 19,7% απάντησε αρνητικά.

3.7.4 Γενικές Ερωτήσεις

Η 4^η και τελευταία ενότητα περιλάμβανε 3 γενικές ερωτήσεις. Η πρώτη ήταν κλειστού τύπου και ζητήθηκε η άποψη του κάθε ατόμου σχετικά με τη χρησιμότητα των μαθηματικών γενικά στη ζωή των ανθρώπων. Το 98,5% των συμμετεχόντων απάντησε θετικά ενώ μόνο το 1,5% απάντησε αρνητικά.

Για λόγους πληρότητας και για να φανεί η αξία και η χρησιμότητα των μαθηματικών στην καθημερινή ζωή των ανθρώπων οι τελευταίες δύο ερωτήσεις του ερωτηματολογίου ήταν ανοιχτού τύπου και παρουσιάζονται στους επόμενους πίνακες.

Η 1^η ήταν η άποψη του καθενός σχετικά με την σύνδεση των μαθηματικών μόνο με αριθμούς, με την εκτέλεση υπολογισμών ή αν είναι κάτι περισσότερο. Η πιο συχνή απάντηση είναι ότι τα μαθηματικά συνδέονται με τον τρόπο σκέψης, τη λογική, εξυπηρετούν στην εγρήγορση του εγκεφάλου, σχετίζονται με την απλή επίλυση προβλημάτων της καθημερινότητας μας και τέλος πως συνδέονται με όλες τις υπόλοιπες επιστήμες. Στη συνέχεια παρουσιάζονται κάποιες από τις απαντήσεις των ερωτώμενων.

Θεωρείτε ότι τα μαθηματικά συνδέονται μόνο με αριθμούς , με την "εκτέλεση υπολογισμών" ή κάτι περισσότερο?

Συνδέονται με έναν διαφορετικό τρόπο θέασης και ερμηνείας της καθημερινότητας. Άλλωστε Μαθηματικά = η γλώσσα της Φύσης

Σχετίζονται με πολλές πτυχές της καθημερινής ζωής, όπως την επίλυση πρακτικών προβλημάτων. Επιπλέον διάφορα φαινόμενα μπορούν να γίνουν πιο απτά και κατανοητά με την γνώση μαθηματικών.

Συνδέονται γενικότερα με την οργάνωση και την δομή, σε μαθαίνουν να ακολουθείς βήματα για να φτάσεις στον στόχο.

Μαθαίνεις να σκέφτεσαι ,να αναλύεις προβλήματα, να εξετάζεις όλες τις παραμέτρους ώστε να φτάσεις στην επίλυση του κάθε προβλήματος.

Είναι απαραίτητα στην καθημερινή μας ζωή διότι τα μαθηματικά εδραιώνουν την τεκμηριωμένη σκέψη και λογική. Μας βοηθάνε να βρίσκουμε μοτίβα γύρω μας, να κατανοούμε τη φύση και τις ανθρώπινες σχέσεις, και μας οδηγούν στην εύρεση λύσεων.

Τα μαθηματικά εκτός από υπολογισμούς μας βοηθούν να μοντελοποιήσουμε και να καταλάβουμε τα φαινόμενα της φύσης. Επίσης συνδέονται με την προσπάθεια κατανόησης του εξωτερικού κόσμου.

Τα μαθηματικά σχετίζονται με κατασκευή δομών σε όλους τους κλάδους της ζωής ακόμη σε φιλοσοφικό επίπεδο σύμφωνα με τον μαθηματικό Γκεντελ.

Συνδέονται με οτιδήποτε σχετίζεται με την εξέλιξη του ανθρώπινου μυαλού σε πολλούς γενικούς και ειδικούς τομείς της παιδείας και της μόρφωσης.

Τα μαθηματικά συνδέονται με ό,τι κάνουμε καθημερινά, καθώς χρειαζόμαστε: την συνδυαστική σκέψη και την αίσθηση του να μην τα παρατάς όταν έχεις να λύσεις ένα πρόβλημα που δεν έχει προφανή λύση.

Σχετίζονται με την λογική, την έκφραση, την τεκμηρίωση, την μουσική, το σχέδιο, συγκροτούν τους κανόνες της αναλυτικής σκέψης και σύνθεσης. Τα Μαθηματικά είναι παντού.

Τα μαθηματικά πέρα της υπολογιστικής ικανότητας που σου προσφέρουν, σου αναπτύσσουν και καλλιεργούν ευρύτερα την πνευματική και κριτική ικανότητα. Συνδέονται άμεσα η έμμεσα στην λήψη αποφάσεων, από την λύση ποικίλων προβλημάτων θεωρητικών η μη έως και την αντιμετώπιση καταστάσεων καθημερινής φύσεως ορθολογιστικά.

Η επιστήμη των μαθηματικών είναι κάτι ιδιαίτερα βαθύτερο... Αρχικά μαθαίνει στον ανθρώπινο νου να σκέφτεται με λογική, να παρατηρεί και να εξελίσσεται.. Ένας άνθρωπος που αντιλαμβάνεται τον τρόπο σκέψης των μαθηματικών γενικότερα μπορεί να εκπαιδευτεί με προσπάθεια βέβαια σε οποιαδήποτε επιστήμη.. Μην ξεχνάμε ότι η αρχή οποιασδήποτε ανακάλυψης - εφεύρεσης οφείλεται στην παρατήρηση με οποία μορφή και αν εφαρμόζεται αυτή.

Τα μαθηματικά δε συνδέονται ιδιαίτερα με την εκτέλεση υπολογισμών. Σίγουρα έχουν σχέση με τους αριθμούς, αλλά τα μαθηματικά είναι περισσότερο ιδέες και λογική, γι' αυτό συνδέονται και με τη φιλοσοφία. Από τη στιγμή που κάποιος συνειδητοποιεί ότι η βάση των μαθηματικών δεν είναι η εκτέλεση υπολογισμών αλλά κάτι πολύ βαθύτερο και σπουδαιότερο, δε μπορεί παρά να κατανοήσει την ομορφιά τους, η οποία πηγάζει από την απόλυτη λογική και την αρμονία τους. Το σπουδαιότερο που μας προσφέρουν τα μαθηματικά είναι η ελπίδα ότι όλα τα προβλήματα έχουν μία λύση και ότι για να καταλήξουμε σε αυτή πρέπει απλά να συνδυάσουμε τις γνώσεις και τις εμπειρίες μας. Ή ότι ακόμα και αν δεν υπάρχει λύση για όλα τα προβλήματα, αξίζει να συνεχίσουμε να προσπαθούμε. Τα μαθηματικά προσφέρουν μια τελειότητα και μια βεβαιότητα τόσο ανακουφιστική σε αυτό το χαοτικό κόσμο που ζούμε.

Η τελευταία ερώτηση ήταν οι εφαρμογές των μαθηματικών στην καθημερινή ζωή των ανθρώπων και κάποιες από τις απαντήσεις εμφανίζονται παρακάτω.

Θεωρείτε ότι τα μαθηματικά έχουν εφαρμογές στην καθημερινή ζωή; Αν ναι, που;
Σε καθημερινές δραστηριότητες που συνδέονται με υπολογισμούς και χρηματικές συναλλαγές. Έχουν εφαρμογή στο 99.9% της ημέρας.
Σε όλους τους τομείς. Υγεία, εκπαίδευση, καθημερινές πρώτες ανάγκες, στις αγορές, στην αποταμίευση, στον εργασιακό χώρο.
Αθλητισμός, ψυχαγωγία, μουσική, πολιτική, ιατρική, πληροφορική, εφαρμογή νομοθεσίας. Ακόμη σε κάθε διάλογο που απαιτεί δομημένα επιχειρήματα και παραγωγή λογικών συμπερασμάτων.
Από τα ψώνια στο σούπερ μάρκετ, τις εκπτώσεις, τα τυχερά παιχνίδια, τα έπιπλα που χωράνε στο σπίτι μας μέχρι στη ψυχολογία μέσω στατιστικής. Τα μαθηματικά είναι γύρω μας.
Στις οικονομικές συναλλαγές, στον αθλητισμό, στο χώρο εργασίας (οικοδόμοι, ελαιοχρωματιστές κτλ.), στα παιχνίδια τύχης, στην έρευνα, στις εκτιμήσεις (υπολογισμών, μήκους, βάρους, όγκου, πλήθους αντικειμένων ή ατόμων κτλ.)
Σούπερ μάρκετ, ψώνια, μαγειρική, εκπτώσεις, μουσική, εκλογές, διακόσμηση & χωροθέτηση επίπλων, παρακολούθηση φορολογικών αλλαγών.
Οπτική και κατασκευή φακών, μελέτη του υποθαλάσσιου περιβάλλοντος, αρχιτεκτονική
Στην διαχείριση των οικονομικών μας, στην χρήση υπολογιστή, στην αντίληψη του χώρου μας, στα αθλήματα, στην διαχείριση του χρόνου μας. Επιπλέον από τον υπολογισμό ενός εμβαδόν μέχρι μία απλή μέθοδο των τριών.
Από τις επενδύσεις και τα δάνεια μέχρι την Αποταμίευση και τον υπολογισμό των εξόδων. Χωρίς μαθηματικά δεν μπορείς να κανείς τίποτα.
Ιατρική, φυσική, μαγειρική, εκλογές, ψυχολογία, κοινωνιολογία, στατιστική, οικονομία κ.α.
Ο, τι χρησιμοποιούμε έχει από πίσω μαθηματικά. Για παράδειγμα στις τηλεπικοινωνίες, οικοδομικές εφαρμογές, οικονομικές δραστηριότητες, ιατρικές εφαρμογές.
Χρησιμοποιούνται σε κάθε λογής υπολογισμό που καλείται να κάνει ένα άτομο. (πχ υπολογισμός χρονικών ή χιλιομετρικών αποστάσεων, αγορά αγαθών κτλ)
Από έναν απλό υπολογισμό (% έκπτωσης) έως και αμυντικό μηχανισμό απέναντι σε παραπλανητικές ειδήσεις.
Σε όλους τους τομείς από λογαριασμούς, έσοδα έξοδα μέχρι τον τρόπο που θα παρκάρεις το αμάξι. Επιπλέον χρειάζονται στο να υπολογίζουμε απλές πράξεις και αποστάσεις με το μυαλό μας (πχ την ώρα τα μέτρα/χιλιόμετρα που διανύουμε)
Ένα παράδειγμα είναι στο πώς η θεωρία αριθμών έχει εφαρμογή στην κρυπτογραφία που χρησιμοποιείται στις τράπεζες για να προστατεύουν λογαριασμούς πελατών

Χρηματιστήριο, μελέτη συμπεριφοράς βακτηριδίων μυκήτων ιών, μελέτη συμπεριφοράς οργανισμών κάτω από φαρμακευτικές αγωγές, εφαρμογή στην κατασκευή έργων με την έννοια των δρόμων, πολυκατοικιών, γεφυρών, πληροφορική, προγραμματισμός, οικονομικά, μετεωρολογία και πολλές ακόμη

Τα μαθηματικά βοηθάνε πολύ στην λήψη αποφάσεων τόσο απλών αλλά και πολύ δύσκολων. Επίσης βοηθάνε στην σωστά δομημένη σκέψη άρα και κατ' επέκταση στην γραφή και ομιλία καθώς και στην εποικοδομητική συζήτηση. Προφανώς βοηθάνε σε υπολογισμούς που μπορεί να χρειαστεί να κάνουμε και τέλος στην καλύτερη κατανόηση πολλών γεγονότων.

Εννοείται πως τα μαθηματικά έχουν εφαρμογές στην καθημερινή ζωή. Θα αποφύγω να αναφερθώ στη χρήση των μαθηματικών στις καθημερινές αγορές ενός ατόμου ή στη συμβολή τους στην ανάπτυξη της τεχνολογίας και των φυσικών επιστημών γιατί τα θεωρώ τετριμμένα. Πιστεύω πως η κατανόηση των μαθηματικών ακόμα και σε σχολικό επίπεδο μπορεί να προσφέρει στο άτομο τη δυνατότητα ανάπτυξης μιας πιο λογικής και αναλυτικής σκέψης ή οποία στη συνέχεια μπορεί να συμβάλει στην καλύτερη κατανόηση και διαχείριση καθημερινών προσωπικών προβλημάτων.

Εννοείται στο να ξέρεις πόσα ρέστα θα πάρεις ή τι θα πληρώσεις όσον αφορά τις πράξεις πολλές φορές υπολογίζουμε ακόμα και ασυναίσθητα το ποσό της έκπτωσης που γίνεται και γενικά για να κάνουμε έναν οικονομικό προϋπολογισμό. Φυσικά, τα μαθηματικά που έχουμε διδαχθεί έχουν διευρύνει τον νου οπότε έχουν βοηθήσει σημαντικά και στην διαμόρφωση της σκέψης μας και το πως επιλύουμε ορισμένα προβλήματα που παρουσιάζονται.

Τα μαθηματικά έχουν σαφώς εφαρμογές στην καθημερινή ζωή, πρωτίστως με τη μορφή υπολογισμών, οι οποίοι είναι απαραίτητοι σε όλες τις πτυχές της ζωής του σύγχρονου ανθρώπου. Εκτός όμως από αυτό, είναι γεγονός ότι τα μαθηματικά συνδέονται με φαινόμενα της φύσης αλλά και με φαινόμενα που εμφανίζονται λόγω της αλληλεπίδρασης των ανθρώπων στις σύγχρονες κοινωνίες, και επομένως μπορούν να μας βοηθήσουν να κατανοήσουμε καλύτερα τον κόσμο στον οποίο ζούμε.

4 ΜΟΝΤΕΛΟ

4.1 Το σύνολο δεδομένων

Το σύνολο δεδομένων που επιλέξαμε αφορά το ερωτηματολόγιο που αναλύθηκε στο προηγούμενο κεφάλαιο και όπου είχε τις παρακάτω ερωτήσεις:

1. Φύλο
2. Ηλικία
3. Κατοικία
4. Επίπεδο Εκπαίδευσης
5. Εργασία
6. Οικονομική Κατάσταση
7. Μορφωτικό επίπεδο γονέων (έστω ένας γονέας)
8. Είδος Εκπαίδευσης
9. Πόσες φορές έχετε δώσει πανελλαδικές εξετάσεις;
10. Πόσες ώρες διαβάζατε κατά μέσο όρο ημερησίως την χρονιά που δώσατε πανελλαδικές εξετάσεις;
11. Κάνατε φροντιστήριο ή βασιστήκατε στην δημόσια εκπαίδευση;
12. Είχατε άγχος κατά την διάρκεια της χρονιάς που δίνατε πανελλαδικές εξετάσεις;
13. Είχατε άγχος τις ημέρες των πανελλαδικών εξετάσεων;
14. Πόσα μόρια γράψατε στις πανελλαδικές εξετάσεις;
15. Πετύχατε Σχολή της Πρώτης Προτίμησης σας;
16. Θετική ή Θεωρητική κατεύθυνση;
17. Τι μέσο όρο είχατε στα θετικά μαθήματα (φυσική, χημεία, βιολογία) στις 3 τάξεις του λυκείου;
18. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Α' Λυκείου;
19. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Β' Λυκείου;
20. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Γ' Λυκείου;
21. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Α' Λυκείου;
22. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Β' Λυκείου;
23. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Γ' Λυκείου;

24. Πόσες ώρες διαβάζατε κατά μέσο όρο ημερησίως ΜΑΘΗΜΑΤΙΚΑ την χρονιά που δώσατε πανελλαδικές εξετάσεις; (μόνο όσοι εξετάστηκαν)

25. Τι βαθμολογία είχατε στα Μαθηματικά Κατεύθυνσης στις Πανελλαδικές; (μόνο όσοι εξετάστηκαν)

26. Είστε ικανοποιημένοι από την βαθμολογία σας στα μαθηματικά;

27. Πόσο εξοικειωμένοι νιώθετε με τα μαθηματικά γενικότερα;

28. Πόσο εξοικειωμένοι νιώθατε με την ύλη των μαθηματικών της Γ' Λυκείου;

29. Χρειάστηκαν τα μαθηματικά στις προπτυχιακές σπουδές σας;

30. Θεωρείτε ότι τα μαθηματικά χρειάζονται γενικά στην ζωή ενός ανθρώπου;

31. Θεωρείτε ότι τα μαθηματικά συνδέονται μόνο με αριθμούς, με την "εκτέλεση υπολογισμών" ή κάτι περισσότερο? (περιγράψτε)

32. Θεωρείτε ότι τα μαθηματικά έχουν εφαρμογές στην καθημερινή ζωή; Αν ναι, πού; (περιγράψτε)

Όπως προαναφέρθηκε οι απαντήσεις καταχωρήθηκαν σε φύλλο EXCEL, και οι ερωτήσεις που έγιναν ήταν κλειστού τύπου, γεγονός που μας βοήθησε να έχουμε συγκεκριμένες απαντήσεις, εκτός από τις 3 τελευταίες που έγιναν για λόγους πληρότητας.

Για την δημιουργία του μοντέλου μας επιλέξαμε συγκεκριμένες ερωτήσεις όπου κάθε κατηγορία ορίστηκε σαν αριθμός από το 0 έως N ανάλογα με τις N κατηγορίες που υπάρχουν σε κάθε ερώτηση. Τα δεδομένα μετασχηματίζονται έτσι ώστε να διευκολύνουν την ανακάλυψη γνώσης. Για παράδειγμα:

- στο πεδίο «φύλλο» που υπήρχαν ως απαντήσεις Γυναίκα, Άντρας δώσαμε 0,1.
- στο πεδίο «Κατοικία» δώσαμε τις αντίστοιχες τιμές για «Αγροτική» = 0, «Ημιαστική» = 1, «Αστική» = 2
- στο πεδίο «Είδος εκπαίδευσης» δόθηκαν οι αντίστοιχες τιμές «Δημόσιο σχολείο» = 0 και «Ιδιωτικό σχολείο» = 1

Οι ερωτήσεις 6, 12 και 13 έχουν ήδη ως απάντηση αριθμούς (από το 1 έως το 5) οπότε δεν χρειάστηκε να γίνει «αλλαγή».

Βασικό σημείο της προεπεξεργασίας των δεδομένων μας είναι να βρούμε εκείνα τα χαρακτηριστικά των μαθητών τα οποία θα επιλέξουμε για να φτιάξουμε το μοντέλο πρόβλεψης. Στόχος αυτής της φάσης είναι να δεδομένα που έχουμε συλλέξει να τα φέρουμε σε μια τέτοια μορφή η οποία να είναι κατάλληλη ώστε να επιτρέπει την εφαρμογή αλγορίθμων εξόρυξης γνώσης. Πρέπει λοιπόν να επιλέξουμε τις ερωτήσεις που είναι οι πιο σημαντικές ώστε ο αλγόριθμος μας να έχει όσο το δυνατόν μεγαλύτερη ακρίβεια.

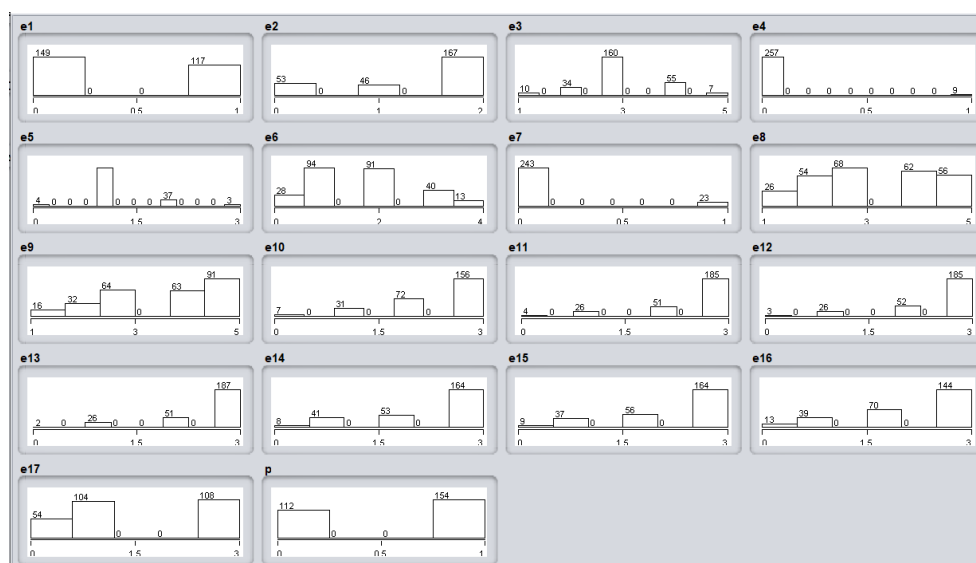
Από τις προηγούμενες ερωτήσεις επιλέχθηκαν οι παρακάτω:

Αρχική Ερώτηση	Τελική Καταχώρηση
1. Φύλλο	e1
3. Κατοικία	e2
6. Οικονομική Κατάσταση	e3
8. Είδος Εκπαίδευσης	e4
9. Πόσες φορές έχετε δώσει πανελλαδικές εξετάσεις;	e5
10. Πόσες ώρες διαβάζατε κατά μέσο όρο ημερησίως την χρονιά που δώσατε πανελλαδικές εξετάσεις;	e6
11. Κάνατε φροντιστήριο ή βασιστήκατε στην δημόσια εκπαίδευση;	e7
12. Είχατε άγχος κατά την διάρκεια της χρονιάς που δίνετε πανελλαδικές εξετάσεις;	e8
13. Είχατε άγχος τις ημέρες των πανελλαδικών εξετάσεων;	e9
17. Τι μέσο όρο είχατε στα θετικά μαθήματα (φυσική, χημεία, βιολογία) στις 3 τάξεις του λυκείου;	e10
18. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Α' Λυκείου;	e11
19. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Β' Λυκείου;	e12
20. Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Γ' Λυκείου;	e13
21. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Α' Λυκείου;	e14
22. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Β' Λυκείου;	e15
23. Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Γ' Λυκείου;	e16
25. Τι βαθμολογία είχατε στα Μαθηματικά Κατεύθυνσης στις Πανελλαδικές;	e17

Με τις παραπάνω στήλες στοχεύουμε να προβλέψουμε :

Αν τελικά θα Πετύχουμε στη Σχολή της Πρώτης Προτίμησης,

η οποία είναι αρχική ερώτηση 15 και καταχωρήθηκε τελικά ως **p**, που είναι ουσιαστικά η στήλη που θέλουμε να προβλέψουμε.



Εικόνα 4-1 Οπτικοποίηση των attributes (weka)

Τελικά το αρχείο εισόδου παίρνει την παρακάτω μορφή:

	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12
0	0	2	0	1	2	0	4	4	3	3	3
0	0	3	0	1	2	0	4	4	3	3	3
0	1	1	0	0	2	0	3	3	0	0	0
0	0	3	0	1	1	0	4	5	3	3	3
0	1	3	0	1	2	0	4	5	1	2	2
1	0	3	0	1	0	0	1	2	0	1	1
1	0	2	0	1	2	0	1	3	2	2	2
1	2	3	0	1	1	0	2	3	3	3	3
0	0	3	0	1	2	0	5	5	2	2	2
0	2	3	0	1	3	0	5	5	2	3	2
1	0	2	0	1	0	0	1	1	1	1	1
1	0	3	0	1	0	1	1	1	2	2	2
0	1	3	0	1	1	0	3	4	2	2	2
1	1	3	0	1	0	0	3	3	1	1	1
1	0	3	0	1	0	0	1	4	1	2	2
1	2	4	0	2	2	0	4	4	1	1	1
1	1	1	0	0	0	0	1	2	1	1	1
1	0	1	0	1	0	0	2	3	2	1	1
0	0	3	0	1	1	0	4	5	1	2	1
1	0	3	0	1	0	0	1	1	1	2	1
0	0	3	0	1	1	0	1	2	3	3	3
0	0	3	0	0	0	1	1	1	0	2	1
1	1	3	0	1	2	0	4	5	1	2	1

Εικόνα 4-2 Τελική Μορφή Αρχείου Εισόδου CSV

4.2 Εφαρμογή Νευρωνικών Δικτύων και ανάγκη Γενετικών Αλγορίθμων

Στόχος μας είναι να δημιουργηθεί ένα μοντέλο που με τα παραπάνω στοιχεία θα εντοπίζει αν μπορεί να πετύχει ή όχι στις εξετάσεις σε σχολή της επιλογής του υποψηφίου. Μια τέτοια μηχανή θα μπορούσε να βοηθήσει τους γονείς ή τους καθηγητές να κάνουν πρόβλεψη για το αν ένας μαθητής θα πάει καλά ή όχι στις πανελλαδικές εξετάσεις, λαμβάνοντας υπόψιν τις βαθμολογίες του (γραπτές και προφορικές) στις τρεις τάξεις του Λυκείου, την μόνιμη κατοικία του , την οικονομική του κατάσταση και το άγχος.

Για να επιτευχθεί αυτό πρέπει να εφαρμοστούν μοντέλα όπως τα νευρωνικά δίκτυα με τον πλέον κατάλληλο τρόπο, ορίζοντας τις βέλτιστες παραμέτρους. Για παράδειγμα αν ορίσουμε ένα νευρωνικό δίκτυο τριών (3) επιπέδων τότε έχουμε να δούμε παραμέτρους όπως τους νευρώνες σε κάθε επίπεδο, τις παραμέτρους learning rate και ορμή (momentum) κ.α. Για να βρούμε τις καλύτερες παραμέτρους πρέπει να κάνουμε πολλές δοκιμές που ουσιαστικά είναι άπειροι συνδυασμοί. Ένας τρόπος για να ξεπεράσουμε ένα τέτοιο πρόβλημα είναι η εφαρμογή γενετικών αλγορίθμων, οι οποίοι αναλύονται στην επόμενη ενότητα. Η ιδέα του συνδυασμού των νευρωνικών δικτύων και ενός γενετικού αλγορίθμου προέκυψε από τις δυνατότητες τους να βρίσκουν λύση εκεί που άλλες παραδοσιακές μέθοδοι αποτυγχάνουν. Από την μία πλευρά, τα ΤΝΔ μιμούνται τις λειτουργίες του ανθρώπινου εγκεφάλου και από την άλλη, οι γενετικοί αλγόριθμοι μιμούνται την διαδικασία της εξέλιξης στην φύση.

Θεωρώντας σαν χρωμοσώματα τις παραμέτρους που πειράζουμε δηλαδή τον αριθμό νευρώνων σε κάθε επίπεδο, το learning rate και το Momentum τότε μπορούμε με χρήση ενός τυχαίου αριθμού και εφαρμογή ενός γενετικού αλγορίθμου N γενεών να έχουμε το νευρωνικό που με συνεχείς μεταλλάξεις και διασταυρώσεις του πληθυσμού θα πετύχει την καλύτερη επίδοση.

4.2.1 Γενετικοί Αλγόριθμοι

Οι Γενετικοί αλγόριθμοι- Genetic Algorithms (ΓΑ) ανήκουν στο κλάδο της επιστήμης υπολογιστών και αποτελούν μια μέθοδο αναζήτησης βέλτιστων λύσεων σε συστήματα που μπορούν να περιγραφούν ως μαθηματικό πρόβλημα και έχουν ως βασική αρχή τη διατήρηση ενός πληθυσμού του προβλήματος με τη μορφή κωδικοποιημένης πληροφορίας και η εξέλιξη του πληθυσμού με την πάροδο του χρόνου.

Οι ΓΑ είναι μια οικογένεια υπολογιστικών μοντέλων εμπνευσμένο από τις αρχές τις βιολογικής εξέλιξης, συγκεκριμένα από την ιδέα της εξέλιξης μέσω γενετικής μετάλλαξης, φυσικής επιλογής (επιβίωση του ισχυρότερου) και διασταύρωσης. Αυτοί οι αλγόριθμοι κωδικοποιούν μία ενδεχόμενη λύση σε ένα συγκεκριμένο πρόβλημα σε ένα απλό χρωμόσωμα, σαν δομή δεδομένων και εφαρμόζει τελεστές ανασυνδυασμού σε αυτές τις δομές κατά τέτοιο τρόπο ώστε να διατηρεί κρίσιμες πληροφορίες.

Οι γενετικοί αλγόριθμοι θεωρούνται συχνά ως βελτιστοποιητές λειτουργιών, παρόλο που το φάσμα των προβλημάτων στα οποία έχουν εφαρμοστεί οι γενετικοί αλγόριθμοι είναι αρκετά ευρύ (Whitley, 1994). Η εκτενής χρησιμοποίηση των ΓΑ ως εργαλείο βελτιστοποίησης είναι εύκολο να δώσει σε κάποιον την εντύπωση ότι οι ΓΑ είναι αποκλειστικά αλγόριθμοι βελτιστοποίησης. Αυτό, όμως δεν ευσταθεί, διότι υπάρχουν πολλές περιπτώσεις, όπου οι ΓΑ αποτυγχάνουν να βρουν μια προφανή βέλτιστη λύση μέσα σε ένα συγκεκριμένο χώρο ψαξίματος (Λυκοθανάσης, 2001).

Η γενική λειτουργία των ΓΑ είναι πως διατηρούν έναν πληθυσμό κωδικοποιημένων πιθανών λύσεων, στη συνέχεια εξελίσσουν τον πληθυσμό εφαρμόζοντας σε αυτόν διάφορες γενετικές διαδικασίες (διαδικασίες επιλογής, αναπαραγωγής, μετάλλαξης), έπειτα δημιουργούν νέο πληθυσμό που αντικαθιστά τον προηγούμενο και τέλος επαναλαμβάνουν τη διαδικασία έως ότου «βρουν λύση».

Πιο αναλυτικά, η εφαρμογή ενός γενετικού αλγορίθμου ξεκινά με ένα πληθυσμός (συνήθως τυχαίων) χρωμοσωμάτων. Ένας στη συνέχεια αξιολογεί αυτές τις δομές και κατανέμει τις αναπαραγωγικές ευκαιρίες με τέτοιο τρόπο ώστε αυτά τα χρωμοσώματα που αντιπροσωπεύουν την καλύτερη λύση στο πρόβλημα-στόχο να δίνουν περισσότερες πιθανότητες «αναπαραγωγής» από αυτά τα χρωμοσώματα τα οποία είναι «φτωχότερες» λύσεις. Οι πιο ικανές λύσεις για ένα συγκεκριμένο πρόβλημα συνεχίζουν να εξελίσσονται και ανασυνδυάζονται τυχαία, μέχρις ότου "επιβιώσουν" οι καλύτερες. Σε κάθε γενιά δηλαδή, οι σχετικά «καλές» λύσεις αναπαράγονται, ενώ οι σχετικά «κακές» αφαιρούνται. Η «καλοσύνη» μιας λύσης ορίζεται συνήθως σε σχέση με τον τρέχοντα πληθυσμό. (Whitley, 1994). Ο διαχωρισμός και η αξιολόγηση των διαφόρων λύσεων γίνεται με την βοήθεια μιας αντικειμενικής συνάρτησης ή συνάρτησης ικανότητας (objective ή fitness function), η οποία παίζει το ρόλο του περιβάλλοντος μέσα στο οποίο εξελίσσεται ο πληθυσμός.

Η ικανότητα αναζήτησης των γενετικών αλγορίθμων αξιοποιείται για την αναζήτηση κατάλληλων κέντρων των συστάδων στον χώρο χαρακτηριστικών έτσι ώστε να βελτιστοποιείται μια μέτρηση ομοιότητας των συστάδων που προκύπτουν. Τα χρωμοσώματα, τα οποία αντιπροσωπεύονται ως συμβολοσειρές (strings) πραγματικών αριθμών, κωδικοποιούν τα κέντρα ενός σταθερού αριθμού συστάδων (Maulik, Ujjwal, and Sanghamitra Bandyopadhyay, 2000). Όπως και στη βιολογία η συμβολοσειρά συνήθως

αναφέρεται ως χρωμόσωμα (chromosome) ενώ τα επιμέρους τμήματα της που κωδικοποιούν κάποιο χαρακτηριστικό ονομάζονται γονίδια (genes).

Στον επόμενο πίνακα φαίνεται η αναλογία του βιολογικού προτύπου με το γενετικό αλγόριθμο:

Βιολογικό πρότυπο	Γενετικοί αλγόριθμοι
Γονίδιο	Παράμετρος (μεταβλητή)
Χρωμόσωμα	Υποψήφια λύση (συμβολοσειρά)
Πληθυσμός	Σύνολο υποψήφιας λύσεων
Πιθανότητα επιβίωσης	Συνάρτηση βελτιστοποίησης
Μετάλλαξη	Τυχαία αναζήτηση λύσης
Διασταύρωση	Σύνθεση λύσεων
Γενιά	Ανακύκλωση

Εικόνα 4-3 Βιολογικό πρότυπο και γενετικοί αλγόριθμοι (Γεωργιλιάκης, 2004)

Στην πραγματική βιολογία, η γενετική εφαρμόζεται σε πληθυσμούς χρωμοσωμάτων που αντιπροσωπεύουν πληθυσμούς οργανισμών οι οποίοι ανταγωνίζονται για να επιβιώσουν. Ανάλογα, στην τεχνητή βιολογία οι γενετικοί τελεστές⁶ εφαρμόζονται σε πληθυσμούς συμβολοσειρών που αντιπροσωπεύουν πληθυσμούς λύσεων οι οποίες επιλέγονται για να ικανοποιήσουν τους σκοπούς του χρήστη.

Τα πλεονεκτήματα των ΓΑ είναι αρκετά, έτσι η χρήση τους σε διάφορες εφαρμογές είναι ελκυστική. Οι πιο σημαντικοί είναι ότι μπορούν να επιλύουν δύσκολα προβλήματα γρήγορα και αξιόπιστα, μπορούν να συνεργαστούν με τα υπάρχοντα μοντέλα και συστήματα, είναι εύκολα επεκτάσιμοι και εξελίξιμοι, μπορούν να συμμετέχουν σε υβριδικές μορφές με άλλες μεθόδους, δεν απαιτούν περιορισμούς στις συναρτήσεις που επεξεργάζονται κ.α. Γενικά οι ΓΑ είναι μέθοδοι που κάνουν ταυτόχρονα εξερεύνηση του χώρου αναζήτησης και εκμετάλλευση της ήδη επεξεργασμένης πληροφορίας (exploration and exploitation).

Με πιο απλά λόγια οι γενετικοί αλγόριθμοι με τυχαίο τρόπο βρίσκουν τη βέλτιστη λύση μέσα από μία σειρά λύσεων. Στην περίπτωση μας βρίσκουν δηλαδή τον καλύτερο συνδυασμό των παραμέτρων του νευρωνικού δικτύου μας.

⁶ Οι γενετικοί τελεστές περιλαμβάνουν το στάδιο της Επιλογής, όπου επιλέγει με κάποιο τρόπο τα πιο κατάλληλα μέλη του πληθυσμού και τα μεταφέρει στο νέο πληθυσμό, την Διασταύρωση όπου συνδυάζει τα στοιχεία δύο χρωμοσωμάτων γονέων ώστε να δημιουργηθούν δύο νέοι απόγονοι κάνοντας ανταλλαγή τα αντίστοιχα κομμάτια από τους γονείς και τέλος την Μετάλλαξη όπου αλλάζει αυθαίρετα ένα ή περισσότερα γονίδια ενός συγκεκριμένου χρωμοσώματος

4.3 Αλγόριθμος

Με βάση τα παραπάνω δημιουργήθηκε κώδικας σε python που είναι ο παρακάτω:

```
import random
import pandas
from sklearn import tree
#import pydotplus
from sklearn.neural_network import MLPClassifier
import matplotlib.pyplot as plt
import matplotlib.image as pltimg

#synartisi pou vgazei tin apodosi tou nevronikou
def MLP(X,y, lr,mom,lev1,lev2,lev3):

    #to synolo trainer
    Xtrain=X[1:200]
    #to synolo test
    Ytrain=y[1:200]

    #montelo nevronikou
    clf = MLPClassifier(hidden_layer_sizes=(lev1,lev2,lev3),momentum=mom,
learning_rate_init=lr,random_state=1, max_iter=300).fit(X, y)
    #apodosi nevronikou
    clf = clf.fit(Xtrain, Ytrain)

    #apotelesma tou score
    Xtest=X[200:266]
    Ytest=y[200:266]
    return (clf.score(Xtest,Ytest))

#to kyrio programma
#diavazoume to arxeio
df = pandas.read_csv("data.csv")

#epilegoume tis stiles pou theloume gia to motnelo mas
features =
['e1','e2','e3','e4','e5','e6','e7','e8','e9','e10','e11','e12','e13','e14','e14','e15','e16','e17']

#pairnoume synolo pou mpainei eisodo sto montelo
X= df[features]
#pairnoume tin stili pou theloume na provlepsoume
y = df['p']

#pinakas me ta xromosomata tou genetiku
crom=[]

#apodoseis tou genetiku
ev=[]
```

```

#genies
MC=50

#plithismos genneas
N=100

#synaritisi diastavrosis
pcross=0.9
#synartisi metallaxis
pmul=0.9

print("start")

#gia kathe gennea
for i in range(0,MC):

    #parinoume tixaies times stis metavlites tou nevronikou
    lr=random.random()
    lev1=random.randrange(2,100)
    lev2=random.randrange(2,100)
    lev3=random.randrange(2,100)
    mom=random.random()

    #pairnoume ta xromosomata
    c=[lr, mom ,lev1, lev2, lev3 ]

    #pairnoume to score
    score=MLP(X,y, lr,mom,lev1,lev2,lev3)

    #ta vazoume stous antistoixous pinakess
    crom.append(c)
    ev.append(score)
    print (i)

#trexoume ton gennetiko
for i in range (0,N):
    print("Gen: ",i)
    ev,crom=zip(*sorted(zip(ev,crom),reverse=True))
    #taxinomoume me vasi tin apodosi tou nevronikou
    ev=list(ev)
    crom=list(crom)

    #kanoume diastavrosi
    if(random.random())<pcross):
        i1=random.randrange(0,int(MC/2))
        i2=random.randrange(0,int(MC/2))

        lr=(crom[i1][0]+crom[i2][0])/2

```

```
mom=(crom[i1][1]+crom[i2][1])/2
lev1=int((crom[i1][2]+crom[i2][2])/2)
lev2=int((crom[i1][3]+crom[i2][3])/2)
lev3=int((crom[i1][4]+crom[i2][4])/2)
c=[lr, mom ,lev1, lev2, lev3 ]

for j in range (0,5):
    crom[MC-1][j]=c[j]
score=MLP(X,y, lr,mom,lev1,lev2,lev3)
ev[MC-1]=score
#kanoume metallaxi
if(random.random())< pmul):
    i1=random.randrange(0,int(MC/2))
    i2=random.randrange(0,5)
    k=[]
    k.append(random.random())
    k.append(random.random())
    k.append(random.randrange(2,100))
    k.append(random.randrange(2,100))
    k.append(random.randrange(2,100))

#topothetoume ta paidia sto plithismo
crom[i1][i2]=k[i2]
i2=random.randrange(0,5)
crom[i1][i2]=k[i2]
score=MLP(X,y, crom[i1][0],crom[i1][1],crom[i1][2],crom[i1][3],crom[i1][4])
ev[MC-1]=score

#ektiposi genneon kai epidosewn
print(crom)
print(ev)

#kalytero nevroniko kai apodosi
print("Best Solution:", crom[0])
print("Evaluate Solution:", ev[0])
```

4.4 Βασικά σημεία Αλγορίθμου

4.4.1 Η συνάρτηση του νευρωνικού δικτύου

```
#synartisi pou vgazei tin apodosi tou nevronikou
def MLP(X,y, lr,mom,lev1,lev2,lev3):

    #to synolo trainer
    Xtrain=X[1:200]

    #to synolo test
    Ytrain=y[1:200]

    #montelo nevronikou
    clf = MLPClassifier(hidden_layer_sizes=(lev1,lev2,lev3),momentum=mom,
learning_rate_init=lr,random_state=1, max_iter=300).fit(X, y)

    #aposdosi nevronikou
    clf = clf.fit(Xtrain, Ytrain)

    #apotelesma tou score
    Xtest=X[200:266]
    Ytest=y[200:266]
    return (clf.score(Xtest,Ytest))
```

Όπως βλέπουμε χρησιμοποιεί τις 500 εγγραφές για σύνολο εκπαίδευσης (train set) και 100 για σύνολο ελέγχου (test set) και δίνει σαν αποτέλεσμα το score του νευρωνικού δηλαδή το ποσοστό των σωστών προβλέψεων. Το Training set αποτελεί το μεγαλύτερο υποσύνολο των δεδομένων, καθώς αυτό χρησιμοποιείται για την εκπαίδευση του δικτύου, ενώ το Testing είναι το υποσύνολο το οποίο δεν λαμβάνει μέρος στην εκπαίδευση, αντιθέτως παραμένει ανεξάρτητο για τη μετέπειτα σύγκριση με τα δεδομένα εξόδου που προβλέπονται από το νευρωνικό.

Στόχος μας είναι να προβλέψουμε την στήλη p, δηλαδή αν ο υποψήφιος με την συμμετοχή του στις Πανελλαδικές εξετάσεις, πέτυχε ή όχι την πρώτη επιλογή Σχολής (Πανεπιστημίου/ΤΕΙ) της προτίμησης του.

4.4.2 Διάβασμα αρχείου και καταχώρηση σε λίστες

```
#to kyrio programma
#diavazoume to arxeio
df = pandas.read_csv("data.csv")

#epilegoume tis stiles pou theloume gia to motnelo mas
features =
['e1','e2','e3','e4','e5','e6','e7','e8','e9','e10','e11','e12','e13','e14','e14','e15','e16','e17']

#pairnoume synolo pou mpainei eisodo sto montelo
X= df[features]

#pairnoume tin stili pou theloume na provlepsoume
y = df['p']
```

Όπως βλέπουμε διαβάζουμε το αρχείο με χρήση της βιβλιοθήκης pandas και παίρνουμε τα σύνολα που θέλουμε. Η βιβλιοθήκη Pandas προσφέρει ένα σύνολο εργαλείων για την διαχείριση και επεξεργασία δομημένων συνόλων δεδομένων και αποτελεί ένα ισχυρό συμπλήρωμα της στοίβας των εργαλείων που προσφέρει η Python. Στην περίπτωση μας ουσιαστικά το X και y έχει τις τιμές από το αρχείο μας για τα αντίστοιχα πεδία.

4.4.3 Δημιουργία αρχικού συνόλου χρωμοσωμάτων

Θεωρήσαμε σαν χρωμοσώματα τις παραμέτρους που πειράζουμε δηλαδή τον αριθμό νευρώνων σε κάθε επίπεδο, ο ρυθμός εκμάθησης (learning rate) και ο συντελεστής ορμής (Momentum) , με την χρήση ενός τυχαίου αριθμού και εφαρμογή ενός γενετικού αλγορίθμου N γενεών να έχουμε το νευρωνικό που με συνεχείς μεταλλάξεις και διασταυρώσεις του πληθυσμού θα πετύχει την καλύτερη επίδοση. Με τον παρακάτω κώδικα δημιουργούμε τα χρωμοσώματα δηλαδή ορίζουμε γενιές MC , τιμές για το learning rate, όπου είναι ιδιαίτερα σημαντικό για την αποτελεσματικότητα της διαδικασίας εκπαίδευσης του δικτύου, το Momentum το οποίο είναι μια ιδιαίτερα σημαντική παράμετρος της εκπαίδευσης , Nodes στο επίπεδο 1, Nodes στο επίπεδο 2, Nodes στο επίπεδο 3 του νευρωνικού μας.

Οι λίστες crom και en είναι οι λίστες με τις τιμές των χρωμοσωμάτων του γενετικού και οι αποδόσεις της εφαρμογής του νευρωνικού για τις αντίστοιχες τιμές των χρωμοσωμάτων, αντίστοιχα. Επίσης το N είναι οι γενιές που θα εξετάσουμε, pcross η πιθανότητα διασταύρωσης, pmul η πιθανότητα μετάλλαξης.

```
#pinakas me ta xromosomata tou genetikou
crom=[]

#apodoseis tou genetikou
ev=[]

#genies
MC=50

#plithismos genneas
N=100

#synaritisi diastavrosis
pcross=0.9
#synartisi metallaxis
pmul=0.9

print("start")

#gia kathe gennea
for i in range(0,MC):

    #parinoume tixaies times stis metavlites tou nevronikou
    lr=random.random()
    lev1=random.randrange(2,100)
    lev2=random.randrange(2,100)
    lev3=random.randrange(2,100)
    mom=random.random()
```



```
#pairnouve ta xromosomata
c=[lr, mom ,lev1, lev2, lev3 ]

#pairnouve to score
score=MLP(X,y, lr,mom,lev1,lev2,lev3)

#ta vazoume stous antistoixous pinakess
crom.append(c)
ev.append(score)

print (i)
```

4.4.4 Γενετικός Αλγόριθμος

Στην συνέχεια εφαρμόζουμε το γενετικό αλγόριθμο:

```
#trexoume ton genetiko
for i in range (0,N):
    print("Gen: ",i)
    ev,crom=zip(*sorted(zip(ev,crom),reverse=True))
    #taxinomoume me vasi tin apodosi tou nevronikou
    ev=list(ev)
    crom=list(crom)

    #kanoume diastavrosi
    if(random.random())<pcross):
        i1=random.randrange(0,int(MC/2))
        i2=random.randrange(0,int(MC/2))

        lr=(crom[i1][0]+crom[i2][0])/2
        mom=(crom[i1][1]+crom[i2][1])/2
        lev1=int((crom[i1][2]+crom[i2][2])/2)
        lev2=int((crom[i1][3]+crom[i2][3])/2)
        lev3=int((crom[i1][4]+crom[i2][4])/2)
        c=[lr, mom ,lev1, lev2, lev3 ]

    for j in range (0,5):
        crom[MC-1][j]=c[j]
        score=MLP(X,y, lr,mom,lev1,lev2,lev3)
        ev[MC-1]=score
    #kanoume metallaxi
    if(random.random())< pmul):
        i1=random.randrange(0,int(MC/2))
        i2=random.randrange(0,5)
```

```
k=[]
k.append(random.random())
k.append(random.random())
k.append(random.randrange(2,100))
k.append(random.randrange(2,100))
k.append(random.randrange(2,100))

#topothetoume ta paidia sto plithismo
crom[i1][i2]=k[i2]
i2=random.randrange(0,5)
crom[i1][i2]=k[i2]
score=MLP(X,y, crom[i1][0],crom[i1][1],crom[i1][2],crom[i1][3],crom[i1][4])
ev[MC-1]=score

#ektiposi genneon kai epidosewn
print(crom)
print(ev)

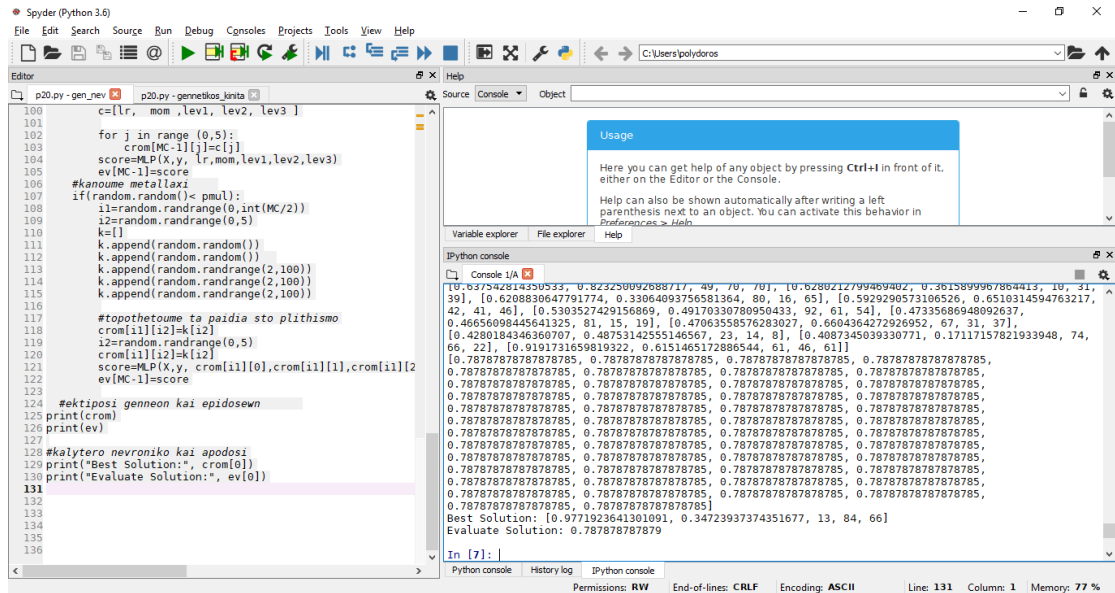
#kalytero nevroniko kai apodosi
print("Best Solution:", crom[0])
print("Evaluate Solution:", ev[0])
```

Όπως βλέπουμε όταν έχουμε διασταύρωση γίνεται για κάθε χρωμόσωμα του παιδιού στα αντίστοιχα στοιχεία του ο μέσος όρος των δύο γονέων. Στη μετάλλαξη έχουμε 2 στοιχεία ενός χρωμοσώματος να αλλάζουν τυχαία.

Για κάθε στοιχείο εφαρμόζεται το νευρωνικό και παίρνουμε την νέα τιμή του score δηλαδή την πιθανότητα να έχουμε σωστή πρόβλεψη.

4.5 Εκτέλεση του προγράμματος

Η εκτέλεση του προγράμματος μας έγινε μέσω του Spyder, ενός ολοκληρωμένου περιβάλλοντος ανάπτυξης ανοιχτού κώδικα για επιστημονικό προγραμματισμό στη γλώσσα Python. Μια εκτέλεση του προγράμματος είναι η παρακάτω:



```
100 c=[r, mom ,lev1, lev2, lev3 ]
101
102 for j in range (0,5):
103     crom[MC-1][j]=c[j]
104     score=MLP(X,y, [r,mom,lev1,lev2,lev3])
105     ev[MC-1]=score
106     #kanoyme meta1lav1
107     if(random.random())< pmu1:
108         i1=random.randrange(0,int(MC/2))
109         i2=random.randrange(0,5)
110         k=i1
111         k.append(random.random())
112         k.append(random.random())
113         k.append(random.randrange(2,100))
114         k.append(random.randrange(2,100))
115         k.append(random.randrange(2,100))
116
117 #topothetoume ta paidia sto plithismo
118 crom[i1][i2]=k[i2]
119 i2=random.randrange(0,5)
120 crom[i1][i2]=k[i2]
121 score=MLP(X,y, crom[i1][0],crom[i1][1],crom[i1][2]
122     ev[MC-1]=score
123
124 #ektiposi geneon kai epidosewn
125 print(crom)
126 print(ev)
127
128 #kalyttero nevroniko kai aposodi
129 print("Best Solution:", crom[0])
130 print("Evaluate Solution:", ev[0])
131
132
133
134
135
136
```

Console Output:

```
Best Solution: [0.971922641301091, 0.34723937374351677, 13, 84, 66]
Evaluate Solution: 0.787878787879
```

Εικόνα 4-4 Τελική εκτέλεση προγράμματος

Όπως βλέπουμε έχουμε την καλύτερη απόδοση κοντά στο 80% (0,78) με τιμές στο νευρωνικό 0.97, 0.34, 13, 84,66. Δεδομένου ότι ο γενετικός αλγόριθμος είναι στοχαστικός , βάζει τυχαίες τιμές και με βάση αυτές βγάζει το αποτέλεσμα , έτσι κάθε φορά θα είναι διαφορετικά. Φυσικά αυξάνοντας τον πληθυσμό και τις γενεές πιθανά να έχουμε και καλύτερα αποτελέσματα αλλά χρειάζονται πολύ ισχυροί υπολογιστικοί πόροι και μεγάλος χρόνος αναμονής.

Το μειονέκτημα των Τεχνητών Νευρωνικών Δικτύων είναι πως η τελική λύση εξαρτάται από τις αρχικές συνθήκες του δικτύου και πως οι διαδικασίες εκμάθησης και επικύρωσης φαίνεται να εξαρτώνται πλήρως από τα αρχικά Βάρη. Έτσι, το τελικό αποτέλεσμα είναι σε μικρό βαθμό ένα “black box”.

Όπως προαναφέρθηκε για κάθε στοιχείο εφαρμόζεται το νευρωνικό και παίρνουμε την νέα τιμή του score δηλαδή την πιθανότητα να έχουμε σωστή πρόβλεψη. Το ποσοστό της απόδοσης μας δείχνει πως η κατοικία , ο βαθμός των διαγωνισμάτων και τον βαθμών τον τετραμήνων στο Λύκειο , το άγχος κατά την διάρκεια της χρονιάς και την ημέρα των εξετάσεων επηρεάζουν την επιτυχία στις πανελλαδικές εξετάσεις.

5 ΕΠΙΛΟΓΟΣ

5.1 Συμπεράσματα

Όπως είδαμε στα προηγούμενα κεφάλαια, μπορούμε έως ένα ποσοστό να προβλέψουμε το μέλλον μιας ή και περισσότερων μελλοντικών τιμών. Δηλαδή, μπορούμε με τιμές από το παρελθόν, συγκεκριμένα συλλέγοντας βαθμολογίες ατόμων που έχουν εξεταστεί πανελλαδικά στο παρελθόν να προβλέψουμε βαθμολογίες μαθητών στο μέλλον. Πολύ σημαντικό ρόλο, βέβαια κατέχει το μοντέλο πρόβλεψης που θα χρησιμοποιήσουμε καθώς δεν υπάρχει πανάκεια. Δεν υπάρχει μία μέθοδος η οποία να είναι αποτελεσματική σε κάθε πρόβλημα.

Η εκπαίδευση αποτελεί ένα πολύ κρίσιμο στοιχείο της κοινωνίας μας. Οι τεχνικές εξόρυξης γνώσης επιτρέπουν μία εξαγωγή γνώσης από τα δεδομένα, προσφέροντας μας έτσι ενδιαφέρων προοπτικές και στον τομέα της εκπαίδευσης. Σε αυτήν την εργασία έγινε μία προσπάθεια να προβλέψουμε την επιτυχία των μαθητών στις πανελλαδικές εξετάσεις λαμβάνοντας υπόψιν κατά κύριο λόγο την βαθμολογία του στο μάθημα των μαθηματικών .

Ο αλγόριθμος που χρησιμοποιήσαμε ανήκει στην κατηγορία της κατηγοριοποίησης και είχε ακρίβεια πρόγνωσης περίπου 80%. Έτσι με την βοήθεια των τεχνικών μηχανικής μάθησης και την χρήση ενός προγνωστικού εργαλείου σαν αυτό που κατασκευάσαμε στα πλαίσια αυτής της εργασίας οι εκπαιδευτικοί , μαθητές και αλλά και γονείς είναι σε θέση να εντοπίσουν εκείνους τους μαθητές, οι οποίοι δεν θα ανταποκριθούν στις τελικές εξετάσεις να γνωρίζουν τις μελλοντικές επιδόσεις τους.

5.2 Μελλοντικές Επεκτάσεις και Βελτιώσεις

Οι πανελλαδικές εξετάσεις αποτελούν κομβικό σημείο στην ζωή του ανθρώπου. Η δημιουργία ενός ιδανικού μοντέλου πρόβλεψης βαθμολογιών θα μπορούσε να βοηθήσει σημαντικά. Όσο περισσότερα είναι τα δεδομένα, τόσο μεγαλύτερη ακρίβεια θα έχει η πρόβλεψη. Σε αυτό θα μπορούσε να βοηθήσει η ηλεκτρονική καταγραφή των επιδόσεων ενός μαθητή κατά την διάρκεια της σχολικής του ζωής.

Για παράδειγμα θα μπορούσε να δημιουργηθεί μία ηλεκτρονική καρτέλα ενός μαθητή που να περιλαμβάνει τις βαθμολογίες των διαγωνισμάτων, των τεστ, αξιολογήσεις λυμένων ασκήσεων, βαθμολογίες τριμήνων/τετραμήνων από το δημοτικό σχολείο έως και την Γ' Λυκείου. Έτσι μέσα σε λίγα χρόνια και δεδομένου των χιλιάδων μαθητών που φοιτούν κάθε χρόνο σε δημόσια και ιδιωτικά σχολεία θα έχουμε ένα σύνολο δεδομένων με εκατομμύρια δεδομένα που θα μπορούν να μας δώσουν ένα μοντέλο με μεγάλη ακρίβεια πρόβλεψης.

6 Βιβλιογραφία

- Ashish Dutt, Maizatul Akmar Ismail & Tutut Herawan. (2016). A systematic Review on Educational Data Mining. *IEEE Access*.
- Cristobal Romero, Sebastian Ventura. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.
- David Hand, Heikki Mannila, Padhraic Smyth. (2001). *Principles of Data Mining*.
- I. Aleksander and H. Morton. (1990). An Introduction to Neural Computing.
- Joost Verbraeken, Matthijs Wolting, Jonathan Katzy , Jeroen Kloppenburg, Tim Verbelen, Jan Rellermeyer. (2020). A Survey on Distributed Machine Learning. *ACM Computing Surveys, Vol. 53, No. 2, Article 30*.
- Kenneth Koedinger, Ryan Baker, Kyle Cunningham, Alida Skogsholm , Brett Leber, John Stamper. (2010). A Data Repository for the EDM community: The PSLC DataShop. *Handbook of educational data mining*.
- Kotsiantis, S. B. (2011). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades.
- Larose, D. (2006). *Data mining methods and models*.
- Lawrence, B. (2017, Νοέμβριος 5). Machine Learning — What it is and why it should interest you. *Towards data science*.
- Maulik, Ujjwal, and Sanghamitra Bandyopadhyay. (2000). Genetic algorithm-based clustering technique. *Pattern recognition*.
- Mitchell, T. (1997). *Machine Learning*.
- Mohammad Saeid Mahdavejad, Mohammadreza Rezvan , Mohammadamin Barekatin , Peyman Adibi , Payam Barnaghi , Amit P. Sheth. (2018). Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks*.
- Nehra, M. (2019, Μάρτιος 13). AL (Artificial Intelligence) VS Machine Learning VS Deep Learning). *Irish tech news*.
- Osmar Zaïane, Jun Luo. (2001). Towards Evaluating Learners' Behaviour in a Web-Based Distance Learning Environment. *Proceedings IEEE International Conference on Advanced Learning Technologies*.
- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. (2018). *Introduction to Data Mining*.
- Pena-Ayala, A. (2013). Educational data mining: A survey and a data mining-based analysis of recent works. *Elsevier*.
- Simon, P. (2013). *Too Big to Ignore: The Business Case for Big Data*.

- Toon Calders, Mykola Pechenizkiy. (2012). Introduction to The Special Section on Educational Data Mining.
- Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*.
- Valero, P. (2017). Ένταξη και αποκλεισμός στην Μαθηματική εκπαίδευση και "κατασκευή" του σύγχρονου πολίτη.
- Vidushi Sharma , Sachin Rai , Anurag Dev. (2012). A Comprehensive Study of Artificial Neural Networks. *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*.
- Βασίλειος Βερούκιος, Βασίλειος Καγκλής , Ηλίας Σταυρόπουλος. (2015). *Η επιστήμη των δεδομένων μέσα απο τη γλώσσα R*. Κάλλιπος.
- Γεωργιλιάκης. (2004). Εφαρμογή Γενετικών Αλγορίθμων στην Παραγωγή Ηλεκτρικής Ενέργειας.
- Γεωργούλη, Α. (2015). *Μηχανική Μάθηση*. Κάλλιπος.
- Δημητριάδης, Σ. (2015). *Θεωρίες Μάθησης & Εκπαίδευσης Λογισμικού*.
- Διαμαντάρας, Κ. (2007). *Τεχνητά Νευρωνικά Δίκτυα* . Κλειδάριθμος.
- Θεοδόση-Κόκκινου, Λ. (2013, Ιούνιος). Τεχνητά Νευρωνικά Δίκτυα και εφαρμογές στα Συστήματα Αυτόματου Ελέγχου. *Διπλωματική Εργασία Παν. Πατρών*.
- Κολοσιώνης, Κ. (2018). Συνδυασμός Τεχνητών Νευρωνικών Δικτύων και Γεωστατιστικών μοντέλων για την βελτιστοποίηση δικτύου γεωτρήσεων παρακολούθησης στάθμης υπόγειου υδροφορέα. *Διπλωματική εργασία ΕΜΠ*.
- Κοτζαπαναγιώτου, Ε. (2019). Χρήση Μεθόδων Μηχανικής Μάθησης για Πρόβλεψη.
- Λυκοθανάσης, Σ. (2001). Γενετικοί Αλγόριθμοι και Εφαρμογές.
- Ντάλλα, Μ. (2009, Οκτώβριος). Εφαρμογή αλγορίθμων επαγωγικού λόγου προγραμματισμού στη σχεσιακή εξόρυξη δεδομένων.
- Σοφianoπούλου, Εμβλωτής, Πίτσια & Καρακολίδης. (2017). Έκθεση Αποτελεσμάτων του Διεθνούς Προγράμματος PISA 2015 για την Αξιολόγηση των Μαθητών στην Ελλάδα.
- Σταυλιώτης, Γ. (2008, Ιούνιος). Εξόρυξη δεδομένων (Data mining) και κατηγορικά δεδομένων. *Διπλωματική εργασία* .
- Στεργιανού, Ι. (2018). *Άγχος και Μαθηματικά: Διερεύνηση των στάσεων των μαθητών*. ΕΑΠ.
- Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996) From Data Mining to Knowledge Discovery in Databases
- Richard J. Roiger (2017) Data Mining: A Tutorial-Based Primer, Second Edition
- I.E. Livieris, K. Drakopoulou, and P. Pintelas (2012) "Predicting students' performance using artificial neural networks", In the Proceedings of the 8th Pan-Hellenic Conference Information and Communication Technology in Education

S. Kotsiantis (2010) "Educational data mining: A case study for predicting dropout-prone students", International Journal of Knowledge Engineering and Soft Data Paradigms, Volume 10,

Ελένη Γολέμη (2010) Μεταπτυχιακή Εργασία , Κρυπτογραφία και Εξόρυξη Δεδομένων

Σταματόπουλος Σάββας , Διπλωματική Εργασία, Πληροφοριακά Συστήματα Εκπαίδευσης – Μελέτη περίπτωσης e-arsakeio

Ντίνου Μαρία (2015) Διπλωματική Εργασία , Από το Λύκειο στο Εθνικό Μετσόβιο Πολυτεχνείο: Έρευνα για το «κενό» στη διδασκαλία των μαθηματικών.

Αγγελόπουλος Α. (2017) Διπλωματική Εργασία , Η χρήση των τεχνικών εξόρυξης από δεδομένα (data-mining) στην ανάλυση ερωτηματολογίων - εφαρμογή στην ανάλυση ερωτηματολογίων για τη διάγνωση του αυτισμού

Σουρής Εμμανουήλ (2016) Διπλωματική Εργασία , « Ο εθισμός στο διαδίκτυο στην πρωτοβάθμια εκπαίδευση και οι επιπτώσεις στην ζωή των μαθητών Ε΄ και ΣΤ΄ Δημοτικού».

Μανιά Αγγελική (2018) Πτυχιακή Εργασία , Εξυπνη Εξόρυξη Δεδομένων από το Διαδίκτυο

ΠΑΡΑΡΤΗΜΑ

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΕΡΕΥΝΑΣ

Προσωπικές Πληροφορίες

1. 1. Φύλο *

Να επισημαίνεται μόνο μία έλλειψη.

- Άνδρας
 Γυναίκα

2. 2. Ηλικία *

Να επισημαίνεται μόνο μία έλλειψη.

- 18-20
 20-25
 25-30
 30-35
 35-40
 >40

3. 3. Κατοικία *

Να επισημαίνεται μόνο μία έλλειψη.

- Αστική Περιοχή
 Ημιαστική Περιοχή
 Αγροτική Περιοχή

4. 4. Επίπεδο Εκπαίδευσης *

Να επισημαίνεται μόνο μία έλλειψη.

- Απόφοιτος/η Γυμνασίου
 Απόφοιτος/η Λυκείου
 Απόφοιτος/η ΑΕΙ/ΤΕΙ
 Κάτοχος Μεταπτυχιακού
 Κάτοχος Διδακτορικού
 Άλλο: _____

5. 5. Εργασία *

Να επισημαίνεται μόνο μία έλλειψη.

- Φοιτητής/Φοιτήτρια
 Άνεργος/Άνεργη
 Δημόσιος Τομέας
 Ιδιωτικός Τομέας
 Αυταπασχόληση
 Άλλο: _____

6. 6. Οικονομική Κατάσταση

Να επισημαίνεται μόνο μία έλλειψη.

- 1 2 3 4 5
Όχι καλή Πολύ καλή

7. 7. Μορφωτικό επίπεδο γονέων (έστω ένας γονέας)

Να επισημαίνεται μόνο μία έλλειψη.

- Απόφοιτοι Γυμνασίου
 Απόφοιτοι Λυκείου
 Πανεπιστημιακής Εκπαίδευσης
 Άλλο: _____

Ερωτήσεις σχετικά με τις Πανελλαδικές Εξετάσεις

8. 8. Είδος Εκπαίδευσης

Να επισημαίνεται μόνο μία έλλειψη.

- Δημόσιο Σχολείο
 Ιδιωτικό Σχολείο

9. 9. Πόσες φορές έχετε δώσει πανελλαδικές εξετάσεις; *

Να επισημαίνεται μόνο μία έλλειψη.

- Καμία
 1
 2
 >2

10. 10. Πόσες ώρες διαβάσατε κατά μέσο όρο ημερησίως την χρονιά που δώσατε πανελλαδικές εξετάσεις;

Να επισημαίνεται μόνο μία έλλειψη.

- 0-2
 2-4
 4-6
 6-8
 >8

11. 11. Κάνατε φροντιστήριο ή βασιστήκατε στην δημόσια εκπαίδευση; *

Να επισημαίνεται μόνο μία έλλειψη.

- Δημόσια Εκπαίδευση
 Φροντιστήριο - Ιδιαίτερα Μαθήματα
 Άλλο: _____

12. 12. Είχατε άγχος κατά την διάρκεια της χρονιάς που δίνετε πανελλαδικές εξετάσεις;

Να επισημαίνεται μόνο μία έλλειψη.

- 1 2 3 4 5
Καθόλου Πολύ

13. 13. Είχατε άγχος τις ημέρες των πανελλαδικών εξετάσεων;

Να επισημαίνεται μόνο μία έλλειψη.

- 1 2 3 4 5
Καθόλου Πολύ

14. 14. Πόσα μόρια γράψατε στις πανελλαδικές εξετάσεις;

Να επισημαίνεται μόνο μία έλλειψη.

- 0-5.000
 5.000-10.000
 10.000-15.000
 15.000-20.000

15. 15. Πετύχατε Σχολή της Πρώτης Προτίμησής σας;

Να επισημαίνεται μόνο μία έλλειψη.

- Ναι
 Όχι

16. 16. Θετική ή Θεωρητική κατεύθυνση;

Να επισημαίνεται μόνο μία έλλειψη.

- Θετική/Τεχνολογική
 Θεωρητική
 Άλλο: _____

Πληροφορίες
Σχετικά με τα
Μαθηματικά

Οι ερωτήσεις που αφορούν τα μαθηματικά της Γ' Λυκείου να απαντηθούν μόνον από όσους είχαν επιλέξει Θετική/Τεχνολογική Κατεύθυνση και εξετάστηκαν σε αυτό το μάθημα.

17. 17.Τι μέσο όρο είχατε στα Θετικά μαθήματα (φυσική, χημεία, βιολογία) στις 3 τάξεις του λυκείου; *

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

18. 18.Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Α' Λυκείου; *

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

19. 19.Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Β' Λυκείου; *

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

20. 20.Τι βαθμολογία είχατε στο σχολείο στα μαθηματικά στην Γ' Λυκείου;

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

Πληροφορίες
Σχετικά με τα
Μαθηματικά

Οι ερωτήσεις που αφορούν τα μαθηματικά της Γ' Λυκείου να απαντηθούν μόνον από όσους είχαν επιλέξει Θετική/Τεχνολογική Κατεύθυνση και εξετάστηκαν σε αυτό το μάθημα.

25. 25. Τι βαθμολογία είχατε στα Μαθηματικά Κατεύθυνσης στις Πανελλαδικές; (μόνο όσοι εξετάστηκαν)

Να επισημαίνεται μόνο μία έλλειψη.

- 0-5
 5-10
 10-15
 15-20

26. 26.Είστε ικανοποιημένοι από την βαθμολογία σας στα μαθηματικά;

Να επισημαίνεται μόνο μία έλλειψη.

- Ναι
 Όχι

27. 27. Πόσο εξοικειωμένοι νιώθετε με τα μαθηματικά γενικότερα; *

Να επισημαίνεται μόνο μία έλλειψη.

- 1 2 3 4 5
καθόλου πολύ

21. 21.Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Α'Λυκείου; *

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

22. 22.Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Β'Λυκείου; *

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

23. 23.Πόσο γράφατε κατά μέσο όρο στα διαγωνίσματα/τεστ στα μαθηματικά στην Γ'Λυκείου;

Να επισημαίνεται μόνο μία έλλειψη.

- <10
 10-15
 15-17
 17-20

24. 24.Πόσες ώρες διαβάσατε κατά μέσο όρο ημερησίως ΜΑΘΗΜΑΤΙΚΑ την χρονιά που δώσατε πανελλαδικές εξετάσεις; (μόνο όσοι εξετάστηκαν)

Να επισημαίνεται μόνο μία έλλειψη.

- 0-2
 2-4
 4-6
 6-8
 >8

28. 28.Πόσο εξοικειωμένοι νιώθατε με την ύλη των μαθηματικών της Γ'Λυκείου;

Να επισημαίνεται μόνο μία έλλειψη.

- 1 2 3 4 5
καθόλου πολύ

29. 29.Χρειάστηκαν τα μαθηματικά στις προπτυχιακές σπουδές σας;

Να επισημαίνεται μόνο μία έλλειψη.

- Ναι
 Όχι

Γενικές Ερωτήσεις

30. 30. Θεωρείτε ότι τα μαθηματικά χρειάζονται γενικά στην ζωή ενός ανθρώπου;

Να επισημαίνεται μόνο μία έλλειψη.

Ναι

Όχι

31. 31. Θεωρείτε ότι τα μαθηματικά συνδέονται μόνο με αριθμούς , με την "εκτέλεση υπολογισμών" ή κάτι περισσότερο? (περιγράψτε)

32. 32. Θεωρείτε ότι τα μαθηματικά έχουν εφαρμογές στην καθημερινή ζωή; Αν ναι, πού; (περιγράψτε)

ΠΡΟΓΡΑΜΜΑ

```
1 import random
2 import pandas
3 from sklearn import tree
4 #import pydotplus
5 from sklearn.neural_network import MLPClassifier
6 import matplotlib.pyplot as plt
7 import matplotlib.image as pltimg
8
9 #synartisi pou vgazei tin apodosi tou nevronikou
10 def MLP(X,y, lr,mom,lev1,lev2,lev3):
11
12     #to synolo trainer
13     Xtrain=X[1:200]
14     #to synolo test
15     Ytrain=y[1:200]
16
17 #montelo nevronikou
18 clf = MLPClassifier(hidden_layer_sizes=(lev1,lev2,lev3),momentum=mom, learning_rate_init=lr,random_state=1, max_iter=300).fit(X, y)
19 #apodosi nevronikou
20 clf = clf.fit(Xtrain, Ytrain)
21
22 #apotelesma tou score
23 Xtest=X[200:266]
24 Ytest=y[200:266]
25 return (clf.score(Xtest,Ytest))
26
27
28 #to kyrio programma
29 #diavazoume to arxeio
30 df = pandas.read_csv("data.csv")
31
32 #epilegoume tis stiles pou theloume gia to motnelo mas
33 features = ['e1','e2','e3','e4','e5','e6','e7','e8','e9','e10','e11','e12','e13','e14','e14','e15','e16','e17']
34
35 #pairnoume synolo pou mpainei eisodo sto montelo
36 X= df[features]
37 #pairnoume tin stili pou theloume na provlepsoume
38 y = df['p']
39
40 #pinakas me ta xromosomata tou genetikou
```

```
41 crom=[]
42
43 #apodoseis tou genetikou
44 ev=[]
45
46 #genies
47 MC=50
48
49 #plithismos genneas
50 N=100
51
52 #synartitisi diastavrosis
53 pcross=0.9
54 #synartitisi metallaxis
55 pmul=0.9
56
57 print("start")
58
59 #gia kathe gennea
60 for i in range(0,MC):
61
62     #parinoume tixaies times stis metavlites tou nevronikou
63     lr=random.random()
64     lev1=random.randrange(2,100)
65     lev2=random.randrange(2,100)
66     lev3=random.randrange(2,100)
67     mom=random.random()
68
69     #pairnoume ta xromosomata
70     c=[lr, mom ,lev1, lev2, lev3 ]
71
72
73     #pairnoume to score
74     score=MLP(X,y, lr,mom,lev1,lev2,lev3)
75
76     #ta vazoume stous antistoixous pinakess
77     crom.append(c)
78     ev.append(score)
79     print (i)
80
```

```

81
82 #trexoume ton gennetiko
83 for i in range (0,N):
84     print("Gen: ",i)
85     ev,crom=zip(*sorted(zip(ev,crom),reverse=True))
86     #taxinomoume me vasi tin apodosi tou nevronikou
87     ev=list(ev)
88     crom=list(crom)
89
90     #kanoume diastavrosi
91     if(random.random()<pcross):
92         i1=random.randrange(0,int(MC/2))
93         i2=random.randrange(0,int(MC/2))
94
95         lr=(crom[i1][0]+crom[i2][0])/2
96         mom=(crom[i1][1]+crom[i2][1])/2
97         lev1=int((crom[i1][2]+crom[i2][2])/2)
98         lev2=int((crom[i1][3]+crom[i2][3])/2)
99         lev3=int((crom[i1][4]+crom[i2][4])/2)
100        c=[lr, mom ,lev1, lev2, lev3 ]
101
102        for j in range (0,5):
103            crom[MC-1][j]=c[j]
104            score=MLP(X,y, lr,mom,lev1,lev2,lev3)
105            ev[MC-1]=score
106        #kanoume metallaxi
107        if(random.random()< pmul):
108            i1=random.randrange(0,int(MC/2))
109            i2=random.randrange(0,5)
110            k=[]
111            k.append(random.random())
112            k.append(random.random())
113            k.append(random.randrange(2,100))
114            k.append(random.randrange(2,100))
115            k.append(random.randrange(2,100))
116
117        #topothetoume ta paidia sto plithismo
118        crom[i1][i2]=k[i2]
119        i2=random.randrange(0,5)
120        crom[i1][i2]=k[i2]
121        score=MLP(X,y, crom[i1][0],crom[i1][1],crom[i1][2],crom[i1][3],crom[i1][4])
122        ev[MC-1]=score
123
124        #ektiposi genneon kai epidosewn
125        print(crom)
126        print(ev)
127
128        #kalytero nevroniko kai apodosi
129        print("Best Solution:", crom[0])
130        print("Evaluate Solution:", ev[0])
131
132
133

```