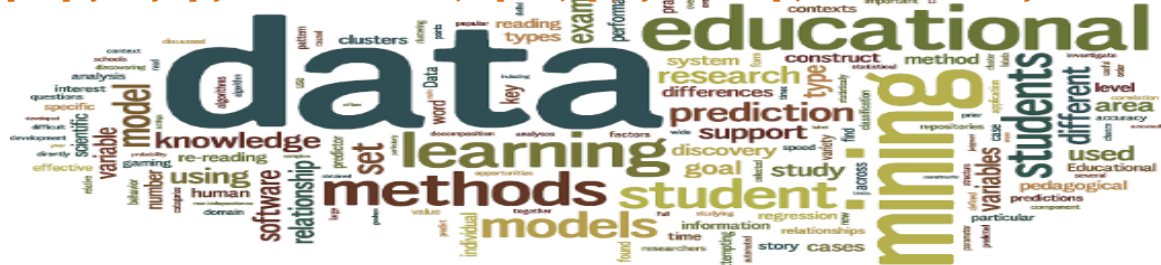




Εφαρμογή μεθόδων εξόρυξης γνώσης από δεδομένα



στην εκπαίδευση. Μελέτη περίπτωσης για τη σχέση του επαγγέλματος των γονιών με τη σχολική επίδοση και τη σχέση της σχολικής επίδοσης με την επιλογή του επαγγέλματος των παιδιών.

14078

Μαλιαρίτση Διονυσία Μαρία

Επιβλέπων καθηγητής:

Μαραγκουδάκης Εμμανουήλ

Τριμελής επιτροπή:

Καρύδα Μαρία, Κοκολάκης Σπυρίδων, Μαραγκουδάκης
Εμμανουήλ

Περίληψη

Η ραγδαία εξάπλωση του internet έχει πολυσήμαντες συνέπειες, μια από αυτές είναι η δημιουργία μεγάλων αποθηκών πληροφοριών που κρύβουν γνώση . Η αξιοποίηση της κρυμμένης γνώσης δίνει σημαντικό πλεονέκτημα ώστε να ληφθούν οι βέλτιστες αποφάσεις σε σημαντικά διλήμματα. Αυτό επιτυγχάνεται με τη χρήση τεχνικών εξόρυξης δεδομένων. Η εξόρυξη δεδομένων (**data mining**), περιλαμβάνει τεχνικές εξαγωγής προτύπων δεδομένων ή περιγραφές δεδομένων από μεγάλες αποθήκες και έχει στόχο την περιγραφή και πρόβλεψη από τα σύνολα αυτά με σκοπό την εξαγωγή χρήσιμης γνώσης.

Η παρούσα μελέτη επιχειρεί να διερευνήσει με τεχνικές εξόρυξης γνώσης από τα δεδομένα αν η σχολική επίδοση της Στ Δημοτικού όπως αυτή εκφράζεται στην αναλυτική βαθμολογία , είναι παράγοντας από τον οποίο μπορεί να γίνει πρόβλεψη για την επαγγελματική επιλογή . Με ένα δεύτερο μοντέλο γίνεται διερεύνηση της σχέσης του επαγγέλματος των γονιών με την επίδοση των μαθητών της Στ Δημοτικού όπως αυτή εκφράζεται με το μέσο όρο της βαθμολογίας τους .

Ως θεωρητικό πλαίσιο επιλέχτηκε το κονστρουβιστικό μοντέλο και το κοινωνικογνωστικό μοντέλο Wisconsin. Το εργαλείο που χρησιμοποιήθηκε για τη συλλογή των δεδομένων είναι το ερωτηματολόγιο και το βιβλίο Μητρώου των μαθητών. Η μέθοδος που χρησιμοποιήθηκε για **περιγραφή** και **πρόβλεψη** είναι η **κατηγοριοποίηση** και η τεχνική των **Δικτύων Bayes**. Η κατηγοριοποίηση εφαρμόζεται σε δεδομένα που έχουν ταξινομηθεί σε συγκεκριμένες κλάσεις με στόχο την εξαγωγή κανόνων ώστε να χρησιμοποιηθούν για κατηγοριοποίηση νέων δεδομένων. Εδώ επιτυγχάνεται με τον αλγόριθμο ID3 για τα δέντρα απόφασης, NN για τα νευρωνικά δίκτυα και KNN . Τα δίκτυα Bayes είναι τεχνική που κάνει κατηγοριοποίηση με περισσότερες από μια κλάσεις και επιτρέπει έτσι την ανακάλυψη περισσότερων κρυμμένων συσχετίσεων καθώς υπάρχει η δυνατότητα απομόνωσης μιας μεταβλητής και παρατήρησης της συμπεριφοράς των υπολοίπων . Δημιουργήθηκαν περιγραφικά μοντέλα και επιχειρήθηκε με μια σειρά πειραμάτων να δημιουργηθούν μοντέλα πρόβλεψης με ακριβή αποτελέσματα

Τα αποτελέσματα δείχνουν συσχέτιση ανάμεσα στην επίδοση και την επιλογή επαγγέλματος και αναδεικνύουν και άλλες ενδιαφέρουσες συσχετίσεις με παράγοντες που σχετίζονται με την επίδοση και το επάγγελμα. Δημιουργείται μοντέλο πρόβλεψης με υψηλή ακρίβεια μόνο για δυο ομάδες επαγγελμάτων. Ως προς το επάγγελμα των γονιών διαπιστώνεται σχέση ως προς τις χαμηλότερες Βαθμολογίες. Το μοντέλο πρόβλεψης εμφανίζει μικρή απόκλιση.

ABSTRACT

The rapid spread of the internet has ambiguous consequences; one of them is the creation of large information stores that hide knowledge. The use of hidden knowledge gives a significant advantage in order to take optimal decisions on important dilemmas. This is accomplished by using techniques of data mining .Data mining includes data models' extraction techniques or descriptions of data from large data stores and aims at the description and prediction of these sets in order to extract useful knowledge.

This study attempts to explore with knowledge mining techniques from the data whether the school performance of sixth grade elementary school students, as expressed in the detailed transcript, is a factor based on which a prediction can be made concerning the choice of profession. A second model is to investigate the relationship of the parents' profession to the performance of the sixth grade elementary school students as expressed by their average score.

As a theoretical framework, the constructivist model and the socio-cognitive Wisconsin model were selected. The tool used for data collection was the questionnaire and the book of Students' Registry. The method used for description and prediction was the categorization/classification and the technique of Bayes Networks. The classification is applied to data registered in specific classes, targeted to extract rules so as to be used to classify new data. This is achieved by the ID3 algorithm for the decision trees, the NN for the neural networks and the KNN. The Bayes networks is a technique that makes classification in more than one class and thus allows the discovery of more hidden correlations, as there is the potential isolation of a variable and observation of the behavior of the rest. Descriptive models were created and a series of experiments was attempted to create prediction models with accurate results.

The results show a correlation between school performance and the choice of profession and highlight other interesting correlations with factors related to school performance and profession. A prediction model with high accuracy is created only for two employment categories. As far as the profession of parents is concerned, a correlation with the lowest score in students' performance is detected. The prediction model shows little deviation.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Μαραγκουδάκη Εμμανουήλ που με την επιστημονική του καθοδήγηση και τις κριτικές του επισημάνσεις, ολοκληρώθηκε αυτή η εργασία. Θα ήθελα να τον ευχαριστήσω θερμά γιατί με τη μεθοδικότητα και τη μεταδοτικότητά που τον διακρίνει, μετέτρεψε δύσκολους στόχους στο να φαίνονται εφικτοί, δημιουργώντας το κατάλληλο κίνητρο για τη συνέχεια.

Θα ήθελα να ευχαριστήσω και τον μικρό μου Άγγελο για την ώριμή του στάση απέναντι στο χρόνο που δεν του αφιέρωνα.

ΠΕΡΙΕΧΟΜΕΝΑ

ABSTRACT	4
ΕΥΧΑΡΙΣΤΙΕΣ	5
Πίνακας σχημάτων	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
Πίνακας πινάκων	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
ΕΙΣΑΓΩΓΗ	10
ΠΡΟΓΕΝΕΣΤΕΡΗ ΕΡΓΑΣΙΑ	19
1.1 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ(EDUCATIONAL DATA MINING)	19
1.2 ΕΡΕΥΝΗΤΙΚΑ ΑΝΤΙΚΕΙΜΕΝΑ- ΕΦΑΡΜΟΖΟΜΕΝΑ ΠΡΟΓΡΑΜΜΑΤΑ EDM.....	20
1.3 ΜΕΘΟΔΟΙ ΚΑΙ ΤΕΧΝΙΚΕΣ	22
ΣΧΟΛΙΚΗ ΕΠΙΔΟΣΗ	24
2.1 ΑΞΙΟΛΟΓΗΣΗ	24
2.1.1 ΜΕΤΡΗΣΗ	24
2.1.2 ΈΝΝΟΙΑ ΚΑΙ ΚΛΙΜΑΚΕΣ ΒΑΘΜΟΛΟΓΗΣΗΣ	25
2.1.3 ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ	27
2.2 ΟΡΙΣΜΟΣ ΣΧΟΛΙΚΗΣ ΕΠΙΔΟΣΗΣ	27
2.2.1 ΠΑΡΑΓΟΝΤΕΣ ΜΕ ΤΟΥΣ ΟΠΟΙΟΥΣ ΣΧΕΤΙΖΕΤΑΙ Η ΣΧΟΛΙΚΗ ΕΠΙΔΟΣΗ.....	28
2.2.2 ΙΣΤΟΡΙΚΗ ΑΝΑΣΚΟΠΗΣΗ.....	29
2.2.3 ΕΙΝΑΙ ΑΝΑΓΚΑΙΑ Η ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΕΠΙΔΟΣΗΣ ΤΩΝ ΜΑΘΗΤΩΝ.....	31
ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΑΝΑΠΤΥΞΗ	32
3.1 ΒΑΣΙΚΕΣ ΘΕΩΡΙΕΣ	32
3.1.1 ΤΥΠΟΛΟΓΙΚΗ ΘΕΩΡΙΑ ΤΟΥ JOHN HOLLART	33
3.1.2 ΑΦΗΓΗΜΑΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ MARK SAVICKAS(2005).....	33
3.1.3 ΘΕΩΡΙΑ ΤΗΣ LINDA COTTFREDSON.....	33
3.1.4 ΘΕΩΡΙΑ ΤΟΥ JOHN KRUMBOLTZ	34
3.1.5 ΚΟΙΝΩΝΙΚΟΨΥΧΟΛΟΓΙΚΟ ΜΟΝΤΕΛΟ WISCONSIN	34
3.2 Η ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΕΠΑΓΓΕΛΜΑΤΩΝ ΣΕ ΟΜΑΔΕΣ.....	35
DATA MINING	38
4.1 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ.....	38
4.1. Α. ΤΟΜΕΙΣ ΣΥΣΧΕΤΙΣΗΣ.....	39
4.1.Β. ΚΑΘΕ ΕΡΓΑΣΙΑ ΑΝΑΚΑΛΥΨΗΣ ΕΙΝΑΙ ΕΞΟΡΥΞΗ;.....	40
4.2 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ.....	41
4.2.1. Κ- ΚΟΝΤΙΝΟΤΕΡΟΙ ΓΕΙΤΟΝΕΣ (K- NEAREST NEIGHBORS- K-NN)	42

4.2.1.Α.	ΑΛΓΟΡΙΘΜΟΣ ΚΝΝ.....	43
4.2.1.Β	ΠΕΡΙΟΡΙΣΜΟΙ.....	44
4.2.2	Νευρωνικά δίκτυα (Neural Networks)	44
4.2.2.Α	ΑΛΓΟΡΙΘΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ ΤΟΥ ΝΕΥΡΩΝΑ PERCEPTRON.....	45
4.2.3	ΔΕΝΤΡΑ ΑΠΟΦΑΣΗΣ (DECISION TREES).....	47
4.2.3.Α.	ΑΛΓΟΡΙΘΜΟΣ ΔΕΝΤΡΟΥ ΑΠΟΦΑΣΗΣ.....	48
4.2.3.Β.	ΓΝΩΡΙΣΜΑΤΑ ΚΑΤΑΣΚΕΥΗΣ ΔΕΝΤΡΟΥ ΑΠΟΦΑΣΗΣ.....	49
4.2.3.Γ.	ΑΛΓΟΡΙΘΜΟΣ ID3	50
4.2.3.Δ.	ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ.....	51
4.3	ΒΑΥΕΣ ΔΙΚΤΥΑ.....	52
4.1.Α	ΑΛΓΟΡΙΘΜΟΣ ΠΑΡΑΓΩΓΗΣ ΤΗΣ ΤΟΠΟΛΟΓΙΑΣ ΕΝΟΣ ΔΙΚΤΥΟΥ ΒΑΥΕΣ.....	54
	ΜΕΘΟΔΟΛΟΓΙΑ	56
5.1	ΣΤΟΧΟΙ.....	56
5.2	ΔΙΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ.....	57
5.3	ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ.....	57
5.4	ΕΡΓΑΛΕΙΑ.....	61
5.4.1	ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ.....	61
5.4.2	ΔΙΚΤΥΑ ΒΑΥΕΣ.....	63
	ΠΕΙΡΑΜΑΤΑ.....	65
6.1	ΠΕΙΡΑΜΑΤΑ ΣΤΑ ΔΙΚΤΥΑ ΒΑΥΕΣ (Α ΜΟΝΤΕΛΟ).....	65
6.1.1	ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ.....	66
6.1.2	ΣΤΑΤΙΣΤΙΚΕΣ ΕΞΑΡΤΗΣΕΙΣ (ΠΡΟΒΛΕΨΗ).....	72
6.2	Β ΜΟΝΤΕΛΟ.....	80
6.2.Α	ΠΕΡΙΓΡΑΦΗ.....	80
6.3	ΠΕΙΡΑΜΑΤΑ RAPIDMINER (Α ΜΟΝΤΕΛΟ).....	84
6.3.1	ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ.....	84
6.3.2	ΠΡΟΒΛΕΨΗ ΚΛΑΣΗΣ.....	86
6.3.3	ΔΙΑΚΡΙΤΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ.....	88
6.3.3.Α	ΟΜΑΔΑ 2- ΟΜΑΔΑ 4- ΟΜΑΔΑ 5	88
6.3.3.Β	ΟΜΑΔΑ 2- ΌΧΙ ΟΜΑΔΑ 2.....	91
6.3.3.Γ	ΟΜΑΔΑ 4 – ΌΧΙ ΟΜΑΔΑ 4.....	93
6.3.3.Δ	ΟΜΑΔΑ 5- ΌΧΙ 5.....	94
6.4	ΠΕΙΡΑΜΑΤΑ RAPIDMINER (Β ΜΟΝΤΕΛΟ).....	99
6.4.1	ΔΕΔΟΜΕΝΑ ΕΙΣΟΔΟΥ.....	99
6.4.2	ΠΡΟΒΛΕΨΗ ΚΛΑΣΗΣ.....	101

6.4.2.A ΠΡΟΒΛΕΨΗ ΚΛΑΣΗΣ	102
6.4.2.B ΠΡΟΒΛΕΨΗ ΚΛΑΣΗΣ ③.....	103
6.4.3 ΚΑΝΟΝΕΣ ΣΥΣΧΕΤΙΣΗΣ (ASSOCIATION RULES)	104
ΣΥΜΠΕΡΑΣΜΑΤΑ	106
ΒΙΒΛΙΟΓΡΑΦΙΑ	111
ΓΛΩΣΣΑΡΙΟΡΩΝ	113
ΠΑΡΑΡΤΗΜΑ	115

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ	Σελίδα
Εικ. 1 ερωτηματολόγιο	59
Εικ.2 RapidMiner	61
Εικ.3 Επάγγελμα -επίδοση	81
Εικ.4 Μαθηματικά –Γλώσσα_ Επάγγελμα	84
Εικ.5 Δέντρο απόφασης ομάδα 2,4,5	85
Εικ.6 Δέντρο απόφασης ομάδα 2-όχι ομάδα 2	87
Εικ.7 Δέντρο απόφασης ομάδα 4 –όχι ομάδα 4	88
Εικ.8 Εξαίρεση μεταβλητής Φύλο	90
Εικ.9 παράμετροι για Δέντρο απόφασης	91
Εικ.10 Παράμετροι για K-NN	92
Εικ.11 Παράμετροι για NN	92
Εικ.12 Δέντρο απόφασης ομάδα5-όχι ομάδα5	92
Εικ.13 Training set	96
Εικ.14 Σώμα εκπαίδευσης 3	98
Εικ.15 Απόκλιση	99

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ	Σελίδα
Σχ.1 Α Μοντέλο	14
Σχ. 2 Β Μοντέλο	15
Σχ. 3 Νευρώνας	45
Σχ. 4 Τοπολογία νευρωνικού δικτύου	46

Σχ. 5 Δέντρο απόφασης	48
Σχ. 6 Δίκτυα Bayes	53
Σχ. 7 Δίκτυα Bayes A Μοντέλο	63
Σχ. 8 Φύλο	64
Σχ. 9 Ταξινόμηση επαγγελματιών	64
Σχ.10 Αναλυτική Βαθμολογία	65
Σχ.11 Προσπάθεια	66
Σχ.12 Βαθμολογία-Αυτοαντίληψη	67
ΣΧ.13 Βαθμολογία- Στόχοι	68
Σχ.14 Φυσικά	69
Σχ.15 Γλώσσα	70
Σχ.16 Μαθηματικά	70
Σχ.17 Βαθμολογία –Ενίσχυση	71
Σχ. 18 Συναισθήματα άντρες –γυναίκες	72
Σχ.19 Φύλο- μαθησιακές δυσκολίες	72
Σχ.20 Φύλο- Ξένη Γλώσσα	73
Σχ.21 Βαθμολογία –συναισθήματα –μαθησιακές	73
Σχ.22 Επαγγελματικό άγχος – ικανότητα	74
Σχ.23 Επαγγελματικό άγχος –Γλώσσα	75
Σχ.24 Σχέσεις	77
Σχ.25 Φύλο B Μοντέλο	77
Σχ. 26 Επάγγελμα γονιών	77
Σχ.27 Δίκτυα Bayes B Μοντέλο	78
Σχ. 28 χαμηλότερη βαθμολογία –επάγγελμα πατέρα	95
Σχ.29 Χαμηλότερη βαθμολογία –επάγγελμα μητέρας	95

ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ	Σελίδα
Πιν. 1 Ηλικιακές ομάδες	64
Πιν.2 Μαθητικό- Επαγγελματικό προφίλ	66
Πιν.3 Παράγοντες επιλογής	67
Πιν. 4 Βαθμολογία – Οικογένεια	67
Πιν. 5 Βαθμολογία –Ικανότητα	68
Πιν. 6 Προτιμήσεις	67

Πιν. 7 Δυσκολίες – Στόχοι	71
Πιν. 8 Training set	82
Πιν. 9 Ακρίβεια δέντρου απόφασης	83
Πιν. 10 Αξιοπιστία KNN	83
Πιν. 11 Αξιοπιστία Neural Net	83
Πιν.12 Ακρίβεια Δέντρου απόφασης ομάδες 2,4,5	85
Πιν.13 Αξιοπιστία K-NN	86
Πιν.14 Αξιοπιστία Neural Net	86
Πιν.15 Αξιοπιστία Δέντρου απόφασης ομάδα2-όχι ομάδα 2	87
Πιν.16 βαρύτητα μεταβλητών	87
Πιν. 17 Αξιοπιστία	88
Πιν.18 Αξιοπιστία Neural net	88
Πιν. 19 Αξιοπιστία Δέντρου απόφασης ομάδα 4-όχι ομάδα 4	89
Πιν.20 Βαρύτητα μεταβλητών	89
Πιν.21 Αξιοπιστία NN ομάδα4-όχι ομάδα4	89
Πιν.22 Αξιοπιστία K-NN ομάδα4-όχι ομάδα4	89
Πιν.23 Weight by information Gain	90
Πιν . 25 Αξιοπιστία K-NN ομάδα 5-όχι ομάδα5	91
Πιν. 26 Αξιοπιστία Neural Net ομάδα 5-οχι ομαδα5	91
Πιν.27 Συνοπτικός Πίνακας	93
Πιν. 28 Ακρίβεια Δέντρου απόφασης	93
Πιν. 29 Training set 2	97
Πίν.30 Ακρίβεια δέντρου απόφασης 2	97
Πιν.31 Linear Regression	99

ΕΙΣΑΓΩΓΗ

Κίνητρα

Απώτερος σκοπός της αξιολόγησης των μαθητών είναι σύμφωνα με το αναλυτικό πρόγραμμα (Κασσωτάκης Μ., 2010) η «Βελτίωση της ποιότητας του εκπαιδευτικού έργου , ο προσδιορισμός της προόδου του μαθητή και η λήψη περαιτέρω πρωτοβουλιών για τη βελτίωσή του». Στο τέλος κάθε τριμήνου όμως οι εκπαιδευτικοί ειδικά της πρωτοβάθμιας εκπαίδευσης, αντιμετωπίζουν με έντονο προβληματισμό την εξής υποχρέωση :

«Πρέπει να δώσω βαθμολογία». Και είναι μια υποχρέωση που προβληματίζει γιατί με ένα βαθμό πρέπει να αποδοθούν : Η απόκτηση γνώσεων ως διαδικασία , ως κατάκτηση και ως ανάπτυξη γνωστικών δεξιοτήτων . Αλλά θα πρέπει να συμπεριληφθούν και οι παράγοντες που σχετίζονται με αυτό: ατομική προσπάθεια, νοητικές ικανότητες, προσωπικά ενδιαφέροντα. Και παράγοντες που ενώ δεν αφορούν τον ίδιο το μαθητή τον επηρεάζουν καταλυτικά : όπως ο βαθμός στήριξης από το οικογενειακό περιβάλλον, η αποτελεσματικότητα της διδασκαλίας, τα κίνητρα μάθησης , οι πολιτισμικοί, οι κοινωνικοί και οι οικονομικοί παράγοντες . «Η βαθμολογία προκαλεί συχνά άγχος στους μαθητές αλλά και στους γονείς. Για ορισμένους, λίγους, αποτελεί κίνητρο για να δουλέψουν. Βέβαια, πολλές φορές οι βαθμοί στο δημοτικό είναι αρκετά υψηλοί και δεν ανταποκρίνονται στην εικόνα του μαθητή. Η βαθμολογία δεν πληροφορεί ακριβώς , με λεπτομέρειες για την επίδοση του μαθητή. Επιπλέον μπορεί να δημιουργήσει και να καλλιεργήσει τον ανταγωνισμό μέσα στην τάξη.» Α.Π Δασκάλα της Στ Δημοτικού.

Ωστόσο το εκπαιδευτικό σύστημα λειτουργεί για όλους με τον ίδιο τρόπο ,ίδιοι διδάσκοντες , ίδια βιβλία, ίδιες ή περίπου ίδιες συνθήκες και όλα αυτά προβάλλονται ως κριτήριο αντικειμενικότητας ως προς την προσφερόμενη εκπαίδευση. Εντούτοις αν ο ομοιογενής τρόπος με τον οποίο προσφέρεται μια γνώση, διασφάλιζε και ομοιογενή αποτελέσματα η έννοια της σχολικής επίδοσης θα αφορούσε και θα οριζόταν από τον τρόπο εκπαίδευσης . Όμως δε συμβαίνει αυτό αφού τα άτομα διαφοροποιούνται σημαντικά καθώς προέρχονται από διαφορετικές κοινωνικές αφετηρίες. Έτσι οι αποδέκτες του ίδιου μηνύματος δεν σημειώνουν και την ίδια σχολική επίδοση αλλά και η ίδια σχολική επίδοση δεν έχει το ίδιο νόημα και απήχηση σε κάθε αποδέκτη μετατρέποντας τον όρο σε ένα πολυσήμαντο αποτέλεσμα που διαμορφώνεται με την επίδραση ενός πλήθους παραγόντων αλλά και μπορεί να διαμορφώσει και πολλούς παράγοντες ακόμα. Το πιο σημαντικό βέβαια είναι η σημασία που δίνει εκείνος που αποδέχεται το αποτέλεσμα , δηλαδή ο μαθητής. Από την άλλη οι σύγχρονες κοινωνικοοικονομικές συνθήκες παρουσιάζουν ιδιαίτερα υψηλά επίπεδα ανεργίας , ειδικά σε άτομα με πτυχίο ανώτατων σπουδών , όπως επισημαίνει η Euro stat , η Ελλάδα είναι η μόνη χώρα από τις 27 της ΕΕ που τα ποσοστά ανεργίας δε

μειώνονται ανάλογα με τα αποκτηθέντα τυπικά προσόντα εφόσον 100.000 πτυχιούχοι σύμφωνα με έρευνα βρίσκονται σε λίστα ανεργίας και επιπλέον 20.000 γιατροί , φαρμακοποιοί , δικηγόροι από τους οποίους 2.200 είναι κάτοχοι master και διδακτορικού .Οι παραπάνω λόγοι κάνουν αναγκαία την εφαρμογή ενός στοχευόμενου και πιο πρώιμου επαγγελματικού προσανατολισμού , ώστε να καλλιεργούνται ρεαλιστικές προσδοκίες.

Η **εξόρυξη δεδομένων** από την άλλη αποτελεί αντικείμενο μελέτης από πολλούς ερευνητές ειδικά τα τελευταία χρόνια με τη ραγδαία αύξηση του όγκου της πληροφορίας .Οι εφαρμογές της εκτείνονται στο Διαδίκτυο, παράδειγμα η μηχανή αναζήτηση της Google, όπου γίνεται αναζήτηση σε τεράστια αποθήκη δεδομένων σε πολύ μικρό χρόνο, αλλά και επιλογή των πιο χρήσιμων πληροφοριών ώστε να προβληθούν αρχικά. Εκτείνονται στην επιστήμη, με σημαντικό παράδειγμα το SKYCAT(SKY Catamaran), σύστημα εξόρυξης δεδομένων που αναλαμβάνει ανάλυση και κατηγοριοποίηση χωρικών αντικειμένων. Η ανίχνευση αυτών των αντικειμένων γίνεται από λήψη εικόνων με την εκτέλεση αλγορίθμων.

Εκτείνονται στο Μάρκετινγκ όπου αναζητούνται απαντήσεις γύρω από τις ανάγκες και τη καταναλωτική συμπεριφορά των πελατών. Με εφαρμογές της εξόρυξης δεδομένων εξάγονται πληροφορίες από μη δομημένα κείμενα για την πρόβλεψη τάσεων σε μετοχές στον τομέα των επενδύσεων. Ακόμη έχει αποτελεσματικές εφαρμογές στην πρόβλεψη απάτης στον τομέα της ασφάλειας και πρόληψης. Στην εκπαίδευση οι εφαρμογές εκτείνονται στο e-learning όπου η χρήση νέων μεθόδων αναζήτησης, έκφρασης και μετάδοσης της γνώσης παρέχει προκλήσεις για ανακάλυψη συσχετίσεων, προτύπων ή πρόβλεψη φαινομένων. Για την εκπαιδευτική έρευνα όμως επειδή τα σχολεία ως τα τελευταία χρόνια διέθεταν αυτόνομο σύστημα καταχώρησης στοιχείων δεν υπήρχαν μεγάλες αποθήκες δεδομένων. Το πανελλήνιο σχολικό δίκτυο που αποτελεί ενιαίο πληροφοριακό σύστημα μπορεί στο μέλλον να αποτελέσει μια αποθήκη δεδομένων από όπου θα εξαχθεί χρήσιμη γνώση.

Συνεισφορά – Στόχοι

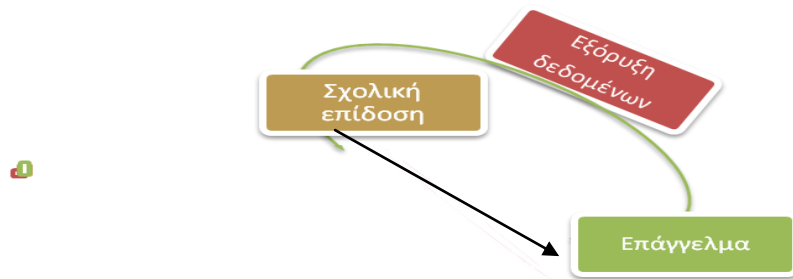
Στον τομέα της εκπαίδευσης συναντώνται δυο περιοχές που ειδικεύονται στη χρήση μεγάλου όγκου δεδομένων η **EDM** και **learning analysis**. Η EDM είναι μια επιστήμη γνώσης και μια πλούσια περιοχή εφαρμογής της εξόρυξης γνώσης, δεδομένου της αυξημένης διαθεσιμότητας εκπαιδευτικών δεδομένων. Θεωρείται ότι ενισχύει την καθοδηγούμενη από τα δεδομένα λήψη αποφάσεων με σκοπό τη βελτίωση της τρέχουσας εκπαιδευτικής πρακτικής και του εκπαιδευτικού υλικού που παρέχονται στο υπό εξέταση

πλαίσιο. (Calders et al.,2011). Το ερευνητικό πεδίο learning analysis εστιάζει στην ερμηνεία δεδομένων, τα οποία παράγονται και συλλέγονται για λογαριασμό των φοιτητών, με σκοπό την αξιολόγηση και την πρόβλεψη της μελλοντικής τους απόδοσης ή τον εντοπισμό προβλημάτων. Ωστόσο μπορεί να χρησιμοποιηθεί και στην αξιολόγηση προγραμμάτων σπουδών και κατάρτισης. Διαφέρουν στην περιοχή έρευνας . Η EDM αναζητά πρότυπα στα δεδομένα και στοχεύει στην ανάπτυξη νέων αλγόριθμων. Η learning Analysis εφαρμόζει γνωστά μοντέλα πρόβλεψης στα εκπαιδευτικά συστήματα. Υπάρχει ένα πλήθος ερευνών στην περιοχή της εκπαίδευσης. Οι έρευνες αυτές αποκαλύπτουν την τάση μελέτης και εκμετάλλευσης δεδομένων που προέρχονται από τα διαδικτυακά συστήματα μάθησης. Εφόσον η διεύρυνση της εξ αποστάσεως εκπαίδευσης, e-learning, κάνει δυνατή την καταγραφή μεγάλου όγκου δεδομένων. Στα πεδία ενδιαφέροντος είναι η ανακάλυψη της επίδοσης των μαθητών (πρόβλεψη) και η ανάλυση μαθησιακών συμπεριφορών (περιγραφή).

Στην Ελληνική βιβλιογραφία δεν αναφέρονται έρευνες που να διεξάγονται με την εφαρμογή τεχνικών εξόρυξης δεδομένων στην **πρωτοβάθμια εκπαίδευση**. Η **συνεισφορά** της έρευνας αυτής είναι ότι γίνεται επεξεργασία ενός μικρού αριθμού δεδομένων με τεχνικές εξόρυξης και δοκιμάζεται η ακρίβεια ενός προβλεπτικού μοντέλου στο χώρο της πρωτοβάθμιας εκπαίδευσης

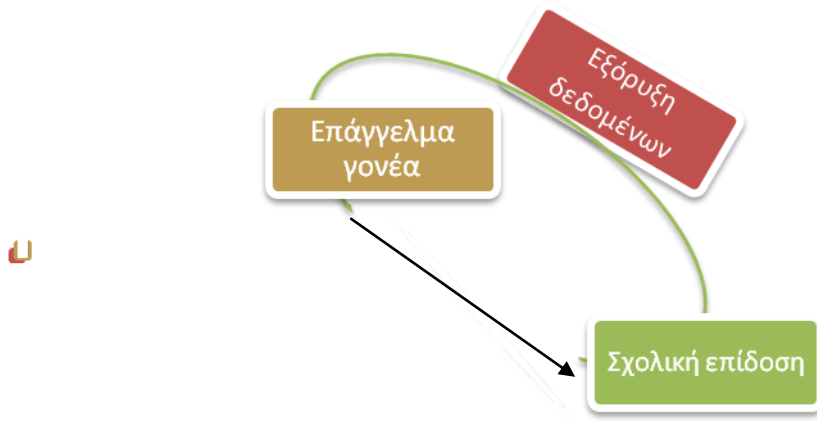
Στην παρούσα έρευνα με αφορμή τα συμπεράσματα που αναφέρθηκαν σε σχέση με την επίδοση και τις επικρατούσες συνθήκες, γίνεται μια ποσοτική και ποιοτική έρευνα , με αιώτερο σκοπό τη διερεύνηση των παρακάτω θεμάτων: Τι είναι αυτό που τελικά μεταφράζεται σε **σχολική επίδοση** και πως ειδικά η βαθμολογία επιδρά στις επαγγελματικές προσδοκίες στην επιλογή και την προσπάθεια επίτευξης στόχων; **Πώς έχουν βιώσει** τα ίδια τα άτομα τη βαθμολογία τους στην Στ Δημοτικού . Ήταν ένα γεγονός που πρόσφερε ανατροφοδότηση και αναπροσαρμογή στόχων ή ήταν απλά ένας βαθμός (**διαμορφωτική αξιολόγηση**). Ανίχνευαν δυνατότητες ή τους προκαλούσε απογοήτευση και αίσθηση αδικίας . Τελικά υπάρχει **συσχέτιση άμεση ή έμμεση με την επιλογή επαγγέλματος** και συναισθηματικής ισορροπίας μέσα σε αυτό;(αίσθηση δημιουργικότητας , επαγγελματικό άγχος). Ποιο ήταν το **μαθητικό προφίλ** των ατόμων που ερωτήθηκαν; Το **επάγγελμα των γονιών** είναι ένας παράγοντας ικανός να οδηγήσει στη δημιουργία μοντέλου πρόβλεψης ως προς τη σχολική επίδοση ώστε να γίνουν παρεμβάσεις σε πρώιμο στάδιο ; Πόσο χρήσιμα μπορεί να είναι τα εργαλεία του **Data Mining** στην εκπαίδευση και πως αυτά μοντελοποιώντας και δημιουργώντας από τα δεδομένα ,εργαλεία πρόβλεψης, μπορούν να διακρίνουν σχέσεις και ικανότητες οδηγώντας σε πρωτοβουλίες ενός περισσότερο στοχευόμενου επαγγελματικού προσανατολισμού ; Η παρούσα διπλωματική ίσως ανοίξει ένα δρόμο προς αυτή την κατεύθυνση ενσωματώνοντας τα σύγχρονα ευφυή συστήματα και τις δυνατότητες που αυτά προσφέρουν στη σύγχρονη εκπαιδευτική έρευνα . Σκοπός αυτής της διπλωματικής είναι να διερευνηθούν, με τις δυνατότητες που προσφέρουν τα εργαλεία του data mining, τα παραπάνω ερωτήματα.

Για το σκοπό αυτό δημιουργήθηκαν δυο μοντέλα . Το **A μοντέλο** ερευνά τη σχέση της σχολικής επίδοσης , όπως αυτή εκφράζεται με την αναλυτική βαθμολογία της Στ (Εκτης) Δημοτικού, με το επάγγελμα που τελικά τα άτομα που πήραν μέρος (60) ακολούθησαν. Αποσκοπώντας στη δημιουργία μοντέλου πρόβλεψης με υψηλά ποσοστά αξιολόγησης.



σχ.1 μοντέλο A

Το Μοντέλο B ερευνά τη σχέση του επαγγέλματος των γονέων με τη σχολική επίδοση των μαθητών της Στ τάξης . Το δείγμα περιλαμβάνει 250 μαθητές στην Στ τάξης συγκεκριμένου σχολείου στην περιοχή Καρλοβάσου κατά τη δεκαετία 2002-2012, αποσκοπώντας στη δημιουργία μοντέλου πρόβλεψης με υψηλά ποσοστά αξιολόγησης



σχ.2 Μοντέλο Β

Α Μοντέλο

Η βιβλιογραφική έρευνα με άξονες τη σχολική εξέλιξη, την επαγγελματική ανάπτυξη και την εξόρυξη δεδομένων οδήγησε στη σύνταξη ενός ερωτηματολογίου . Οι ερωτήσεις που συμπεριλήφθηκαν εξυπηρετούσαν τους παρακάτω **στόχους**:

- Να διερευνηθεί πιθανή σχέση ανάμεσα στη σχολική επίδοση της Στ δημοτικού και της επιλογής επαγγέλματος.
- Να διερευνηθεί η δυναμική που ασκεί η σχολική επίδοση ως βαθμολογία στους μαθητές και το οικογενειακό περιβάλλον τους.(διαμορφωτική αξιολόγηση, αυτορρύθμιση της γνώσης , επίδοση και ικανότητα)
- Να διερευνηθούν εξατομικευμένα χαρακτηριστικά (επαγγελματικό προφίλ , μαθητικό προφίλ) των ατόμων που συμμετείχαν.
- Να αξιολογηθεί μοντέλο πρόβλεψης ως προς την ικανότητα να προβλέπει το επάγγελμα από την επίδοση σε συγκεκριμένα μαθήματα της Στ Δημοτικού.

Υλοποίηση

Τα **ερωτήματα** που εξυπηρετούν τους παραπάνω στόχους συνοψίζονται ως εξής:

- Ήταν η σχολική επίδοσή μου αντιπροσωπευτική των ικανοτήτων ;

- Είχε η βαθμολογία την έννοια της διαμορφωτικής αξιολόγησης;
- Ποια η αντιμετώπιση της επίδοσης από το οικογενειακό περιβάλλον;
- Η σχολική επίδοση επηρέασε την επιλογή του επαγγέλματος;
- Το επάγγελμά αντικατοπτρίζει την εκπλήρωση των στόχων ;
- Ποιο ήταν το μαθητικό προφίλ;
- Ποιο το επαγγελματικό προφίλ;

Δημιουργήθηκαν λοιπόν 7 ομάδες διερευνητικών ερωτήσεων , 4 ομάδες που αφορούσαν δημογραφικά ζητήματα και 1 που αφορούσε την αναλυτική βαθμολογία στην Στ δημοτικού. Στο κεφάλαιο 5 που αφορά την μεθοδολογία παραθέτονται οι ερωτήσεις. Το ερωτηματολόγιο κατασκευάστηκε και στάλθηκε ηλεκτρονικά σε 100 άτομα. Τελικά απάντησαν τα 60.

Η επεξεργασία των δεδομένων έγινε με δίκτυα Bayes. Γενικά με τον όρο Bayesian ανάλυση εννοούμε μεθόδους που χρησιμοποιούν μοντέλα πιθανοτήτων για ποσότητες που γνωρίζουμε και για άλλες που θέλουμε να μάθουμε. Ειδικά τα δίκτυα Bayes είναι γραφικά μοντέλα που κωδικοποιούν πιθανολογικές σχέσεις μεταξύ μεταβλητών. Η επιλογή της συγκεκριμένης μεθόδου ανάλυσης έγινε επειδή πέρα από τη στατιστική επεξεργασία, παρέχεται η δυνατότητα επεξηγηματικής ανάλυσης αφού απεικονίζεται η σχέση των μεταβλητών και οι αλλαγές που συμβαίνουν, με την απομόνωση μιας μεταβλητής, στις άλλες.

Ενδεικτικά σύμφωνα με τα στατιστικά στοιχεία η σχολική επίδοση εμφανίζεται αντιπροσωπευτική των ικανοτήτων σε ποσοστό 59% πολύ και πάρα πολύ. Γενικά επικρατεί η άποψη ότι η βαθμολογία βοήθησε στο να καταλάβει ο ερωτώμενος που ήταν πραγματικά καλός .

Το οικογενειακό περιβάλλον δεν είχε αδιάφορη στάση απέναντι στη βαθμολογία . Ήταν αφορμή για επαινετικά σχόλια σε ποσοστό 72%. Ενώ για ένα 9% ήταν αφορμή για καβγάδες. Θεωρούν ότι η επίδοση ήταν ένας παράγοντας που καλλιέργησε επαγγελματικές φιλοδοξίες σε ποσοστό 34% πολύ και 38% λίγο.

Γενικά το μεγαλύτερο ποσοστό των ατόμων που συμμετείχαν είχαν καλύτερη επίδοση στη Γλώσσα και ήταν και στην προτίμησή τους καλύτερο ως μάθημα σε ποσοστό 72%. Στα μαθηματικά είχαν ελλείψεις και χαμηλή βαθμολογία Το δείγμα άνηκε κυρίως σε ποσοστό 51% στην **ομάδα 2** που περιλαμβάνει άτομα που ασκούν επιστημονικά, καλλιτεχνικά και συναφή επαγγέλματα. Εμφανίζονται ενδιαφέρουσες συσχετίσεις και πιθανολογικές σχέσεις που αναλύονται στο σχετικό κεφάλαιο π.χ.: Αν κάποιος έχει 10 στο μάθημα της Γλώσσας και κάτω από 9 στα μαθηματικά έχει 53% πιθανότητες να ακολουθήσει κάποιο από τα επαγγέλματα της ομάδας 2 .

Στη συνέχεια κατασκευάστηκε μοντέλο **κατηγοριοποίησης** με σκοπό : **περιγραφικό** . Ως επεξηγηματικό εργαλείο για τη διάκριση των διαφορετικών κατηγοριών . **Προβλεπτικό** .Ως μοντέλο πρόβλεψης για το επάγγελμα που πιθανό να ακολουθήσει κάποιος που δεν ανήκει στην ομάδα του δείγματος και έχει μια

συγκεκριμένη αναλυτική βαθμολογία. Τα πειράματα που έγιναν και αναλύονται στο αντίστοιχο κεφάλαιο κατέληξαν : Το σύνολο εκπαίδευσης που εμφάνισε τη μεγαλύτερη αξιοπιστία ήταν αυτό που η κλάση πρόβλεψης είχε τη διωνυμική τιμή ομάδα 4- όχι ομάδα 4 με 83,33% στα δέντρα απόφασης με τον αλγόριθμο ID3 και στα Νευρωνικά δίκτυα με 80% validation Με υψηλή αξιοπιστία 85% εμφανίζεται και το σύνολο εκπαίδευσης που η κλάση πρόβλεψης έχει τις διωνυμικές τιμές ομάδα 5 –όχι ομάδα 5 με Νευρωνικά δίκτυα . (Αναλυτική παρουσίαση στο κεφάλαιο)

B Μοντέλο

Τα δεδομένα για το B μοντέλο συλλέχθηκαν από το βιβλίο μητρώου των μαθητών του 1^{ου} Δημ. Σχολείου Καρλοβάσου και αφορούν 250 μαθητές που φοίτησαν στο σχολείο αυτό στην Στ´ (Εκτη Δημοτικού) τη δεκαετία 2002-2012. Από το αρχείο Excel που αποθηκεύτηκαν αρχικά μεταφέρθηκαν σε αρχείο CSV και με το εργαλείο Genie 2.4 αναζητήθηκαν συσχετίσεις στα δίκτυα Bayes. Στα δίκτυα Bayes εμφανίζεται πιθανολογική σχέση με το προσανατολισμένο βέλος από το επάγγελμα του πατέρα στο επάγγελμα της μητέρας όπου αν ο πατέρας ανήκει στην ομάδα 2 η μητέρα ανήκει επίσης στην ομάδα 2 ή ασχολείται με οικιακά. Αν ο πατέρας στην ομάδα 4 στην ίδια ομάδα ανήκει και η μητέρα. Πιθανολογείται λοιπόν μια προσπάθεια διατήρησης του κοινωνικοοικονομικού στάτους με παρόμοια επαγγέλματα στην οικογένεια. Από το επάγγελμα της μητέρας ξεκινά προσανατολισμένο βέλος προς τη μητρική γλώσσα . Αν το επάγγελμα της μητέρας ανήκει στην ομάδα 9 (ανειδίκευτος εργάτης) κατά 63% η μητρική γλώσσα είναι η Αλβανική. Οι υπόλοιπες συσχετίσεις αφορούν ομάδες μαθημάτων . Στη συνέχεια εκπαιδευτήκε σύστημα πρόβλεψης και έγιναν κάποια πειράματα με εναλλαγές ως προς την κλάση πρόβλεψης , άλλοτε γλώσσα , άλλοτε μαθηματικά ως ονοματικές κλάσεις . Τελικά το σώμα εκπαίδευσης με τη μεγαλύτερη ακρίβεια είναι αυτό με μεταβλητές το επάγγελμα της μητέρας, το επάγγελμα του πατέρα, το φύλλο, τη μητρική γλώσσα και με κλάση πρόβλεψης το μέσο όρο της αναλυτικής βαθμολογίας

Διάρθρωση εργασίας

Η εργασία είναι χωρισμένη σε δυο μέρη :το Θεωρητικό και το πειραματικό. Το θεωρητικό κομμάτι περιλαμβάνει τα κεφάλαια 1^ο , 2^ο , 3^ο , 4^ο . Το πειραματικό κομμάτι αναλύεται στο 6^ο κεφάλαιο. Πιο συγκεκριμένα:

Στο 1^ο Κεφάλαιο γίνεται αναφορά στο ιδιαίτερο ερευνητικό πεδίο της εξόρυξης δεδομένων , το Educational Data Mining, EDM. Αναφέρονται οι εξειδικευμένες τεχνικές της εξόρυξης γνώσης από τις αποθήκες εκπαιδευτικών δεδομένων και παρουσιάζονται συγκεκριμένες έρευνες και αντικείμενα.

Στο 2^ο κεφάλαιο γίνεται βιβλιογραφική έρευνα γύρω από την έννοια της σχολικής επίδοσης και αναλύονται έννοιες που σχετίζονται με αυτή, μέτρηση, αξιολόγηση, γνωστική διαδικασία.

Στο 3^ο κεφάλαιο παρουσιάζονται θεωρίες γύρω από την επαγγελματική ανάπτυξη του ατόμου.

Στο 4^ο κεφάλαιο αναλύονται οι τεχνικές Data Mining που χρησιμοποιήθηκαν στα πειράματα και γίνεται μια συνοπτική παρουσίαση των δικτύων Bayes.

Στο 5^ο κεφάλαιο περιγράφεται η μεθοδολογία της έρευνας , εργαλεία , τρόποι και τεχνικές άντλησης και ανάλυσης των δεδομένων.

Στο 6^ο κεφάλαιο περιγράφονται τα πειράματα και τα περιγραφικά και προβλεπτικά μοντέλα που εξάγονται, καθώς και η απόδοσή τους.

Το 7^ο κεφάλαιο περιλαμβάνει γενικά συμπεράσματα και κάποιες έρευνες που ενώ έχουν διεξαχθεί με παραδοσιακές μεθόδους στηρίζουν τα συμπεράσματα.

Κεφάλαιο 1

ΠΡΟΓΕΝΕΣΤΕΡΗ ΕΡΓΑΣΙΑ

1.1 Εξόρυξη δεδομένων στην εκπαίδευση(Educational data mining)

«Για πρώτη φορά στην ιστορία της ανθρωπότητας, παρέχεται η δυνατότητα να έχει ο καθένας τη γνώση που επιθυμεί» (Wiley Green & Soang, 2012). Η διαπίστωση αυτή συνοψίζει την έκρηξη της εξ αποστάσεως μάθησης που τα τελευταία χρόνια, έφερε επανάσταση στην εκπαίδευση. Η χρήση νέων μεθόδων έκφρασης, μετάδοσης και αναζήτησης της γνώσης, είχε συνέπεια τη δημιουργία μεγάλων αποθηκών πληροφοριών, οι οποίες αξίζει να ερευνηθούν ώστε να ανακαλυφθεί κρυμμένη και χρήσιμη γνώση. Γεγονός που οδήγησε στην ανάπτυξη τεχνικών για την ανάλυση αυτών των δεδομένων.

EDM educational data mining είναι ένα ερευνητικό πεδίο που εφαρμόζει εξόρυξη δεδομένων, μηχανική μάθηση και στατιστική, σε πληροφορίες από εκπαιδευτικά περιβάλλοντα: πανεπιστήμια, ευφυή συστήματα διδασκαλίας. Στόχος είναι η ανακάλυψη νέων ιδεών για το πώς μαθαίνουν οι άνθρωποι, συμβάλλοντας έτσι στην εκπαιδευτική ψυχολογία και τις επιστήμες της μάθησης.

ΙΣΤΟΡΙΑ : Από το 2000 ως το 2007 διεξήχθησαν έρευνες που στόχευαν στην ανάπτυξη αυτών των τεχνικών. Τα πρώτα αποτελέσματα παρουσιάζονται σε διεθνές συνέδριο, το 2008, στο Μόντρεαλ. Το αυξημένο ενδιαφέρον των ερευνητών οδήγησε στην έκδοση περιοδικού το 2009. Πρόκειται για την **Journal of educational data mining (JEDM)**, ένα διεθνές διεπιστημονικό φόρουμ της έρευνας σχετικά με την ανάλυση δεδομένων στην εκπαίδευση με στόχο να δώσει απαντήσεις σε εκπαιδευτικούς, να εξερευνήσει δεδομένα από ρυθμίσεις κ.α.. Περιλαμβάνει άρθρα που αφορούν :

- Διεργασίες που ακολουθούνται για την ανάλυση εκπαιδευτικών δεδομένων,
- Ενσωμάτωση εξόρυξης δεδομένων με παιδαγωγικές θεωρίες.
- Προγράμματα βελτίωσης εκπαιδευτικών λογισμικών.
- Βελτίωση, κατανόησης αναπαράστασης χώρου των μαθητών.

Το 2011 ιδρύεται το international educational data mining society, το National center for education statistics NCES αποτελούν κόμβους ανταλλαγής των αποτελεσμάτων των ερευνητών.

Για τον Ιούνιο του 2015 έχει προγραμματιστεί το διεθνές συνέδριο της EDM στη Μαδρίτη και μερικά από τα θέματα που θα συζητηθούν είναι: «αξιολόγηση της αποτελεσματικότητας του προγράμματος

σπουδών και των παρεμβάσεων», «ανίχνευση της συναισθηματικής κατάστασης των φοιτητών» «παροχή ανατροφοδότησης δασκάλων, σπουδαστών με τις τεχνικές της EDM», «εξόρυξη δεδομένων από εκπαιδευτικά παιχνίδια και διερευνητική μάθηση» κ.α.

1.2 Ερευνητικά αντικείμενα- Εφαρμοζόμενα προγράμματα EDM

Η EDM έχει εφαρμοστεί σε πολλές εκπαιδευτικές έρευνες και αντικείμενα, με σκοπό τη βελτίωση της διαδικασίας της μάθησης. Κεντρικός άξονας των ερευνητικών αντικειμένων είναι η **μοντελοποίηση του μαθητή**. Το μοντέλο αυτό περιλαμβάνει πληροφορίες για δεξιότητες, κίνητρα μάθησης, μεταγνώση, εμπειρίες, γνωστική ανάπτυξη, μαθησιακά αποτελέσματα κ.α. Γύρω από το πρότυπο του μαθητή αναζητούνται τα εξής

Πρόβλεψη της επίδοσης και μαθησιακών αποτελεσμάτων. Στηρίζεται στη μελέτη δεδομένων που πλαισιώνουν το μαθησιακό αντικείμενο. **Ανάλυση συμπεριφοράς εκπαιδευόμενου.** Αφορά το μαθησιακό προφίλ και τη λήψη αποφάσεων για εφαρμογή κατάλληλης μεθοδολογίας στη διδασκαλία. **Επικοινωνία με εμπλεκόμενους.** Αφορά την παροχή ανατροφοδότησης στους εκπαιδευτικούς για την μαθησιακή πορεία του εκπαιδευόμενου σε πραγματικό χρόνο, με σκοπό τη λήψη αποφάσεων για τη βελτίωση της διαδικασίας της μάθησης (Romero C., Ventura S., 2010). **Περιοχή Δομημένης ανάλυσης** Το αντικείμενο είναι η δόμηση και βελτίωση του περιεχομένου μιας εκπαιδευτικής διαδικασίας. Γίνεται εφικτό με την πρόβλεψη της μαθησιακής επίδοσης κάθε εκπαιδευόμενου μέσω των αποθηκευμένων μαθησιακών επιδόσεων. **Διατήρηση και βελτίωση μαθημάτων.** Το αντικείμενο εδώ αφορά το περιεχόμενο και τη δομή των μαθημάτων και οι πληροφορίες λαμβάνονται από πληροφορίες για το λειτουργικό ή χρηστικό αντίκτυπο των μαθημάτων στους μαθητές. **Μελέτη επιπτώσεων των διαφορετικών τύπων παιδαγωγικής υποστήριξης.** Αναλύεται η χρήση διαδικτυακών και παιδαγωγικών λογισμικών για την υποστήριξη της εκπαιδευτικής διαδικασίας. **Προώθηση επιστημονικής Γνώσης.** Αφορά τη δόμηση, ανακάλυψη ή βελτίωση μοντέλων του μαθητή ή της μαθησιακής διαδικασίας

Μερικές σημαντικές έρευνες της EDM αναφέρονται στη συνέχεια:

Η πρόβλεψη της επίδοσης των σπουδαστών είναι ένα πρόβλημα το οποίο έχει κερδίσει την προσοχή της εκπαιδευτικής έρευνας στην εξόρυξη δεδομένων. Η παραγοντοποίηση μητρώων είναι μια τεχνική που αφορά την κατάτμηση πινάκων δεδομένων , έχει επιδείξει τις καλύτερες προβλέψεις. Η έρευνα των Nguyen Thai –Nghe et al, εκ μέρους του Πανεπιστημίου Hidesheim συσχέτισε με την τεχνική της Κατάτμησης Μητρώων το στόχο , την επιτυχία του στόχου αλλά και την πρόσκτηση δεξιοτήτων που απαιτούνται για το στόχο. (Nguyen Thai-Nghe, Drumont L., et al, 2011)

Στον Καναδά εφαρμόζεται πρόγραμμα πρόβλεψης σε περίπου 225.000, φοιτητές για την εκδοχή εγκατάλειψης σπουδών. Το πρόγραμμα ονομάζεται DEWS Dropout Early Warning System του Wisconsin και πρόκειται για μια ευέλικτη σειρά λογισμικών που μπορεί να προβλέψει την εγκατάλειψη των σπουδών, στις βαθμολογίες 6-9, αλλά και να περιγράψει συνθήκες που προαναγγέλλουν μια τέτοια εξέλιξη. (Knowles E. Jared, 2015)

Στόχος της επόμενης έρευνας ήταν η εξατομίκευση στην ανάκτηση πληροφοριών στα πλαίσια της εκπαίδευσης. Δημιουργήθηκε μοντέλο πρόβλεψης από τη βάση δεδομένων των ψηφιακών βιβλιοθηκών. Διερευνήθηκε η δυνατότητα ενσωμάτωσης των πληροφοριών που συνοδεύουν τις εγγραφές των φοιτητών ώστε να παραχθούν εξατομικευμένα αποτελέσματα αναζήτησης. Προτάθηκε, εφαρμόστηκε και αξιολογήθηκε με πολύ καλά αποτελέσματα, η εφαρμογή με το όνομα PersoLib. (Alaofi M., Rumantir G., 2015)

Από τους Zimmerman J. Brodersen K. διεξήχθη έρευνα σε 171 αρχεία σπουδαστών της πληροφορικής και την ανάλυση 81 μεταβλητών με στόχο την δημιουργία μοντέλου πρόβλεψης δεικτών απόδοσης για τους προπτυχιακούς φοιτητές. Η ανάλυση έδειξε καλή αξιολόγηση για ενιαίο σύνολο δυνατοτήτων των σπουδαστών. Ειδικά υπήρχε ακρίβεια πρόβλεψης για το τρίτο έτος σπουδών. Απώτερος σκοπός η εφαρμογή μεθοδολογίας που θα προσφέρει καλύτερη ποιότητα σπουδών στην τριτοβάθμια εκπαίδευση. (Zimmerman J, Brodersen K., Heinimann H., Buhmann J., 2015)

Ανάλυση δεδομένων (Συσταδοποίηση), που αφορούσαν τη χρήση στρατηγικών που χρησιμοποιούν οι σπουδαστές για να ανέβουν επίπεδο σε εκπαιδευτικά παιχνίδια και διεξήχθη στην Καλιφόρνια, έδειξε ότι με τη χρήση τεχνικών εξόρυξης δεδομένων μπορούν να αναβαθμιστούν τέτοιου είδους παιχνίδια με τέτοιο τρόπο ώστε η χρήση τους να μην επηρεάζει αρνητικά την επίδοση των φοιτητών. (Kerr Deirdre, 2015)

Την επιλογή όμως της κατάλληλης τεχνικής και μεθοδολογίας, την ορίζουν τα ίδια τα δεδομένα. Πιο συγκεκριμένα όταν τα δεδομένα σχετίζονται με τη συμπεριφορά και το γνωστικό στυλ των φοιτητών και σα στόχος τίθεται η πρόβλεψη της επίδοσης τότε η τεχνική που επιλέγεται είναι η **κατηγοριοποίηση**. Αν τα δεδομένα είναι οι τελικοί βαθμοί και τελικός στόχος η αναγνώριση και ταξινόμηση των δραστηριοτήτων τους απέναντι σε αυτή την επίδοση, η τεχνική που επιλέγεται είναι η **παλινδρόμηση**. Στην περίπτωση δεδομένων που αφορούν βαθμολογία σε επιμέρους μαθήματα και ο στόχος είναι η πρόβλεψη της τελικής βαθμολογίας τότε οι τεχνικές που χρησιμοποιούνται είναι **κατηγοριοποίηση, νευρωνικά δίκτυα**. Αν τα δεδομένα σχετίζονται με την προσπάθεια που καταβάλλεται για την ολοκλήρωση μιας εργασίας και ο στόχος που τίθεται είναι η ανάλυση των τρόπων που οι εκπαιδευόμενοι χρησιμοποιούν τις πηγές, η τεχνική που χρησιμοποιείται είναι οι **κανόνες συσχέτισης** (association rules). (Bousbia N. et al, 2010)

1.3 Μέθοδοι και τεχνικές

Για τη μελέτη των παραπάνω αντικειμένων χρησιμοποιούνται οι μέθοδοι που χρησιμοποιούνται και στο ευρύτερο πεδίο της εξόρυξης δεδομένων και αναλύονται στο σχετικό κεφάλαιο . Κατηγοριοποίηση, Συσταδοποίηση, κανόνες συσχέτισης. Ωστόσο το πεδίο της EDM αποτελεί ένα ιδιαίτερο πεδίο με χαρακτηριστικά που απαιτούν διαφορετική ανάλυση των δεδομένων. Για το λόγο αυτό οι ερευνητές διεύρυναν τις παραπάνω τεχνικές . Οι μέθοδοι που χρησιμοποιούνται συνοψίζονται σε 5 κατηγορίες

Πρόβλεψη (Prediction) Στόχος είναι να αναπτυχθεί ένα πρότυπο- μοντέλο που μπορεί να συμπεράνει κάποιες πτυχές του προβλήματος με την είσοδο κάποιων άλλων μεταβλητών οι οποίες διαθέτουν συγκεκριμένα γνωρίσματα- ετικέτες. Η πρόβλεψη στην εκπαιδευτική εξόρυξη δεδομένων έχει δύο χρήσεις. Σε κάποιες περιπτώσεις , όπως τα προγράμματα που κάνουν πρόβλεψη για την έκβαση των σπουδών των φοιτητών, μελετούν τα χαρακτηριστικά γνωρίσματα του προτύπου που είναι σημαντικά για την πρόβλεψη. Σε άλλες περιπτώσεις όμως η μεταβλητή μπορεί να αλλάξει συμπεριφορά όταν μετατραπεί σε πρότυπο, όπως για παράδειγμα συμβαίνει σε μεταβλητές που αφορούν συναισθήματα, τότε η πρόβλεψη αφορά την αξία της μεταβλητής που χρησιμοποιείται. Η πρόβλεψη γίνεται κυρίως με κατηγοριοποίηση (classification), παλινδρόμηση (regression) όπου από ανεξάρτητες μεταβλητές προβλέπουν τις τιμές μιας εξαρτημένης. Και μέτρηση πυκνότητας (density estimation) .

Συσταδοποίηση(Clustering). Πρόκειται για ανακάλυψη στιγμιότυπων που ομαδοποιούνται και κατά συνέπεια χωρίζουν το σύνολο δεδομένων σε υποσύνολα. Όπως για παράδειγμα την ομαδοποίηση μαθητών ανάλογα με την επίδοσή τους.

Σχεσιακή εξόρυξη(Relationship mining). Πρόκειται για την αναζήτηση συσχετίσεων μεταξύ των μεταβλητών (Bayes Network) ή την αναζήτηση κανόνων συσχέτισης που τις διέπουν. Για παράδειγμα η ανακάλυψη σχέσεων ανάμεσα στις διαδικτυακές δραστηριότητες των μαθητών και της τελικής τους επίδοσης.(Romero C., Ventura S., 2010)

Απόσταξη δεδομένων για ανθρώπινη κρίση(distillation of data for human judgment). Πρόκειται για την αναπαράσταση των δεδομένων ώστε να εντοπίζονται οι χρήσιμες πληροφορίες για τη λήψη αποφάσεων.

Ανακάλυψη με μοντέλα (Discovery with models). Χρήση μοντέλων με σκοπό τη βαθύτερη κατανόηση φαινομένων. Για παράδειγμα ένα μοντέλο χρησιμοποιείται για την ανακάλυψη συσχετίσεων ανάμεσα στους μαθητές και το μαθητικό προφίλ τους.

Ανίχνευση ακραίων τιμών(outlier detection). Ανακάλυψη μέσα σε ένα σύνολο δεδομένων σημείων που παρουσιάζουν ακραίες τιμές σε σχέση με τα υπόλοιπα.

Ανάλυση κοινωνικών δικτύων (Social network analysis) .Αφορά τις συσχετίσεις που δημιουργούνται σε μια συνεργατική ή άλλη ομάδα.

Εξόρυξη γνώσης (process Mining). Διακρίνεται σε τρεις φάσεις. Α) έλεγχος συμμόρφωσης, Β)ανακάλυψη μοντέλου και Γ) επέκταση μοντέλου. Για παράδειγμα από τους βαθμούς ,την επιλογή μαθημάτων , τις παρακολουθήσεις μπορεί να αποτυπωθεί η συμπεριφορά του εκπαιδευόμενου.

Εξόρυξη από δεδομένα κειμένου (Text Mining). Πρόκειται για ανακάλυψη χρήσιμων και ενδιαφερόντων προτύπων , τάσεων κ.α. μέσα από ένα μη δομημένο κείμενο (μηνύματα συνομιλίας από ηλεκτρονική αλληλογραφία, περιοχές συζητήσεων, φόρουμ, ιστοσελίδες κ.α.). Περιλαμβάνει κατηγοριοποίηση κειμένου, ανάλυση συναισθημάτων, σύνοψη αρχείων, παραγωγή μοντέλων.

Ιχνηλάτηση της γνώσης (Knowledge tracking). Πραγματεύεται την πορεία της γνωστικής κατάκτησης του μαθητή και χρησιμοποιείται για πρόβλεψη συμπεριφοράς μέσα από την πορεία κατάκτησης της γνώσης.

Παραγοντοποίηση μητρώων(Matrix factorization) Πρόκειται για τη διάσπαση ενός πίνακα δεδομένων σε επιμέρους πίνακες ώστε να αναλυθούν κάποια χαρακτηριστικά που παρουσιάζουν ενδιαφέρον ιδιαίτερα. Για παράδειγμα έστω πίνακας που περιέχει δοκιμασίες και αποτελέσματα εκπαιδευόμενου . Ο πίνακας αυτός μπορεί να διασπαστεί σε δυο πίνακες δεξιοτήτων και τιμών αντίστοιχα . Σε μια τέτοια διάσπαση στηρίχτηκε η ερευνητική εργασία του Nguyen Thai-Nghe το 2011 όπως προαναφέρθηκε.

ΚΕΦΑΛΑΙΟ 2

ΣΧΟΛΙΚΗ ΕΠΙΔΟΣΗ

2.1 Αξιολόγηση

Αρχικά θα πρέπει να διασαφηνιστεί η έννοια της αξιολόγησης αφού ο τρόπος κυρίως αξιολόγησης της σχολικής επίδοσης (βαθμολογία) αφορά την παρούσα εργασία. Η αξιολόγηση γενικότερα είναι η απόδοση ορισμένης αξίας σε αντικείμενα, πρόσωπα, καταστάσεις ή γεγονότα. Αξία που αποδίδεται με βάση συμφωνημένα κριτήρια με την εφαρμογή μιας κατάλληλης μεθόδου αποτίμησης. Η αξία αυτή συνήθως αφορά την α) απόδοση ορισμένης θετικής ή αρνητικής ιδιότητας σε ότι αξιολογείται, β) στο αποτέλεσμα που προκύπτει μετά τη σύγκριση αυτού που αξιολογείται με κάποιο ομοειδές αντικείμενο πρόσωπο ή κατάσταση ως προς ένα ή περισσότερα ορισμένα χαρακτηριστικά και γ) στο βαθμό επίτευξης ενός αρχικού και συγκεκριμένου στόχου. (Κασσωτάκης Μ., 2013)

Στους Stufflebeam(1971) και Alkin((Alkin M.C., 1974)) οφείλεται ο ορισμός που χρησιμοποιείται μέχρι σήμερα. Η αξιολόγηση είναι περισσότερο μια διαδικασία παρά ένα αποτέλεσμα, μια διαδικασία συλλογής πληροφοριών για την επιλογή εναλλακτικών λύσεων και για τη λήψη αποφάσεων. Λειτουργεί δηλαδή ως ένας μηχανισμός ανατροφοδότησης του σχεδίου ενέργειας με σκοπό τη συνεχή βελτίωσή του.

Η αξιολόγηση στο σχολείο διακρίνεται στη **διαγνωστική αξιολόγηση (diagnostic)** η γίνεται στην αρχή της διδακτικής διαδικασίας και στοχεύει στην ανίχνευση δυσκολιών και σε αναπροσαρμογή των στόχων αν αυτό είναι αναγκαίο. Στη **διαμορφωτική αξιολόγηση** που είναι η αξιολόγηση που γίνεται κατά τη διάρκεια της διδακτικής διαδικασίας. Αποσκοπεί στον έλεγχο της πορείας της διαδικασίας και σε βελτιωτικές ενέργειες αν αυτές απαιτούνται. Σύμφωνα με τις έρευνες των Black, William, Brookhart (2005) (Brookhart S.M, 2008) η διαμορφωτική αξιολόγηση βελτιώνει την επίδοση, προαπαιτεί όμως σαφήνεια στη διατύπωση των διδακτικών στόχων σε ατομικό επίπεδο ή σε επίπεδο τάξης. Στην **αθροιστική ή τελική αξιολόγηση** που έχει διαπιστωτικό χαρακτήρα και πληροφορεί το μαθητή για το επιδιωκόμενο συνολικό επίτευγμα.

2.1.1 Μέτρηση

Η **μέτρηση** είναι μια άλλη έννοια αλληλένδετη με την αξιολόγηση αλλά όχι ταυτόσημη. Η μέτρηση προηγείται της αξιολόγησης και πρόκειται είτε για σύγκριση του αντικειμένου με μια συγκεκριμένη σταθερή μονάδα είτε για τη συχνότητα εμφάνισης ενός χαρακτηριστικού ή φαινομένου σε προκαθορισμένο χρονικό διάστημα. Η επιστημονική φύση της μέτρησης την καθιστά ακριβή. Δε συμβαίνει το ίδιο με την αξιολόγηση όπου παρά τις προσπάθειες να εξαλειφθεί ο υποκειμενικός παράγοντας, συχνά δεν είναι αυτό εφικτό

. Ωστόσο η μέτρηση στην εκπαίδευση δε χαρακτηρίζεται από την ακρίβεια των θετικών επιστημών αφού ο βαθμός αντικατοπτρίζει μια σχετική κατάσταση. Για παράδειγμα εκτός από μεταβλητές που αφορούν εκπαιδευτικές δαπάνες κ.α. που σχετίζονται με ακριβείς μετρήσεις στην περίπτωση της απόδοσης βαθμού σε ένα τεστ ή στη βαθμολογία του τριμήνου το 0 δε σημαίνει μηδενικές γνώσεις αλλά και το 5 με άριστα το 10 δε σημαίνει κατοχή των μισών γνώσεων ενώ συχνά η ίδια βαθμολογία δύο μαθητών σε ένα μάθημα δε σημαίνει και ίδια γνώση ή δεξιότητα αφού η ικανότητα κατανόησης ενός κειμένου μπορεί για παράδειγμα να αντισταθμίζεται με την ικανότητα στην ανάκληση πληροφοριών ή τη σαφήνεια του προφορικού λόγου και το αποτέλεσμα της μέτρησης- βαθμολογίας, να είναι -9-.

Η μέτρηση λοιπόν στην Εκπαίδευση έχει την έννοια της «απόδοσης στα αντικείμενα, στα πρόσωπα και στα πράγματα αριθμητικών τιμών κατά τέτοιο τρόπο, ώστε οι σχέσεις που υπάρχουν ανάμεσα στις τιμές αυτές να αντιστοιχούν στις σχέσεις που υπάρχουν μεταξύ εκείνων στα οποία αποδίδονται» (Eysenck et al. 1975 :639, αναφορά Κασσωτάκης, 2013) Εξάλλου ο Cangelosi θεωρεί τη μέτρηση στην εκπαίδευση ως διαδικασία συλλογής δεδομένων μέσω της εμπειρικής παρατήρησης. (Cangelosi . J, 1990) Οι Belanoff και Dickson (Belanoff P.& Dickson M., 1991)θεωρούν κατά τον ίδιο τρόπο τη μέτρηση στην εκπαίδευση ως διαδικασία ποσοτικοποίησης της επίδοσης ή ως απόδοση αριθμών σε αυτή. Και ο Airasian (1994) θεωρεί τη μέτρηση στην εκπαίδευση ως διαδικασία ποσοτικοποίησης και απονομής αριθμών σε επιδόσεις, ορισμός που αντικατοπτρίζει τη βαθμολογία.(αναφορά: (Κασσωτάκης M., 2013) Με αυτή την έννοια η βαθμολογία είναι μια μέτρηση αφού με τους βαθμούς εκφράζεται η σχέση ανάμεσα στη συγκεκριμένη επίδοση ενός μαθητή και του επιπέδου επίτευξης των επιδιωκόμενων στόχων.

2.1.2 Έννοια και κλίμακες βαθμολόγησης

Η αξιολόγηση λοιπόν είναι μια διαδικασία αποτίμησης ενός πράγματος ή μιας κατάστασης (Ματσαγγούρας Η., 2001)). Για να επιτευχθεί απαραίτητη προϋπόθεση είναι τα κριτήρια και η συγκεκριμένη διαδικασία. Το αποτέλεσμα της αξιολόγησης είναι η έκφραση της αποτίμησης με τη βοήθεια τυπικής ή άτυπης βαθμολογικής κλίμακας. Όταν η έκφραση αυτού του αποτελέσματος της

αξιολόγησης γίνεται με τη βοήθεια τυπικής βαθμολογικής κλίμακας τότε μιλάμε για βαθμολογία. Ωστόσο η έννοια της βαθμολογίας και της αξιολόγησης παρά το ότι συνδέονται , χωρίς αξιολόγηση δεν είναι δυνατή η βαθμολόγηση, δεν αποτελούν ταυτόσημες έννοιες.

Η αξιολόγηση επιτυγχάνεται με διάφορους τρόπου και είδη κλιμάκων. Οι πιο γνωστές και αυτές που εφαρμόζονται στην πρωτοβάθμια εκπαίδευση είναι: Η **αριθμητική κλίμακα** που εφαρμόζεται μέχρι σήμερα στην ελληνική εκπαίδευση για τις δυο τελευταίες τάξεις του δημοτικού. Βασικό χαρακτηριστικό των αριθμητικών κλιμάκων γενικότερα είναι ο ορισμός μιας βάσης , ενός ανώτερου βαθμού που είναι το άριστα και οι ενδιάμεσες διαβαθμίσεις που επιτρέπουν την άμεση σύγκριση των μαθητών ,την ιεραρχική κατάταξη και τελικά τη στατιστική επεξεργασία των δεδομένων. Στο δημοτικό χρησιμοποιείται η δεκάβαθμη κλίμακα 1-10 , γεγονός που δεν επιτρέπει λεπτές διαφοροποιήσεις αλλά βοηθάει στην αποφυγή λάθους καθώς είναι γεγονός ότι τα κριτήρια που μπορούν να εφαρμοστούν αλλά και οι διαδικασίες αξιολόγησης της σχολικής επίδοσης των μαθητών στην Ε´ και Στ δημοτικού δεν επιτρέπουν την λεπτή διακριτικότητα που θα ισοδυναμούσε μια μεγαλύτερη κλίμακα π.χ. 1-20 .

Η **κλίμακα γραμμάτων** εφαρμόστηκε στο ελληνικό σχολείο τη δεκαετία του 1980 και τα γράμματα της αλφαβήτας αντικαθιστούν τους αριθμούς . Ωστόσο οι πληροφορίες που δίνει για το μαθητή η κλίμακα των γραμμάτων έχει τους ίδιους περιορισμούς με την αριθμητική . Το 7 ή το Γ δεν δίνει στην ουσία καμιά πληροφορία για τα συγκεκριμένα προβλήματα ή δυσκολίες του μαθητή περιορίζοντας έτσι την έννοια της αξιολόγησης ως ιεραρχικής κατάταξης, στερώντας της την πιο σημαντική και χρήσιμη αξία που είναι η ανατροφοδότηση.

Η **φραστική κλίμακα** αντικαθιστά τα γράμματα και τους αριθμούς με φράσεις . «Σχεδόν καλά», «Καλά», «Πολύ καλά», «Άριστα» . Δεν επιτρέπει τη στατιστική επεξεργασία και όπως και οι προηγούμενες αδυνατεί να προσφέρει ανατροφοδότηση στους εμπλεκόμενους στη διδακτική διαδικασία. Η **περιγραφική κλίμακα** περιγράφει αναλυτικά την επίδοση του μαθητή στους τομείς που συγκροτούν τη συνολική του εικόνα μέσα στην τάξη και έξω από αυτήν. Συνεκτιμώνται ακαδημαϊκή επίδοση, προσοχή ,ενδιαφέρον και προσπάθεια αλλά και ανάληψη πρωτοβουλιών, εργασία χωρίς συνεχή στήριξη ,ομαδικό πνεύμα κ.α. . Δεν πρόκειται για ιεραρχική κλίμακα αλλά για ελεύθερη έκφραση αποτιμήσεων. Ωστόσο επειδή στην καθημερινή διαδικασία ο δάσκαλος δεν έχει τον απαιτούμενο χρόνο και την εξειδίκευση για συστηματική παρατήρηση και καταγραφή εφαρμόζεται σύμφωνα με το Π.Δ 409/94 και 8/95 δελτίο περιγραφικής αξιολόγησης που χρησιμοποιείται παράλληλα με την αριθμητική βαθμολόγηση στις τάξεις Ε και Στ ενώ αποκλειστικά στην Α´ και Β´ που εφαρμόζεται μόνο περιγραφική αξιολόγηση.

2.1.3 Μέθοδοι αξιολόγησης

Η αξιολόγηση απαιτεί τη συγκέντρωση δεδομένων. Οι ερευνητές χωρίζουν τις μεθόδους συλλογής σχετικών δεδομένων σε δυο κατηγορίες :

A)Στις εξεταστικές διαδικασίες γραπτές ή προφορικές , καθημερινές και περιοδικές , ατομικές ή ομαδικές.

B)Στην παρατήρηση που μπορεί να είναι άτυπη ή συστηματική. Η συστηματική γίνεται με προγραμματισμό και καθορισμένη τεχνική . Η άτυπη σχετίζεται με την αξιολόγηση συμπεριφορών σχετικών με τη μαθησιακή διαδικασία και αποτελεί περισσότερο περιγραφική αποτίμηση του προφίλ του μαθητή . Η ανάληψη πρωτοβουλιών , το ενδιαφέρον, η προσοχή , η ανάληψη πρωτοβουλιών, οι συνεργατικές δεξιότητες , η εργασία χωρίς στήριξη κ.α. αποτελούν αντικείμενα της άτυπης παρατήρησης.

Έρευνα σχετική με τον τρόπο που αξιολογούν τελικά οι εκπαιδευτικοί κατέληξε στο ότι το μεγαλύτερο ποσοστό των εκπαιδευτικών αξιολογεί κάθε μαθητή διαπιστώνοντας και την πιο μικρή πρόοδο και επιβραβεύοντάς την ώστε και ο «αδύναμος» μαθητής να μπορεί να βιώσει την επιτυχία.

2.2 Ορισμός σχολικής επίδοσης

Αυτό που μετράται και αξιολογείται στη σχολική τάξη είναι η επίδοση των μαθητών . Αλλά σε τι αναφέρεται τελικά η επίδοση. Η επίδοση των μαθητών αναφέρεται στη διαδικασία απόκτησης γνώσεων , την κατάκτηση γνώσεων και την ανάπτυξη γνωστικών δεξιοτήτων . (Κολιάδης Ε. , 2002) Οι γνωστικές δεξιότητες αφορούν γενικές δεξιότητες και τις εξειδικευμένες δεξιότητες της κριτικής σκέψης.

Οι γενικές δεξιότητες αναφέρονται στη **δηλωτική γνώση (declarative knowledge)** δηλαδή στη γνώση που σχετίζεται με γεγονότα και πληροφορίες π.χ. ημερομηνίες , χαρακτηριστικά ενός έργου –πλοκή πλαίσιο –υπόθεση, είναι το τι γνωρίζει κάποιος (**Knowing what**). Περιλαμβάνει **τις λεκτικές πληροφορίες και γνώσεις** σύμφωνα με την ταξινόμηση των ικανοτήτων του Gagne. Η δηλωτική γνώση είναι η προϋπόθεση για την ανάπτυξη και την καλλιέργεια της ανθρώπινης σκέψης αφού διευκολύνει το άτομο στην εκμάθηση των κοινωνικών ρόλων και του πολιτισμού, ενώ συμβάλει στη δόμηση των προσωπικών σχέσεων . Πραγματώνεται με νοηματική μάθηση –βιβλία , παραδόσεις ,ομιλίες – και με μηχανιστική μάθηση ,δηλαδή την εκμάθηση λεκτικού μαθησιακού υλικού με γραμμικό τρόπο ,ανάκληση πληροφοριών με συγκεκριμένη σειρά και διάταξη και με ζευγαρωτή μάθηση όπου η μάθηση γίνεται με λεκτικές μονάδες συνδεδεμένες με συνειρμικό τρόπο.

Οι γενικές δεξιότητες αναφέρονται επίσης στη **διαδικαστική γνώση** που αφορά τη διαδικασία (**Knowing how**) , τον τρόπο δηλαδή πραγματοποίησης μιας δραστηριότητας . Η κατάταξη εννοιών η γνώση κανόνων και η επίλυση προβλημάτων εντάσσονται σε αυτό το πλαίσιο. Το είδος της γνώσης που αφορά την ικανότητα του ατόμου να γνωρίζει πότε και γιατί (**knowing when and why**) χρησιμοποιεί τη δηλωτική και διαδικαστική γνώση , λέγεται **υποθετική γνώση**.

Οι δεξιότητες που σχετίζονται τόσο με τη δηλωτική όσο και με τη διαδικαστική γνώση κατακτώνται γενικά μέσα από στρατηγικές και τακτικές μάθησης κατά την καθημερινή διδακτική διαδικασία. ΟΙ πολύ σημαντικές εξειδικευμένες δεξιότητες της **κριτικής σκέψης** αναφέρονται στη χρήση συλλογισμών , στον προσδιορισμό ουσιωδών χαρακτηριστικών στα πράγματα , στον εντοπισμό ομοιοτήτων και διαφορών , στην περιγραφή , στην ταξινόμηση πραγμάτων και γεγονότων και στην κατανόηση.

2.2.1 Παράγοντες με τους οποίους σχετίζεται η σχολική επίδοση

Η επίδοση λοιπόν των μαθητών σχετίζεται με

- Την ατομική τους προσπάθεια
- Με τις νοητικές τους δεξιότητες (αντιληπτική ικανότητα, νοημοσύνη , μεταγνωστικές δεξιότητες)
- Με τις ιδιαίτερες κλίσεις και τα προσωπικά τους ενδιαφέροντα.
- Με την αποτελεσματική διδασκαλία και τα παρεχόμενα εκπαιδευτικά μέσα.
- Με τη βοήθεια και τη στήριξη της οικογένειας .
- Με τις προσδοκίες της οικογένειας.

Μελέτη των Gigliotti και Brookover το 1975 σε μαθητές της Έκτης τάξης Δημοτικού έδειξε υψηλή συσχέτιση της σχολικής επίδοσης και των προσδοκιών που είχαν ήδη διαμορφώσει οι γονείς. Ωστόσο η έρευνα των Howell και Frese το 1979 που αναζητούσε συσχέτιση αντίστροφα ανάμεσα στη σχολική επίδοση και τη διαμόρφωση των προσδοκιών των γονιών , έδειξε ότι η σχολική επίδοση ασκούσε ελάχιστη επιρροή στις προσδοκίες των γονιών . Οι γονείς λοιπόν διαμορφώνουν τις προσδοκίες για τα παιδιά τους σύμφωνα με τις δικές τους ανάγκες και αντιλήψεις.(Φλουρής, 2004)

- Με ψυχοσωματικούς παράγοντες (συναισθηματική ισορροπία, αυτοεκτίμηση, κίνητρα ,δύναμη της θέλησης).

- Με κοινωνικούς ,πολιτισμικούς και οικονομικούς παράγοντες που αφορούν το περιβάλλον του μαθητή.

2.2.2 Ιστορική ανασκόπηση

Κατά το 19^ο και το μεγαλύτερο μέρος του 20ου αιώνα η διδασκαλία στην πρωτοβάθμια εκπαίδευση χαρακτηρίζεται από συντηρητισμό και είναι δασκαλοκεντρική.

Πριν από το 1836 επειδή δεν υπήρχε συγκεκριμένη κλίμακα αξιολόγησης , ο δάσκαλος χρησιμοποιούσε δικό του σύστημα βαθμολόγησης . Το 1836 επικράτησε στη δημοτική εκπαίδευση της Ελλάδας εξάβαθμη ποιοτική κλίμακα.

Το 1842 η επικράτηση της αλληλοδιδακτικής μεθόδου διδασκαλίας , όπου οι μεγαλύτεροι μαθητές διδάσκουν μικρότερους, εισήγαγε την τρίβαθμη αξιολογική κλίμακα ως εξής: 4 για τον άριστο, 3 για τον καλό και 2 για τον μέτριο. Το 1896 η τρίβαθμη κλίμακα αντικαταστάθηκε με την τετράβαθμη βαθμολογική κλίμακα : 4 άριστα, 3 λίαν καλώς, 2 καλώς και 1 ενδεώς . Η βάση ήταν το 2. Η κλίμακα αφορούσε όλες τις τάξεις.

Το 1912 επανέρχεται η εξάβαθμη κλίμακα :6 =άριστα, 5=λίαν καλώς, 4=καλώς, 3=μετρίως, 2=μετριότατα και 1=κακώς . Η βάση ήταν το 4. (οι σχετικές πληροφορίες αντλούνται από Κασσωτάκη ,685-700, και από εγκυκλίους της εποχής στα βιβλία εισερχομένων εγγράφων του δημοτικού σχολείου)

Το 1929 τελικά θεσπίστηκε και διατηρήθηκε ως το 1980, για τους μαθητές των δημοτικών η κλίμακα 0-10 όπου 10=άριστα, 8-9 =λίαν καλώς, 6-7 =καλώς , 5= σχεδόν καλώς, 3-4= μετρίως , 1-2=μετριότατα και 0=κακώς. Ως βάση ορίζεται το 5.

Με το προεδρικό διάταγμα ΠΔ462/91 θεσπίστηκε αξιολογική κλίμακα με χαρακτηρισμούς Α άριστα, Β πολύ καλά και Γ καλά , για την Πρώτη και Δευτέρα δημοτικού και για τις άλλες τάξεις θεσμοθετείται αριθμητική και λεκτική κλίμακα με :Άριστα (9-10), Πολύ καλά (7-8), Καλά (5-6) και σχεδόν καλά <5.

Από το 1995 ορίστηκε για την Πρώτη Α και Δευτέρα Β περιγραφική αξιολόγηση που δίνεται προφορικά (πίνακας από παράρτημα) , ενώ για τις Τρίτη Γ και Τετάρτη Δ εκτός από την περιγραφική αξιολόγηση αποδίδεται γραπτά και σύμφωνα με ποιοτική κλίμακα, αξιολόγησης με τους χαρακτηρισμούς: Άριστα (Α) , Πολύ καλά (Β), καλά (Γ) και Σχεδόν καλά (Δ). Με Σχεδόν καλά βαθμολογούνται μαθητές με σοβαρές μαθησιακές δυσκολίες και εφαρμόζεται σε αυτούς ενισχυτική διδασκαλία. Για τις δυο τελευταίες τάξεις Ε και Στ εκτός από περιγραφική αξιολόγηση θεσπίστηκε η εφαρμογή λεκτικής και αριθμητικής κλίμακας

Άριστα(9-10), πολύ καλά(7-8), Καλά (5-6), και Σχεδόν καλά. Το σύστημα βαθμολόγησης διατηρείται μέχρι σήμερα στην πρωτοβάθμια εκπαίδευση.

Με το Π.Δ 8/1995

1^ο, 2^ο και 4^ο άρθρο του νόμου Π.Δ 8/1995 ορίζει:

Άρθρο 4 Κλίμακα βαθμολογίας Άρθρο 2 Διαδικασία Αξιολόγησης 1. Η αξιολόγηση του μαθητή κατά τη διάρκεια της φοίτησής του στο Δημοτικό Σχολείο γίνεται από το δάσκαλο ή τους δασκάλους της τάξης και στηρίζεται: α. Στην καθημερινή προφορική εξέταση και την όλη συμμετοχή του μαθητή στη διδακτική μαθησιακή διαδικασία και στις άλλες σχολικές δραστηριότητες. β. Στα αποτελέσματα της επίδοσής του στα κριτήρια αξιολόγησης, που αποτελούν οργανικό στοιχείο του σχολικού προγράμματος και περιλαμβάνονται στο διδακτικό υλικό. Στα κριτήρια που απευθύνονται στους μαθητές των δύο ανώτερων τάξεων μπορούν να προστίθενται συνθετότερης μορφής ερωτήματα, τα οποία είναι δυνατόν να αναφέρονται σε περισσότερες της μιας γενικές ενότητες. γ. Στα αποτελέσματα των εργασιών, τις οποίες πραγματοποιεί ο μαθητής στο σχολείο ή στο σπίτι.

3. Στις τάξεις Ε' και ΣΤ' εκτός από την περιγραφική αξιολόγηση χρησιμοποιείται κλίμακα βαθμολογίας που είναι λεκτική και αριθμητική και έχει ως εξής: Άριστα (9-10), Πολύ Καλά (7-8), Καλά (5-6), Σχεδόν Καλά (1-4). Στο βιβλίο Μητρώου και Προόδου η βαθμολογία καταχωρείται με αριθμητικό σύμβολο. Με "Σχεδόν Καλά" βαθμολογούνται όσοι εμφανίζουν σοβαρές μαθησιακές δυσκολίες. Για τους μαθητές αυτούς εφαρμόζονται προγράμματα ενισχυτικής διδασκαλίας σύμφωνα με τις κείμενες διατάξεις. Αν κατά την εξαγωγή του Μέσου Όρου της ετήσιας επίδοσης του μαθητή κατά μάθημα όσο και του Γενικού Μέσου Όρου όλων των μαθημάτων προκύπτει κλάσμα ίσο ή μεγαλύτερο από το μισό της ακεραίας μονάδας, τότε αυτό λογίζεται ως ακέραια μονάδα και προστίθεται στο ακέραιο μέρος του Μέσου Όρου.

Άρθρο 6 Έναρξη

1. Η ισχύς του παρόντος Προεδρικού Διατάγματος αρχίζει από τη δημοσίευση του στην Εφημερίδα της Κυβερνήσεως.

2. Από την ίδια ημερομηνία καταργούνται οι διατάξεις των άρθρων 1, 2, 3 και 4 του Π.Δ. 462/91 (ΦΕΚ 171τΑ).

2.2.3 Είναι αναγκαία η αξιολόγηση της επίδοσης των μαθητών.

Για το θέμα αυτό υπάρχουν δυο αντικρουόμενες πλευρές.

Η πλευρά που υποστηρίζει την αναγκαιότητα του συστηματικού ελέγχου της προόδου των μαθητών και στηρίζεται σε κοινωνικούς λόγους και παιδαγωγικούς: Το εκπαιδευτικό σύστημα δεν είναι δυνατό να έχει σα σκοπό τη διάδοση μιας γενικής κουλτούρας ,όταν όλο και περισσότεροι εισάγονται στο πανεπιστήμιο προσδοκώντας μια επαγγελματική δραστηριότητα «υψηλής μόρφωσης» . Σε αυτά τα πλαίσια η ατομική αξιολόγηση είναι αποδεικτικό της μόρφωσης που ο καθένας έχει δεχτεί. (Reuchlin M., 1974). Σύμφωνα με την άποψη αυτή, ο μαθητής καταβάλλει προσπάθεια και πρέπει να ενημερώνεται για το αποτέλεσμα. Το αποτέλεσμα της αξιολόγησης είναι ένα ενισχυτικό κίνητρο. Η διαμόρφωση της προσωπικής αντίληψης για το επίπεδο των ικανοτήτων γίνεται μέσα από τις επιδόσεις σε κάθε μάθημα. Οι διδάσκοντες παρακολουθούν την πρόοδο των μαθητών. Οι γονείς ενημερώνονται για την πρόοδο των μαθητών . Με βάση την επίδοση των μαθητών αξιολογούνται μέτρα που λαμβάνει η πολιτεία και η καταλληλότητα των αναλυτικών προγραμμάτων.

Η αντίθετη άποψη αμφισβητεί τα παραπάνω. Θεωρούν ότι τελικά οι αξίες που καλλιεργεί το σχολείο με τις διαδικασίες αξιολόγησης είναι ποσοτικές αξίες. Ενώ η προσωπική ανάπτυξη δεν είναι μια μετρήσιμη οντότητα. Επιπλέον οι μαθητές τείνουν να αξιολογούν μόνο ότι είναι μετρήσιμο..

Εντούτοις: Η αξιολόγηση της επίδοσης των μαθητών έστω με ερωτήσεις σε ένα γνωστικό αντικείμενο, δίνει τη δυνατότητα στους μαθητές να επεξεργάζονται καλύτερα πληροφορίες που πρέπει να θυμούνται. Αποτελέσματα ερευνών δείχνουν ότι μαθητές των οποίων η επίδοση αξιολογείται θυμούνται περισσότερο από τους μαθητές που η επίδοσή τους δεν αξιολογείται (Κολιάδης 2002). Μεγάλη μερίδα των εκπαιδευτικών συμφωνεί στο ότι το πρόβλημα δεν είναι η ίδια η αξιολόγηση αλλά ο τρόπος εφαρμογής της.

Κεφάλαιο 3

ΕΠΑΓΓΕΛΜΑΤΙΚΗ ΑΝΑΠΤΥΞΗ

3.1 Βασικές θεωρίες

Οι βασικές θεωρίες της επαγγελματικής ανάπτυξης συμφωνούν στο ότι η επαγγελματική ανάπτυξη του ατόμου ακολουθεί εξελικτική πορεία με αρχή, διάρκεια και τέλος και στο ότι διαμορφώνεται από ένα συνδυασμό παραγόντων. Η επιλογή δεν είναι αποτέλεσμα στιγμιαίας απόφασης αλλά μακροχρόνια διαδικασία που συμβαδίζει με την ψυχολογική εξέλιξη. Ωστόσο υπάρχει διαχωρισμός ως προς τη βαρύτητα που αποδίδουν στους παράγοντες διαμόρφωσης της. Οι θεωρίες επιλογής χωρίζονται σε **μη ψυχολογικές** που δίνουν έμφαση σε εξωτερικούς παράγοντες :

- ✚ Θεωρία της τύχης (Hopson & Hayes, 1968)
- ✚ Οικονομολογικές θεωρίες (Κάντας & Χατζή , 1991)
- ✚ Κοινωνιολογικές –Πολιτιστικές θεωρίες (Hotchkiss & Borrow, 1

Σε **ψυχολογικές** που δίνουν έμφαση στο ρόλο του ίδιου του ατόμου ενώ οι εξωτερικοί παράγοντες ασκούν έμμεση επιρροή:

- ✚ Θεωρία χαρακτηριστικών και παραγόντων (Parsons 1999)
- ✚ Ψυχοδυναμικές θεωρίες (Segal, 1961)
- ✚ Θεωρίες αναγκών(Hoprock, 1976)
- ✚ Τυπολογική θεωρία (Holland)
- ✚ Εξελικτικές θεωρίες (Ginzberg , 1951) (Savickas & son, 2005)
- ✚ Θεωρίες μάθησης (O' Hara , 1968) (Miller, 1968)

Σε **γενικές συστημικές** που υποστηρίζουν πολλές κατηγορίες παραγόντων:

- ✚ Θεωρία Blaw (1956)
- ✚ Θεωρία ζωντανών συστημάτων (Ford, 1987)
- ✚ Θεωρία αντιστοίχισης (Lofouist David, 1991)
- ✚ Θεωρία της Linda Cottfredson,(1981)

Στη συνέχεια γίνεται μια αναφορά στις πιο σημαντικές θεωρίες από κάθε κατηγορία που έχουν τεκμηριωθεί με σχετικές έρευνες και σε αυτές στηρίζεται το θεωρητικό πλαίσιο αυτής της έρευνας ως

3.1.1 Τυπολογική θεωρία του John Holland .

Επικεντρώνεται στους παράγοντες που επηρεάζουν τις επιλογές σε ένα δεδομένο χρονικό και τοπικό πλαίσιο. Αυτοί οι παράγοντες είναι :

A) η προσωπικότητα του ατόμου που ταξινομείται σε έξι τύπους :ρεαλιστικός, ερευνητικός, καλλιτεχνικός, κοινωνικός, επιχειρησιακός, συμβατικός.

B) Τα περιβάλλοντα. Οι προσωπικότητες των ατόμων σύμφωνα με τη θεωρία του Holland, έλκονται από περιβάλλοντα του ίδιου ή όμοιου τύπου.

3.1.2Αφηγηματική προσέγγιση του Mark Savickas(2005)

Η αφηγηματική προσέγγιση δέχεται ότι τα άτομα μπορούν να δημιουργήσουν μια επαγγελματική σταδιοδρομία αλλά σε καμιά περίπτωση δεν μπορούν να εξακριβωθούν χαρακτηριστικά που θα οδηγούσαν στην επιλογή του ταιριαστού για τον καθένα επαγγέλματος. Ο Savickas υποστηρίζει ότι «οι γρήγορες αλλαγές του 21^{ου} αιώνα δεν εξασφαλίζουν αντιστοιχία ανάμεσα στο άτομο και το περιβάλλον», « Οι καριέρες χτίζονται καθώς τα άτομα κάνουν επιλογές που εκφράζουν την αυτοαντίληψη τους και τους στόχους τους στην κοινωνική πραγματικότητα της ζωής». Η κινητήρια δύναμη έκφρασης της αυτοαντίληψης και ανάληψης ρόλου, είναι **οι ατομικές εμπειρίες** .

3.1.3 Θεωρία της Linda Cottfredson

Η προσέγγιση αυτή διακρίνει τέσσερις φάσεις στην επαγγελματική ανάπτυξη που αντιστοιχούν στον προσανατολισμό του ατόμου προς:

- τη δύναμη , στην ηλικία 3-5 ετών

- το ρόλο του φύλου, στην ηλικία των 6-8 ετών
- τις κοινωνικές αξίες , στην ηλικία 9-13
- τον εσωτερικό εαυτό , 14 και άνω

Για κάθε ηλικία υπάρχει διαφορετική επιρροή των στοιχείων αυτών. Ωστόσο η ελκυστικότητα που ασκεί ένα επάγγελμα σε ένα άτομο εξαρτάται από το επίπεδο της αίγλης που έχει αυτό αλλά και από **τη θέση του ως προς το φύλο.**

Δίνεται ιδιαίτερη έμφαση στην αυτοαντίληψη του ατόμου ως προς το φύλο και στη διαδικασία που αυτή τελικά οδηγεί στην επιλογή επαγγέλματος. Οι φιλοδοξίες μετατρέπονται σε επιλογές που τελικά ανταποκρίνονται στον προσωπικό τύπο φύλου και στις προσωπικές ανάγκες για κοινωνική δύναμη-αίγλη. Ο συμβιβασμός λειτουργεί εις βάρος του επαγγέλματος και του ατόμου.

3.1.4 Θεωρία του John Krumboltz

Η θεωρία της κοινωνικής μάθησης του Krumboltz είναι επηρεασμένη από την κοινωνικογνωστική προσέγγιση του Bandura και εξετάζει την ανάπτυξη της σταδιοδρομίας μέσα από τις αποκτηθείσες συμπεριφορές. Θεωρεί την επιρροή στην επιλογή του επαγγέλματος τεσσάρων καθοριστικών παραγόντων: των γενετικών χαρακτηριστικών, των περιβαλλοντικών συνθηκών , των μαθησιακών εμπειριών και την απόκτηση δεξιοτήτων

3.1.5 Κοινωνικοψυχολογικό μοντέλο Wisconsin

Οι μελέτες που εξετάζουν τις σχέσεις μεταξύ του περιβάλλοντος προέλευσης, των ατομικών χαρακτηριστικών και των φιλοδοξιών των μαθητών, καταλήγουν στο συμπέρασμα ότι «το κοινωνικοοικονομικό υπόβαθρο και η ικανότητα επηρεάζουν τις φιλοδοξίες για το σχολείο και την καριέρα, εξαιτίας της μετατροπής τους σε σχολική επίδοση και σε κοινωνική υποστήριξη από τους σημαντικούς άλλους» (Μουσταΐρας, 2004).

Από τα παραπάνω προκύπτει ότι το γενικότερο οικογενειακό και κοινωνικό περιβάλλον από το οποίο προέρχεται και στο οποίο ζει ένας μαθητής επηρεάζει σοβαρά την εκπαιδευτική του πορεία και κατά συνέπεια και το εκπαιδευτικό και επαγγελματικό του επίτευγμα. Βέβαια η επίδραση αυτή δεν είναι άμεση αλλά γίνεται διαμέσου διαφόρων ατομικών και προσωπικών χαρακτηριστικών του μαθητή.

Σε σχέση με την κοινωνική διαστρωμάτωση και την ανισότητα στην εκπαίδευση, κυριαρχούν δύο προσεγγίσεις, η δομολειτουργική και η συγκρουσιακή. Σύμφωνα με την πρώτη, αίτιο της άνισης σχολικής επιτυχίας είναι οι έμφυτες, γενετικά καθορισμένες, ικανότητες των ατόμων, ενώ για τη δεύτερη βάση της σχολικής αποτυχίας είναι το κοινωνικό και οικογενειακό περιβάλλον και αιτία της σχολικής ανισότητας η κοινωνική ανισότητα.

Και οι δύο πλευρές έχουν επικεντρώσει την προσοχή τους στο να δείξουν ότι η διαδικασία της κοινωνικής διαλογής μέσω της εκπαίδευσης είναι μια διαδικασία επιτεύγματος (οι δομολειτουργιστές) ή απόδοσης (οι συγκρουσιακοί). Οι δομολειτουργιστές επιχειρούν να δείξουν ότι η επίτευξη κοινωνικού επιπέδου είναι κυρίως επίτευγμα και αποτέλεσμα «κοινωνικοποίησης» και ψυχολογικών κυρίως μηχανισμών, ενώ οι συγκρουσιακοί υποστηρίζουν ότι είναι αποτέλεσμα απόδοσης και δομικών περιορισμών που επιβάλλει το σύστημα σε ορισμένες κοινωνικές ομάδες.

Στο δομολειτουργικό εξηγητικό σχήμα, το πιο γνωστό ίσως μοντέλο επίτευξης κοινωνικού στάτους είναι το μοντέλο του Wisconsin, σύμφωνα με το οποίο τα χαρακτηριστικά προέλευσης μεταφράζονται σε διαφορετική επιτυχία κοινωνικοοικονομικού στάτους, μόνο αφού πρώτα έχουν μεταμορφωθεί σε ψυχολογικές μεταβλητές και **ατομική επίδοση**. Επιπλέον, τα σχολεία επιλέγουν, προωθούν και αμείβουν μαθητές με βάση ατομικά χαρακτηριστικά και επιτεύγματα όπως φιλοδοξίες και ικανότητα, ενώ η εκπαιδευτική επιτυχία είναι ο κύριος παράγοντας της επαγγελματικής και κοινωνικής επιτυχίας.

3.2 Η ταξινόμηση των επαγγελματιών σε ομάδες

Η ταξινόμηση των επαγγελματιών σε ομάδες θεωρήθηκε αναγκαία για την καλύτερη επεξεργασία των δεδομένων των ερωτηματολογίων και στηρίζεται :

- στην διεθνώς ενδεδειγμένη κωδικοποίηση και ταξινόμηση των ειδικοτήτων και επαγγελματιών International Classification of Occupations (ISCO-88) και
- στην ελληνική της τροποποίησή της προκειμένου να ανταποκρίνεται πλήρως στην ελληνική αγορά εργασίας, την ΣΤΕΠ-92 (dide.mag.sch.gr)

Η ομαδοποίηση των ειδικοτήτων σε επίπεδο χαρακτηριστικών είναι σύμφωνη και συνάδει απόλυτα με τη μεθοδολογία που χρησιμοποιείται από την Ευρωπαϊκή Στατιστική Υπηρεσία και πιο συγκεκριμένα τη μεθοδολογία υλοποίησης της έρευνας για τις εργασιακές συνθήκες στην Αγορά Εργασίας (European Working Conditions Survey7). (**οικονομική και κοινωνική επιτροπή**) (**ΟΚΕ**)

Μεγάλη Ομάδα 1: Μέλη των βουλευόμενων σωμάτων ανώτερα διοικητικά και διευθυντικά στελέχη του δημόσιου και του ιδιωτικού τομέα Περιλαμβάνονται τα πρόσωπα τα οποία νομοθετούν και καθορίζουν την κυβερνητική πολιτική και είναι υπεύθυνα για την εφαρμογή της σε κεντρικό, περιφερειακό και τοπικό επίπεδο π.χ. υπουργοί, βουλευτές, γενικοί γραμματείς, δήμαρχοι, κοινοτάρχες κλπ. Περιλαμβάνονται, επίσης, τα διευθυντικά στελέχη δημόσιων και ιδιωτικών υπηρεσιών, επιχειρήσεων και οργανισμών

Μεγάλη Ομάδα 2: Πρόσωπα που ασκούν επιστημονικά, καλλιτεχνικά και συναφή επαγγέλματα

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται κατά κανόνα ανώτατο επίπεδο εκπαίδευσης, γνώσεων και εμπειρίας στους τομείς των επιστημών, της διδασκαλίας και της καλλιτεχνικής δημιουργίας, π.χ. φυσικοί, μηχανικοί, ιατροί, νομικοί, καθηγητές, καλλιτέχνες κλπ.

Μεγάλη Ομάδα 3: Τεχνολόγοι, τεχνικοί βοηθοί και ασκούντες συναφή επαγγέλματα

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτούνται τεχνικές γνώσεις και εμπειρία σε όλους τους τομείς των επιστημών. Κατά κανόνα είναι απόφοιτοι τεχνολογικών εκπαιδευτικών ιδρυμάτων (ΤΕΙ) ή άλλων ειδικών δημόσιων ή ιδιωτικών σχολών και εργάζονται συνήθως υπό την καθοδήγηση και επίβλεψη προσώπων που ταξινομούνται στις μεγάλες ομάδες 1 και 2.

Μεγάλη Ομάδα 4: Υπάλληλοι γραφείου και ασκούντες συναφή επαγγέλματα

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται γνώση, εξειδίκευση και εμπειρία για την εκτέλεση διαφόρων εργασιών όπως : γραμματειακή υποστήριξη, εγγραφή υπολογισμός και εισαγωγή αριθμητικών δεδομένων, καθήκοντα εξυπηρέτησης πελατών, ταχυδρομικές και ταμειακές υπηρεσίες κλπ.

Μεγάλη Ομάδα 5: Απασχολούμενοι στην παροχή υπηρεσιών και πωλητές σε καταστήματα και υπαίθριες αγορές

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται γνώση, εξειδίκευση και εμπειρία για την παροχή υπηρεσιών σε νοικοκυριά και σε πρόσωπα ή απασχολούνται με την τροφοδοσία, την προστασία προσώπων και περιουσίας, καθώς και με την πώληση αγαθών σε καταστήματα ή υπαίθριες αγορές.

Μεγάλη Ομάδα 6: Ειδικευμένοι γεωργοί, κτηνοτρόφοι, δασοκόμοι και αλιείς

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται γνώση, εξειδίκευση και εμπειρία για την καλλιέργεια της γης, την εκτροφή ζώων, την αλιεία ανοικτής θαλάσσης και την ιχθυοκαλλιέργεια, τη δασοκομία και τη θήρα.

Μεγάλη Ομάδα 7: Ειδικευμένοι τεχνίτες και ασκούντες συναφή τεχνικά επαγγέλματα

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται γνώση, εξειδίκευση και εμπειρία, όπως για την εξόρυξη ορυκτών, την κατασκευή κτιρίων και άλλων δομικών έργων, την παραγωγή ή κατασκευή διαφόρων προϊόντων, καθώς και προϊόντων χειροτεχνίας, πράγμα το οποίο προϋποθέτει πλήρη γνώση της παραγωγικής διαδικασίας, των εργαλείων που χρησιμοποιούνται, των πρώτων υλών και του προορισμού

των τελικών προϊόντων. Οι ανωτέρω εργασίες εκτελούνται με το χέρι ή με τη χρήση χειροκίνητων ή άλλων απλών εργαλείων

Μεγάλη Ομάδα 8: Χειριστές σταθερών βιομηχανικών εγκαταστάσεων, μηχανημάτων και εξοπλισμού και συναρμολογητές (μονταδόροι)

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται γνώση, εμπειρία και εξειδίκευση για τη ρύθμιση και το χειρισμό σταθερών εγκαταστάσεων και συχνά υψηλού αυτοματισμού μηχανημάτων εξόρυξης, επεξεργασίας και παραγωγής ή συναρμολόγησης προϊόντων. Περιλαμβάνονται, επίσης, οι οδηγοί μεταφορικών μέσων, καθώς και οι χειριστές γεωργικών, χωματουργικών και άλλων κινητών μηχανημάτων.

Μεγάλη Ομάδα 9: Ανειδίκευτοι εργάτες, χειρόνακτες και μικροεπαγγελματίες

Περιλαμβάνονται όσοι ασκούν επαγγέλματα, για τα οποία απαιτείται ικανότητα για την άσκηση απλών και συνηθισμένων εργασιών, για τις οποίες χρειάζεται κυρίως σωματική προσπάθεια και περιορισμένη ειδίκευση και πρωτοβουλία, π.χ. φύλαξη ή καθάρισμα κτιρίων και κατοικιών, πώληση προϊόντων στο δρόμο, απλές χειρωνακτικές εργασίες στη γεωργία, στις κατασκευές, σε εργοστάσια, σε εμπορικά καταστήματα κλπ.

Μεγάλη Ομάδα 0: Ένοπλες δυνάμεις

Περιλαμβάνονται οι μόνιμοι στρατιωτικοί, οι υπηρετούντες κανονική θητεία, καθώς και όσοι ασκούν πολιτικά επαγγέλματα, αλλά ανήκουν στις ένοπλες δυνάμεις, π.χ. ιατροί, μηχανικοί, νοσηλευτικό προσωπικό, μάγειροι κλπ

ΚΕΦΑΛΑΙΟ 4

DATA MINING

4.1 Εξόρυξη δεδομένων

Τα τελευταία χρόνια έχουν σημειωθεί σημαντικές εξελίξεις στη συλλογή δεδομένων διευκολύνοντας έτσι την αποθήκευση ενός μεγάλου όγκου δεδομένων σε διάφορους τομείς . Η εξέλιξη αυτή καλλιεργεί τη στάση της χωρίς διάκριση συλλογής ,ενώ η βαρύτητα δίνεται στην ανάλυση και στην ανακάλυψη ενδιαφερόντων σημείων.

Ένα πεδίο έρευνας , η **ανακάλυψη γνώσης** μέσα από βάσεις δεδομένων είναι το πεδίο που έχει αναλάβει την ανακάλυψη της βαρύτητας, δηλαδή αναζητά χρήσιμες πληροφορίες μέσα από τα δεδομένα(KDD Knowledge Discovery in Databases).

Τα **δεδομένα** ορίζονται ως σύνολα παραδειγμάτων ή στιγμιοτύπων ενός προβλήματος σε μια βάση δεδομένων. Όπως μια συλλογή εγγραφών από τα δεδομένα μιας τράπεζας όπου κάθε εγγραφή περιλαμβάνει γνωρίσματα (χρέος , εισόδημα κ.α.). Ένα υποσύνολο από αυτά τα σύνολα εκφραζόμενο σε μια συγκεκριμένη γλώσσα αποτελεί το **πρότυπο** . Τα πρότυπα προκύπτουν από τη διαδικασία ανακάλυψης γνώσης και θα πρέπει να ισχύουν με κάποιο βαθμό βεβαιότητας για κάθε νέο άγνωστο στιγμιότυπο του προβλήματος. Τα πρότυπα λοιπόν πρέπει να είναι ακριβή και χρήσιμα. Δηλαδή να οδηγούν σε κάποιες χρήσιμες ενέργειες.

Η εξόρυξη γνώσης περιλαμβάνει τις διαδικασίες και τα μέσα εξαγωγής προτύπων από το σύνολο των δεδομένων. Ορίζεται ως διαδικασία ανακάλυψης νέων πιθανώς κρυμμένων προτύπων και μοντέλων με αυτόματο ή ημιαυτόματο τρόπο με σκοπό την περιγραφή των δεδομένων αλλά και τη πρόβλεψη και εξήγηση νέων δεδομένων.

Χρησιμοποιώντας σαν θεμέλιο λίθο τις περιοχές ανάλυσης δεδομένων που ήδη υπήρχαν και αναπτύσσοντας και εντελώς νέα πεδία ,όπως αυτό της ανάλυσης συσχέτισης (association analysis), εξαπλώθηκε το πεδίο της **εξόρυξης δεδομένων** ανατρέποντας τους περιορισμούς που έθεταν οι προηγούμενοι μέθοδοι ανάλυσης καθώς ακόμα και στην περίπτωση ενός μικρού όγκου δεδομένων οι ερωτήσεις που θέτονται δεν μπορούσαν να απαντηθούν με τα εργαλεία παλαιότερων μεθόδων. Η εξόρυξη δεδομένων είναι λοιπόν η τεχνολογία που συνδυάζει σύγχρονους αλγόριθμους επεξεργασίας και τεχνικές μοντελοποίησης από την τεχνητή νοημοσύνη και μοντελοποίηση προτύπων από τη μηχανική μάθηση, με τα παραδοσιακά εργαλεία ανάλυσης δεδομένων όπως τη δειγματοληψία –

εκτίμηση- έλεγχο υποθέσεων από τη στατιστική. Χρησιμοποιείται για παράδειγμα σε εφαρμογές επιχειρηματικής ευφυίας όπως το στοχευμένο μαρκετινγκ ή η ανίχνευση απάτης . Σε εφαρμογές της ιατρικής στον τομέα της μοριακής βιολογίας όπου συλλέγεται μεγάλος όγκος γενωμικών δεδομένων με σκοπό την κατανόηση της δομής και της λειτουργίας των γονιδίων.

Σε εφαρμογές της επιστήμης για την κατανόηση συσχετίσεων που αφορούν την ατμόσφαιρα της Γης με την ανάλυση δεδομένων από τις παρατηρήσεις που καταγράφουν οι δορυφόροι που κινούνται γύρω από τη Γη.)

«Η εξόρυξη δεδομένων λοιπόν είναι η διαδικασία αυτόματης ανακάλυψης χρήσιμων πληροφοριών μέσα από μεγάλες δεξαμενές δεδομένων» (P-N Tan, M. Steinbach and V. Kumar, 2010).

4.1. α. Τομείς συσχέτισης

Αναφέρθηκε παραπάνω ο όρος **Μηχανική Μάθηση** και κρίνεται σκόπιμο να διευκρινιστεί η σχέση της με την εξόρυξη γνώσης . Η Μηχανική Μάθηση είναι ένας τομέας της Τεχνητής Νοημοσύνης που ασχολείται με τη δημιουργία προγραμμάτων που μαθαίνουν. Στην εξόρυξη γνώσης η Μηχανική Μάθηση χρησιμοποιείται για πρόβλεψη και κατηγοριοποίηση. Πιο συγκεκριμένα στην περίπτωση της πρόβλεψης ο υπολογιστής κάνει μια πρόβλεψη και με βάση την ανάδραση που λαβαίνει για την ορθότητα της πρόβλεψης , αποκτά τη γνώση ώστε σε μια μελλοντική νέα περίπτωση να κάνει μια νέα πρόβλεψη. Όταν με τη μηχανική μάθηση επιδιώκεται η κατηγοριοποίηση , κατασκευάζεται μοντέλο που αναπαριστά τα δεδομένα . Ένα δείγμα από τη βάση Δεδομένων εκπαιδεύει το σύστημα ώστε αυτό να είναι σε θέση να εκτελεί την κατηγοριοποίηση με επιτυχία(φάση εκπαίδευσης) . Στη συνέχεια το σύστημα εφαρμόζεται στη γενική βάση δεδομένων ώστε να εκτελεί την κατηγοριοποίηση(φάση ελέγχου .

Ένας άλλος τομέας που σχετίζεται με την εξόρυξη δεδομένων είναι η **τεχνητή νοημοσύνη**. «Τεχνητή νοημοσύνη είναι ο τομέας της επιστήμης των υπολογιστών που ασχολείται με τη σχεδίαση ευφυών συστημάτων , δηλαδή με συστήματα που επιδεικνύουν χαρακτηριστικά που σχετίζονται με τη νοημοσύνη στην ανθρώπινη συμπεριφορά»Barr Feigenbaum, (Russell S. & Norvig P., 2000).Σκοπός της τεχνητής νοημοσύνης είναι η εξαγωγή λογικών συμπερασμάτων από δεδομένα που δεν έχουν υποστεί επεξεργασία. Τα νευρωνικά δίκτυα που αναλύονται παρακάτω και τα δέντρα απόφασης είναι εργαλεία από το χώρο της τεχνητής νοημοσύνης.

Η **Στατιστική** αποτελεί θεμέλιο λίθο της εξόρυξης δεδομένων, αφού μεγάλο μέρος της ερευνητικής βάσης της στηρίζεται στη στατιστική (ανάλυση συστάδων κ.α.)

4.1.β. Κάθε εργασία ανακάλυψης είναι εξόρυξη;

Εφαρμόζοντας τεχνικές που ερευνούν σε βάθος τα δεδομένα αναζητούν νέα και χρήσιμα πρότυπα που σε άλλη περίπτωση ανάλυσης θα παρέμεναν άγνωστα. Ωστόσο δεν αποτελεί κάθε εργασία ανακάλυψης πληροφοριών εξόρυξη δεδομένων. Η εξόρυξη δεδομένων είναι επέκταση στα συστήματα ανάκτησης πληροφοριών και γενικά χωρίζεται σε δυο κατηγορίες με κριτήριο το στόχο των εργασιών:

Προγνωστικές εργασίες (predictive tasks). Ο στόχος σε αυτή την περίπτωση είναι η πρόβλεψη της τιμής ενός συγκεκριμένου χαρακτηριστικού (target) με βάση τις τιμές άλλων χαρακτηριστικών. Το χαρακτηριστικό που προβλέπεται είναι η εξαρτημένη μεταβλητή (dependent variable) ενώ τα χαρακτηριστικά που χρησιμοποιούνται για να γίνει η πρόβλεψη λέγονται ανεξάρτητες μεταβλητές (independent variables).

Περιγραφικές εργασίες (descriptive tasks) Στόχος στις περιγραφικές εργασίες είναι η διερεύνηση βασικών σχέσεων στα δεδομένα με εξαγωγή υποδειγμάτων συσχετίσεων, συστάδων, τάσεων, τροχιών κ.α.

Τα παραπάνω γίνονται εφικτά με τη χρήση αλγόριθμων. Οι αλγόριθμοι αυτοί επιτελούν δυο λειτουργίες :

Περιγράφουν το μοντέλο, τη λειτουργία του και τους βασικούς στόχους κατηγοριοποίησης ή συσταδοποίησης. Η λειτουργία του μοντέλου καθορίζει τις βασικές εργασίες που θα γίνουν. Εξίσου σημαντικός όμως είναι και ο τρόπος αναπαράστασης του μοντέλου ώστε να έχει μια κατανοητή από τον ανθρώπινο νου, δομή. (Δέντρα απόφασης, Νευρωνικά δίκτυα)

Αξιολογούν το μοντέλο, ως προς την εγκυρότητα των προτύπων και την αξιολόγηση της ακρίβειας, της χρησιμότητας και της δυνατότητας κατανόησης του μοντέλου. Η αξιολόγηση γίνεται βάσει κάποιων κριτηρίων αξιολόγησης.

Οι μέθοδοι που χρησιμοποιούνται για την επίτευξη των στόχων της εξόρυξης δεδομένων (data mining) εκτελούν κατά την εφαρμογή τους ένα σύνολο εργασιών οι οποίες συνοψίζονται στα εξής:

Κατηγοριοποίηση (Classification)

Συσταδοποίηση (Clustering)

Εξαγωγή κανόνων συσχέτισης (association rules)

Πρόβλεψη (Prediction)

4.2 Κατηγοριοποίηση

Η κατηγοριοποίηση είναι η πιο δημοφιλής τεχνική εξόρυξης . Παραδείγματα όπου γίνεται χρήση συστημάτων κατηγοριοποίησης είναι τα συστήματα έγκρισης δανείων , συστήματα ιατρικών διαγνώσεων κ.α. Γενικά στην εκτέλεση της κατηγοριοποίησης απαιτείται γνώση των δεδομένων. Κατά την κατηγοριοποίηση ένα σύνολο εγγραφών με συγκεκριμένες ιδιότητες , εκ των οποίων μία αποτελεί την κλάση , αποτελεί το σώμα εκπαίδευσης για την εύρεση μοντέλου πρόβλεψης. Ένα σύνολο αξιολόγησης χρησιμοποιείται για να διαπιστωθεί η ακρίβεια του μοντέλου. Τα δεδομένα εισόδου αποτελούν ένα σύνολο εγγραφών. Κάθε εγγραφή ονομάζεται **στιγμιότυπο** και ορίζεται ως (χ, ψ) όπου χ είναι το σύνολο των χαρακτηριστικών του και ψ το ειδικό χαρακτηριστικό που είναι και η ετικέτα της κατηγορίας γι αυτό και πρέπει να είναι διακριτό. Πιο ειδικά η κατηγοριοποίηση (classification) είναι μια διαδικασία εκμάθησης μιας συνάρτησης- στόχου f , η οποία απεικονίζει ένα σύνολο δεδομένων σε προκαθορισμένες ομάδες που ονομάζονται κλάσεις ή κατηγορίες. Το μοντέλο κατηγοριοποίησης μπορεί να είναι αντίστοιχα με το στόχο περιγραφικό ή προγνωστικό. Ενώ το **περιγραφικό** αποτελεί εργαλείο περιγραφής και διάκρισης των αντικειμένων από διαφορετικές κατηγορίες , **το προγνωστικό** προβλέπει την ετικέτα της κατηγορίας άγνωστων εγγραφών. Αποδίδει ετικέτα κάθε φορά που δέχεται σύνολο χαρακτηριστικών άγνωστης εγγραφής. Πρακτικά η επίλυση προβλημάτων κατηγοριοποίησης περιλαμβάνει δυο βασικά στάδια. Ένα **επαγωγικό** για την κατασκευή ενός μοντέλου κατηγοριοποίησης από τα δεδομένα και ένα **συμπερασματικό** για την εφαρμογή του μοντέλου στα δείγματα ελέγχου. Το πρώτο περιλαμβάνει τη δημιουργία ενός μοντέλου από την αξιολόγηση και την ανάλυση των δεδομένων εκπαίδευσης . Αυτό το στάδιο έχει σαν είσοδο τα δεδομένα εκπαίδευσης και σαν έξοδο τον ορισμό του αναπτυσσόμενου μοντέλου. Το μοντέλο αυτό μπορεί να κατηγοριοποιεί τα δεδομένα εκπαίδευσης με όσο το δυνατό μεγαλύτερη ακρίβεια.

Παραδείγματα των τεχνικών που χρησιμοποιούνται στο στάδιο αυτό είναι οι κατηγοριοποιητές δέντρων απόφασης , τα νευρωνικά δίκτυα, οι κατηγοριοποιητές Bayes. Κάθε τεχνική χρησιμοποιεί έναν αλγόριθμο μάθησης ώστε να γίνει δυνατός ο εντοπισμός εκείνου του μοντέλου που αποδίδει με μεγαλύτερη ακρίβεια τη σχέση του συνόλου των χαρακτηριστικών με την ετικέτα κατηγορίας των δεδομένων εισόδου. Πρέπει δηλαδή ο αλγόριθμος να ταιριάζει με τα δεδομένα εισόδου αλλά και να προβλέπει σωστά την ετικέτα για κάθε καινούρια εγγραφή που είναι και το δεύτερο στάδιο της κατηγοριοποίησης.

Για να επιλυθεί λοιπόν ένα πρόβλημα κατηγοριοποίησης πρέπει να δοθεί ένα **σύνολο εκπαίδευσης**. Το σύνολο εκπαίδευσης χρησιμοποιείται για να κατασκευάσει ένα μοντέλο κατηγοριοποίησης που θα εφαρμοστεί σε ένα **σύνολο ελέγχου (test set)** που περιλαμβάνει εγγραφές με ετικέτες κατηγορίας

άγνωστες. Η απόδοση του μοντέλου κατηγοριοποίησης καταγράφεται σε μετρήσεις που τοποθετούνται σε πίνακα γνωστό ως **μήτρα σύγχυσης (Confusion matrix)** και αφορούν τις εγγραφές που έχουν προβλεφθεί σωστά και λανθασμένα από το μοντέλο. Για τη σύγκριση της απόδοσης διαφορετικών μοντέλων χρησιμοποιείται ένα **μέτρο απόδοσης** και συνοψίζει τις παραπάνω πληροφορίες σε έναν αριθμό ώστε να είναι ευκολότερη η σύγκριση και λέγεται **ακρίβεια (accuracy)** . Για παράδειγμα για ένα δυαδικό πρόβλημα ορίζεται : **training set**) που αποτελείται από εγγραφές με γνωστές ετικέτες κατηγορίας.

$$\text{Accuracy} = \text{NC} / \text{T}$$

όπου NC είναι το πλήθος των σωστών προβλέψεων και T είναι το συνολικό πλήθος των προβλέψεων.

Η απόδοση ενός μοντέλου εκφράζεται με βάση το ρυθμό σφάλματος (**error rate**) ως εξής:

$$\text{Error rate} = \text{NW} / \text{T}$$

Όπου NW είναι το πλήθος των λανθασμένων προβλέψεων και T το συνολικό πλήθος των προβλέψεων. Στόχος βέβαια είναι η αναζήτηση από τον αλγόριθμο μοντέλων που επιτυγχάνουν μεγαλύτερη ακρίβεια ή με τον ίδιο τρόπο μικρότερο ρυθμό σφάλματος , όταν εφαρμόζονται σε ένα σύνολο ελέγχου.

Διακρίνονται τα εξής είδη κατηγοριών **αλγορίθμων κατηγοριοποίησης**:

- Αλγόριθμοι κατηγοριοποίησης βασισμένοι στην **απόσταση**.
- Αλγόριθμοι βασισμένοι στα **δέντρα απόφασης** .
- Αλγόριθμοι βασισμένοι στα **Νευρωνικά Δίκτυα** .
- Αλγόριθμοι βασισμένοι στους **κανόνες**.

Θα αναλυθούν οι αλγόριθμοι που θα χρησιμοποιηθούν στα πειράματα αυτής την έρευνα.

4.2.1. K- κοντινότεροι γείτονες (K- nearest neighbors- K-NN)

Ο KNN είναι αλγόριθμος που στηρίζεται στην **απόσταση**. Η βασική ιδέα των αλγορίθμων που βασίζονται στην απόσταση είναι ότι η απεικόνιση κάθε στοιχείου του συνόλου δεδομένων που βρίσκεται

στην ίδια κατηγορία εμφανίζει μικρότερη απόσταση από την απεικόνιση των στοιχείων της κατηγορίας του από οποιαδήποτε άλλη απεικόνιση που ανήκει σε διαφορετική κατηγορία.

Η τεχνική την οποία υποστηρίζει ο KNN αλγόριθμος ανήκει στους **απρόθυμους μαθητές (lazy learners)** καθώς δε μαθαίνει το μοντέλο αλλά προϋποθέτει την κατηγοριοποίηση για κάθε στοιχείο. Το κάθε σημείο κατηγοριοποιείται με βάση τις ετικέτες κατηγορίας των γειτόνων, τα ίδια τα δεδομένα έτσι αποτελούν το μοντέλο κατηγοριοποίησης. Για κάθε νέο στοιχείο είναι απαραίτητο να δοθεί η απόστασή του από κάθε στοιχείο του συνόλου εκπαίδευσης. Στη συνέχεια αφού αναζητηθούν οι K κοντινότερες εκχωρήσεις δηλαδή όλα τα δείγματα εκπαίδευσης που είναι σχετικά όμοια με τα χαρακτηριστικά του δείγματος ελέγχου, μόνο αυτά θα χρησιμοποιηθούν για να δοθεί η ετικέτα κατηγορίας του δείγματος ελέγχου. Το νέο στοιχείο θα τοποθετηθεί στη συνέχεια στην κατηγορία που περιέχει τα περισσότερα στοιχεία από το σύνολο των κοντινότερων.

Οι k -πλησιέστεροι γείτονες ενός δείγματος z είναι τα k σημεία που είναι πλησιέστερα στο z . Τι συμβαίνει όμως όταν οι γείτονες έχουν περισσότερες από μια ετικέτες; Στην περίπτωση αυτή στο σημείο αποδίδεται ίδια ετικέτα κατηγορία με τους κοντινότερους γείτονες που πλειοψηφούν. Αν δεν υπάρχει πλειοψηφία τότε επιλέγεται τυχαία μια ετικέτα που κατηγοριοποιεί το σημείο δεδομένων. Ο αλγόριθμος κατηγοριοποίησης των K -πλησιέστερων γειτόνων υπολογίζει την απόσταση ανάμεσα σε κάθε δείγμα ελέγχου $z = (x, y)$ και σε όλα τα δείγματα εκπαίδευσης για να καθορίσει τη λίστα των πλησιέστερων γειτόνων.

4.2.1.α. Αλγόριθμος KNN

1. Έστω K το πλήθος των πλησιέστερων γειτόνων και D το σύνολο των εγγραφών εκπαίδευσης.
2. **Για κάθε** δείγμα ελέγχου $z=(x,y)$ **επανάλαβε**
3. **Υπολόγισε** το $d(x',x)$ την απόσταση ανάμεσα στο z $(x,y) \in D$.
4. $D_z \subset D$, το σύνολο των k πλησιέστερων δειγμάτων του z .
5. $Y = v \ \Sigma (x_i, y_i) \in D_z \ I(v=y_i)$
6. **Τέλος για κάθε** (P-N Tan, M. Steinbach and V. Kumar, 2010) (σελ.249)

Όταν δημιουργηθεί η λίστα με τους πλησιέστερους γείτονες, το δείγμα ελέγχου κατηγοριοποιείται με βάση την κατηγορία πλειοψηφίας των κοντινότερων γειτόνων.

$$\text{Majority Voting: } y' = \operatorname{argmax} \sum I(v = y_i)$$

$$v = (x_i, y_i) \in D_z$$

όπου v είναι μια ετικέτα κατηγορίας, y_i η ετικέτα κατηγορίας για έναν από τους πλησιέστερους γείτονες και $I()$ μια συνάρτηση δείκτης, η οποία επιστρέφει την τιμή **1** αν είναι αληθής και **0** αν δεν είναι.

4.2.1.β Περιορισμοί

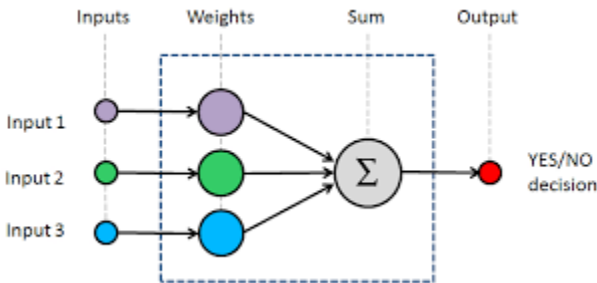
Οι κατηγοριοποιητές πλησιέστερου γείτονα κάνουν προβλέψεις χρησιμοποιώντας τοπικές πληροφορίες, όταν οι τιμές του K είναι μικρές, είναι πολύ επιρρεπείς στο θόρυβο. Επίσης μπορούν να παράγουν αυθαίρετα σχηματισμένα όρια απόφασης. Γεγονός που σημαίνει μια πιο ευέλικτη αναπαράσταση του μοντέλου αλλά και υψηλή μεταβλητότητα λόγω της σύνθεσης δειγμάτων εκπαίδευσης. Ωστόσο αν αυξηθεί το πλήθος των γειτόνων μπορεί να μειωθεί η μεταβλητότητα. Υπάρχει η εκδοχή αν δε γίνει προεπεξεργασία των δεδομένων και δε χρησιμοποιηθεί το κατάλληλο μέτρο εγγύτητας, οι κατηγοριοποιητές κοντινότερου γείτονα μπορεί να παράγουν λανθασμένες προβλέψεις.

4.2.2 Νευρωνικά δίκτυα (Neural Networks)

Τα τεχνητά νευρωνικά δίκτυα στηρίζονται στη λειτουργία των νευρώνων του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος περιέχει περίπου 100 δισεκατομμύρια νευρώνες. Οι νευρώνες είναι κύτταρα που λαμβάνουν πληροφορίες και τις μεταδίδουν σε άλλα κύτταρα με ηλεκτροχημικές ώσεις. Αποτελούνται από τέσσερα κύρια τμήματα: το κυτταρικό σώμα, τους δενδρίτες, το νευράξονα και τις **συνάψεις**. Οι συνάψεις είναι τα σημεία επικοινωνίας και μεταβίβασης των πληροφοριών ανάμεσα στα νευρωνικά κυκλώματα.

Με τρόπο ανάλογο ένα τεχνητό νευρωνικό δίκτυο, αποτελείται από ένα σύνολο διασυνδεδεμένων κόμβων και προσανατολισμένων βελών. Ένας νευρώνας αποτελεί αντίστοιχα με τον ανθρώπινο νευρώνα μια μονάδα επεξεργασίας πληροφορίας με βασικά στοιχεία α) σύνολο από συνάψεις ή συνδετικούς κρίκους β) έναν αθροιστή και γ) μια συνάρτηση ενεργοποίησης.

Το απλούστερο είναι ο νευρώνας **perceptron** και χρησιμοποιείται στην επίλυση προβλημάτων κατηγοριοποίησης. Ο perceptron αποτελείται από δύο τύπους κόμβων. Από τους κόμβους εισόδου που αναπαριστούν τα χαρακτηριστικά εισόδου και έναν κόμβο εξόδου που αναπαριστά την έξοδο του μοντέλου. Οι κόμβοι αυτοί συνδέονται ως είσοδος και έξοδος με έναν σταθμισμένο σύνδεσμο.



σχ.3 Νευρώνας

Αν x_1, x_2, \dots, x_n οι εισόδοι και y η έξοδος, ο νευρώνας υπολογίζει την τιμή εξόδου αθροίζοντας τις εισόδους και αφού αφαιρέσει το *μεροληπτικό παράγοντα* t εξετάζει το πρόσημο $+$ ή $-$ του αθροίσματος.

$$y = \text{sign}(w_1x_1 + w_2x_2 + \dots + w_nx_n - t)$$

Όπου w_1, w_2, w_n είναι τα βάρη των συνδέσμων εισόδου και x_1, x_2, x_n οι τιμές των χαρακτηριστικών εισόδου, η συνάρτηση sign (activation function) του νευρώνα εξόδου δίνει $+1$ αν το όρισμά της είναι θετικό και -1 αν είναι αρνητικό. Για να χρησιμοποιηθεί ένα νευρωνικό δίκτυο πρέπει πρώτα να εκπαιδευτεί. Κατά την εκπαίδευση του μοντέλου αυτού οι παράμετροι βάρους w ρυθμίζονται μέχρι οι έξοδοι να είναι συνεπείς με τις πραγματικές εξόδους των δειγμάτων εκπαίδευσης. Η μάθηση πραγματοποιείται με τη χρήση αλγορίθμων εκπαίδευσης. Γενικά η διαδικασία μάθησης συνοψίζεται στα εξής βήματα:

- Διέγερση του νευρωνικού δικτύου από το περιβάλλον
- Διαδικασία αλλαγών ως συνέπεια της διέγερσης
- Απάντηση με έναν νέο τρόπο.

4.2.2.a Αλγόριθμος εκπαίδευσης του νευρώνα perceptron

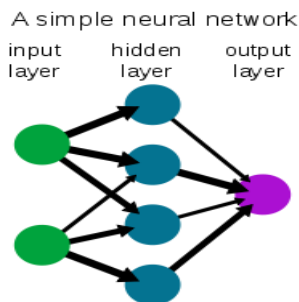
1. Έστω $D = \{(x_i, y_i) | i=1, 2, \dots, N\}$ το σύνολο των δειγμάτων εκπαίδευσης
2. Αρχικοποίησε το διάνυσμα βαρών με τυχαίες τιμές, w .
3. **Επανάλαβε**
4. Για κάθε δείγμα εκπαίδευσης $(x_i, y_i) \in D$ **επανάβαλε**.
5. Υπολόγισε την προβλεπόμενη έξοδο $y_i^{(n)}$

6. Για κάθε βάρος w_j επανέβαλε.
7. Ενημέρωσε το βάρος, $w_j^{(n+1)} = w_j^{(n)} + \lambda(y_i - y_i^{(n)})x_{ij}$
8. Τέλος για κάθε
9. Τέλος για κάθε
10. Μέχρι η συνθήκη τερματισμού να υλοποιηθεί. (P-N Tan, M. Steinbach and V. Kumar, 2010)

Όπου λ είναι η παράμετρος ρυθμός εκπαίδευσης, $w^{(n)}$ η παράμετρος βάρους και x_{ij} η τιμή του χαρακτηριστικού j του δείγματος εκπαίδευσης x_i .

Ωστόσο ενώ ο αλγόριθμος αυτός λειτουργεί καλά σε γραμμικά διαχωρίσιμα προβλήματα κατηγοριοποίησης, αποτυγχάνει όμως σε μη διαχωρίσιμα δεδομένα. Στην περίπτωση αυτή επιτυγχάνουν τα νευρωνικά δίκτυα πολλών επιπέδων που έχουν μια πιο πολύπλοκη δομή. Η πολυπλοκότητα αυτή προκύπτει από την ύπαρξη στο δίκτυο περισσότερων ενδιάμεσων επιπέδων ανάμεσα στα επίπεδα εισόδου και εξόδου (hidden layers) που φέρουν κατά συνέπεια ενσωματωμένους κρυφούς κόμβους (hidden nodes). Έτσι μπορεί να υπάρχουν συνδέσεις με κόμβους στο ίδιο επίπεδο ή από το ένα επίπεδο σε άλλο.

Τυπικά οι νευρώνες σε κάθε επίπεδο έχουν σαν εισόδους τα σήματα εξόδου του προηγούμενου επιπέδου. Τα δίκτυα αυτά ονομάζονται Perceptrons πολλών επιπέδων (MLPs)



σχ.4 Τοπολογία Νευρωνικού δικτύου

Η πολυπλοκότητα μπορεί να αφορά και τη χρήση από το δίκτυο διαφορετικών τύπων συναρτήσεων ενεργοποίησης εκτός από συναρτήσεις πρόσημου (sign function). Γραμμική, σιγμοειδής και υπερβολική εφαπτόμενη. Αυτή η πολυπλοκότητα στη δομή επιτρέπει τη μοντελοποίηση εξίσου πολύπλοκων σχέσεων μεταξύ μεταβλητών εισόδου και εξόδου.

Γενικά πριν εκπαιδευτεί το νευρωνικό δίκτυο για την εκμάθηση μιας ενέργειας κατηγοριοποίησης πρέπει : να καθορίζεται το πλήθος των κόμβων στο επίπεδο εισόδου, να καθορίζεται το πλήθος των κόμβων στο επίπεδο εξόδου να ορίζεται το πλήθος επιπέδων κρυφών κόμβων, να αρχικοποιούνται τα βάρη, οι ελλειπείς τιμές να αφαιρούνται ή να αντικαθίστανται με άλλες πιθανές τιμές.

Και στα νευρωνικά δίκτυα πολλών επιπέδων ο στόχος του αλγόριθμου εκπαίδευσης είναι ένα σύνολο βαρών που θα μικραίνει στο ελάχιστο το συνολικό άθροισμα των τετραγωνικών σφαλμάτων. Όταν γίνεται χρήση γραμμικής συνάρτησης το ολικό ελάχιστο βρίσκεται $E(w)=1/2\sum(y_i-\hat{y}_i)^2$

Όταν όμως η έξοδος λόγω της επιλογής της συνάρτησης είναι μη γραμμική, η διαδικασία είναι περισσότερο πολύπλοκη $w_j \leftarrow w_j - \lambda \theta E(w)/\theta w_j$

Όπου λ =βαθμός εκμάθησης. Για τους κρυφούς κόμβους είναι δύσκολο να εκτιμηθεί ο όρος σφάλματος $\theta E/\theta w_j$ χωρίς να είναι γνωστές οι τιμές εξόδων τους. Η τεχνική για την επίλυση του προβλήματος αυτού είναι η **οπισθοδιάδοση (back-propagation)**. Σε κάθε επανάληψη του αλγόριθμου υπάρχουν δυο φάσεις: η εμπρόσθια και η οπίσθια. Στην εμπρόσθια τα βάρη της προηγούμενης επανάληψης χρησιμοποιούνται για τον υπολογισμό της τιμής εξόδου κάθε νευρώνα στο δίκτυο. Στην οπίσθια συμβαίνει το αντίθετο.

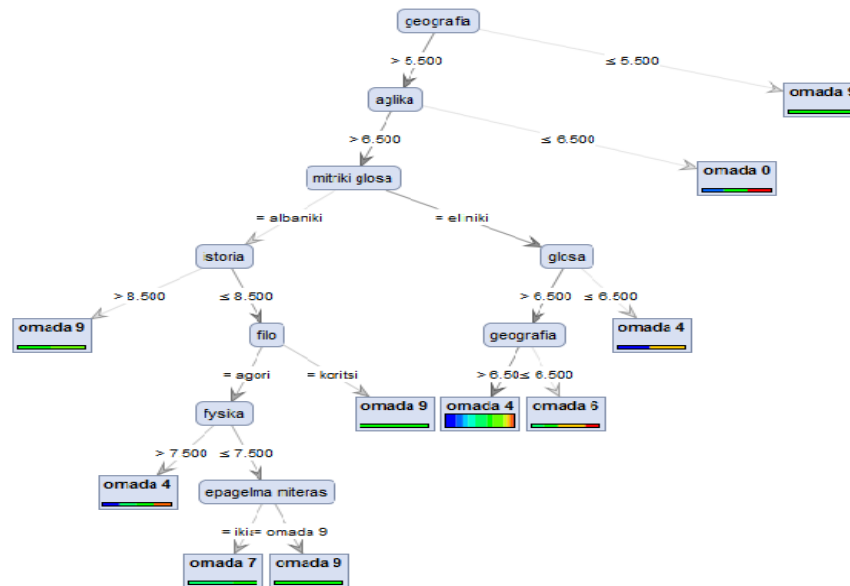
Γενικά τα νευρωνικά δίκτυα με ένα τουλάχιστον κρυφό κόμβο αποτελούν οικουμενικές τεχνικές προσέγγισης, μπορούν να χρησιμοποιηθούν για να προσεγγίσουν όλες τις συναρτήσεις στόχους (282,2010). Όμως « η εκπαίδευση ενός νευρωνικού δικτύου είναι μια χρονοβόρα διαδικασία, ειδικότερα όταν το πλήθος των κρυφών κόμβων είναι μεγάλο. Ωστόσο είναι δυνατή η ταχύτερη κατηγοριοποίηση των δειγμάτων ελέγχου» (P-N Tan, M. Steinbach and V. Kumar, 2010)

4.2.3 Δέντρα απόφασης (decision trees)

Το δέντρο απόφασης είναι μια γραφική απεικόνιση με τη μορφή δέντρου που περιλαμβάνει όλες τις πιθανές αποφάσεις, τις πιθανές επιρροές και τα πιθανά αποτελέσματα.

Ένα δέντρο απόφασης είναι γενικά ένα δέντρο του οποίου η **ρίζα** και κάθε εσωτερικός κόμβος έχει χαρακτηριστεί με μια ερώτηση. Η απάντηση στη σχετική ερώτηση βρίσκεται στα τόξα που προέρχονται από κάθε κόμβο. Κάθε **φύλλο** αντιπροσωπεύει μια πρόβλεψη της λύσης του υπό εξέταση προβλήματος. Στα προβλήματα κατηγοριοποίησης η πρόβλεψη είναι η κατηγορία της πλειάδας που εξετάζεται.

Πιο ειδικά κάθε δέντρο απόφασης έχει τρεις κόμβους. Τον κόμβο που δεν έχει εισερχόμενες ακμές ενώ έχει μηδέν ή n εξερχόμενες και λέγεται **ρίζα**, τους **εσωτερικούς κόμβους** με μια εισερχόμενη ακμή και 2 έως n εξερχόμενες και τα **φύλλα ή τερματικούς κόμβους** που έχουν μία εισερχόμενη ακμή και καμιά εξερχόμενη



σχ.5 Δέντρο απόφασης

Σε κάθε κόμβο φύλλο δίνεται μια ετικέτα κατηγορίας ενώ στη ρίζα και στους εσωτερικούς κόμβους περιέχονται συνθήκες ελέγχου χαρακτηριστικών που κατηγοριοποιούν τις εγγραφές. Οι αλγόριθμοι που χρησιμοποιούνται παρέχουν συνθήκες ελέγχου για τα διαφορετικού τύπου χαρακτηριστικά.

Για **δυαδικά** χαρακτηριστικά η συνθήκη ελέγχου παράγει δυο πιθανά αποτελέσματα ενώ για **ονομαστικά** χαρακτηριστικά ανάλογα με τις διακριτές τιμές η συνθήκη ελέγχου παράγει διαχωρισμό n κατευθύνσεων.

Η κατηγοριοποίηση μιας εγγραφής ελέγχου αρχίζει από τον κόμβο ρίζα. Αφού εφαρμοστεί η συνθήκη ελέγχου στην εγγραφή, από το αποτέλεσμα του ελέγχου ακολουθεί η κατάλληλη διακλάδωση. Και είτε οδηγείται σε έναν εσωτερικό κόμβο που εφαρμόζεται νέα συνθήκη ελέγχου είτε σε ένα φύλλο και τότε η ετικέτα κατηγορίας που σχετίζεται με το φύλλο αποδίδεται στη συνέχεια στην εγγραφή.

4.2.3.α. Αλγόριθμος δέντρου απόφασης

Ο σκελετός του αλγόριθμου επαγωγής ενός δέντρου απόφασης, treeGrowth,

TreeGrowth (E,F)

1. **If** stopping_cond(E,F)=true then
2. leaf=create Node()
3. Leaf. Label=Classify(E)

4. Return leaf.
5. **else**
6. root=create Node()
7. root.test-cond=find_best_split(E,F)
8. let $V=\{u \mid u \text{ is a possible outcome of root.test-cond}\}$
9. **for** each $u \in V$ do
10. $E_u=\{e \mid \text{root.test_cond}(e)=u \text{ and } e \in E\}$
11. Child=TreeGrowth (E_u,F).
12. add child as descendent of root and label the edge(root-child)as u
13. **end for**
14. **end if**
15. return root (Dunham , 2004)

4.2.3.β. Γνωρίσματα κατασκευής δέντρου απόφασης

Ακόμα και αν το μέγεθος του συνόλου εκπαίδευσης είναι πολύ μεγάλο οι τεχνικές που αναπτύσσονται είναι υπολογιστικά φθηνές με αποτέλεσμα η κατασκευή των μοντέλων να είναι αρκετά γρήγορη. Επιπλέον η κατηγοριοποίηση μιας εγγραφής ελέγχου από τη στιγμή που δημιουργείται το δέντρο είναι πολύ γρήγορη.

Η ερμηνεία ειδικά στα μικρά δέντρα απόφασης είναι εύκολη και σχετικά ακριβής. Οι αλγόριθμοι δέντρων απόφασης παρουσιάζουν ανθεκτικότητα στην παρουσία θορύβου.

Η ακρίβεια των αποτελεσμάτων δε φαίνεται να επηρεάζεται από περιττά χαρακτηριστικά. Ωστόσο αν υπάρχουν χαρακτηριστικά άχρηστα για την εργασία της κατηγοριοποίησης τότε η πιθανά τυχαία επιλογή μερικών από αυτά , μπορεί να οδηγήσει στην ανάπτυξη ενός πολύ μεγαλύτερου απ ότι χρειάζεται δέντρου.

Επειδή οι αλγόριθμοι των δέντρων απόφασης βασίζονται στο «διαίρει και βασίλευε» η ίδια συνθήκη ελέγχου μπορεί να εφαρμοστεί και σε διαφορετικά τμήματα του χώρου που καταλαμβάνει το δέντρο με αποτέλεσμα να δημιουργούνται επαναλαμβανόμενα υποδέντρα που κάνουν το δέντρο πολύπλοκο και δύσκολο στην ερμηνεία.

4.2.3.γ. Αλγόριθμος ID3

Ο αλγόριθμος ID3 στηρίζεται στην επιστημονική αρχή της «οικονομίας» του Γουλιέλμου Όκαμ, σύμφωνα με την οποία «Δοθέντων δυο μοντέλων με τα ίδια σφάλματα γενίκευσης, το απλούστερο μοντέλο είναι προτιμότερο από το πιο πολύπλοκο» *ξυράφι του Occam* (P-N Tan, M. Steinbach and V. Kumar, 2010) Ο ID3 είναι ένας άπληστος αλγόριθμος που υποστηρίζει την ιδέα ότι η επιλογή των χαρακτηριστικών διάσπασης σχετίζεται με το μεγαλύτερο **κέρδος πληροφορίας**. Το κέρδος αυτό σημαίνει ερωτήσεις που απορρίπτουν μεγάλο μέρος του χώρου αναζήτησης, με την ελαχιστοποίηση του αριθμού των συγκρίσεων. Βασικό στάδιο του αλγόριθμου είναι η επιλογή της ανεξάρτητης μεταβλητής πάνω στην οποία θα συνεχιστεί η ανάπτυξη του δέντρου. Για την επιλογή αυτή απαιτείται μηχανισμός που καθοδηγεί προς την αναζήτηση του καλύτερου δέντρου. Η συνάρτηση αξιολόγησης είναι το **κέρδος πληροφορίας** που βασίζεται στο μέγεθος της **Εντροπίας**.

Έστω S ένα σύνολο δεδομένων. Η Εντροπία χαρακτηρίζει το βαθμό αβεβαιότητας

$$E(S) = \sum p_i \cdot \log_2 1/p_i$$

Όπου p_1, p_2, \dots, p_i οι πιθανότητες του κάθε ενδεχομένου που περιλαμβάνεται στο σύνολο. Π.χ.

Αν κουτί με N σχήματα.

$N \cdot p$ τρίγωνα και $N(1-p)$ τετράγωνα

Αν είναι όλα τρίγωνα ή όλα τετράγωνα τότε η Εντροπία=0

Αν $p=50\%$ τότε η Εντροπία=1 και είναι η μέγιστη.

Information Gain

Έστω A χαρακτηριστικό σε S σύνολο δεδομένων. Κέρδος είναι το πόση πληροφορία φέρει το χαρακτηριστικό.

m

$$G(S, A) = E(S) - \sum_{i=1}^m f_s(A_i) \cdot E(SA_i)$$

i=1

Όπου E = συνάρτηση Εντροπίας., m πλήθος τιμών που το A παίρνει στο $SF(A_i)$ ποσοστό αντικειμένων στο S που παίρνουν τιμή A_i , S_A το υποσύνολο του S όπου η τιμή του A είναι A_i .

Μια γενική περιγραφή του αλγόριθμου είναι η παρακάτω:

- Επιλογή του κατάλληλου χαρακτηριστικού για έλεγχο στη ρίζα.
- Για κάθε δυνατή τιμή του χαρακτηριστικού δημιουργούνται οι αντίστοιχοι κόμβοι. Τα δεδομένα θα μοιραστούν στους νέους κόμβους ανάλογα με την τιμή που έχουν για το χαρακτηριστικό που ελέγχεται στη ρίζα.
- Για κάθε νέο κόμβο επαναλαμβάνεται η ίδια διαδικασία.
- Ένας κόμβος γίνεται τερματικός όταν όλα τα δεδομένα που ανήκουν σε αυτόν ανήκουν στην ίδια κατηγορία και τότε λέγεται **αμιγής κόμβος (pure node)**. Τερματικός είναι και ένας κόμβος που σε κάποιο βάθος τελειώνουν τα προς έλεγχο χαρακτηριστικά και τότε η τιμή του συμπίπτει με την κατηγορία στην οποία ανήκει η πλειοψηφία των δεδομένων του κόμβου αυτού.

4.2.3.δ. Ανάλυση συσχέτισης

Η ανάλυση συσχέτισης έχει στόχο την ανακάλυψη κρυμμένων συσχετίσεων μεταξύ των χαρακτηριστικών μιας βάσης δεδομένων. Με την ανάλυση συσχέτισης αναζητούνται κανόνες για την ποσοτικοποίηση των σχέσεων μεταξύ δύο ή περισσότερων χαρακτηριστικών της βάσης δεδομένων. Οι κανόνες αυτοί ονομάζονται κανόνες συσχέτισης (association rules).

Χαρακτηριστικό των κανόνων συσχέτισης είναι το κατώφλι στήριξης (support threshold) που αναγνωρίζει τα στοιχεία που εμφανίζονται συχνά σε ένα σύνολο δεδομένων και το κατώφλι εμπιστοσύνης (confidence threshold) που ορίζεται σαν η υπό συνθήκη πιθανότητα της εμφάνισης (conditional probability) ενός στοιχείου σε μια διαδικασία όταν ένα άλλο στοιχείο εμφανίζεται.

4.3 Bayes Δίκτυα

Τα δίκτυα Bayes παρέχουν ένα ισχυρό πιθανολογικό μηχανισμό για συμπερασμό υπό συνθήκες αβεβαιότητας. Λόγω της δομής τους μπορούν να αναπαραστήσουν πεποιθήσεις και γνώση σχετικά με την εξεταζόμενη θεματική περιοχή.

Τα δίκτυα Bayes είναι γραφικά μοντέλα που κωδικοποιούν πιθανολογικές σχέσεις μεταξύ μεταβλητών που παρουσιάζουν ενδιαφέρον. Είναι σύνθετα διαγράμματα που παρουσιάζουν με δομημένο τρόπο το σύνολο των δεδομένων περιγράφοντας σχέσεις αιτίας αποτελέσματος μεταξύ των βασικών μεταβλητών ενώ παρουσιάζουν σε ποσοστά το βαθμό στον οποίο είναι πιθανό να επηρεάσει η μία μεταβλητή την άλλη.

Αν και όπως προαναφέρθηκε υπάρχουν αρκετοί τρόποι ανάλυσης και αναπαράστασης των δεδομένων για τα δεδομένα της παρούσας έρευνας επιλέχθηκε η τεχνική των δικτύων Bayes επειδή επιτρέπουν τη γνώση αιτιωδών σχέσεων μεταξύ των μεταβλητών που μπορεί να οδηγήσουν σε πρόβλεψη αλλά επιτρέπουν και μια πιο **επεξηγηματική ανάλυση** αφού μπορεί να παρατηρηθεί το κατά πόσο επηρεάζει η απομόνωση μιας μεταβλητής, τις υπόλοιπες.

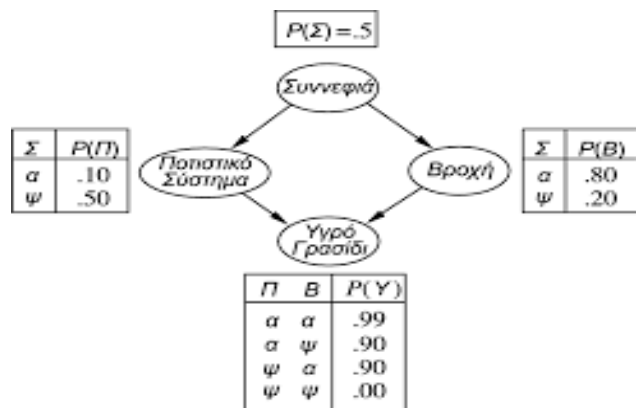
Τα βασικά στοιχεία ενός δικτύου Bayes είναι :

- η κωδικοποίηση των σχέσεων εξάρτησης ανάμεσα σε ένα σύνολο μεταβλητών αποτυπώνεται σε έναν μη κυκλικό κατευθυνόμενο γράφο (directed acyclic graph-dag)
- Η συσχέτιση κάθε κόμβου με τους άμεσους κόμβους –γονείς , αποδίδεται με έναν πίνακα πιθανοτήτων.

Πιο συγκεκριμένα σε ένα δίκτυο Bayes για σύνολο μεταβλητών υπάρχει μια δομή δικτύου S , η οποία υποδηλώνει το σύνολο των εξαρτήσεων μεταξύ των μεταβλητών του συνόλου. Επίσης δηλώνει από ένα σύνολο P , τις τοπικές κατανομές πιθανοτήτων που συσχετίζονται με κάθε μια από αυτές τις μεταβλητές. Το X και το P ορίζουν τη συνδυασμένη κατανομή πιθανότητας για το ίδιο το X .

Η δομή του δικτύου S είναι ένας μη κυκλικός γράφος, δηλαδή γράφος με κατευθυνόμενα τόξα σε κάθε σύνδεσμο που λέγονται ακμές ο οποίος δεν παρουσιάζει κλειστά μονοπάτια (κατευθυνόμενους κύκλους). Ο γράφος αυτός έχει κόμβους που συσχετίζονται ένας προς ένα με το σύνολο των μεταβλητών. X_i είναι μια μεταβλητή και άρα ένας κόμβος του δικτύου, $Parents(X_i)$ οι γονείς του κόμβου X_i στο S . Οι συσχετιζόμενες μεταβλητές ενώνονται με ακμές ενώ η έλλειψη ακμών μεταξύ δυο μεταβλητών του δικτύου , δηλώνει υπό συνθήκη ανεξαρτησία μεταξύ τους

$$P(X_1 \dots X_n) = \prod_{i=1} P(X_i | \text{parents}(X_i))$$



σχ. 6 δίκτυα Bayes

Κάθε κόμβος στο γράφο αναπαριστά μια μεταβλητή και κάθε βέλος δείχνει τη σχέση εξάρτησης ανάμεσα στο ζευγάρι μεταβλητών. Ο προσανατολισμένο κόμβος από Συννεφιά στο βροχή δείχνει ότι το Συννεφιά είναι ο γονέας του κόμβου Βροχή και ο κόμβος Υγρό γρασίδι αντίστοιχα είναι παιδί του κόμβου Βροχή. Γενικά αν στο δίκτυο υπάρχει ένα προσανατολισμένο μονοπάτι από έναν κόμβο A σε έναν κόμβο B τότε ο A είναι πρόγονος του B και ο B είναι απόγονος του A. Πιο συγκεκριμένα Το ενδεχόμενο το γρασίδι να είναι υγρό εξαρτάται από το ενδεχόμενο «Ποτιστικό σύστημα» και το ενδεχόμενο «Βροχή». Αν συμβαίνουν και τα δύο ο επισυναπτόμενος πίνακας κατανομών πιθανότητας δίνει το ποσοστό το γρασίδι να είναι βρεγμένο 99%. Αν συμβαίνει το ένα από τα δύο, τότε η πιθανότητα να είναι το γρασίδι βρεγμένο είναι 90%. Αν κανένα από τα δύο ενδεχόμενα δε συμβαίνει οι πιθανότητες να είναι βρεγμένο το γρασίδι είναι 0%. Τα δύο ενδεχόμενα με τη σειρά τους επηρεάζονται από το αν έχει ή όχι συννεφιά. Η πιθανότητα να ανοίξει το ποτιστικό αν έχει συννεφιά είναι 10% και 50% να μην ανοίξει. Η πιθανότητα να βρέξει αν έχει συννεφιά είναι 80% και να μη βρέξει 20%.

Όπως αναφέρθηκε και προηγουμένως κάθε κόμβος συνδέεται επίσης με έναν πίνακα πιθανοτήτων ο οποίος καταγράφει μόνο την **εκ των προτέρων πιθανότητα P(X)** για έναν κόμβο X που δεν έχει γονείς. Καταγράφει την **υπό συνθήκη πιθανότητα P(X | Y)** για έναν κόμβο X που έχει έναν μόνο γονέα. Ενώ για κόμβους με πολλαπλούς γονείς (Y1, Y2, ..., Yk) περιέχει την υπό συνθήκη πιθανότητα P(X | Y1, Y2, ..., Yk). Εκφράζει την πιθανότητα του να βρίσκεται μια μεταβλητή X σε μια κατάσταση χ

,αν ο πατέρας Y βρίσκεται στην κατάσταση Y_1 και ο πατέρας Y_k στην κατάσταση Y_k . Οι κόμβοι που δεν έχουν γονέα, έχουν κι αυτές πίνακα υπό συνθήκη πιθανοτήτων για κάθε κατάσταση κόμβου.

4.1.α Αλγόριθμος παραγωγής της τοπολογίας ενός δικτύου Bayes

Η διαδικασία παραγωγής της τοπολογίας ενός δικτύου Bayes συνοψίζεται στον παρακάτω αλγόριθμο:

1. Έστω $T=(X_1,X_2,\dots,X_k)$ μια ολική διάταξη του συνόλου των μεταβλητών.
2. **Για** $j=1$ ως k **επανάλαβε**
3. Έστω ότι $X_{T(j)}$ είναι η μεγαλύτερη στη διάταξη μεταβλητή στο T .
4. Έστω ότι το $\pi(X_{T(j)}) = \{X_{T(1)},X_{T(2)},\dots,X_{T(j-1)}\}$ δηλώνει το σύνολο των μεταβλητών που προηγούνται του $X_{T(j)}$
5. Αφαίρεσε τις μεταβλητές από το $\pi(X_{T(j)})$ που δεν επηρεάζουν το X_j (χρησιμοποιώντας εκ των προτέρων γνώση).
6. Σχεδίασε ένα βέλος μεταξύ του $X_{T(j)}$ και των υπόλοιπων μεταβλητών του $\pi(X_{T(j)})$.
7. Τέλος (Dunham , 2004)

Χαρακτηριστικό είναι ότι ο αλγόριθμος αυτός αποτρέπει τη σύνδεση των ψηλότερων με χαμηλότερους στη διάταξη κόμβους με αποτέλεσμα να μην υπάρχουν κύκλοι στην τοπολογία. Επιλέχθηκε το γραφικό αυτό μοντέλο για την ανάλυση των δεδομένων επειδή κωδικοποιεί εξαρτήσεις μεταξύ όλων των μεταβλητών , είναι ευέλικτο στο χειρισμό καταχωρήσεων ακόμα κι αν τα δεδομένα παρουσιάζουν ελλείψεις , δίνει τη δυνατότητα αξιολόγησης αιτιωδών σχέσεων που βοηθούν στην κατανόηση των αποτελεσμάτων του αρχικού στόχου μέσα από ένα πολυεστιακό πρίσμα .

4.1.β Δομή των Δικτύων Bayes

Τα δίκτυα Bayes είναι μοντέλα που απεικονίζονται γραφικά και έχουν τη μορφή ενός άκυκλου γράφου. Κάθε μεταβλητή αναπαριστάται με ένα κόμβο, ενώ οι σχέσεις των μεταβλητών σημειώνονται με προσανατολισμένα βέλη. Κάθε **κόμβος** αναπαρίσταται γραφικά με ένα οβάλ σχήμα το οποίο έχει κάποια ετικέτα.

Μια **πλευρά** (edge) δηλώνει μια σχέση επιρροής ανάμεσα σε δυο κόμβους. Απεικονίζεται με προσανατολισμένο βέλος του οποίου ο προσανατολισμός δείχνει και την κατεύθυνση της άμεσης

επιρροής. Ο τρόπος , το ποσοστό που ένας κόμβος επηρεάζει έναν άλλον αποτυπώνεται στον πίνακα των υπό συνθήκη πιθανοτήτων.

Όταν δυο κόμβοι συνδέονται με μια πλευρά, ο κόμβος από τον οποίο ξεκινάει το βέλος λέγεται **πατέρας**(parent), ενώ ο κόμβος στον οποίο καταλήγει ονομάζεται **παιδί**(Child).

Οι τιμές που μπορεί να πάρει μια μεταβλητή ονομάζονται **καταστάσεις** της μεταβλητής.

Σε κάθε κόμβο αντιστοιχεί ένας πίνακας υπό συνθήκη πιθανοτήτων. Οι πιθανότητες αυτές στηρίζονται σε προηγούμενη γνώση. Η υπό συνθήκη πιθανότητα εκφράζει, όπως αναφέρθηκε και παραπάνω, την πιθανότητα να βρίσκεται μια μεταβλητή X σε μια κατάσταση χ , όταν ο πατέρας $P1$ βρίσκεται στην κατάσταση $\rho1$, ο πατέρας $P2$ στην κατάσταση $\rho2$ και ο πατέρας Pn στην κατάσταση ρn .

Κεφάλαιο 5

ΜΕΘΟΔΟΛΟΓΙΑ

Γενικός σκοπός της έρευνας αυτής είναι η μελέτη περίπτωσης για τη σχέση της σχολικής επίδοσης με την επιλογή του επαγγέλματος και τη σχέση του επαγγέλματος των γονιών με τη σχολική επίδοση των μαθητών. Για το σκοπό αυτό δημιουργήθηκαν δυο μοντέλα . Το πρώτο, **μοντέλο Α**, ερευνά τη σχέση της σχολικής επίδοσης με την επιλογή επαγγέλματος , εφαρμόζοντας μεθόδους εξόρυξης δεδομένων. Το δεύτερο, **μοντέλο Β**, ερευνά τη σχέση του επαγγέλματος των γονιών με τη σχολική επίδοση των μαθητών, εφαρμόζοντας μεθόδους εξόρυξης δεδομένων.

5.1 Στόχοι

Α μοντέλο

Το Α μοντέλο ερευνά τη σχέση της σχολικής επίδοσης με την επιλογή του επαγγέλματος, για τη διερεύνηση αυτή τέθηκαν, σε σχέση και με το θεωρητικό πλαίσιο της έρευνας οι εξής επιμέρους στόχοι:

Να διερευνηθεί πιθανή σχέση ανάμεσα στη σχολική επίδοση της Στ δημοτικού και της επιλογής επαγγέλματος. Η σχολική επίδοση όπως εκφράζεται με τη βαθμολογία στην ΣΤ΄ δημοτικού επηρέασε ή καλλιέργησε κάποιες φιλοδοξίες για ένα συγκεκριμένο επαγγελματικό τομέα;

Να διερευνηθεί η δυναμική που ασκεί η σχολική επίδοση ως βαθμολογία στους μαθητές και το οικογενειακό περιβάλλον τους. Η βαθμολογία είναι ένα σημαντικό γεγονός για τους ίδιους τους μαθητές και το οικογενειακό τους περιβάλλον; Ποια είναι τα μηνύματα που λαμβάνονται και ποιες οι επιδράσεις;

Να διερευνηθούν εξατομικευμένα χαρακτηριστικά (επαγγελματικό προφίλ , μαθητικό προφίλ) των ατόμων που συμμετείχαν. Αφού περιγραφεί το προφίλ των ενηλίκων αυτών θα αναζητηθούν σχέσεις ανάμεσα στο επαγγελματικό και το μαθητικό προφίλ.

Να αξιολογηθεί μοντέλο πρόβλεψης ως προς την ικανότητα να προβλέπει το επάγγελμα από την επίδοση σε συγκεκριμένα μαθήματα της Στ Δημοτικού με απώτερο σκοπό έναν πιο στοχευόμενο επαγγελματικό προσανατολισμό.

Β μοντέλο

Να διερευνηθεί πιθανή σχέση του επαγγέλματος των γονιών με την επίδοση στην Στ Δημοτικού. Το επάγγελμα των γονιών σχετίζεται με την επίδοση των μαθητών στην τελευταία τάξη του Δημοτικού;

Να αξιολογηθεί μοντέλο πρόβλεψης ως προς την ικανότητα να προβλέπει την επίδοση από το επάγγελμα των γονιών με απώτερο σκοπό την έγκαιρη παρέμβαση.

5.2 Διερευνητικά ερωτήματα

A μοντέλο

Τα ερωτήματα που διερευνήθηκαν και εξυπηρετούσαν τους παραπάνω στόχους είναι:

- Επηρέασε η σχολική επίδοση την επιλογή επαγγέλματος;
- Είναι η σχολική επίδοση αντιπροσωπευτική των ικανοτήτων;
- Έχει η βαθμολογία την έννοια της διαμορφωτικής αξιολόγησης;
- Ποια η αντιμετώπιση από το οικογενειακό περιβάλλον;
- Ποιο είναι το επαγγελματικό προφίλ του δείγματος;
- Ποιο ήταν το μαθητικό του προφίλ;
- Ποιες οι πιθανολογικές σχέσεις των απαντήσεων;
- Ποιο σύνολο εκπαίδευσης είναι πιο αξιόπιστο για τη δημιουργία ενός μοντέλου πρόβλεψης;

B μοντέλο

- Επηρεάζει το επάγγελμα των γονιών τη σχολική επίδοση;
- Ποιες πιθανολογικές σχέσεις δημιουργούνται ;
- Υπάρχει δυνατότητα δημιουργίας ενός αξιόπιστου μοντέλου εκπαίδευσης που θα οδηγεί στην πρόβλεψη της επίδοσης ανάλογα με το επάγγελμα των γονιών;

5.3 Συλλογή δεδομένων

A Μοντέλο

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Με βάση τα παραπάνω διερευνητικά ερωτήματα συντάχτηκε ερωτηματολόγιο στον ιστότοπο <https://drive.google.com/drive/my-drive> που είχε την παρακάτω μορφή (Οι ερωτήσεις που συμπεριλήφθηκαν ήταν κλειστού τύπου και η πολλαπλών επιλογών έγιναν σύμφωνα με την κλίμακα Likert με κωδικοποίηση 1 – 5):

Form Settings

- Show progress bar at the bottom of form pages
- Only allow one response per person (requires login) ?
- Shuffle question order ?

Page 1 of 1

Ερωτηματολόγιο συνάφειας σχολικής επίδοσης και επιλογής ε

Το παρόν ερωτηματολόγιο δημιουργήθηκε στο πλαίσιο της μεταπτυχιακής εργασίας της Μαλιαρίτση Διονυσίας μεταπτυχιακής φοιτήτρια του τμήματος πληροφοριακών συστημάτων του πανεπιστημίου Αιγαίου και Δασκάλας στο επάγγελμα. Τα δεδομένα που θα συλλέγονται θα χρησιμοποιηθούν για την εξαγωγή χρήσιμης γνώσης σχετικά με τη σχολική επίδοση και το επάγγελμα. Το ερωτηματολόγιο αποτελείται από 12 κύριες ερωτήσεις και απαιτεί λιγότερο από 5 λεπτά να συμπληρωθεί. Δεν είναι ονομαστικό και δεν απαιτείται η συμπλήρωση προσωπικών δεδομένων. Σας ευχαριστώ εκ των προτέρων για τη συμμετοχή σας.

2. Ηλικία

3. Σε ποια από τις παρακάτω ομάδες ανήκει το επάγγελμά σας



- Μέλη των βουλευομένων σωμάτων
- Ανώτερα διοικητικά και διευθυντικά στελέχη του ιδιωτικού και δημόσιου τομέα
- Πρόσωπα που ασκούν επιστημονικά, καλλιτεχνικά και συναφή επαγγέλματα
- Τεχνολόγοι, τεχνικοί βοηθοί
- Υπάλληλοι γραφείου
- Απασχολούμενοι στην παροχή υπηρεσιών και πωλητές
- Ειδικευμένοι γεωργοί –κτηνοτρόφοι
- Ειδικευμένοι τεχνίτες
- Χειριστές σταθερών βιομηχανικών εγκαταστάσεων μηχανών
- Ανείδικευτοι εργάτες, χειρωνακτές
- Ένοπλες δυνάμεις

4. Ποιο είναι το επάγγελμά σας

5. Η βαθμολογία μου στο δημοτικό*

	Πολύ συχνά	Συχνά	Κάποιες φορές	Σπάνια	Ποτέ
Μου προκαλούσε άγχος	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ήταν ένα γεγονός που με χαροποιούσε	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ήταν ένα αδιάφορο γεγονός	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Με βοηθούσε να βάζω στόχους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Με αποθάρρυνε από το να βάζω στόχους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ήταν η αφορμή για καβγάδες στο σπίτι	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ήταν η αφορμή για επαινετικά σχόλια από τους γονείς	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είχε αδιάφορη αντιμετώπιση από το σπίτι	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

δ. Πιστεύω ότι η βαθμολογία μου στο δημοτικό



	Πάρα πολύ	Πολύ	Λίγο	Ελάχιστα	καθόλου
Με αδικούσε σε κάποια μαθήματα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Υπερτιμούσε τις ικανότητές μου σε κάποιο μάθημα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Στο μάθημα που είχα την υψηλότερη βαθμολογία ήμουν πραγματικά καλός	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Με βοήθησε να καταλάβω που ήμουν πραγματικά καλός	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου καλλιέργησε φιλοδοξίες για την επαγγελματική μου επιλογή	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Με αποθάρρυνε από φιλοδοξίες για την επαγγελματική μου επιλογή	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Δεν έπαιξε κανένα ρόλο στην επιλογή επαγγέλματος	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

ε. Στην επιλογή του επαγγέλματος μου ήταν καθοριστικός παράγοντας*

	Πάρα πολύ	Πολύ	Λίγο	Ελάχιστα	Καθόλου
Η παρότρυνση της οικογένειάς μου.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Οι επιδόσεις μου σε ένα μάθημα .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η τελική μου βαθμολογία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Η παρότρυνση ενός άλλου προσώπου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Οικονομικοί λόγοι	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Συγκυρίες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

β. Στο επάγγελμά μου*

	Πάρα Πολύ	Πολύ	Λίγο	Ελάχιστα	Καθόλου
Νιώθω δημιουργικός	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Νιώθω χαρούμενος	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου προκαλεί άγχος	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου προκαλεί ανασφάλεια	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

κατέκτησα τους στόχους μου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είμαι καλός σε αυτό που κάνω	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Εκμεταλεύομαι τις ικανότητές που έχω	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Έχω ικανότητες που μένουν ανεκμετάλλευτες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9.Ως μαθητής*

	Πάρα πολύ	Πολύ	Λίγο	Ελάχιστα	καθόλου
Κατέβαλα μεγάλη προσπάθεια στην προετοιμασία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Κατέβαλα μικρή προσπάθεια στην προετοιμασία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Χρειάζομαι ενίσχυση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είχα την ενίσχυση που χρειαζόμουν	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Δεν είχα καμιά ενίσχυση	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είχα μαθησιακές δυσκολίες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10.Στο δημοτικό*

	Πολύ συχνά	Συχνά	Κάποιες φορές	Σπάνια	καθόλου
Μου άρεσε η Γλώσσα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου άρεσαν τα μαθηματικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου άρεσαν τα καλλιτεχνικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου άρεσε η Μουσική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου άρεσε η Γυμναστική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μου άρεσαν οι Φυσικές επιστήμες	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είχα ελλείψεις στη Γλώσσα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Είχα ελλείψεις στα Μαθηματικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αντιμετώπιζα δυσκολίες στις σχέσεις με τους συμμαθητές μου	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αντιμετώπιζα δυσκολίες στις σχέσεις μου με τους δασκάλους	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11.Τα Ελληνικά είναι η μητρική μου γλώσσα

12.Η βαθμολογία μου στην Στ Δημοτικού ήταν*

	10	09	08	07	06	05
Γλώσσα	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μαθηματικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Φυσικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Γεωγραφία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ιστορία	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Εικαστικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Μουσική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Γυμναστική	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Αγγλικά	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Add item

εικ. 1 Ερωτηματολόγιο

Τα ερωτηματολόγια στάλθηκαν ηλεκτρονικά σε 100 άτομα που διαμένουν σε διάφορες περιοχές της Ελλάδας και ηλικιακά κάλυπταν τα έτη 25-45. Από τους 100 απάντησαν οι 60. Επιλέχθηκε ο ηλεκτρονικός τρόπος λόγω της ευκολίας στην σύνταξη και την αποστολή και σε πιο μακρινές περιοχές, σε συνδυασμό με την ευκολία στην επεξεργασία των απαντήσεων αφού υπάρχει η δυνατότητα άμεσης αποθήκευσης σε αρχείο Excel . Θεωρώ όμως ότι ήταν και ανασταλτικός παράγοντας στη συμπλήρωσή του η μη άμεση προσωπική επαφή γεγονός στο οποίο αποδίδεται ότι απάντησαν μόνο οι 60. Ο κάθε ερωτώμενος ήταν διασφαλισμένο ότι μπορούσε να συμπληρώσει το ερωτηματολόγιο μόνο μια φορά.

Το live form του ερωτηματολογίου βρίσκεται στη διεύθυνση:

https://docs.google.com/forms/d/1y42exCAt3la1lxnnFqfGriHyLtCAC-LcUkS-e-KxKFY/viewform?c=0&w=1&usp=mail_form_link

B Μοντέλο

Το Μητρώο είναι το επίσημο βιβλίο στο οποίο εγγράφονται τα στοιχεία των μαθητών του σχολείου μόνο μια φορά κατά τη διάρκεια της φοίτησής τους στο σχολείο. (ΥΠΕΠΘ, 1998). Οι μαθητές εγγράφονται κατά τάξη σε αλφαβητική σειρά (πρώτα το επώνυμο). Τα στοιχεία των 250 μαθητών : επάγγελμα μητέρας –πατέρα και αναλυτική βαθμολογία στην Έκτη (ΣΤ΄) τάξη κατά μάθημα , που ήταν αναγκαία για το B μοντέλο της έρευνας εξήχθησαν από το βιβλίο μητρώου χωρίς το ονοματεπώνυμο για λόγους διαφύλαξης των προσωπικών δεδομένων των μαθητών

5.4 Εργαλεία

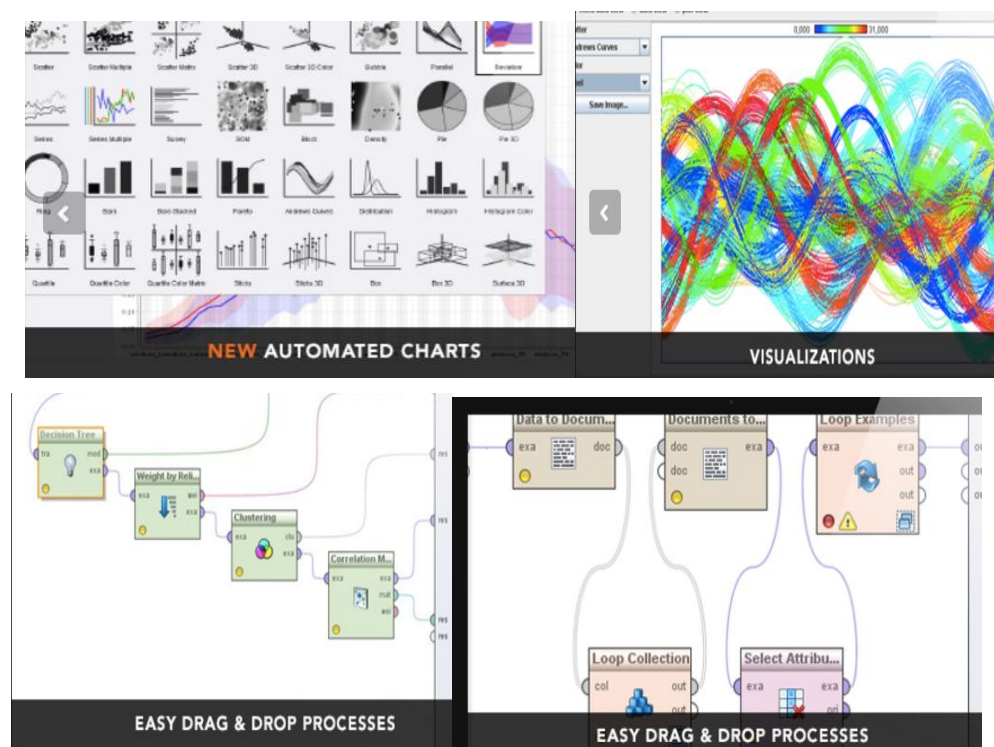
Η ανάλυση των δεδομένων έγινε σε δύο στάδια. Αρχικά **ανάλυση της συμπεριφοράς** των ερωτηθέντων και στη συνέχεια **πρόβλεψη**. Η τεχνική που επιλέχθηκε είναι η **κατηγοριοποίηση** και τα δίκτυα Bayes επειδή παρέχουν δυνατότητα δημιουργίας σεναρίων πρόβλεψης με το συσχετισμό των μεταβλητών. Στη συνέχεια γίνεται μια αναφορά των εργαλείων που χρησιμοποιήθηκαν σε κάθε περίπτωση.

5.4.1 Κατηγοριοποίηση

RapidMiner

Το rapidminer είναι ένα λογισμικό δωρεάν διαθέσιμο στην αρχική του έκδοση , από την επίσημη ιστοσελίδα <http://rapidminer-i.com/> . Είναι μια πλατφόρμα ευρέως διαδεδομένη και περιέχει ένα ολοκληρωμένο περιβάλλον για τη μηχανική μάθηση και την εξόρυξη δεδομένων και κειμένου . Χρησιμοποιείται τόσο για επιστημονικές όσο και για επιχειρηματικές εφαρμογές ενώ διαθέτει υποστήριξη για όλα τα στάδια της διαδικασίας εξόρυξης δεδομένων .Το περιβάλλον της πλατφόρμας είναι εύχρηστο και υπάρχει υποστήριξη με ειδικά video στο Tube όπου επιστήμονες εξηγούν τον τρόπο λειτουργίας ανάλογα με το σκοπό : “loading text Info: “Processing text in Rapid miner”. “Text mining using Rapid miner” ...

Μερικές από τις δυνατότητες παρουσιάζονται από το επίσημο site στις εικόνες:



Εικ.2 RapidMiner

Είναι γραμμένο σε γλώσσα προγραμματισμού Java ενώ διαθέτει εσωτερική xml αναπαράσταση ώστε να είναι προσβάσιμη η τυποποιημένη μορφή ανταλλαγής εξόρυξη δεδομένων στα πειράματα που τελούνται.

ΙΣΤΟΡΙΑ

Η αρχική ονομασία του RapidMiner έως το 2006, ήταν YALE (Yet Another Learning Environment) και ξεκίνησε στο πανεπιστήμιο του Dortmund στο τμήμα Τεχνητής Νοημοσύνης το 2001, από τους Ralf Klinkenberg, Ingo Mierswa και Simon Fischer .

Το 2006 όμως οι Ralf Klinkenberg και Ingo Mierswa ίδρυσαν την εταιρεία Rapid-I που είναι κυρίως υπεύθυνη με τη βοήθεια 50 περίπου προγραμματιστών παγκοσμίως για την εξέλιξη του RapidMiner.

Από τα εργαλεία που χρησιμοποιήθηκαν για πραγματικά έργα (δημοσκόπηση της εφημερίδας, για την εξόρυξη δεδομένων, KDnuggets, 2010) το περισσότερο διαδεδομένο ήταν το rapidMiner. Διανέμεται υπό την άδεια χρήσης λογισμικών ανοικτού κώδικα AGPL και στεγάζεται στο Source Forge (δωρεάν open source ανάπτυξης λογισμικού) από το 2004.

Το rapid miner χρησιμοποιήθηκε εδώ στα πειράματα κατηγοριοποίησης με τη δημιουργία δέντρων απόφασης και εφαρμογή αλγορίθμων KNN και Neural net με σκοπό να εξαχθούν συμπεράσματα από τα δεδομένα των ερωτηματολογίων, αλλά και να αξιολογηθούν αυτά ως προς τις δυνατότητες μιας προβλεπτικής διαδικασίας.

5.4.2 Δίκτυα Bayes

GeNIe 2.0

Το GeNIe 2.0 (Graphical network interface) είναι ένα λογισμικό που δημιουργήθηκε στο Πανεπιστήμιο του Pittsburgh για τη Microsoft. Η εγκατάστασή του είναι δωρεάν .Πρόκειται για μια γραφική απεικόνιση πιθανολογικών ενώσεων μεταξύ των δεδομένων.

Πιο ειδικά χρησιμοποιείται για την ανάλυση των δικτύων Bayes διευκολύνοντας τη συμπερίληψη θορυβωδών ή ελλιπών δεδομένων και οδηγώντας σε αποτελεσματικές προβλέψεις πιθανοτήτων ακόμα και σε πολύπλοκα συστήματα.

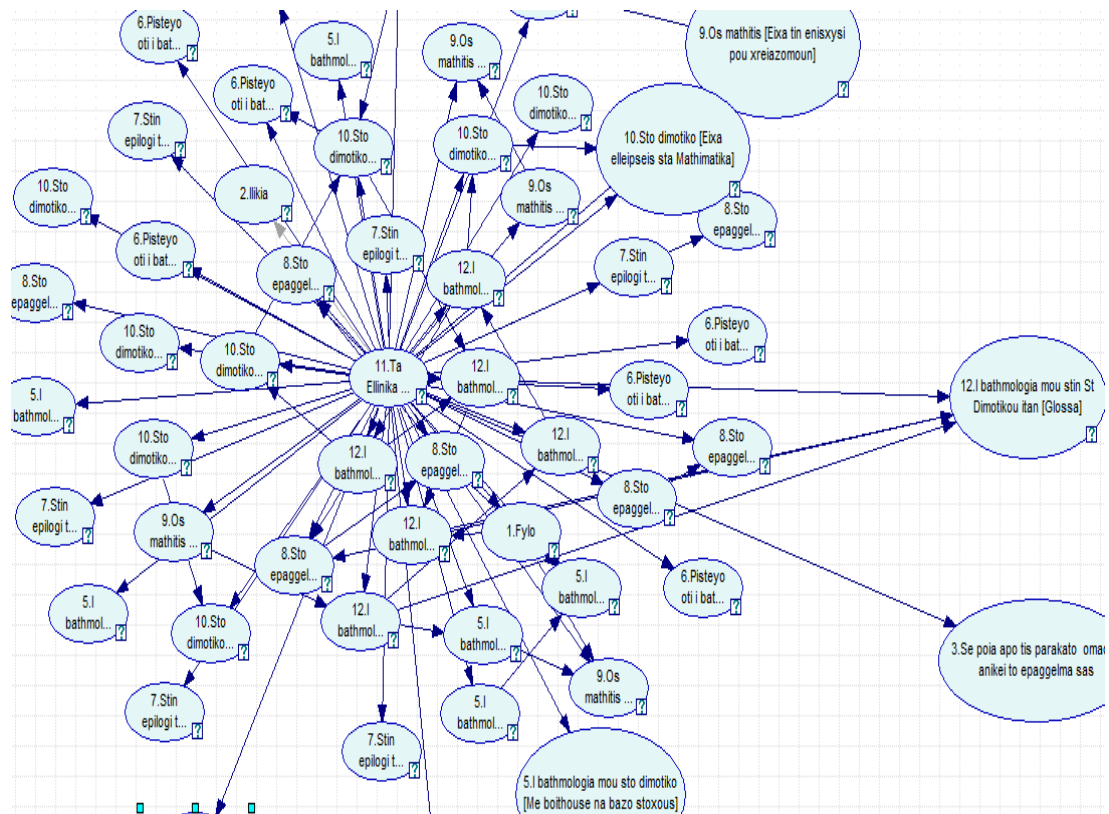
Επιπλέον η γραφική διεπαφή, διευκολύνει τη σφαιρική αντίληψη του δικτύου. Οι απαντήσεις των ερωτηματολογίων περιείχαν πολλά επίπεδα πληροφοριών εκτός από τη σχολική επίδοση καθώς έπρεπε να διερευνηθεί και ο τρόπος που οι απόψεις των ερωτηθέντων σχετικά με την επίδοση- βαθμολογία της ΣΤ δημοτικού σχετίζονται με την τωρινή τους επαγγελματική σταδιοδρομία και γενικά τις επιρροές που αυτή έχει δεχτεί.

Όλες οι απαντήσεις αναλύθηκαν , χρησιμοποιώντας GeNIe σε ένα πλαίσιο πιθανοτήτων και συσχετίσεων που κατέληξαν σε σημαντικά συμπεράσματα που αναλύονται στη συνέχεια.

Κεφάλαιο 6

ΠΕΙΡΑΜΑΤΑ

6.1 Πειράματα στα Δίκτυα Bayes (Α Μοντέλο)



σχ. 7 Δίκτυα Bayes .Μοντέλο Α

Ο γενικός σκοπός της εφαρμογής των δικτύων Bayes είναι η «έξυπνη» διαχείριση των ερωτηματολογίων. Το πλεονέκτημά τους είναι ότι επιτρέπουν τη γνώση των αιτιωδών σχέσεων μεταξύ των μεταβλητών, γεγονός που είναι πολύ σημαντικό όταν επιχειρείται επεξηγηματική ανάλυση δεδομένων και βαθύτερη κατανόηση του προβλήματος. Επίσης η γνώση των αιτιωδών σχέσεων επιτρέπει τη διατύπωση προβλέψεων. Τα δεδομένα που είναι οι απαντήσεις των ερωτηματολογίων μεταφέρθηκαν στο GeNIE. Μεταβλητές είναι οι ερωτήσεις. Κάθε κόμβος είναι μια ερώτηση. Οι αριθμοί σηματοδοτούν τις ομάδες ερωτήσεων σύμφωνα με τα διερευνητικά ερωτήματα. Με τον τρόπο αυτό πραγματοποιείται **γραφική**

απεικόνιση των ερωτηματολογίων. Η μελέτη των αιτιωδών σχέσεων, των πινάκων και των πιθανολογικών σχέσεων κατέληξε στα παρακάτω συμπεράσματα:

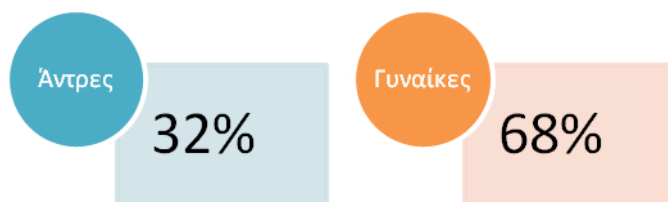
6.1.1 Περιγραφική ανάλυση των δεδομένων

Η ανάλυση των δεδομένων έδειξε για το δείγμα των 60 ατόμων τα παρακάτω στατιστικά στοιχεία :Το 79% του δείγματος ανήκει στην ηλικιακή ομάδα 25-45 σε ποσοστό 89% και 18-25 σε ποσοστό 11%

ηλικιακές ομάδες		
35-45	25-35	18-25
79%	11%	11%

Πιν.1 ηλικιακές ομάδες

Το φύλο του δείγματος είναι κατά 32% άντρες και 68% γυναίκες

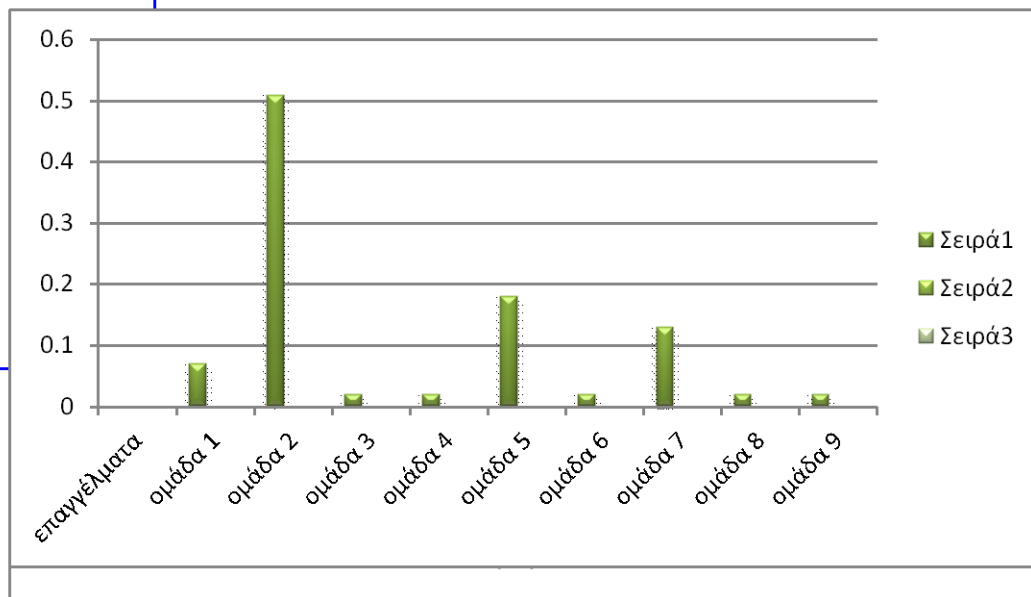


σχ.8 φύλο

Το 51% του δείγματος ανήκε σε πρόσωπα που ασκούν επιστημονικά , καλλιτεχνικά και συναφή επαγγέλματα (ομάδα 2) και στις άλλες ομάδες ως εξής:

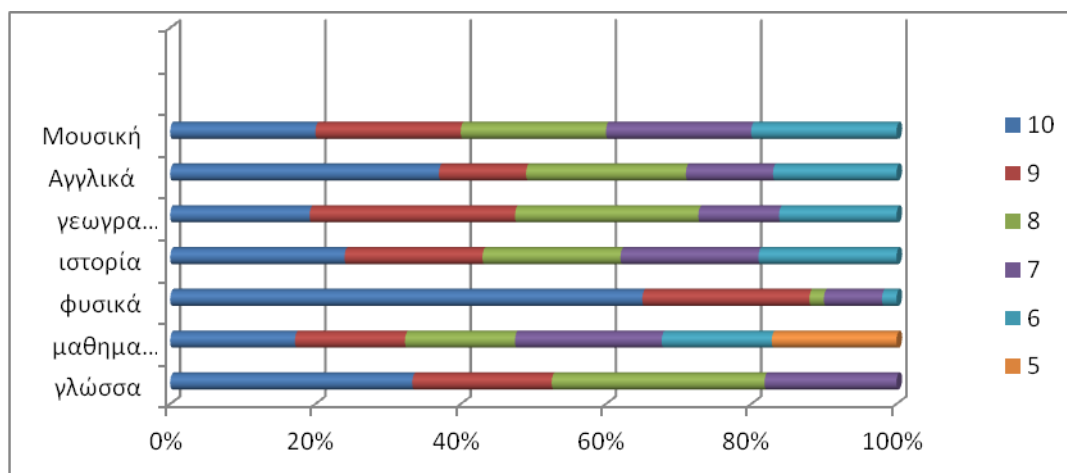
επαγγέλμ ποσοστό

- ομάδα 1
- ομάδα 2
- ομάδα 3
- ομάδα 4
- ομάδα 5
- ομάδα 6
- ομάδα 7
- ομάδα 8
- ομάδα 9



σχ.9 ταξινόμηση επαγγελμάτων δείγματος

Φαίνεται ότι από το δείγμα την καλύτερη επίδοση στην Στ δημοτικού είχαν στο μάθημα της Φυσικής , της Γλώσσας και των Αγγλικών ενώ στα μαθηματικά , τη Γεωγραφία και τη Μουσική φαίνεται ότι υστερούν . Επίσης στα Μαθηματικά εμφανίζεται και η χαμηλότερη βαθμολογία.



σχ.10 αναλυτική βαθμολογία

Γενικά:

Το 54% (36% πολύ, 18% πάρα πολύ) πιστεύει ότι η βαθμολογία στο δημοτικό έπαιξε κάποιο ρόλο στην επιλογή Επαγγέλματος.

Κατά 43% θεωρούν ότι η βαθμολογία δεν υπερτιμούσε τις ικανότητές τους.

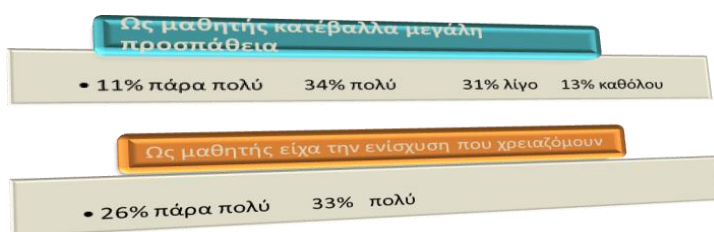
Για το 43% η βαθμολογία ήταν συχνά αφορμή για επαινετικά σχόλια.

Για το 9% (συχνά) και για το 19% (σπάνια) ήταν αφορμή για καβγάδες.

Για το 36% η βαθμολογία βοηθούσε στο να θέτει στόχους ως μαθητής.

Ενώ το 43% πολύ και 16% πάρα πολύ, θεωρούν ότι η βαθμολογία στο δημοτικό βοήθησε να καταλάβουν που ήταν πραγματικά καλοί.

Ειδικότερα : Η ανασφάλεια στο επάγγελμα εμφανίζεται σε ποσοστό λίγο 40% και 29% καθόλου. Ως μαθητής είχα την ενίσχυση που χρειαζόμουν σε ποσοστό πάρα πολύ 26% και πολύ 33%. Ως μαθητής κατέβαλα μεγάλη προσπάθεια σε ποσοστό 34% πολύ , 11% πάρα πολύ ενώ το λίγο 31% και καθόλου 13%



σχ.11 προσπάθεια

ΜΑΘΗΤΙΚΟ ΠΡΟΦΙΛ ⑨	πολύ	λίγο	οχι	ΕΠΑΓΓΕΛΜΑΤΙΚΟ ΠΡΟΦΙΛ ⑧	πολύ	λίγο	όχι
Μαθησιακές δυσκολίες	9%	30%	61%	Εκμεταλλεύομαι τις Ικανότητές μου	83%		
Ενίσχυση που χρειαζόταν	20%	59%	21%	Επίτευξη στόχων	69%	13%	8%
Ενίσχυση που είχε	59%	30%	11%	Άγχος	32%	24%	45%
Προσπάθεια	45%	31%	13%	Ανασφάλεια	31%	40%	29%
Δυσκολίες με δασκάλους	23%	49%	28%	Αίσθηση δημιουργικότητας	80%		
Δυσκολίες με συμμαθητές	33%	38%	29%	Είμαι καλός σε αυτό που κάνω	88%		

Πιν. 2 Μαθητικό- επαγγελματικό προφίλ

Οικονομικοί λόγοι στην επιλογή του επαγγέλματος έπαιξαν πολύ κατά 26% και πάρα πολύ κατά 9%. Στην επιλογή επαγγέλματος έπαιξε ρόλο η επίδοση κατά 26% λίγο και 26% πολύ.

Στην επιλογή του επαγγέλματος φαίνεται ότι καθοριστικό ρόλο έπαιξε η τελική βαθμολογία 36%πολύ , 18% πάρα πολύ . Καθόλου απάντησαν ποσοστό 21%.

Ωστόσο οι επιδόσεις σε ένα μάθημα πιστεύουν ότι επηρέασε στην τελική επιλογή το 44% .

Ο παράγοντας από τους παραπάνω που επηρεάζει λιγότερο είναι η παρότρυνση ενός τρίτου προσώπου. (36% καθόλου) αλλά και οι οικονομικοί λόγοι (34% καθόλου).

ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΟΔΗΓΗΣΑΝ ΣΤΗΝ ΕΠΙΛΟΓΗ ΤΟΥ ΕΠΑΓΓΕΛΜΑΤΟΣ ⑦	Πολύ-πάρα πολύ	Ελάχιστα-λίγο	καθόλου
ΣΧΟΛΙΚΗ ΕΠΙΔΟΣΗ	44%	19%	21%
ΟΙΚΟΝΟΜΙΚΟΙ ΛΟΓΟΙ	35%	30%	34%
ΠΑΡΟΤΡΥΝΣΗ ΠΡΟΣΩΠΟΥ	24%	40%	36%
ΠΑΡΟΤΡΥΝΣΗ ΟΙΚΟΓΕΝΕΙΑΣ	32%	40%	18%
ΣΥΓΚΥΡΙΕΣ	49%	27%	24%
ΟΙ ΕΠΙΔΟΣΕΙΣ ΜΟΥ ΣΕ ΕΝΑ ΜΑΘΗΜΑ	44%	37%	19%

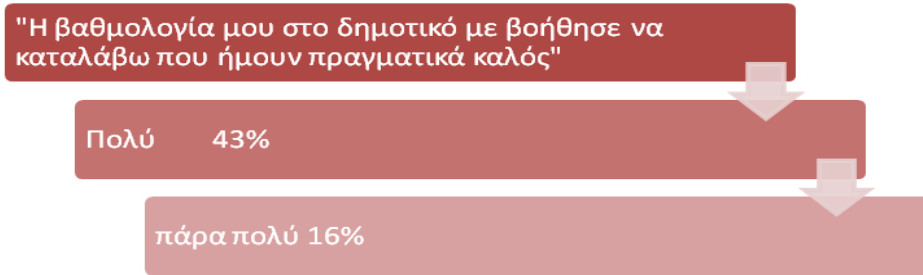
Πιν. 3 Παράγοντες επιλογής

Η βαθμολογία σε ποσοστό 61% ποτέ δεν είχε αδιάφορη αντιμετώπιση από το σπίτι.

ΒΑΘΜΟΛΟΓΙΑ ΟΙΚΟΓΕΝΕΙΑΚΟ ΠΕΡΙΒΑΛΛΟΝ ⑤	ΚΑΙ	Συχνά- πολύ συχνά	Κάποιες φορές-σπάνια	ποτέ
Έπαινοι		72%	21%	6%
Καβγάδες		9%	34%	58%
Αδιαφορία		4%	31%	61%

Πιν. 4 βαθμολογία-οικογένεια

Η βαθμολογία δεν ήταν ένα αδιάφορο γεγονός στο οικογενειακό περιβάλλον αλλά προκαλούσε θετικές ή αρνητικές αντιδράσεις . Για το συγκεκριμένο δείγμα οι θετικές ήταν οι περισσότερες.

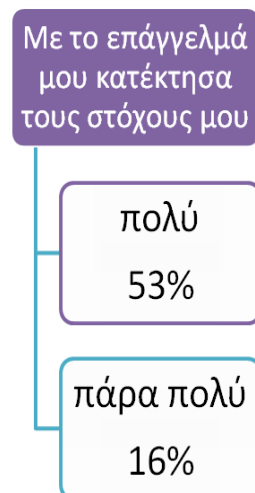


σχ.12 βαθμολογία- αυτοαντίληψη

Η βαθμολογία μου στο δημοτικό με βοήθησε σε ποσοστό 43% πολύ και 16% πάρα πολύ να καταλάβω που ήμουν πραγματικά καλός.

Το δείγμα πιστεύει ότι η βαθμολογία καλλιέργησε λίγο 38% φιλοδοξίες για το επάγγελμα έως ελάχιστα 23%.

Σε ποσοστό 53% **πολύ** και **16%** **πάρα πολύ** ,αισθάνονται ότι με το επάγγελμά τους κατέκτησαν τους στόχους τους



σχ.13 βαθμολογία - στόχοι

Η βαθμολογία ήταν αντιπροσωπευτική των ικανοτήτων ⑥	<i>Πολύ-πάρα πολύ</i>	<i>Λίγο</i>	<i>καθόλου</i>
Η βαθμολογία με αδικούσε	3%	54%	43%
Η βαθμολογία με υπερτιμούσε		56%	43%
Στο μάθημα με το μεγαλύτερο βαθμό είχα ικανότητα	89%	8%	3%
Με βοήθησε να καταλάβω που ήμουν καλός	59%	32%	9%
Η βαθμολογία με αποθάρρυνε από το να θέτω στόχους	23%	51%	26%
Η βαθμολογία μου καλλιέργησε φιλοδοξίες για ένα επάγγελμα	27%	61%	13%

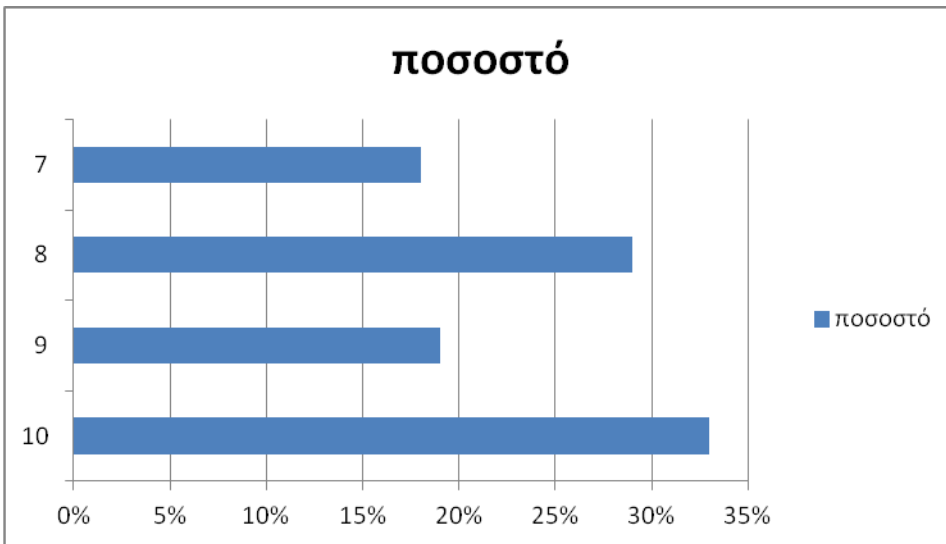
Πιν.5 βαθμολογία-ικανότητα

Το δείγμα σε ποσοστό 34% **καθόλου** και 28% **σπάνια** είχε ελλείψεις στη Γλώσσα. Στα μαθηματικά 34% **Συχνά** και 20% **κάποιες φορές**.

ΠΡΟΤΙΜΗΣΕΙΣ ΜΑΘΗΜΑΤΩΝ ⑩	<i>Συχνά-πολύ συχνά</i>	<i>Κάποιες φορές -Σπάνια</i>	<i>Καθόλου</i>
Μου άρεσαν τα Μαθηματικά	35%	24%	41%
Μου άρεσε η Γλώσσα	73%	25%	2%
Μου άρεσαν οι Φυσικές Επιστήμες	51%	42%	4%
Μου άρεσαν τα καλλιτεχνικά	57%	39%	4%

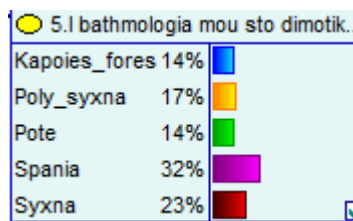
Πιν.6 Προτιμήσεις

Το δείγμα παρουσίασε βαθμολογία στη Γλώσσα άριστα 10 ,σε ποσοστό 33% και πολύ καλά 8 σε ποσοστό 29%



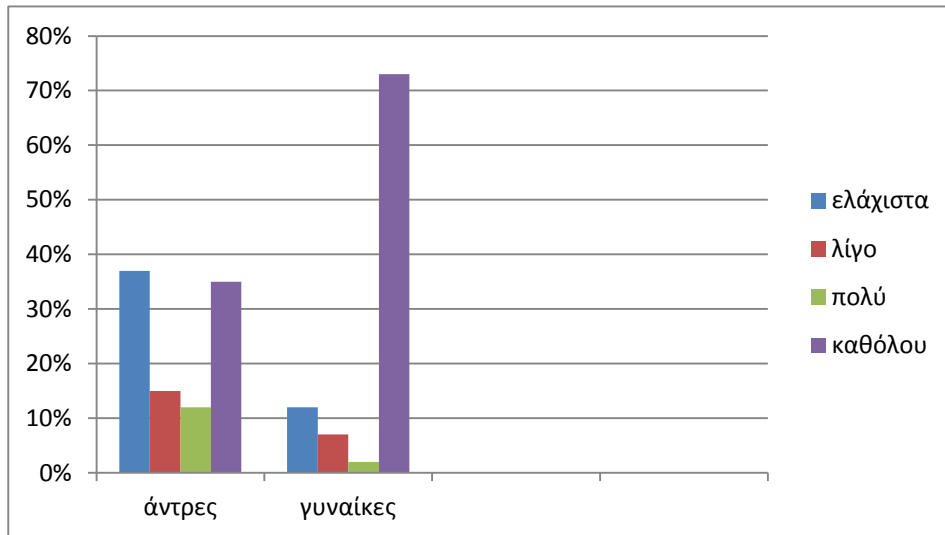
σχ.13 βαθμολογία στο μάθημα της Γλώσσας

Οι γυναίκες εμφανίζονται περισσότερο χαρούμενες με τη βαθμολογία τους σε ποσοστό 34% Συχνά και 28% κάποιες φορές. Οι άντρες εμφανίζονται σπάνια χαρούμενοι με τη βαθμολογία τους σε ποσοστό 32%



σχ.18 Συναισθήματα άντρες γυναίκες

Οι γυναίκες του δείγματος φαίνεται ότι σε ποσοστό 73% δεν εμφάνιζαν μαθησιακές δυσκολίες.



σχ.19 Φύλο, μαθησιακές δυσκολίες

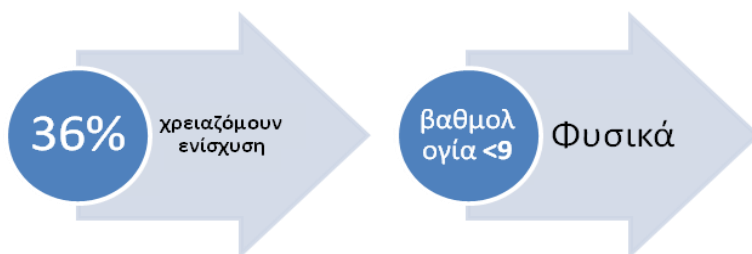
Οι άντρες δεν έχουν μαθησιακές δυσκολίες σε ποσοστό 35%..

6.1.2 Στατιστικές εξαρτήσεις (πρόβλεψη)

Οι ερωτήσεις του ερωτηματολογίου αποτελούν τις **μεταβλητές** στο γράφο των **Δικτύων Bayes**(παράρτημα). Οι ομάδες των ερωτήσεων που αφορούν το ίδιο διερευνητικό ερώτημα, διακρίνονται από τον αριθμό της ερώτησης. Μετά από παρατήρηση του γράφου, εντοπίστηκαν οι κόμβοι που ως **πατέρες** ή **απόγονοι** συνδέονταν με προσανατολισμένα βέλη. Οι απαντήσεις των ερωτήσεων αποτελούν τις **καταστάσεις** των μεταβλητών. Μετά τον εντοπισμό των κόμβων που αλληλεπιδρούν, απομονώθηκε από τον συνοδευτικό πίνακα πιθανοτήτων κάθε κόμβου, μία κατάσταση και παρατηρήθηκαν οι αλλαγές που αυτή προκαλεί στις συσχετιζόμενες μεταβλητές. Προέκυψαν σενάρια που επιβεβαιώνονται ή απορρίπτονται, δίνοντας έτσι απαντήσεις αλλά και διευρύνοντας τα διερευνητικά ερωτήματα.

Σενάριο : Να διερευνηθεί αν μαθητές που χρειάζονταν ενίσχυση (εξατομικευμένη βοήθεια) είχαν σε κάποιο μάθημα τη χαμηλότερη βαθμολογία και αυτό τους αποθάρρυνε από το να θέτουν στόχους. Ποιο είναι το επάγγελμα που ακολούθησαν.

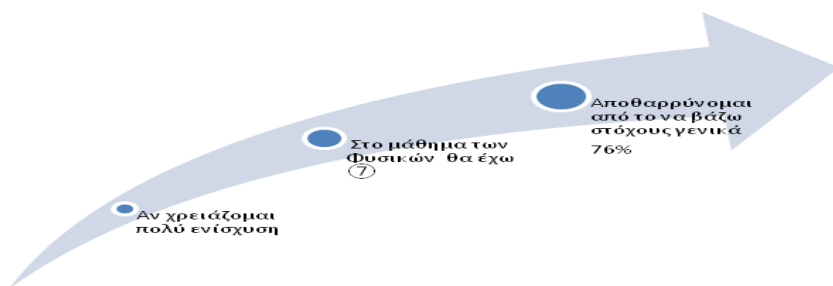
Δεδομένου ότι κάποιος, ως μαθητής, χρειάζεται ενίσχυση προβλέπεται ότι θα έχει κάτω από 9 στο μάθημα της Φυσικής ενώ όσοι δε χρειάζονται καθόλου ενίσχυση θα έχουν 10 στο μάθημα αυτό στην ΣΤ Δημοτικού



σχ. 14 ενίσχυση- μάθημα φυσικής

Υπάρχει προσανατολισμένο μονοπάτι από το μάθημα των «Φυσικών» στον κόμβο «χρειάζομαι ενίσχυση» και «η βαθμολογία μου με αποθάρρυνε από το να θέτω στόχους» :Δεδομένου ότι ως μαθητής δε χρειαζόταν καθόλου ενίσχυση , τότε θα έχει 10 στα Φυσικά σε ποσοστό 81%. Ενώ δεδομένου του 7 ως βαθμολογία στα Φυσικά αυτό το γεγονός θα τον αποθαρρύνει από το να θέτει στόχους σε ποσοστό 76%.

Αν το επάγγελμα που ακολουθήθηκε ανήκει στην ομάδα 3 υπάρχουν πιθανότητες 36% η βαθμολογία ως μαθητής να αποθάρρυνε από το να θέτονται στόχοι.



σχ.17 ενίσχυση -βαθμολογία

Σενάριο : Οι προτιμήσεις σε ένα μάθημα σχετίζονται με τη βαθμολογία σε αυτό; Αν κάποιος εκφράζει ιδιαίτερη προτίμηση στο μάθημα της Γλώσσας, αυτό φαίνεται ότι δεν επηρεάζει τη βαθμολογία του στο μάθημα της Γλώσσας. Το ίδιο δεν ισχύει με τα Μαθηματικά που φαίνεται ότι ο παράγοντας του αν μου αρέσουν ή όχι δημιουργεί ελλείψεις. Δεδομένου ότι σε κάποιον δεν αρέσει το μάθημα των μαθηματικών τότε σε ποσοστό 69% θα έχει ελλείψεις.



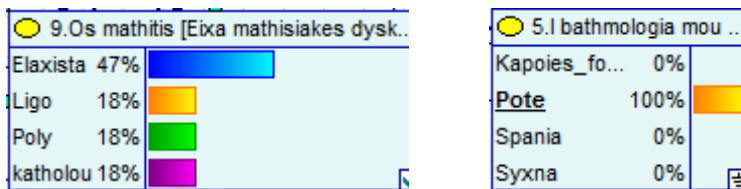
σχ. 16 Μαθηματικά-ελλείψεις

Σενάριο Να διερευνηθεί ο συσχετισμός στη βαθμολογία των μαθημάτων σε συνδυασμό με το επάγγελμα που επιλέχτηκε.

Αν κάποιος έχει 5 ή 6 στο μάθημα των Φυσικών τότε έχει πιθανότητες 74% να έχει και 5 στα Μαθηματικά. Αν έχει 8 στα Φυσικά τότε σε ποσοστό 91% θα έχει 10 στα μαθηματικά. Αν έχει 10 στο μάθημα της Φυσικής και 10 στο μάθημα της Γλώσσας τότε μπορεί στα μαθηματικά να έχει χαμηλότερη βαθμολογία και το επάγγελμα που τελικά θα ακολουθήσει ανήκει στα επαγγέλματα της ομάδας 2(πρόσωπα που ασκούν επιστημονικά, καλλιτεχνικά και συναφή επαγγέλματα)

Σενάριο 3: Ποιος ο ρόλος της βαθμολογίας ως προς τους στόχους που θέτουν άτομα με μαθησιακές δυσκολίες. Ποιο επάγγελμα ακολούθησαν.

Αν κάποιος έχει μαθησιακές δυσκολίες η βαθμολογία του δρα ενισχυτικά και ενθαρρύνει ως προς το να θέτει στόχους . Σε μαθητές χωρίς μαθησιακές δυσκολίες κάποιες φορές η βαθμολογία δρα αποθαρρυντικά με πιθανή αιτία τις υψηλές προσδοκίες.

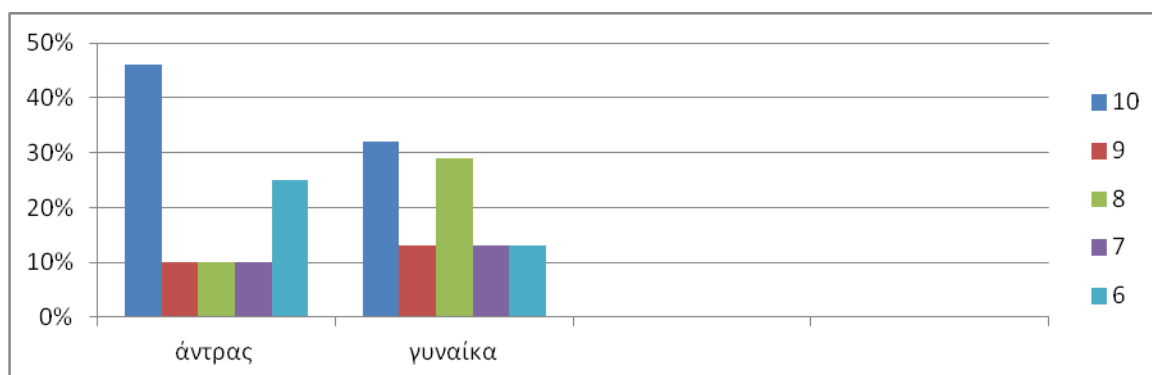


Πιν. 7 Δυσκολίες - στόχοι –πρόβλεψη

Ως προς το επάγγελμα δεδομένου του ότι κάποιος ανήκει στις ομάδες 1, 2, 4 και 5 οι πιθανότητες να είχε μαθησιακές δυσκολίες είναι μόνο 9%.

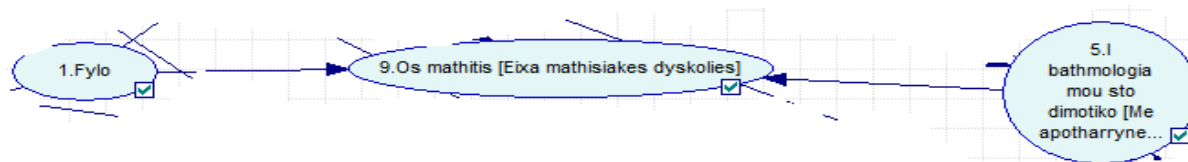
Σενάριο : Υπάρχει επίδοση καλύτερη ή χειρότερη ανάλογα με το φύλο;

Με δεδομένο το ότι είναι άντρας έχει 46% πιθανότητες να έχει 10 στην Αγγλική Γλώσσα, ενώ το να έχει βαθμολογία 6, εμφανίζει πιθανότητες 25%. Αν είναι γυναίκα έχει 32% πιθανότητες να έχει 10 και 29% να έχει 8 στο αντίστοιχο μάθημα.



σχ.20 Φύλο, ξένη γλώσσα

Σενάριο 5 Το φύλο σχετίζεται με τις μαθησιακές δυσκολίες. Με δεδομένο το ότι κάποιος είναι άντρας έχει πιθανότητες 72% να μην έχει μαθησιακές δυσκολίες .



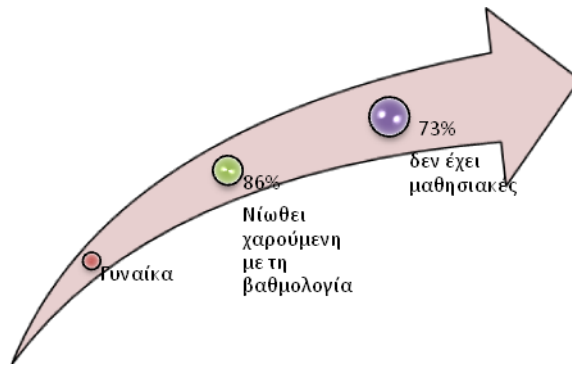
σχ.22 διεπαφές μεταβλητών

Αν είναι γυναίκα έχει περισσότερες πιθανότητες 93% να μην έχει μαθησιακές δυσκολίες.

Σενάριο : Να διερευνηθεί αν το φύλο σχετίζεται με τα συναισθήματα που προκαλεί η βαθμολογία και με τις μαθησιακές δυσκολίες:

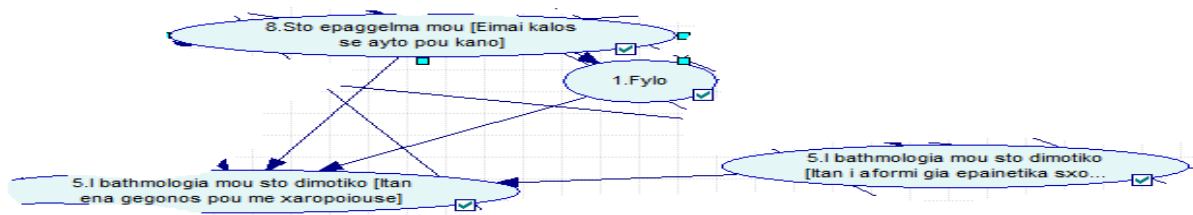
Αν είμαι άντρας Σπάνια θα νιώθω χαρούμενος με τη βαθμολογία μου σε ποσοστό 94%.και θα έχω μαθησιακές δυσκολίες 64%.

Αν είμαι γυναίκα έχω πιθανότητες 86% να νιώθω χαρούμενη με τη βαθμολογία μου και να μην έχω μαθησιακές δυσκολίες κατά 73%



σχ.23 Βαθμολογία συναισθήματα- μαθησιακές δυσκολίες

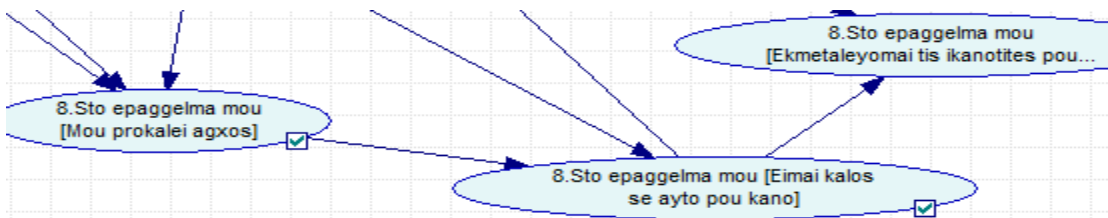
Σενάριο : Να διερευνηθεί αν τα επαινετικά σχόλια των γονιών επηρεάζουν τα συναισθήματα για τη βαθμολογία. Αν κάποιος εισέπραττε θετικά σχόλια για τη βαθμολογία του και ένιωθε χαρούμενος ποιο επάγγελμα ακολούθησε;



σχ.23 διεπαφές μεταβλητών

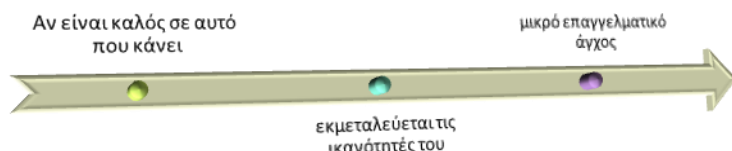
Δεδομένου του ότι κάποιος έχει επαινετικά σχόλια από τους γονείς του, έχει πιθανότητες 73% να είναι η βαθμολογία του ένα ευχάριστο γεγονός. Το επάγγελμα που θα ακολουθήσει είναι κατά 51% κάποιο επάγγελμα της ομάδας 2 (πρόσωπα που ασκούν επιστημονικά, καλλιτεχνικά επαγγέλματα) ή κατά 18% κάποιο επάγγελμα της ομάδας 4 (υπάλληλοι γραφείου) ή κατά 18% στην ομάδα 5(επαγγέλματα παροχής υπηρεσιών)

Σενάριο: Να διερευνηθεί αν κάποιος που πιστεύει ότι είναι καλός σε αυτό που κάνει, νιώθει ότι εκμεταλλεύεται τις ικανότητές του έχει άγχος στο επάγγελμά του. Ποιο είναι το επάγγελμά του και ποια η βαθμολογία του στη Γλώσσα και τα μαθηματικά.



σχ.24 διεπαφές μεταβλητών

Με δεδομένο το ότι κάποιος πιστεύει ότι είναι καλός σε αυτό που κάνει τότε υπάρχουν 90% πιθανότητες να αισθάνεται και ότι εκμεταλλεύεται τις ικανότητές του και να έχει σε μικρό ποσοστό επαγγελματικό άγχος. (Το 49% ελάχιστα). Η βαθμολογία του στη Γλώσσα ως μαθητής τότε θα είναι 10 σε ποσοστό 45%. Στα μαθηματικά η βαθμολογία του μπορεί να είναι 5, 6, 7, 8, 9 , ή 10 με ίσο ποσοστό για το καθένα 17%. Το επάγγελμα που θα ακολουθήσει κατά 53% , θα ανήκει στην ομάδα 2.

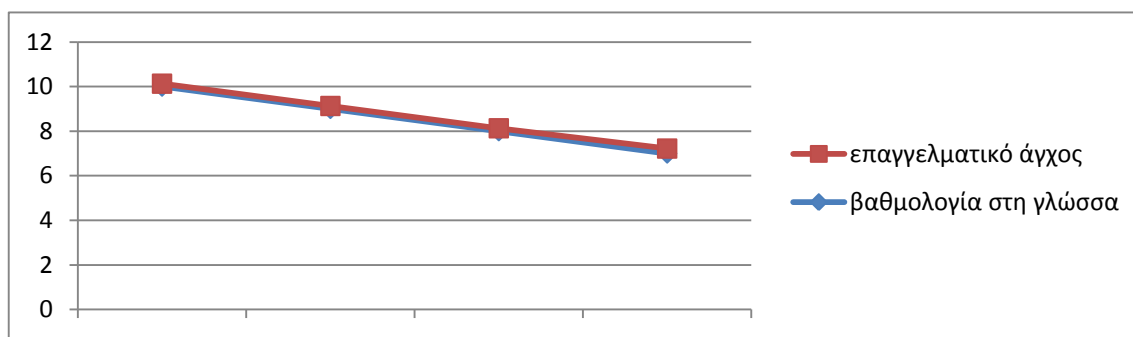


σχ.25 Επαγγελματικό άγχος -ικανότητα

Αν στο επάγγελμα κάποιος νιώθει δημιουργικός , τότε οι πιθανότητες να νιώθει και χαρούμενος είναι **99%**.

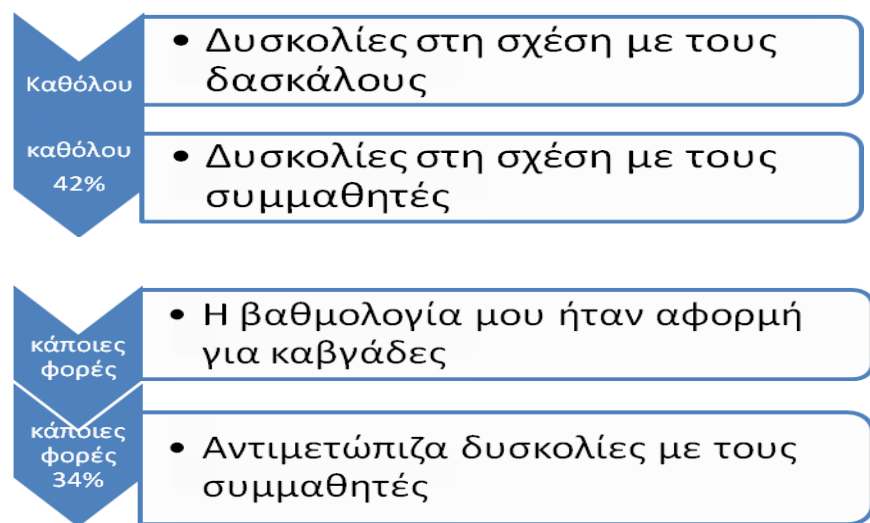
Αν δε νιώθει δημιουργικός τότε οι πιθανότητες να μη νιώθει και χαρούμενος είναι 84%.

Ύσενάριο : Να διερευνηθεί αν η βαθμολογία σε κάποιο μάθημα σχετίζεται με το επαγγελματικό άγχος. Με δεδομένο το άριστα, 10, στη Γλώσσα εμφανίζεται μικρό ποσοστό άγχους στο επάγγελμα. Όσο χαμηλώνει η βαθμολογία αυξάνεται το άγχος στο χώρο του επαγγέλματος. Δηλαδή κάποιος που ως μαθητής είχε 10 στο μάθημα της Γλώσσας, ως ενήλικας, στο επάγγελμά του δεν έχει άγχος.



σχ. 26 Επαγγελματικό άγχος -Γλώσσα

Σενάριο : Να διερευνηθεί αν επηρεάζουν την άποψη για τη βαθμολογία οι σχέσεις του με δασκάλους και συμμαθητές.



Σχ.27 Σχέσεις

Αν ως μαθητής κάποιος αντιμετωπίζει δυσκολίες με τους συμμαθητές και αν η βαθμολογία του ήταν αφορμή για καβγάδες στο οικογενειακό του περιβάλλον τότε είναι πολύ πιθανό να έχει την άποψη ότι η βαθμολογία στο δημοτικό «τον αδικούσε» (74%). Αντίστροφα μαθητής που έχει την αίσθηση της αδικίας στη βαθμολογία είναι πολύ πιθανό να έχει και αρνητική αντιμετώπιση της βαθμολογίας του από το σπίτι και να έχει και χαμηλή βαθμολογία στο μάθημα της Γλώσσας

Οι γενικές δυσκολίες στις διαπροσωπικές σχέσεις με τους συμμαθητές μου, φαίνεται ότι επηρεάζει την άποψη που έχω σχετικά με το αν με αδικούσε η βαθμολογία μου σε ποσοστό 74% .

Αν κάποιος ως μαθητής είχε κάποιες δυσκολίες στις σχέσεις με τους δασκάλους τότε έχει και δυσκολίες στις σχέσεις με τους συμμαθητές (73%). Ωστόσο όταν οι δυσκολίες είναι συχνές αυτό δε φαίνεται να επηρεάζει πολύ στις σχέσεις με τους συμμαθητές καθώς εμφανίζεται ποσοστό 61% να έχει δυσκολίες σπάνια με τους συμμαθητές όταν έχει συχνές δυσκολίες με τους δασκάλους.

Σενάριο : Να διερευνηθεί αν συγκεκριμένη βαθμολογία σε κάποιο μάθημα σημαίνει και συγκεκριμένη επαγγελματική επιλογή και αν το φύλο επιδρά.

Αν έχω 10 στη Γλώσσα το επάγγελμά μου, υπάρχουν 53% πιθανότητες, να ανήκει στην ομάδα 2 «πρόσωπα που ασκούν επιστημονικά καλλιτεχνικά και συναφή επαγγέλματα» . Στα Φυσικά η βαθμολογία μου θα είναι 9 με 69% πιθανότητα. Είμαι γυναίκα και στα μαθηματικά η βαθμολογία μου με πιθανότητες 60% θα είναι κάτω από 9.

Αν έχω 8 στη Γλώσσα (40%) θα ανήκω στην ίδια ομάδα 2 με πιθανότητες 51% . Στα μαθηματικά η βαθμολογία μου και σε αυτή την περίπτωση θα είναι κάτω από 8 και τότε θα είμαι άντρας.

6.2 Β Μοντέλο

Το Β Μοντέλο αφορά τη διερεύνηση συσχετίσεων ανάμεσα στο επάγγελμα των γονιών και τη σχολική επίδοση (βαθμολογία) των μαθητών της Στ Δημοτικού και τη δημιουργία μοντέλου πρόβλεψης με μικρή απόκλιση και κατά συνέπεια μεγάλη αξιοπιστία.

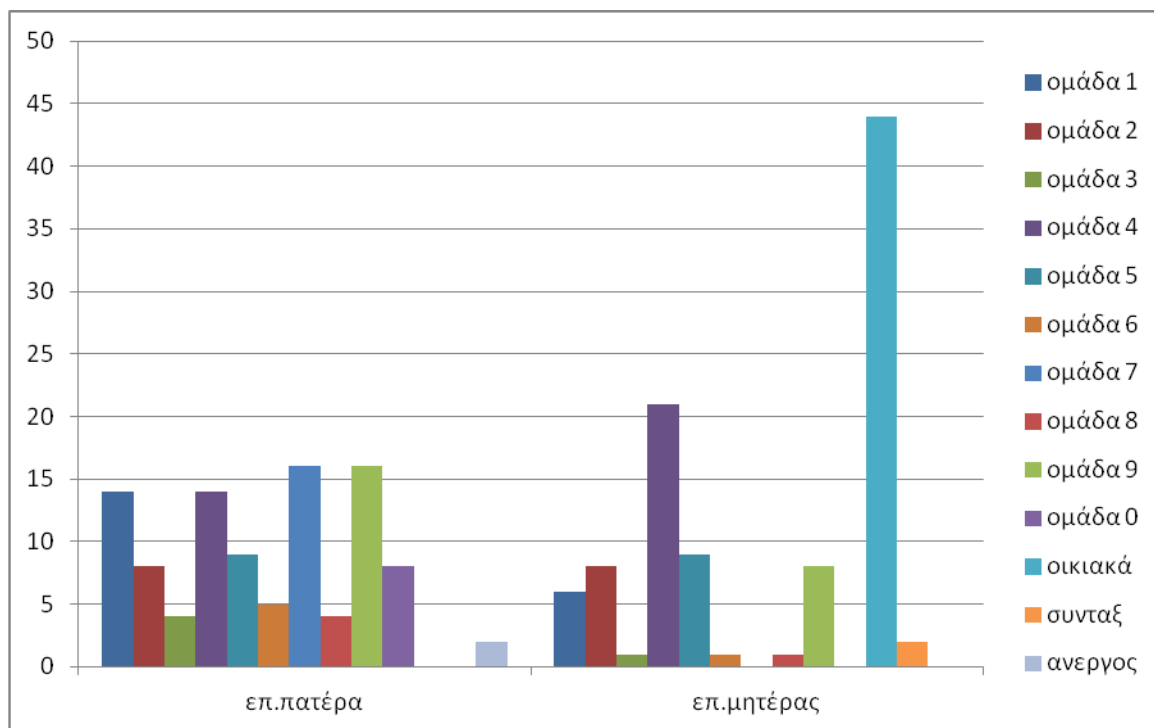
6.2.α Περιγραφή

Τα δεδομένα καλύπτουν τα έτη 2002- 2010 (δεν υπάρχουν δεδομένα για το σχολικό έτος 2007) Αφορά **250** μαθητές συγκεκριμένου σχολείου της περιοχής Καρλοβάσου (1^ο Νέου Καρλοβάσου) και τη βαθμολογία τους στην Στ (έκτη) Δημοτικού κατά τα σχολικά έτη που Προαναφέρθηκαν όπως αυτά καταγράφονται στο βιβλίο Μητρώου των μαθητών. Ως προς το φύλο είναι σε ποσοστό 50% αγόρια , 50% κορίτσια



Σχ. 28 Φύλο Β Μοντέλο

Ως προς το επάγγελμα της μητέρας σε ποσοστό 44% ασχολούνται με οικιακά και κατά 21% ανήκουν σε επαγγέλματα που αφορούν παροχή υπηρεσιών (ομάδα 4)

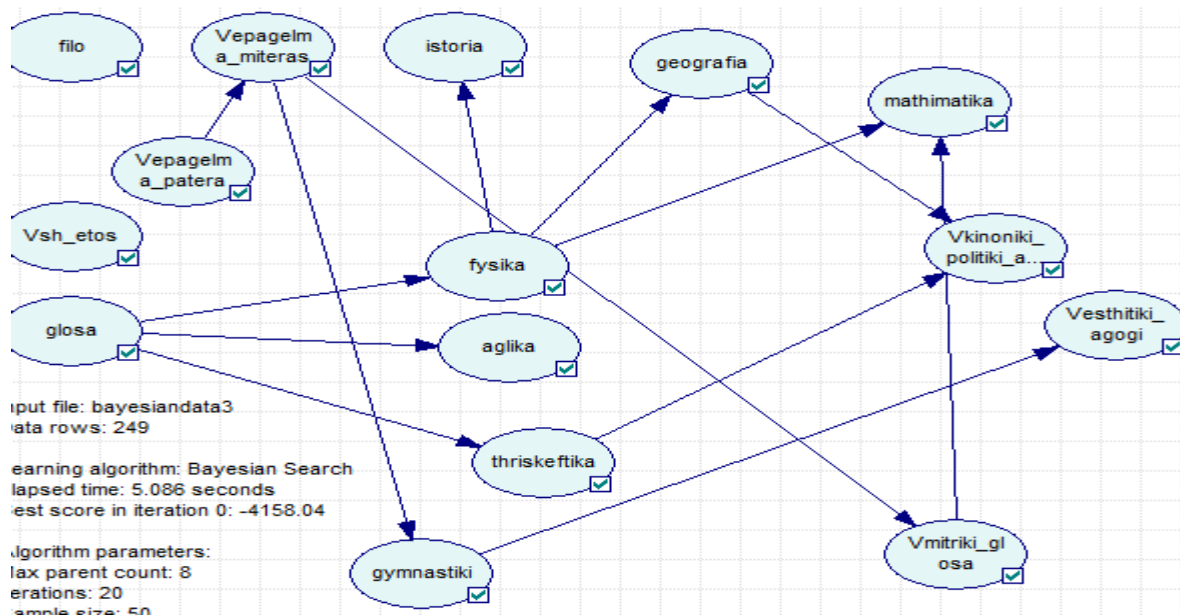


Σχ. 29 Επάγγελμα γονιών

Οι υπόλοιπες ομάδες έχουν χαμηλά ποσοστά . 7% είναι πρόσωπα που ασκούν επιστημονικά ,καλλιτεχνικά και συναφή επαγγέλματα ,8% είναι ανειδίκευτοι εργάτες ή χειρωνάκτες, 6% είναι μέλη βουλευόμενων σωμάτων ή ανώτερα διευθυντικά στελέχη. Ανά 1% ανήκουν στις υπόλοιπες ομάδες . Επιπλέον υπάρχουν σε ποσοστό 2% συνταξιούχοι .

Στους πατέρες παρατηρείται περισσότερο ισότιμη κατανομή ως προς τις ομάδες των επαγγελμάτων :16% είναι ειδικευμένοι τεχνίτες,16% είναι ανειδίκευτοι εργάτες ,χειρωνάκτες,14% είναι ανώτερα διευθυντικά στελέχη,14% είναι υπάλληλοι γραφείου,9% απασχολούνται στην παροχή υπηρεσιών ή είναι πωλητές,8% είναι στις ένοπλες δυνάμεις,5% είναι ειδικευμένοι γεωργοί,4% είναι τεχνολόγοι, τεχνικοί βοηθοί, 4% επίσης είναι χειριστές μηχανημάτων. Το 2% είναι άνεργοι .

Χρησιμοποιήθηκαν δίκτυα Bayes για την ανίχνευση συσχετίσεων και πρόβλεψη.

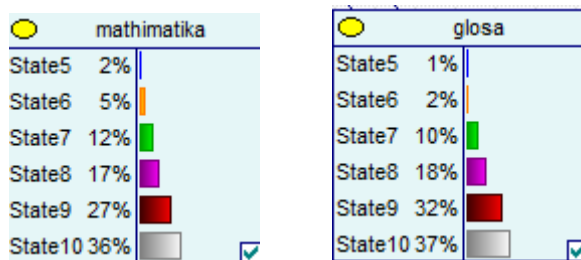


Σχ. 30 Δίκτυα Bayes B Μοντέλο

Οι κόμβοι απεικονίζουν μεταβλητές που για το συγκεκριμένο γράφο είναι το επάγγελμα των γονιών και τα μαθήματα της ΣΤ Δημοτικού. Οι καταστάσεις των μεταβλητών είναι ο μέσος όρος της βαθμολογίας των τριών τριμήνων, για το κάθε μάθημα. Τα προσανατολισμένα βέλη αναπαριστούν συσχετίσεις. Αφού εντοπίστηκαν οι κόμβοι που αλληλεπιδρούν δημιουργήθηκαν σενάρια ως εξής:

Σενάριο: Να διερευνηθεί αν τα επαγγέλματα της μητέρας και του πατέρα σχετίζονται και σε κάθε περίπτωση ποια είναι η βαθμολογία στα βασικά μαθήματα Γλώσσα, Μαθηματικά.

Αν ο πατέρας κάνει επάγγελμα που ανήκει στην ομάδα 9 –ανειδίκευτος εργάτης –χειρωνάκτης- τότε το επάγγελμα της μητέρας είναι κατά 53% οικιακά και κατά 40% ανειδίκευτος εργάτης-χειρωνάκτης. Τότε η βαθμολογία του μαθητή στα μαθήματα της Γλώσσας και των μαθηματικών θα είναι

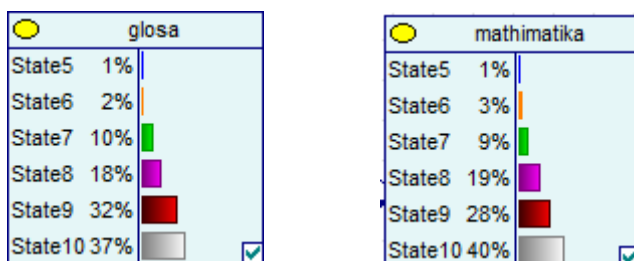


πιν.8 επισυναπτόμενοι πίνακες

Αν το επάγγελμα του πατέρα είναι ειδικευμένος γεωργός –ομάδα 5- τότε η μητέρα ανήκει στην ομάδα 5 κατά 34% ή ασχολείται με οικιακά σε ποσοστό 42%.

Αν ο πατέρας είναι υπάλληλος γραφείου τότε η μητέρα είναι υπάλληλος γραφείου σε ποσοστό 38% ή ασχολείται με οικιακά σε ποσοστό 41%.

Αν ο πατέρας είναι πρόσωπο που ασκεί επιστημονικά ,καλλιτεχνικά και συναφή επαγγέλματα ομάδα2 τότε και η μητέρα ασχολείται με το ίδιο επάγγελμα σε ποσοστό 29% ή είναι υπάλληλος γραφείου σε ποσοστό 29%. Ο επισυναπτόμενος πίνακας των πιθανοτήτων έχει την εξής μορφή για Γλώσσα και Μαθηματικά.



πιν.9 σχετικοί επισυναπτόμενοι πίνακες

Αν ο πατέρας ανήκει στην ομάδα 1 –άνωτερα διευθυντικά στελέχη – τότε και η μητέρα ανήκει στο ίδιο σε ποσοστό 40% ή ασχολείται με οικιακά σε ποσοστό 32%. Τότε η βαθμολογία του μαθητή στα Μαθηματικά θα είναι 10, σε ποσοστό 40%, 9 σε ποσοστό 28%, 8 σε ποσοστό 19%. Η βαθμολογία της Γλώσσας θα είναι 10 με πιθανότητες 37%.

Αν το επάγγελμα του πατέρα είναι στις ένοπλες δυνάμεις η μητέρα ασχολείται με οικιακά σε ποσοστό 64%. Τότε ο μαθητής θα έχει 10 στη Γλώσσα σε ποσοστό 37%, 9 σε ποσοστό 32%, 8 σε ποσοστό 18%. Στα Μαθηματικά οι πιθανότητες είναι 39% να έχει 10, 28% να έχει 9 και 19% να έχει 8.

Αν ο πατέρας είναι άνεργος η μητέρα ασχολείται με οικιακά κατά 82%, τότε ο μαθητής έχει 37% πιθανότητες να έχει 10 στη Γλώσσα και 39% να έχει 10 στα Μαθηματικά. Έχει 32% πιθανότητες να έχει 9 στη Γλώσσα και 28% να έχει 9 στα Μαθηματικά.

Σενάριο: Να διερευνηθεί αν υπάρχει συσχέτιση ανάμεσα στο επάγγελμα της μητέρας και στη μητρική γλώσσα του μαθητή .Αν η μητέρα ανήκει στην ομάδα 9 είναι ανειδίκευτη εργάτρια ή κάνει χειρωνακτική εργασία τότε οι περισσότερες πιθανότητες 64% είναι η μητρική γλώσσα του μαθητή να είναι η Αλβανική.

Σενάριο Να διερευνηθεί αν το φύλο του μαθητή ασκεί κάποια επιρροή. Το φύλο δεν φαίνεται να ασκεί κάποια επιρροή.(παράρτημα)

Η συνέχεια της έρευνας και η παρακολούθηση της επίδοσης των ίδιων μαθητών και στο Γυμνάσιο και πιθανά η επαγγελματική τους επιλογή θα παρείχε ενδιαφέροντα αποτελέσματα για το θέμα αυτό . Ωστόσο υπερβαίνει τα πλαίσια της συγκεκριμένης διπλωματικής

6.3 ΠΕΙΡΑΜΑΤΑ rapidminer (Α Μοντέλο)

Η τεχνική της κατηγοριοποίησης επιλέχθηκε με σκοπό τη δημιουργία μοντέλων πρόβλεψης με τη μεγαλύτερη δυνατή αξιοπιστία. Για το Α Μοντέλο εκπαιδεύεται σύστημα με δεδομένα εισόδου την αναλυτική βαθμολογία που είχαν οι ερωτηθέντες στην ΣΤ Δημοτικού και έξοδο το επάγγελμα που ακολούθησαν, ώστε να μπορεί να προβλέψει το επάγγελμα από τη βαθμολογία σε συγκεκριμένα μαθήματα.

Αρχικά από τα δεδομένα των ερωτηματολογίων, χρησιμοποιήθηκαν $n=60$ παραδείγματα, δημιουργήθηκε αρχείο excel όπου έγινε επεξεργασία των απαντήσεων του ερωτηματολογίου και διατηρήθηκαν τα σημεία ενδιαφέροντος.

6.3.1 Δεδομένα εισόδου

Τα **δεδομένα εισόδου** αφορούν τη βαθμολογία στην Στ δημοτικού των ατόμων που ερωτήθηκαν στα μαθήματα : Γλώσσα , Μαθηματικά , Ιστορία , Φυσικά, Γεωγραφία, Γυμναστική, Εικαστικά , Μουσική, Αγγλικά

Η **έξοδος** ήταν το επάγγελμα με τις εξής ετικέτες : ομάδα 1, ομάδα 2, ομάδα 3, ομάδα 4, ομάδα 5, ομάδα 0. Οι ομάδες 6, 7, 8, 9 δεν είχαν δεδομένα

Πιο συγκεκριμένα: Το αρχείο σε μορφή excel που λαμβάνεται από τη βάση δεδομένων και συγχρόνως αποτελεί το σώμα εκπαίδευσης (**training set**) έχει τα ακόλουθα δεδομένα

Παράμετροι (attributes)

Fylo (φύλο) αναφέρεται στο φύλο των ερωτηθέντων και λαμβάνει δυο τιμές *antras, gynaiika*

Glosa (Γλώσσα) αναφέρεται στη βαθμολογία στο μάθημα της Γλώσσας και λαμβάνει ακέραιες τιμές από 5 έως 10.

Mathimatika (μαθηματικά): αναφέρεται στη βαθμολογία στο μάθημα των μαθηματικών και λαμβάνει τιμές από 5 έως 10.

Istoria (Ιστορία): αναφέρεται στη βαθμολογία στο μάθημα της Ιστορίας και λαμβάνει τιμές από 5 έως 10.

Geografia (Γεωγραφία): αναφέρεται στη βαθμολογία στο μάθημα της γεωγραφίας και λαμβάνει τιμές από 5 έως 10.

Fysika (Φυσικά): αναφέρεται στη βαθμολογία στο μάθημα των φυσικών επιστημών και λαμβάνει τιμές από 5 έως 10.

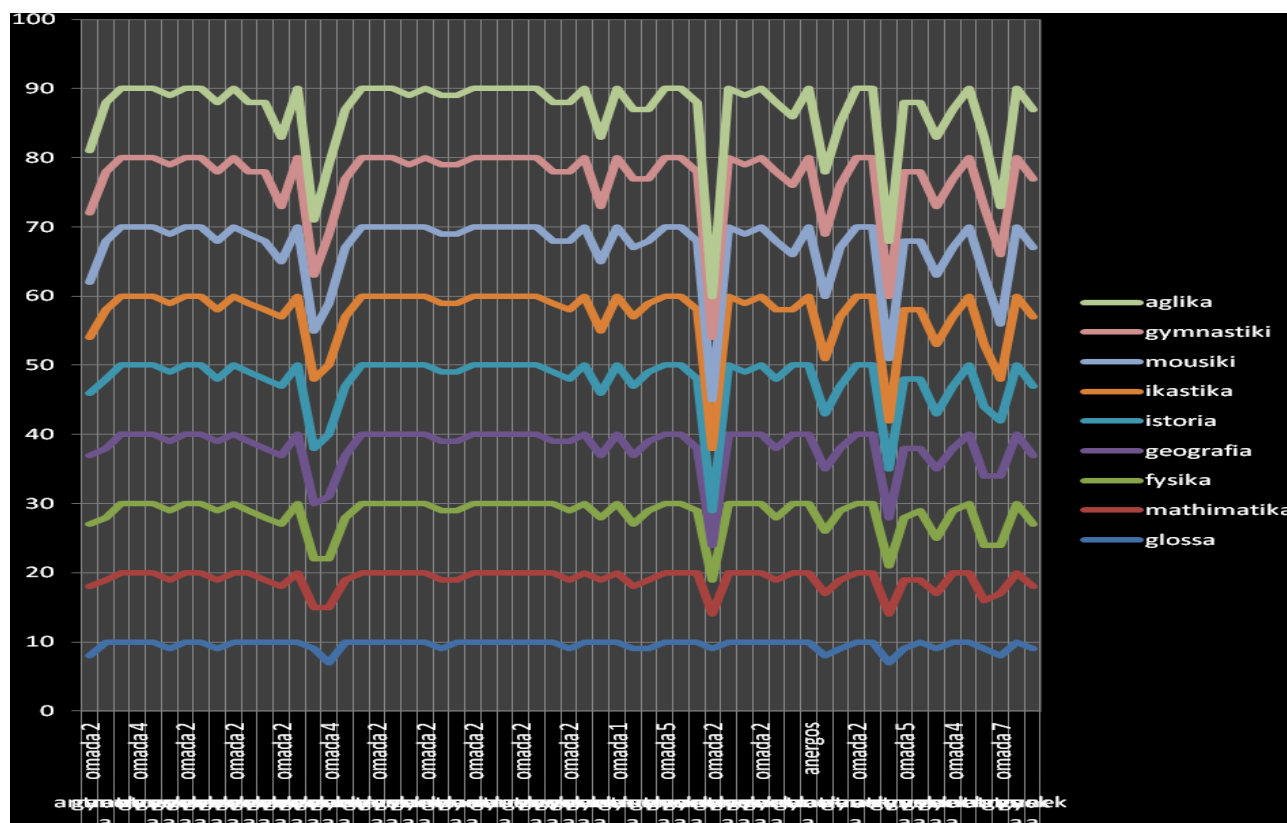
Ikastika (Εικαστικά): αναφέρεται στη βαθμολογία του μαθήματος των εικαστικών και λαμβάνει τιμές από 5 έως 10

Gymnastiki (Γυμναστική): αναφέρεται στη βαθμολογία στο μάθημα των εικαστικών και λαμβάνει τιμές από 5 έως 10.

Mousiki (Μουσική): αναφέρεται στη βαθμολογία στο μάθημα της Μουσικής και λαμβάνει τιμές από 5 έως 10.

Aglika (Αγγλικά): αναφέρεται στη βαθμολογία στο μάθημα των Αγγλικών και λαμβάνει τιμές από 5 έως 10.

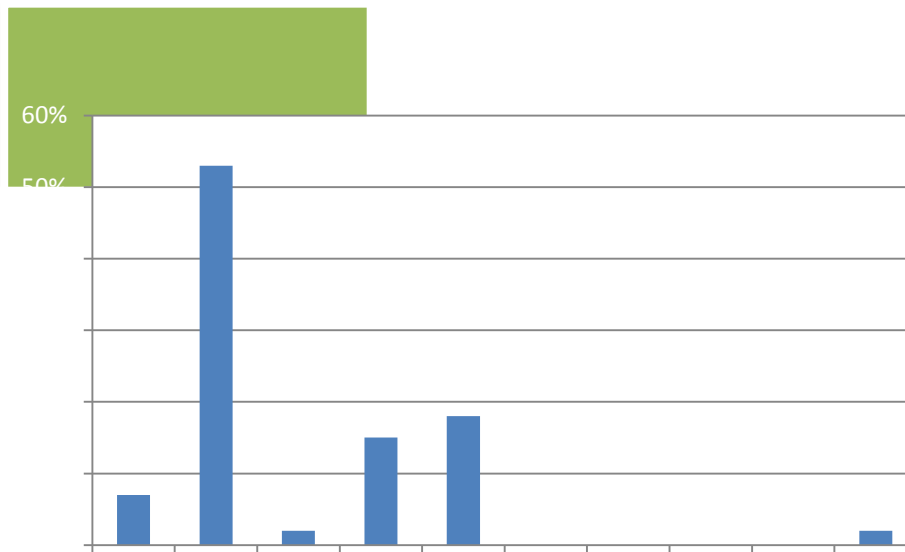
Συνοπτικά οι παράμετροι και οι τιμές τους απεικονίζονται στην παρακάτω εικόνα.



Εικ. 3 Επάγγελμα - επίδοση

6.3.2 Πρόβλεψη κλάσης

Η κλάση είναι ονομαστική αναφέρεται στο επάγγελμα των ατόμων που ερωτήθηκαν και έχει τις εξής τιμές: omada 1, omada 2, omada 3, omada 4, omada 5, omada 7, omada 0, anergos σε ποσοστό:



σχ. ποσοστό επαγγελματιών

Το **training set** σε αυτή την περίπτωση έχει τη μορφή:

ExampleSet (60 examples, 1 special attribute, 10 regular attributes)										Filter (60 / 60 examples):		
Row No.	epagelma	fylo	glossa	mathimatika	fysika	geografia	istoria	ikastika	mousiki	gymnastiki		
1	omada 2	antras	8	10	9	10	9	8	8	10	9	
2	omada 2	gyneka	10	9	9	10	10	10	10	10	10	
3	omada 5	antras	10	10	10	10	10	10	10	10	10	
4	omada 4	antras	10	10	10	10	10	10	10	10	10	
5	omada 2	gyneka	10	10	10	10	10	10	10	10	10	
6	omada 2	gyneka	9	10	10	10	10	10	10	10	10	
7	omada 2	gyneka	10	10	10	10	10	10	10	10	10	
8	omada 3	gyneka	10	10	10	10	10	10	10	10	10	
9	omada 2	gyneka	9	10	10	10	9	10	10	10	10	
10	omada 2	gyneka	10	10	10	10	10	10	10	10	10	
11	omada 2	gyneka	10	10	9	10	10	10	10	9	10	
12	omada 2	gyneka	10	9	9	10	10	10	10	10	10	
13	omada 2	gyneka	10	8	9	10	10	10	8	8	10	
14	omada 5	gyneka	10	10	10	10	10	10	10	10	10	
15	omada 4	gyneka	9	6	7	8	8	10	7	8	8	
16	omada 4	gyneka	7	8	7	9	9	10	9	10	10	
17	omada 2	gyneka	10	9	9	9	10	10	10	10	10	
18	omada 2	antras	10	10	10	10	10	10	10	10	10	

Πιν.10 training set

Με **decision tree** εμφανίζεται αξιοπιστία 50% . Ποσοστό που θεωρείται μικρό.

accuracy: 50.00% +/- 7.45% (mikro: 50.00%)									
	true omada 2	true omada 5	true omada 4	true omada 3	true omada 1	true omada 0	true anergos	true omada 7	class p
pred. omada 2	30	7	9	1	4	1	1	1	55.56%
pred. omada 5	1	0	1	0	0	0	0	0	0.00%
pred. omada 4	1	1	0	0	0	0	0	0	0.00%
pred. omada 3	0	0	0	0	0	0	0	0	0.00%
pred. omada 1	1	1	0	0	0	0	0	0	0.00%
pred. omada 0	0	0	0	0	0	0	0	0	0.00%
pred. anergos	0	0	0	0	0	0	0	0	0.00%
pred. omada 7	0	0	0	0	0	0	0	0	0.00%
class recall	90.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Πιν.11 Ακρίβεια δέντρου απόφασης

Και με **K-NN** εμφανίζεται μικρό ποσοστό αξιοπιστίας επίσης 50%

accuracy: 50.00% +/- 7.45% (mikro: 50.00%)									
	true omada 2	true omada 5	true omada 4	true omada 3	true omada 1	true omada 0	true anergos	true omada 7	class p
pred. omada 2	28	6	6	1	3	1	1	1	59.57%
pred. omada 5	2	1	3	0	1	0	0	0	14.29%
pred. omada 4	2	1	1	0	0	0	0	0	25.00%
pred. omada 3	0	0	0	0	0	0	0	0	0.00%
pred. omada 1	1	0	0	0	0	0	0	0	0.00%
pred. omada 0	0	0	0	0	0	0	0	0	0.00%
pred. anergos	0	0	0	0	0	0	0	0	0.00%
pred. omada 7	0	1	0	0	0	0	0	0	0.00%
class recall	84.85%	11.11%	10.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Πιν.12 Αξιοπιστία KNN

Ωστόσο με Neural Net το ποσοστό αξιοπιστίας φτάνει ως το 55% που θεωρείται επίσης μικρό για να τρέξει μια διαδικασία πρόβλεψης.

accuracy: 55.00% +/- 7.64% (mikro: 55.00%)									
	true omada 2	true omada 5	true omada 4	true omada 3	true omada 1	true omada 0	true anergos	true omada 7	class p
pred. omada	31	8	10	1	2	1	1	0	57.41%
pred. omada	0	0	0	0	0	0	0	0	0.00%
pred. omada	2	0	0	0	0	0	0	0	0.00%
pred. omada	0	0	0	0	0	0	0	0	0.00%
pred. omada	0	1	0	0	2	0	0	0	66.67%
pred. omada	0	0	0	0	0	0	0	1	0.00%
pred. anergos	0	0	0	0	0	0	0	0	0.00%
pred. omada	0	0	0	0	0	0	0	0	0.00%
class recall	93.94%	0.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	

Πιν.13 Αξιοπιστία Neural Net

6.3.3 Διακριτοποίηση δεδομένων

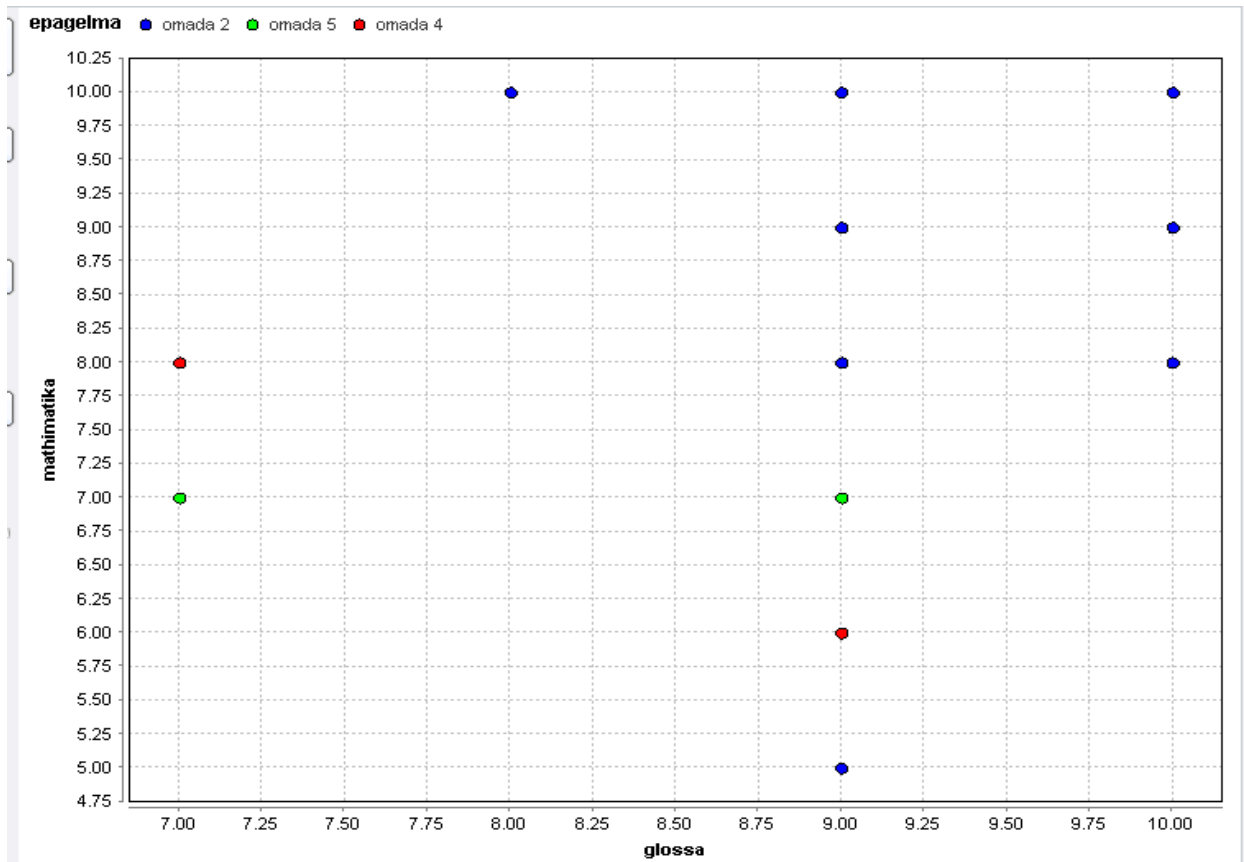
Επειδή το κατηγορικό χαρακτηριστικό Επάγγελμα , περιείχε πολλές τιμές , θεωρήθηκε επικερδές για τη συγκεκριμένη εργασία εξόρυξης , ο συνδυασμός ορισμένων από αυτές τις τιμές ώστε να μειωθεί το πλήθος των κατηγοριών. Για τη διακριτοποίηση του κατηγορικού χαρακτηριστικού επάγγελμα επιλέχθηκαν τα χαρακτηριστικά με τα μεγαλύτερα ποσοστά : omada 4, omada 2, omada 5 .

Στη συνέχεια έγιναν πειράματα με δυαδικοποίηση των δεδομένων: omada 4- oxi omada 4, omada 2- oxi omada 2, omada 5- oxi omada 5.

6.3.3.a ομάδα 2- ομάδα 4- ομάδα 5

Στη συνέχεια έγινε αφαίρεση των πολύ σπάνιων τιμών της κλάσης πρόβλεψης. Αφαιρέθηκε η ομάδα 7, άνεργος, ομάδα 1, ομάδα 3, ομάδα 7, ομάδα 0. Παρέμειναν οι τιμές : ομάδα 2, ομάδα 4, ομάδα 5

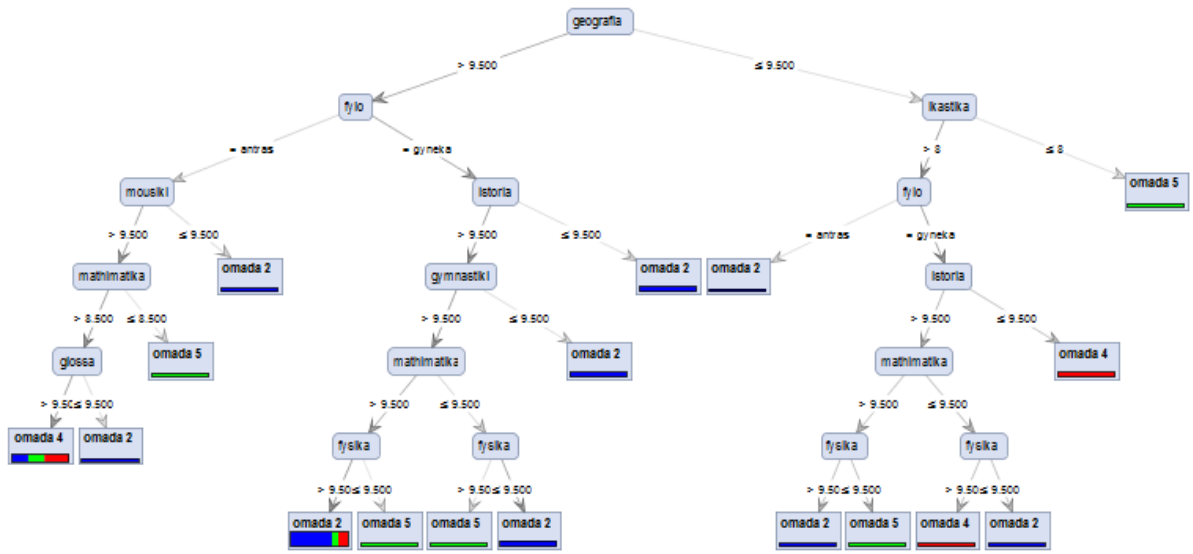
Στο training set οι υψηλές βαθμολογίες στη Γλώσσα και τα Μαθηματικά αφορούν την omada 2,



Εικ.4 Μαθηματικά -Γλώσσα- Επάγγελμα

Decision tree

Το καλύτερο χαρακτηριστικό διάσπασης επιλέγεται πρώτο ώστε να γίνει η ρίζα του δέντρου. Το μάθημα της Γεωγραφίας επιλέγεται στη συγκεκριμένη περίπτωση.



Εικ.5 Δέντρο απόφασης ομάδα2,4,5

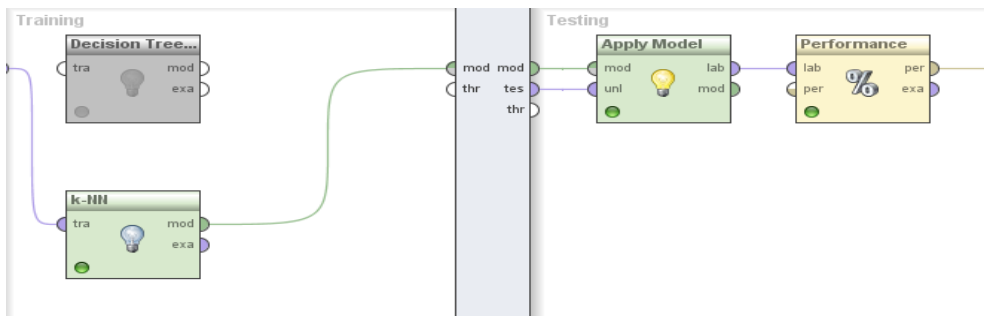
Η αξιοπιστία σε αυτή την περίπτωση εμφανίζει υψηλότερα ποσοστά 57,67 % αλλά όχι κανονποιητικά

accuracy: 57.67% +/- 11.36% (mikro: 57.69%)

	true omada 2	true omada 5	true omada 4	class precision
pred. omada 2	30	8	10	62.50%
pred. omada 5	1	0	0	0.00%
pred. omada 4	2	1	0	0.00%
class recall	90.91%	0.00%	0.00%	

Πιν. 14 Ακρίβεια Δέντρου απόφασης ομάδες 2,4,5

Με K-NN η αξιοπιστία ανεβαίνει λίγο στο 59.67%



accuracy: 59.67% +/- 18.28% (mikro: 59.62%)				
	true omada 2	true omada 5	true omada 4	class precision
pred. omada 2	29	6	6	70.73%
pred. omada 5	2	1	3	16.67%
pred. omada 4	2	2	1	20.00%
class recall	87.88%	11.11%	10.00%	

Πιν. 15 Αξιοπιστία K-NN

Με **Neural Net** παραμένει 59.67%

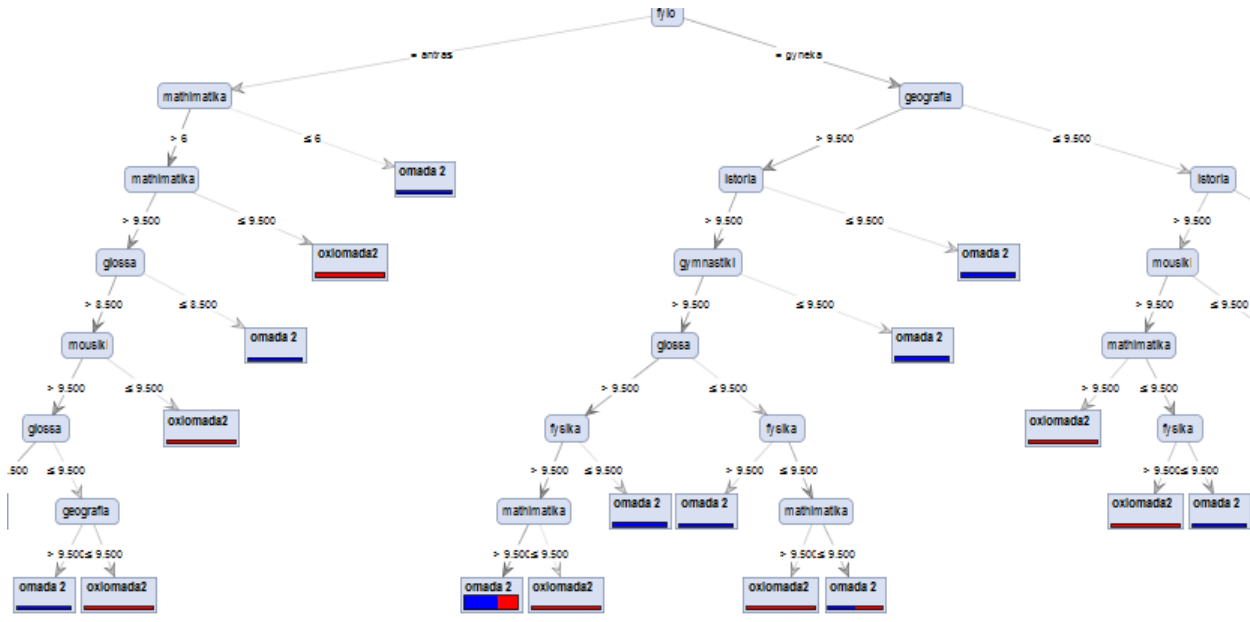
accuracy: 59.67% +/- 3.79% (mikro: 59.62%)				
	true omada 2	true omada 5	true omada 4	class precision
pred. omada 2	31	8	10	63.27%
pred. omada 5	1	0	0	0.00%
pred. omada 4	1	1	0	0.00%
class recall	93.94%	0.00%	0.00%	

Πιν. 16 Αξιοπιστία Neural Net

6.3.3.β ομάδα 2- όχι ομάδα 2

Με στόχο την αύξηση της ακρίβειας στην πρόβλεψη της κλάσης , πειραματιστήκαμε δίνοντας δυωνυμικές τιμές στην κλάση πρόβλεψης . Διατηρήσαμε αρχικά την ομάδα 2 και δώσαμε σε όλες τις άλλες την τιμή όχι ομάδα 2

Το decision tree εμφανίζει χαρακτηριστικό διάσπασης το fylo



Εικ.6 Δέντρο απόφασης ομάδα 2-όχι ομάδα

Ωστόσο η αξιοπιστία σε αυτή την περίπτωση είναι μικρή 53.33%

accuracy: 53.33% +/- 19.44% (mikro: 53.33%)

	true omada 2	true oxiomada2	class precision
pred. omada 2	23	20	53.49%
pred. oxiomada2	8	9	52.94%
class recall	74.19%	31.03%	

Πιν. 17 Αξιοπιστία Δέντρου απόφασης ομάδα 2 –όχι ομάδα 2

Με **weight** by SVM

attribute	weight
ikastika	-0.418
gymnastiki	-0.293
glossa	-0.265
geografia	-0.184
mathimatika	-0.057
fysika	-0.013
mousiki	0.007
aglika	0.114
istoria	0.223

Πίν. 18 Βαρύτητα μεταβλητών

Εμφανίζεται το μάθημα της Ιστορίας και των Αγγλικών στοιχείο μεγαλύτερης βαρύτητας

Όμως και εδώ η αξιοπιστία είναι μικρή με **K-NN** 46.67%

accuracy: 46.67% +/- 16.33% (mikro: 46.67%)			
	true omada 2	true oxiomada2	class precision
pred. omada 2	9	10	47.37%
pred. oxiomada2	22	19	46.34%
class recall	29.03%	65.52%	

Πιν. 19 Αξιοπιστία

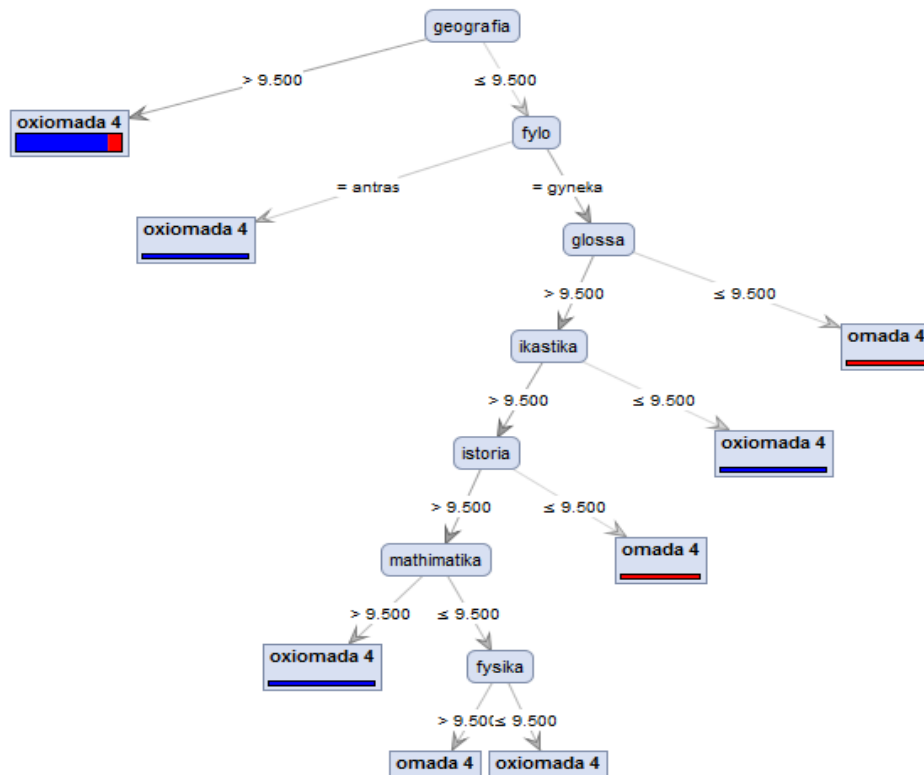
Και με **Neural Net**

accuracy: 46.67% +/- 12.47% (mikro: 46.67%)			
	true omada 2	true oxiomada2	class precision
pred. omada 2	18	19	48.65%
pred. oxiomada2	13	10	43.48%
class recall	58.06%	34.48%	

Πιν. 20 Αξιοπιστία NN

6.3.3.γ ομάδα 4 – όχι ομάδα 4

Στη συνέχεια διατηρήσαμε την τιμή της κλάσης πρόβλεψης ομάδα 4 και αποδώσαμε σε όλες τις άλλες τιμές την τιμή όχι ομάδα 4. Το decision tree έχει ως παράμετρο διαχωρισμού, το μάθημα



Εικ.7 Δέντρο απόφασης ομάδα 4- όχι ομάδα

Η αξιοπιστία για το **decision tree** είναι **83.33%**

accuracy: 83.33% +/- 0.00% (mikro: 83.33%)			
	true oxiomada 4	true omada 4	class precision
pred. oxiomada 4	50	10	83.33%
pred. omada 4	0	0	0.00%
class recall	100.00%	0.00%	

Πιν. 21 Αξιοπιστία Δέντρου απόφασης ομάδα 4-όχι ομάδα 4

Με weight by SVM το σπουδαιότερο διαχωριστικό στοιχείο είναι το μάθημα των Εικαστικών

attribute	weight
aglika	-0.011
mousiki	-0.011
fysika	0.000
istoria	0.000
glossa	0.000
mathimatika	0.000
geografia	0.001
gymnastiki	0.008
ikastika	0.067

Πιν.22 Βαρύτητα μεταβλητών

Η αξιοπιστία με Neural Net

accuracy: 80.00% +/- 6.67% (mikro: 80.00%)			
	true oxiomada 4	true omada 4	class precision
pred. oxiomada 4	48	10	82.76%
pred. omada 4	2	0	0.00%
class recall	96.00%	0.00%	

Πιν.23 Αξιοπιστία NN ομάδα4-όχι ομάδα4

Ενώ με K-NN 78.33%

accuracy: 78.33% +/- 10.67% (mikro: 78.33%)			
	true oxiomada 4	true omada 4	class precision
pred. oxiomada 4	47	10	82.46%
pred. omada 4	3	0	0.00%
class recall	94.00%	0.00%	

Πιν.24 Αξιοπιστία K-NN ομάδα4-όχι ομάδα4

6.3.3.δ ομάδα 5- όχι 5

Τέλος διατηρήσαμε την τιμή της κλάσης πρόβλεψης omada 5 και αντικαταστήσαμε

όλες τις άλλες τιμές με την τιμή oxí omada 5.

Με Select attributes αφαιρέσαμε το Fylo



Εικ.8 Εξαίρεση μεταβλητής Φύλο

Επειδή στο ζύγισμα που έγινε με weight by information Gain εμφάνισε μικρή βαρύτητα

attribute	weight
aglika	0.008
fysika	0.012
fylo	0.012
gymnastiki	0.012
glossa	0.017
geografia	0.017
istoria	0.017
ikastika	0.017
mousiki	0.025
mathimatika	0.036

Πιν.25 Weight by information Gain

Οι παράμετροι με τη μεγαλύτερη βαρύτητα όσες είναι πιο κοντά στο 1. Στον παραπάνω πίνακα φαίνεται ότι το μάθημα της Μουσικής και των Μαθηματικών είναι σημαντική παράμετρος διαχωρισμού.

Ωστόσο με τον weight by SVM οι παράμετροι εμφανίζουν διαφορετική βαρύτητα με το Μάθημα της Μουσικής να έχει την μεγαλύτερη.

attribute	weight
ikastika	-0.080
geografia	-0.051
mathimatika	-0.034
glossa	0.001
aglika	0.010
istoria	0.028
fysika	0.033
gymnastiki	0.035
mousiki	0.091

πιν.26 βαρύτητα παραμέτρων

Ωστόσο η αξιοπιστία είναι μικρότερη 76%

AttributeWeights (Weight by Information Gain) x PerformanceVector (Performance) x

ion

Table View Plot View

accuracy: 76.67% +/- 11.06% (mikro: 76.67%)

	true oxiomada5	true omada 5	class precision
pred. oxiomada5	45	8	84.91%
pred. omada 5	6	1	14.29%
class recall	88.24%	11.11%	

itic)

πιν.27 Αξιοπιστία

Με K-NN η αξιοπιστία είναι με τον ίδιο τρόπο

accuracy: 76.67% +/- 11.06% (mikro: 76.67%)

	true oxiomada5	true omada 5	class precision
red. oxiomada5	45	8	84.91%
red. omada 5	6	1	14.29%
lass recall	88.24%	11.11%	

Πιν.28 Αξιοπιστία K-NN ομάδα 5-όχι ομάδα5

Στη συνέχεια έγινε προσπάθεια για την εύρεση μεγαλύτερης αξιοπιστίας με **Optimize Parameters** (5 population size ,Fournament fraction 0.25 ,Crossover prob 0.9)

Οι παράμετροι που επιλέχθηκαν είναι για το decision tree :

Select Parameters: **configure operator**
Configure this operator by means of a Wizard.

Operators

- Validation (X-Validation)
- Decision Tree (2) (Decision Tree)
- k-NN (k-NN)
- Neural Net (Neural Net)
- Apply Model (Apply Model)
- Performance (Performance)

Parameters

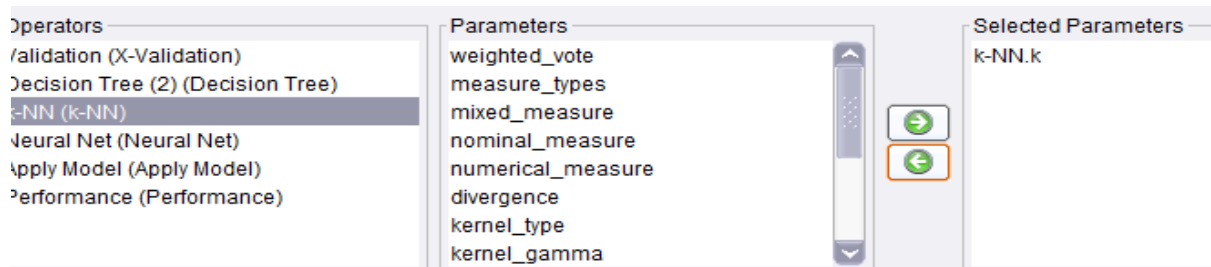
- criterion
- apply_pruning
- confidence
- apply_prepruning
- number_of_prepruning_alternatives

Selected Parameters

- Decision Tree (2).maximal_depth
- Decision Tree (2).minimal_gain
- Decision Tree (2).minimal_leaf_size
- Decision Tree (2).minimal_size_for_split

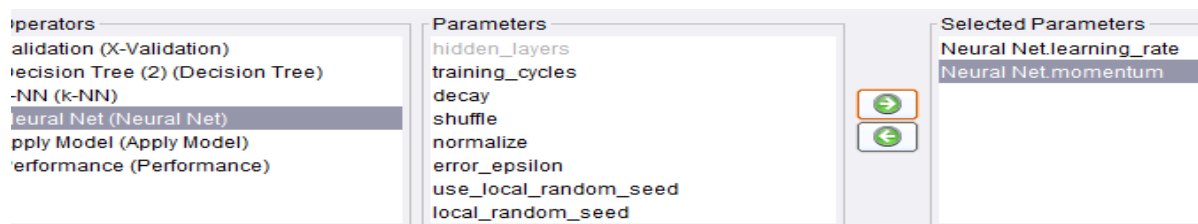
Εικ. 9 παράμετροι για Δέντρο απόφασης

Για τον K-NN



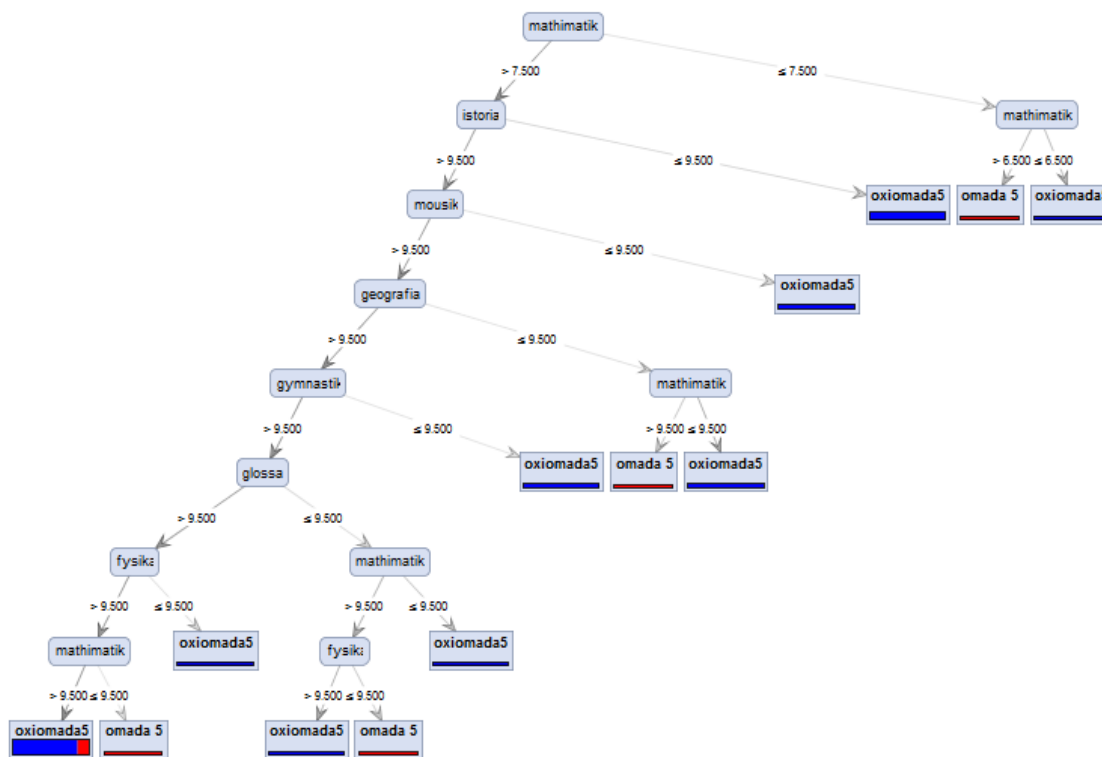
Εικ.10 Παράμετροι για K-NN

Για Neural Net



Εικ.11 παράμετροι για NN

Το decision tree έχει την εξής μορφή με καλύτερο σημείο διάσπασης που επιλέγεται ως ρίζα του δέντρου το μάθημα των μαθηματικών .



Εικ. 12 Δέντρο απόφασης ομάδα5-όχι ομάδα5

Η ακρίβεια του Δέντρου απόφασης είναι 76%. Η αξιοπιστία της πρόβλεψης με Neural Net , χρήση Weight By SVM και optimize parameters με τον τρόπο που είδαμε αγγίζει το 85%

accuracy: 85.00% +/- 5.00% (mikro: 85.00%)			
	true oxiomada5	true omada 5	class precision
pred. oxiomada5	51	9	85.00%
pred. omada 5	0	0	0.00%
class recall	100.00%	0.00%	

Πιν. 29 Αξιοπιστία Neural Net ομάδα 5-οχι ομάδα5

Παρακάτω παρουσιάζεται συνοπτικός πίνακας με τα υψηλότερα ποσοστά αξιοπιστίας

V validation Training set	<i>Decision tree</i>	<i>K-NN</i>	<i>Neural Net</i>
Όλες οι ομάδες	50.00%	50.00%	55.00%
ομάδα 2, 4 ,5	57.67%	59.67%	59.67%
ομάδα4-όχι ομάδα4	83.33%	78.33%	80%
ομάδα 5-όχι ομάδα 5	76.67%	76,67%	85%
Ομάδα2-όχι ομάδα 2	53.33%	46.67%	46.67%

Πιν.30 Συνοπτικός Πίνακας

Όπως φαίνεται στον παραπάνω πίνακα: το σώμα εκπαίδευσης(training set) με τη μεγαλύτερη ακρίβεια (accuracy) είναι αυτό που η κλάση πρόβλεψης έχει τις δυωνυμικές τιμές ομάδα 5- όχι ομάδα 5 με τον αλγόριθμο Neural Net και αυτό που η κλάση πρόβλεψης έχει τις τιμές ομάδα 4- όχι ομάδα 4 με τον αλγόριθμο του decision tree ID3.

Άρα η ακρίβεια των μοντέλων **ομάδα 5-όχι ομάδα 5** και **ομάδα 4- όχι ομάδα 4** είναι αποδεκτή και μπορεί να χρησιμοποιηθεί για να κατηγοριοποιήσει μελλοντικά δείγματα δεδομένων.

6.4 Πειράματα RapidMiner (B Μοντέλο)

Για το B Μοντέλο, με την κατηγοριοποίηση εκπαιδεύεται μοντέλο πρόβλεψης της σχολικής επίδοσης στην ΣΤ Δημοτικού, με δεδομένο το επάγγελμα των γονιών. Επίσης αναζητούνται κανόνες συσχέτισης.

Αρχικά από τα δεδομένα του βιβλίου μητρώου ,χρησιμοποιήθηκαν n=250 παραδείγματα, δημιουργήθηκε αρχείο excel όπου έγινε επεξεργασία και διατηρήθηκαν τα σημεία ενδιαφέροντος.

6.4.1 Δεδομένα εισόδου

Τα δεδομένα εισόδου αφορούν το επάγγελμα της μητέρας , το επάγγελμα του πατέρα , το Φύλο τη Μητρική Γλώσσα και το χαρακτηρισμό στα μαθήματα Γλώσσα και Μαθηματικά της Στ Δημοτικού

Πιο συγκεκριμένα αρχείο σε μορφή excel που λαμβάνεται από τη βάση δεδομένων και συγχρόνως αποτελεί το σώμα εκπαίδευσης (training set) έχει τα ακόλουθα δεδομένα

Παράμετροι(attributes):

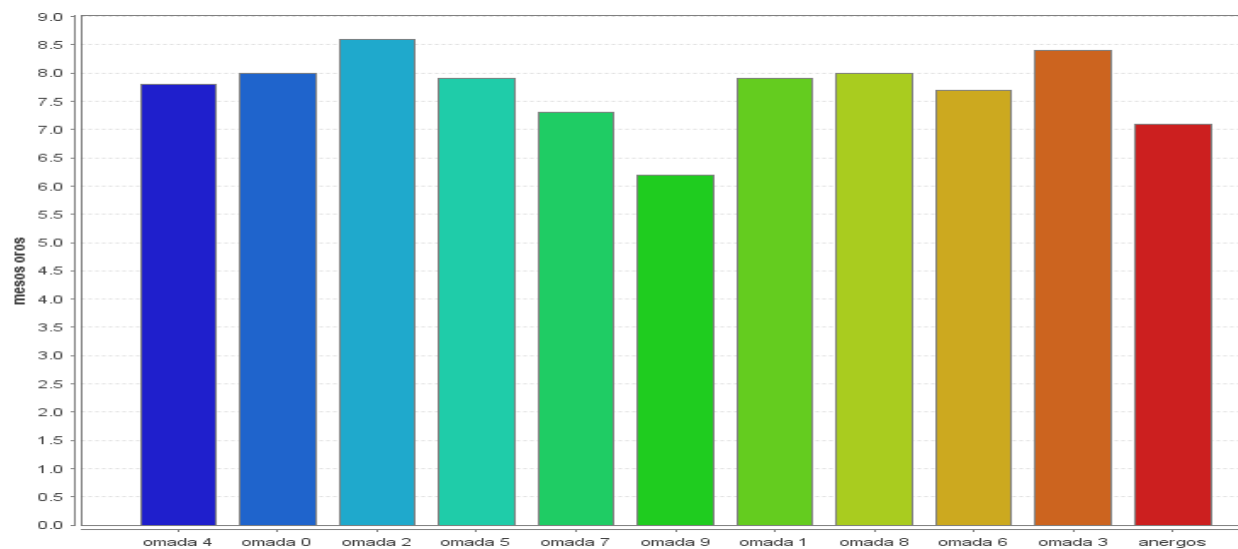
Fylo (φύλο) αναφέρεται στο φύλο των μαθητών και λαμβάνει δυο τιμές *αγόρι, κορίτσι*

Επάγγελμα μητέρας αναφέρεται στο επάγγελμα της μητέρας και έχει ετικέτες :ομάδα 1 , ομάδα 2, ομάδα 3, ομάδα 4, ομάδα 5, ομάδα 6, ομάδα 7, ομάδα 8, ομάδα 9 , ομάδα 0, συνταξιούχος, οικιακά.

Επάγγελμα πατέρα αναφέρεται στο επάγγελμα του πατέρα και έχει **ετικέτες** : ομάδα 1,ομάδα2, ομάδα 3, ομάδα 4, ομάδα 5, ομάδα 6, ομάδα 7, ομάδα 8, ομάδα 9, ομάδα 0, άνεργος, συνταξιούχος.

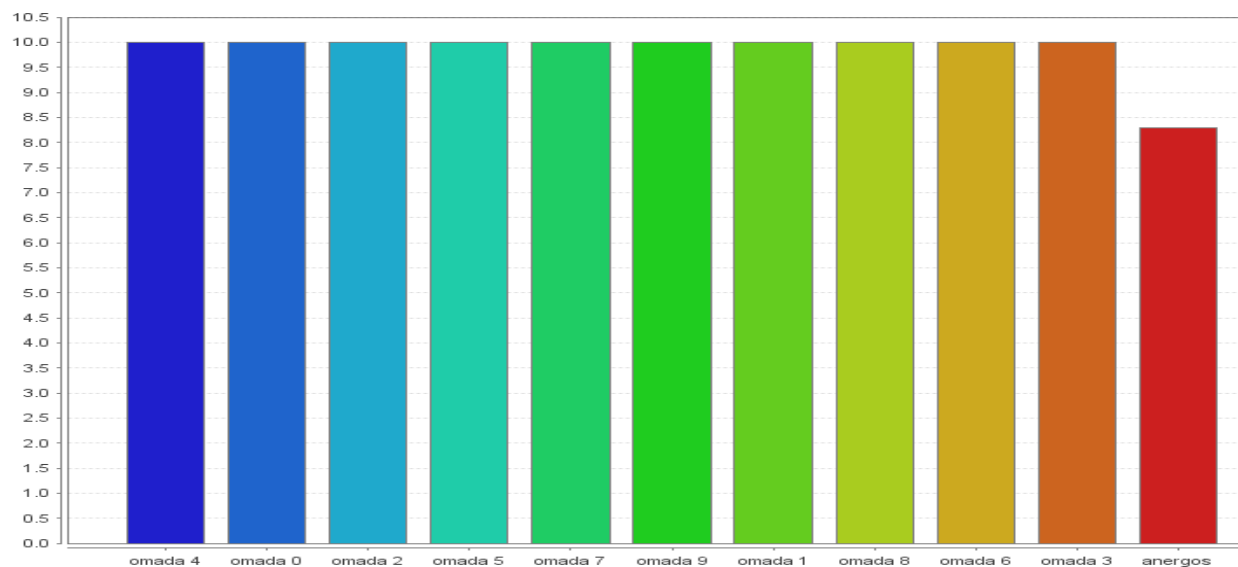
Μητρική Γλώσσα αναφέρεται στη μητρική γλώσσα του μαθητή και εδώ παίρνει ετικέτες :Ελληνική , Αλβανική

ΠΕΡΙΓΡΑΦΗ:

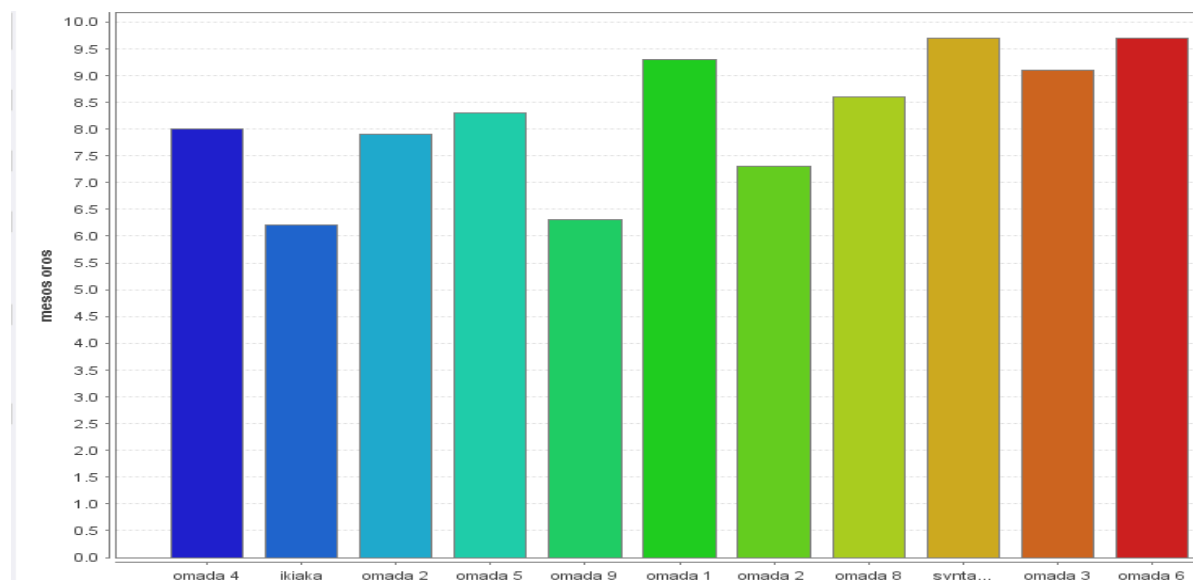


Σχ. 28 χαμηλότερη βαθμολογία –επάγγελμα πατέρα

Ο μικρότερος μέσος όρος στη βαθμολογία αφορά μαθητές που το επάγγελμα του πατέρα ανήκει στην ομάδα 9 (ανειδίκευτος εργάτης) και στην ομάδα άνεργος.



Αλλά και στην υψηλότερη βαθμολογία η χαμηλότερη είναι στην ομάδα άνεργος Ως προς το επάγγελμα της μητέρας χαμηλότερες βαθμολογίες παρατηρούνται όταν ανήκουν στην ομάδα 9 ή στην ομάδα οικιακά.



Σχ.29 Χαμηλότερη βαθμολογία –επάγγελμα μητέρας

6.4.2 Πρόβλεψη κλάσης

Η κλάση είναι ονομαστική αναφέρεται στον χαρακτηρισμό που συνοδεύει την αριθμητική

βαθμολογία των μαθητών του δείγματος της Στ τάξης αρχικά στο μάθημα των Μαθηματικών .

Το training set σε αυτή την περίπτωση έχει τη μορφή

label	Polynomial	0	Least	Most	Values
mathimatika	Polynomial	0	kala (8)	arista (171)	arista (171), poli kala (
filo	Polynomial	0	koritsi (123)	agori (126)	agori (126), koritsi (12
epagelma patera	Polynomial	0	anergos (4)	omada 7 (39)	omada 7 (39), omada 8
epagelma miteras	Polynomial	0	omada 8 (1)	ikiaka (114)	ikiaka (114), omada 4
glosa	Polynomial	0	kala (7)	arista (172)	arista (172), poli kala (
mitriki glosa	Polynomial	0	albaniki (23)	eliniki (226)	eliniki (226), albaniki (2

Εικ. 13 Training set

Το decision tree εμφανίζει αξιοπιστία 86,37%.

accuracy: 86.37% +/- 5.09% (mikro: 86.35%)				
	true arista	true kala	true poli kala	class precision
pred. arista	158	1	14	91.33%
pred. kala	0	4	3	57.14%
pred. poli kala	13	3	53	76.81%
class recall	92.40%	50.00%	75.71%	

Πιν. 31 Ακρίβεια Δέντρου απόφασης

6.4.2.a Πρόβλεψη κλάσης

Στη συνέχεια αντικαταστάθηκε η κλάση πρόβλεψης με το μάθημα της Γλώσσας :

label		Least	Most	Values
glosa	Polynomial 0	kala (7)	arista (172)	arista (172), poli kala (
filo	Polynomial 0	koritsi (123)	agori (126)	agori (126), koritsi (12
epagelma patera	Polynomial 0	anergos (4)	omada 7 (39)	omada 7 (39), omada 9
epagelma miteras	Polynomial 0	omada 8 (1)	ikiaka (114)	ikiaka (114), omada 4
mitriki glosa	Polynomial 0	albaniki (23)	eliniki (226)	eliniki (226), albaniki (2
mathimatika	Polynomial 0	kala (8)	arista (171)	arista (171), poli kala (

Πιν. 32 Training set 2

Το ποσοστό αξιοπιστίας με decision tree είναι:

accuracy: 84.70% +/- 6.29% (mikro: 84.74%)				
	true arista	true kala	true poli kala	class precision
pred. arista	158	0	14	91.86%
pred. kala	0	0	3	0.00%
pred. poli kala	14	7	53	71.62%
class recall	91.86%	0.00%	75.71%	

Πιν.33 Ακρίβεια δέντρου απόφασης 2

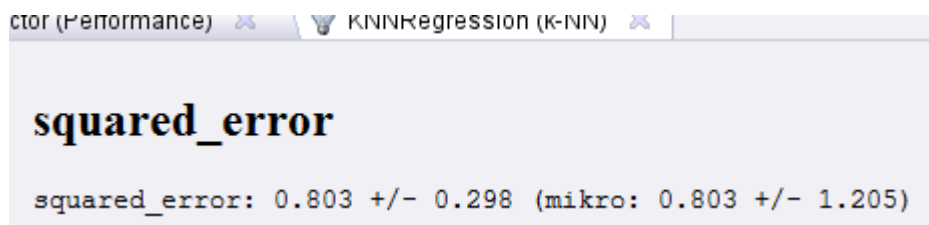
6.4.2.β Πρόβλεψη κλάσης ③

Στη συνέχεια αντικαταστάθηκε η κλάση πρόβλεψης με το μέσο όρο των μαθημάτων της ΣΤ Δημοτικού. Κατά συνέπεια η κλάση είναι αριθμητική.

ExampleSet (249 examples, 1 special attribute, 4 regular attributes)					
Row No.	mesos oros	filo	epagelma p...	epagelma ...	mitriki glo:
1	9.400	agori	omada 4	omada 4	eliniki
2	9.900	agori	omada 0	ikiaka	eliniki
3	10	agori	omada 2	ikiaka	eliniki
4	9.500	agori	omada 5	ikiaka	eliniki
5	10	agori	omada 2	omada 2	eliniki
6	9.100	agori	omada 7	ikiaka	eliniki
7	9.800	agori	omada 7	omada 4	eliniki
8	6.200	agori	omada 9	ikiaka	albaniki
9	10	agori	omada 9	omada 5	eliniki
10	10	agori	omada 5	omada 5	eliniki
11	10	agori	omada 9	ikiaka	eliniki
12	10	agori	omada 5	omada 5	eliniki
13	8.900	agori	omada 4	omada 5	eliniki
14	10	agori	omada 2	omada 2	eliniki
15	9.300	agori	omada 9	ikiaka	eliniki
16	9.800	koritsi	omada 7	ikiaka	eliniki
17	8.600	koritsi	omada 9	ikiaka	eliniki
18	8.700	koritsi	omada 1	ikiaka	eliniki
19	8.600	koritsi	omada 5	omada 5	eliniki
20	9.500	koritsi	omada 5	omada 5	eliniki

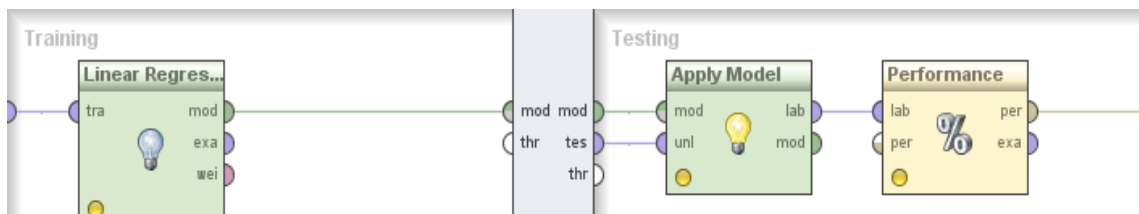
Εικ .14 Σώμα εκπαίδευσης 3

Με KNN η αξιοπιστία είναι:



Εικ.15 Αξιοπιστία με KNN

Στη συνέχεια έγινε προσπάθεια πρόβλεψης με την ίδια κλάση αλλά με linear Regression



Attribute	Coefficient	Std. Error	Std. Coeffici...	Tolerance	t-Stat	p-Value	Code
filo	0.385	0.094	0.250	0.991	4.081	0.000	****
epagelma p:	-0.034	0.018	-0.118	0.991	-1.924	0.066	*
(Intercept)	9.167	0.093	?	?	98.188	0	****

Πιν.34 Linear Regression

Το training set με τη μεγαλύτερη αξιοπιστία είναι αυτό που η κλάση πρόβλεψης είναι ο μέσος όρος της βαθμολογίας (αριθμητική κλάση) Με linear Regression η πρόβλεψη εμφανίζει μικρή απόκλιση

```

view x PerformanceVector (Performance) x LinearRegression (Linear Regression) x

PerformanceVector

PerformanceVector:
root_mean_squared_error: 0.687 +/- 0.077 (mikro: 0.692 +/- 0.000)
squared_error: 0.478 +/- 0.115 (mikro: 0.479 +/- 0.644)

```

```

ctor (Performance) x LinearRegression (Linear Regression) x

squared_error

squared_error: 0.478 +/- 0.115 (mikro: 0.479 +/- 0.644)

```

Εικ.16 .Απόκλιση

6.4.3 Κανόνες συσχέτισης (Association Rules)

Στο σύνολο των δεδομένων αναζητούμε κανόνες που προβλέπουν την εμφάνιση ενός στοιχείου με βάση την εμφάνιση άλλων στοιχείων. Μετατρέπονται τα αριθμητικά στοιχεία με τον operator Numerical to Binominal σε true false. True όπου υπάρχει 1 false όπου υπάρχει 0.

ExampleSet (Nominal to Binominal)														
leSet (249 examples, 0 special attributes, 31 regular attributes)											Filter (249 / 249 examples):			all
os oro...	filo = agori	filo = koritsi	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	epagelma p...	
ε	true	false	true	false	false	false	false	false	false	false	false	false	false	
ε	true	false	false	true	false	false	false	false	false	false	false	false	false	

πιν.35 μετατροπή αριθμητικών στοιχείων

Στη συνέχεια εφαρμόζεται ο FP Growth:

Και οι κανόνες συσχέτισης δίνουν τα παρακάτω αποτελέσματα:

Create Association Rules:

No.	Premises	Conclusion	Support	Confid...	LaPla...	Gain	p-s	Lift	Convi...
1	filo = agori	mitriki glosa = eliniki	0.446	0.881	0.960	-0.566	-0.013	0.971	0.776
2	epagelma miteras = ikiaka	mitriki glosa = eliniki	0.422	0.921	0.975	-0.494	0.006	1.015	1.170
3	filo = koritsi	mitriki glosa = eliniki	0.462	0.935	0.978	-0.526	0.013	1.030	1.420
4	filo = koritsi, epagelma miteras = ikiaka	mitriki glosa = eliniki	0.229	0.950	0.990	-0.253	0.010	1.047	1.847

Πιν.36 κανόνες συσχέτισης

AssociationRules

Association Rules

```
[filo = agori] --> [mitriki glosa = eliniki] (confidence: 0.881)
[epagelma miteras = ikiaka] --> [mitriki glosa = eliniki] (confidence: 0.921)
[filo = koritsi] --> [mitriki glosa = eliniki] (confidence: 0.935)
[filo = koritsi, epagelma miteras = ikiaka] --> [mitriki glosa = eliniki] (confidence: 0.950)
```

Οι κανόνες συσχέτισης αφορούν τη μητρική γλώσσα, το επάγγελμα της μητέρας όταν είναι οικιακά και το φύλο. Ως προς το επάγγελμα δεν αναφέρονται κανόνες συσχέτισης.

Κεφάλαιο 7

ΣΥΜΠΕΡΑΣΜΑΤΑ

Συμπεράσματα στήριξη από άλλες έρευνες

Σύμφωνα με το Classer (Glaser R., 1990) «Η κοινωνία χωρίζεται σε άτομα που ταυτίζονται με την επιτυχία και σε αυτά που ταυτίζονται με την αποτυχία». σε αυτό συμβάλει το εκπαιδευτικό σύστημα ,εφόσον η βαθμολογία και οι έπαινοι χρησιμοποιούνται ως εξωτερική αμοιβή. Με τον τρόπο αυτό χρησιμοποιούν τη σχολική επίδοση ως «τεκμήριο της προσωπικής τους αξίας» (Burns,1982)

Με βάση τις πληροφορίες που αποκτούν οι μαθητές από τις επιδόσεις τους στα διάφορα μαθήματα διαμορφώνουν προσωπική αντίληψη για το επίπεδο των ικανοτήτων τους στους αντίστοιχους τομείς. Η πληροφόρηση αυτή ,σύμφωνα με παλαιότερες έρευνες στο εξωτερικό (Leon , 1977) , (Vernon, 1957) και στην Ελλάδα (Μιχαηλίδης-Ναυάρος , 1966) και σύγχρονες (Κασσωτάκης , 2004), αποτελεί παράγοντα που επηρεάζει σημαντικά στη λήψη αποφάσεων από τους μαθητές , ως αναφορά την επιλογή τους.

Το κοινωνικοοικονομικό υπόβαθρο και η ικανότητα επηρεάζουν τις φιλοδοξίες για το σχολείο και κατ'επέκταση την επαγγελματική φιλοδοξία. Η ικανότητα και το κοινωνικοοικονομικό υπόβαθρο εκφράζονται μέσα από τη σχολική επίδοση ενώ ενισχύονται από την κοινωνική υποστήριξη ή απαξίωση που αυτή προσφέρει . (Κοινωνικοψυχολογικό μοντέλο Wisconsin)

Σύμφωνα με τον Burns (1982) η αντίληψη της ικανότητας του μαθητή είναι ένας δείκτης της αντίληψης του μαθητή για το προηγούμενο σχολικό ιστορικό του σε σχέση με την ανατροφοδότηση που του παρέχεται από τους βαθμούς , τα τεστ , τους δασκάλους, τους γονείς και τους συμμαθητές για τη σχολική του εργασία. Όσο πιο πολλές ενδείξεις δίνονται στο μαθητή τόσο πιο μεγάλο είναι το ενδεχόμενο να λειτουργήσει ο τρόπος που τις φιλτράρει ως πρόβλεψη για τη μελλοντική του πορεία, σχολική ή επαγγελματική. Εκτός αν συμβεί κάτι απρόοπτο στο μαθητή που θα ανατρέψει αυτή την πορεία. (αναφέρεται Φλουρής, 2004)

Σύμφωνα με έρευνες των Marshalls & Tucker , 1992, :Η ακαδημαϊκή επίδοση είναι άμεσα συναρτώμενη με την επαγγελματική ανάπτυξη. Στο ίδιο συμπέρασμα καταλήγουν και έρευνες των Cysbers & Henderson το 1994 και των Resnick & Wirt,1996.(αναφέρονται: (Φλουρής Γ., 2004)

Μαθητές που σημειώνουν υψηλότερες εκπαιδευτικές επιδόσεις από τους συμμαθητές τους παρουσιάζουν υψηλές επαγγελματικές φιλοδοξίες (Δημητρόπουλος Ε., 2003).Αντίθετα οι μαθητές με χαμηλή επίδοση δεν ενθαρρύνονται στην ανάπτυξη ικανοτήτων και πρωτοβουλιών και κατά συνέπεια , χαμηλώνουν τον πήχη των επαγγελματικών τους φιλοδοξιών . (Weinstein C.E & Mayes R.E, 1986)

Έρευνες στην Ελλάδα τεκμηριώνει την επιλογή των σπουδών με βάση τη βαθμολογία στα διάφορα μαθήματα. (Κασσωτάκης Μ., 2013) Η βαθμολογία, σύμφωνα με την παραπάνω έρευνα, η λειτουργεί ως σύστημα πληροφόρησης του μαθητή και της οικογένειάς του, σχετικά με τις ικανότητες και το μαθησιακό επίπεδο αλλά και επέκταση σε μηχανισμούς κοινωνικής κατανομής και επιλογής (Κασσωτάκης Μ., 2013).

Στην παρούσα διπλωματική ερευνάται αν η ικανότητα, όπως την θεωρούν τα ίδια τα άτομα ως ενήλικες και όπως αυτή εκφράζεται με την αριθμητική βαθμολογία της Στ Δημοτικού σχετίζεται με την επιλογή επαγγέλματος . Ερευνάται η δυνατότητα με τα μέσα που παρέχουν τα ευφυή συστήματα, σχετικής πρόβλεψης, εκπαιδευοντας με τα δεδομένα που συνελέγησαν ένα αξιόπιστο μοντέλο κατηγοριοποίησης με σκοπό την πρόβλεψη.

Σε ένα δεύτερο επίπεδο από το κοινωνικοοικονομικό υπόβαθρο απομονώνεται ο παράγοντας επάγγελμα γονιών και ερευνάται η σχέση του επαγγέλματος των γονιών με τη σχολική επίδοση όπως αυτή εκφράζεται με τη βαθμολογία της Στ Δημοτικού. Ερευνάται παράλληλα η αξιοπιστία σχετικού μοντέλου πρόβλεψης .

Τα ίδια τα άτομα που ρωτήθηκαν πιστεύουν ότι στην επιλογή του επαγγέλματος έπαιξε ρόλο η επίδοση στην Στ δημοτικού βαθμολογία σε ένα μάθημα επηρέασε την τελική επιλογή. Η βαθμολογία βοήθησε σε ποσοστό να καταλάβουν που ήταν πραγματικά καλοί.

Διαπιστώθηκε ότι η βαθμολογία ήταν ένα γεγονός που επηρέαζε το οικογενειακό περιβάλλον. Θετικά στην πλειοψηφία των περιπτώσεων που είχαν Άριστη βαθμολογία.

Γενικά η στάση των ατόμων που συμμετείχαν απέναντι στη βαθμολογία τους συμβαδίζει με την ίδια τη βαθμολογία αλλά και με τη στάση που συντηρούν μέχρι σήμερα και απέναντι στο επάγγελμά τους.

Στην πλειοψηφία έχουν ένα επάγγελμα υψηλού κοινωνικού γοήτρου , υψηλή βαθμολογία και αίσθηση δημιουργικότητας στο επάγγελμα. Αίσθηση δικαιοσύνης στη βαθμολογία και ανατροφοδότηση για την ικανότητα .Αν εξετάσουμε μεμονωμένα τις απαντήσεις των ερωτηματολογίων διαπιστώνεται ότι η χαμηλή βαθμολογία είναι αυτή που συνοδεύεται από αρνητικά σχόλια του οικογενειακού περιβάλλοντος και αίσθημα ανικανοποίητου και στο χώρο του επαγγέλματος σήμερα.

Διαπιστώθηκε ότι τα άτομα που το επάγγελμά τους άνηκε στην ομάδα 2 , ασκούσαν δηλαδή επιστημονικά , καλλιτεχνικά και συναφή επαγγέλματα είχαν καλύτερη επίδοση στο Μάθημα της Γλώσσας και ήταν και το πρώτο στην προτίμησή τους μάθημα , ενώ στο μάθημα των Μαθηματικών εμφάνιζαν ελλείψεις και είχαν χαμηλότερη βαθμολογία από το μάθημα της Γλώσσας.

Το μοντέλο κατηγοριοποίησης εμφανίζει την υψηλότερη αξιοπιστία όταν γίνεται διακριτοποίηση της κλάσης πρόβλεψης «επάγγελμα» με δυαδική ετικέτα ομάδα 4 – όχι ομάδα 4 με τον αλγόριθμο του decision tree, ID3. Για την ομάδα 4(υπάλληλοι γραφείου) το μάθημα διαχωρισμού ήταν η Γεωγραφία. Επίσης το μοντέλο με κλάση πρόβλεψης «επάγγελμα» και ετικέτες ομάδα 5 –όχι ομάδα 5 παρουσίασε αποδεκτή ακρίβεια με Neural Net.

Τα παραπάνω μοντέλα μόνο μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση μελλοντικών δειγμάτων δεδομένων.

Επομένως: Η επίδοση ασκεί επιρροή .

Αποτελεί πληροφορία για τις ικανότητες του μαθητή , τόσο για τον ίδιο , όσο και για το οικογενειακό του περιβάλλον.

Καλλιεργεί στάσεις απέναντι στη σχολική εργασία που διατηρούνται και στο χώρο του επαγγέλματος.

Η καλή βαθμολογία έχει θετικές επιδράσεις. Ο βαθμός των αρνητικών επιδράσεων της χαμηλής βαθμολογίας θέλει διερεύνηση.

Τα άτομα που πήραν μέρος και επαγγελματικά άνηκαν στην ομάδα 2 δηλαδή ασκούσαν επιστημονικά , καλλιτεχνικά και συναφή επαγγέλματα και έδειξαν σαφή προτίμηση ως μαθητές στο μάθημα της Γλώσσας , στο οποίο είχαν και την υψηλότερη βαθμολογία. Από την άλλη στο μάθημα των Μαθηματικών υπήρχαν ελλείψεις και χαμηλή βαθμολογία.

Η διαθεματικότητα στο δημοτικό η οποία ενισχύεται και με τα βιβλία που διδάσκονται σήμερα ίσως είναι η λύση προκειμένου να ενισχύονται οι κλίσεις και τα ενδιαφέροντα και να καλλιεργούνται δεξιότητες μάθησης.

Το μοντέλο πρόβλεψης με τη μεγαλύτερη ακρίβεια ήταν αυτό που είχε κλάση το επάγγελμα με ετικέτες ομάδα 4-όχι ομάδα 4 με εφαρμογή ID3. Για την ομάδα 4 λοιπόν , υπάλληλοι γραφείου, η Γεωγραφία είναι το μάθημα διαχωρισμού . Αν έχουν στα Μαθηματικά και στα Φυσικά μεγαλύτερο βαθμό από 9,5 ανήκουν στην ομάδα 4. Το φύλο φαίνεται ότι παίζει ρόλο . Οι γυναίκες με μικρότερο ή ίσο βαθμό με 9,5 στη Γλώσσα , Ιστορία και Εικαστικά ανήκουν στην ομάδα 4.

Στη διεθνή βιβλιογραφία δίνεται ιδιαίτερη βαρύτητα στην επίδραση των γονέων στη σχολική επίδοση. Οι έρευνες καταλήγουν στο συμπέρασμα ότι το συνολικό οικογενειακό περιβάλλον συνεισφέρει περισσότερο στην επίδοση των παιδιών από τις μεταβλητές σχολείο και εκπαιδευτικοί. (Majorbanks 1971, 1981,1983, Keeves 1972).(αναφέρεται: Φλουρής Γ., 2004)

Έρευνα σε 39 σχολεία της Γερμανίας σε 341 μαθητές της Ε΄και ΣΤ΄ καταλήγει στο ότι « η αρχική επίδοση είναι η πιο ισχυρή πρόβλεψη της τελικής επίδοσης». Δίνεται έμφαση στα κίνητρα που επηρεάζουν καθοριστικά την γνωστική ανάπτυξη.

Σε έρευνα του πανεπιστημίου Κρήτης το 1983 με δείγμα 425 μαθητών ενώ διαπιστώθηκε θετική συσχέτιση ανάμεσα στη σχολική επίδοση της Ε και ΣΤ με το μορφωτικό επίπεδο του πατέρα και ειδικά με το επάγγελμα του. Ωστόσο δε διαπιστώθηκε συσχέτιση ανάμεσα στη σχολική επίδοση και τις επαγγελματικές φιλοδοξίες. Διαπιστώθηκε ότι οι επαγγελματικές φιλοδοξίες κατευθύνονται από την οικογένεια σε επαγγέλματα υψηλού κοινωνικού γοήτρου.

Από τους 250 μαθητές που η βαθμολογία και το επάγγελμα των γονιών τους χρησιμοποιήθηκαν ως δεδομένα για το Β Μοντέλο , υπάρχει ένας σημαντικός διαχωρισμός. Οι μαθητές των οποίων το επάγγελμα του πατέρα είναι στην ομάδα άνεργος εμφανίζουν και τη χαμηλότερη βαθμολογία. Τη χαμηλότερη βαθμολογία έχουν και μαθητές των οποίων το επάγγελμα της μητέρας ανήκει στην ομάδα 9 , δηλαδή στην κατηγορία «ανειδίκευτος εργάτης». Εδώ βέβαια παρεμβαίνουν και άλλοι παράγοντες , όπως η ενίσχυση που ο χρόνος που οι συνθήκες επιτρέπουν να αφιερώσει ο γονιός κ.ά.

Ωστόσο μπορεί να αποτελεί μια χρήσιμη γνώση ώστε να δημιουργείται ένα προφίλ του μαθητή το οποίο θα ενισχύεται με παρότρυνση για δανεισμό βιβλίων και άλλες δραστηριότητες που θα ενισχύσουν το υπόβαθρον μαθητών που το οικογενειακό περιβάλλον δεν είναι ανάλογα υποστηρικτικό.

Ως προς τις τεχνικές της εξόρυξης δεδομένων , αποτελούν εργαλεία που στο μέλλον θα κυριαρχήσουν στην εκπαιδευτική έρευνα και στην Ελλάδα καθώς παρέχουν εργαλεία και δυνατότητες πρόβλεψης που δε διαθέτουν οι παραδοσιακοί τρόποι ανάλυσης δεδομένων. Η έρευνα στην εκπαίδευση με χρήση τεχνικών εξόρυξης στην Ελλάδα παρατηρείται κυρίως σε επίπεδο διπλωματικών και διδακτορικών διατριβών και αφορά κυρίως την τριτοβάθμια εκπαίδευση.

Προεκτάσεις-Προτάσεις

Το σύστημα διαχείρισης διαδικτυακών μαθημάτων **moodle** έχει αναπτύξει πέρα από τα βασικά στατιστικά στοιχεία που προσφέρει, από το 2012 ειδικά **modules** που θα αποτελέσουν αντικείμενο έρευνας και εξαγωγής χρήσιμης γνώσης στο πεδίο της EDM.

Στις χαμηλότερες βαθμίδες, πρωτοβάθμια και δευτεροβάθμια εκπαίδευση η συλλογή των δεδομένων βρίσκει εμπόδιο τα προσωπικά δεδομένα. Ωστόσο καθώς στο my school καταχωρούνται πλήθος δεδομένων, Θα μπορούσε να εξεταστεί ανάλογα η δυνατότητα πρόσβασης σε μια μεγάλη αποθήκη πληροφοριών από την οποία μπορεί να εξαχθεί χρήσιμη γνώση στον τομέα της εκπαίδευσης.

Βέβαια ένα ακόμη εμπόδιο είναι η κατάρτιση στις τεχνικές της εξόρυξης δεδομένων και στις δυνατότητες που προσφέρουν. Σεμινάρια κατάρτισης θα μπορούσαν να κατευθύνουν τους εκπαιδευτικούς που ενδιαφέρονται για καλύτερη διαχείριση της μαθησιακής διαδικασίας. Για παράδειγμα θα μπορούσε να γίνει πρόβλεψη του βαθμού δυσκολίας ενός διδακτικού αντικειμένου ώστε να ακολουθηθεί η κατάλληλη μεθοδολογία ανάλογα με το επίπεδο κατάκτησης δεξιοτήτων και στρατηγικών των μαθητών.

Στον τομέα του επαγγελματικού προσανατολισμού ίσως θα ήταν χρήσιμη η πρόβλεψη της ζήτησης ενός επαγγέλματος σε βάθος δεκαετίας ώστε να παρέχεται η συμβουλευτική στήριξη και στοχευόμενος επαγγελματικός προσανατολισμός.

Βέβαια ο ατομικός παράγοντας που κάνει κάθε περίπτωση ξεχωριστή κάνει τα συμπεράσματα μη γενικεύσιμα , όμως μπορούν να ανιχνευτούν τάσεις που θα μετατραπούν σε δυνατότητες πάντα με σεβασμό στην ανθρώπινη ιδιαιτερότητα.

Η σχολική επίδοση είναι ένα πολυπαραγοντικό αποτέλεσμα και τελικά η διερεύνηση του ενός ή του άλλου παράγοντα ανεξάρτητα οδηγεί σε μη ακριβή συμπεράσματα. Ωστόσο σαν αποτέλεσμα ασκεί μεγάλη επιρροή , όπως διαπιστώθηκε . Το ερώτημα που τίθεται είναι αν το άτομο πείθεται για την ικανότητά του περισσότερο από το να συνειδητοποιεί την πραγματική ικανότητα ή το αν η βαθμολογία λειτουργεί ως αυτοεκπληρούμενη προφητεία για εκείνον που τη λαμβάνει. Οι ερευνητές συμφωνούν στο ότι η πρόβλεψη της επίδοσης είναι ένα πεδίο που μπορεί να αποβεί πολύ χρήσιμο για την έγκαιρη ανατροφοδότηση που θα βοηθήσει τους μαθητές να βελτιώσουν τις δεξιότητές τους ενώ θα μειώσει το άγχος της διδασκαλίας και της αξιολόγησης της επιτυχίας της από την πλευρά των εκπαιδευτικών.(Thai-Nghe et al. 2011)

Σε περιπτώσεις μαθησιακών δυσκολιών οι τεχνικές της εξόρυξης δεδομένων θα μπορούσαν να έχουν εφαρμογή στην ανάλυση κειμένων(text mining) με τη δημιουργία μοντέλων πρόβλεψης για τη δυσλεξία , δυσορθογραφία, δυσαριθμησία κ.α.

Γενικά η αξία των ερευνών που αφορούν την εκπαίδευση βρίσκεται στα ερωτήματα που δημιουργούν πέρα από τα προβλήματα που τελικά επιλύουν.

Βιβλιογραφία

- Alaofi M. , Rumantir G. (2015). Personalisation of Generic Library search results using student enrolment information. *JEDM* .
- Alkin M.C. (1974). *Evaluation theory development*. Los Angeles: CSE.
- Amand Rajaramon, Jeff Ullman. (n.d.). Mining of Massive datasets.
- Bardick AD Bernes K.B . (2005). *The need for career planing*. canada: Journal of counseling.
- Belanoff P.& Dickson M. (1991). *Portfolio in teacher education*. Portsmouth: Boynton/Cook Publisher.
- Blustein L. D. . (2006). *The psychology of working*. London: Laurence Erlbaum.
- Brookhart S.M . (2008). *How to give effective feedback to your student*. N.Y: Columbia University.
- Cangelosi . J. (1990). *Designing tests for evaluating achievement*. N.Y : Addison-Wiley.
- Dunham . (2004). *Εισαγωγικά και προηγμένα θέματα εξόρυξης γνώσης από δεδομένα*. (Θ. Βερούκιος, Μεταφρ.) Αθήνα: Νέες Τεχνολογίες.
- Glaser R. (1990). *Testing and Assesment*. Pitsburg: University of Pitsburg.
- Graff H. (2012, Μάρτιος 15). The development of student grading is a story of trial and error.
- Hartigan J.A. (1983). *Bayes theory springer series in statistics* . CRC: Chapman and Hall.
- Hunt D. Blackledge. (1995). *Κοινωνιολογία της εκπαίδευσης*. Αθήνα : Έκφραση.
- Kerr Deirdre. (2015). Using data mining resultes improve educational video game design. *JEDM* .
- Knowles E. Jared. (2015). Building an Accurate Statewide. *JEDM* .
- Nguyen Thai-Nghe, Drumont L.,et al. (2011). multi-relational factorization models for predicting students performance. *JEDM* .
- P-N Tan, M. Steinbach and V. Kumar. (2010). *introduction to data mining*. (Σ. Βερούκιος B., Μεταφρ.) Αθήνα: Τζιόλα.
- Romero C., Ventura S. (2010, April). Educational data mining. *Man Cybern* , σσ. 601-618.
- Russell S. & Norvig P. (2005). *Τεχνητή νοημοσύνη, μια σύγχρονη προσέγγιση*. Αθήνα: Κλειδάριθμος.
- Savickas M. (2005). the theory and practice of carrear construction. *career development and counseling* , σσ. 42-70.
- Swanson L.J, Fouand A.N. (2014). *Career Theory and practice*. L.A: Sage.
- Weinstein C.E & Mayes R.E. (1986). *The teaching of learning strategies in handbook of reasearch on teaching*. N.Y: Macmillan.

- Wiley Green & Soang. (2012). Edm. *JEDM* .
- Zimmerman J, Brodersen K., Heinemann H., Buhmann J. (2015). A model based approach to predicting graduate level performance using indicator of undergraduate level performance. *JEDM* .
- Βερύκιος & Θεοδωρίδης . (2004). *Εισαγωγικά και προηγμένα θέματα εξόρυξης γνώσης από δεδομένα*. Αθήνα: Νέες Τεχνολογίες.
- Βρεττός Ι. & Καψάλης Α. . (1997). *Αναλυτικά προγράμματα*. Αθήνα: Γραφικές Τέχνες.
- Γ., Φ. (1989). *Αυτοσυναίσθημα , επίδοση και επαγγελματικές φιλοδοξίες*. Ηράκλειο: Φλουρής.
- Δημητρόπουλος Ε. (2003). *Εκπαιδευτική αξιολόγηση. Η αξιολόγηση του μαθητή*. Αθήνα : Γρηγόρης.
- Κασσωτάκης Μ. (2013). *Η αξιολόγηση της επίδοσης των μαθητών*. Αθήνα: Γρηγόρης.
- Κασσωτάκης Μ. (2003). *Συμβουλευτική και επαγγελματικός προσανατολισμός. Θεωρία και πράξη*. Αθήνα: Τηπωθήτω.
- Κολιάδης Ε. . (2002). *Γνωστική ψυχολογία, γνωστική νευροεπιστήμη και εκπαιδευτική πράξη*. Αθήνα : Κολιάδης.
- Κωνσταντινίδης Α. (2000). *Εκπαιδευτικές και επαγγελματικές επιλογές των νέων. Φύλο, σχολική επίδοση*. Θεσσαλονίκη: Τμήμα ψυχολογίας.
- Κωνσταντίνου χ. (2007). *Η αξιολόγηση της επίδοσης του μαθητή ως παιδαγωγική, λογική και σχολική πρακτική*. Αθήνα: Gutenberg.
- Ματσαγγούρας Η. (2001). *Η σχολική τάξη*. Αθήνα: Gutenberg.
- Ματσαγγούρας Η.Γ. (2002). *Η διαθεματικότητα στη σχολική γνώση*. Αθήνα: Γρηγόρης.
- Παπαιωάννου Τ. (2000). *Εισαγωγή στις πιθανότητες*. Αθήνα: Σταμούλης.
- Ταγκάλου Α. (2009). Διερεύνηση του επαγγελματικού προσανατολισμού σε μαθητές και μαθήτριες της Στ δημοτικού. *Διδακτορική διατριβή* . (Π. Αθηνών, Επμ.) Αθήνα.
- Φλουρής Γ. (2004). *Αυτοαντίληψη , σχολική επίδοση και επίδραση των γονέων*. Αθήνα : Γρηγόρης.
- Χαλκίδη Μ., Βαζιργιάνης Μ. (2005). *Εξόρυξη γνώσης από βάση δεδομένων και τον παγκόσμιο ιστό*. Αθήνα : Δαρδάνος.

Γλωσσάρι όρων

A

Αλγόριθμοι εξόρυξης : περιγράφουν, αξιολογούν το μοντέλο

Αλγόριθμος ID3:

Αλγόριθμος που χρησιμοποιείται

Για την παραγωγή ενός δέντρου απόφασης.

Ακρίβεια:

Βαθμός συμφωνίας μετρήσεων που

γίνονται με διαφορετικές μεθόδους.

Αναλυτική βαθμολογία:

Η βαθμολογία με την οποία αξιολογείται η επίδοση σε κάθε μάθημα.

B

Βιβλίο μητρώου:

Επίσημο βιβλίο του σχολείου

Στο οποίο καταχωρούνται τα προσωπικά στοιχεία των μαθητών.

Γ

Δ

Δεδομένα εισόδου:

Ένα σύνολο εγγραφών

Δηλωτική γνώση:

(Declarative Knowledge) σύνολο

γενικών πληροφοριών και

τεκμηριωμένων γνώσεων για ιδέες ,

αντικείμενα , γεγονότα.

K

Κανόνες συσχέτισης:

(association rules) Ανακάλυψη κρυμμένων

Συσχετίσεων μεταξύ των γνωρισμάτων ενός

Συνόλου δεδομένων.

Κατηγοριοποίηση:

Εργασία εκμάθησης μιας συνάρτησης

στόχου F η οποία απεικονίζει κάθε

σύνολο χαρακτηριστικών σε μια από

Διακριτοποίηση δεδομένων:

Ο μετασχηματισμός ενός συνεχούς χαρακτηριστικού σε κατηγορικό.

Διαδικαστική γνώση:

(procedural Knowledge)

Η γνώση που απαιτείται για

να εκτελέσει το άτομο διαδοχικές

δραστηριότητες.

Δίκτυα Bayes:

Κατηγοριοποιητές που

κάνουν αποτίμηση πιθανοτήτων.

E

Εξόρυξη δεδομένων:

Ευφυείς μέθοδοι για ενδιαφέροντα πρότυπα γνώσης.

Εγγραφή:

Στιγμιότυπο, χαρακτηρίζεται

από (χ, ψ) όπου χ =σύνολο χαρακτηριστικών

ψ =ειδικό χαρακτηριστικό , ετικέτα.

Z

H

Θ

I

τις προκαθορισμένες ετικέτες.

Κλάση:

Κλίμακα Likert:

Κλίμακα εκτίμησης απόψεων ή

Συμπεριφοράς.

Λ

M

Μαθητικό προφίλ:

Μήτρα σύγχυσης:

Πίνακας στον οποίο καταγράφεται η απόδοση του μοντέλου κατηγοριοποίησης.

N

Νευρωνικά δίκτυα:

Χρησιμοποιούνται για την κατασκευή μοντέλου κατηγοριοποίησης -πρόβλεψης

Ξ

O

O.K.E: Οικονομική και κοινωνική Επιτροπή. Θεσμικό όργανο της Ευρωπαϊκής Ένωσης.

Π

Προεδρικό διάταγμα:

Θέσπισμα που εκδίδεται από τον πρόεδρο της Δημοκρατίας.

Προσανατολισμένο βέλος:

Στα δίκτυα Bayes βέλος που ορίζει την κατεύθυνση της σχέσης γονέας –παιδί

P

Σ

ΣΤ':

Έκτη Δημοτικού

Συνάψεις:

Σχισμές που λειτουργούν ως εξειδικευμένος χώρος επικοινωνίας ανάμεσα στις κομβικές απολήξεις των νευρώνων.

Σύνολο ελέγχου:

Εγγραφές με ετικέτες άγνωστες κατηγορίες

Συστημικές θεωρίες:

Θεωρίες που υποστηρίζουν πολλές κατηγορίες παραγόντων.

T

Y

Υποθετική γνώση:

Η ικανότητα του ατόμου να γνωρίζει πότε και γιατί θα χρησιμοποιήσει δηλωτική ή διαδικαστική

ΠΑΡΑΡΤΗΜΑ

Ερωτηματολόγιο

Ερωτηματολόγιο συνάφειας σχολικής επίδοσης και επιλογής επαγγέλματος

Το παρόν ερωτηματολόγιο δημιουργήθηκε στο πλαίσιο μεταπτυχιακής εργασίας του τμήματος πληροφοριακών συστημάτων του πανεπιστημίου Αιγαίου. Τα δεδομένα που θα συλλεγούν θα χρησιμοποιηθούν για την εξαγωγή χρήσιμης γνώσης σχετικά με τη σχολική επίδοση και το επάγγελμα.

Το ερωτηματολόγιο αποτελείται από ερωτήσεις και απαιτεί λιγότερο από 5 λεπτά να συμπληρωθεί, δεν είναι ονομαστικό και δεν απαιτείται η συμπλήρωση προσωπικών δεδομένων. Σας ευχαριστούμε εκ των προτέρων για τη συμμετοχή σας.

Φύλο

Ηλικία....

Επάγγελμα

Βάλτε **X** στην ομάδα την οποία ανήκει το επάγγελμά σας.



Ομάδα 1	Μέλη των βουλευομένων σωμάτων Ανώτερα διοικητικά και διευθυντικά στελέχη του ιδιωτικού και δημόσιου τομέα	
Ομάδα 2	Πρόσωπα που ασκούν επιστημονικά, καλλιτεχνικά και συναφή επαγγέλματα	
Ομάδα 3	Τεχνολόγοι, τεχνικοί βοηθοί	
Ομάδα 4	Υπάλληλοι γραφείου	
Ομάδα 5	Απασχολούμενοι στην παροχή υπηρεσιών και πωλητές	
Ομάδα 6	Ειδικευμένοι γεωργοί –κτηνοτρόφοι	
Ομάδα 7	Ειδικευμένοι τεχνίτες	
Ομάδα 8	Χειριστές σταθερών βιομηχανικών εγκαταστάσεων μηχανών	
Ομάδα 9	Ανειδίκευτοι εργάτες ,χειρωνάκτες	
Ομάδα 0	Ένοπλες δυνάμεις	

<p>-Σχολική επίδοση και ικανότητα(αυτορρύθμιση της γνώσης – διαμορφωτική αξιολόγηση-ανάδειξη ενδοατομικών διαφορών)-</p> <p>Παρακαλώ βάλτε X στη στήλη που σας εκφράζει</p>	Πολύ	συχνά	Κάποιες	Σπάνια-
--	-------------	--------------	----------------	----------------



		συχνά		φορές	ποτέ
1.	Η βαθμολογία μου στο δημοτικό:				
A.	Μου προκαλούσε άγχος				
B.	Ήταν ένα γεγονός που με χαροποιούσε				
Γ.	Ήταν ένα αδιάφορο γεγονός				
Δ.	Με βοηθούσε να βάζω στόχους				
Ε.	Με αποθάρρυνε από το να βάζω στόχους				
Στ.	Ήταν η αφορμή για καβγάδες στο σπίτι				
Z.	Ήταν η αφορμή για έπαινετικά σχόλια από τους γονείς				
H.	Είχε αδιάφορη αντιμετώπιση από το σπίτι				
Θ.	Με αδικούσε σε κάποια μαθήματα				
I.	Υπερτιμούσε τις ικανότητές μου σε κάποιο μάθημα				
K.	Στο μάθημα που είχα την υψηλότερη βαθμολογία ήμουν πραγματικά καλός				
Λ.	Με βοήθησε να καταλάβω που ήμουν πραγματικά καλός				
M.	Μου καλλιέργησε φιλοδοξίες για την επαγγελματική μου επιλογή				
N.	Με αποθάρρυνε από φιλοδοξίες για την επαγγελματική μου επιλογή				



Ε.	Δεν έπαιξε κανένα ρόλο στην επιλογή επαγγέλματος				
----	--	--	--	--	--

	Επάγγελμα -σχολική επίδοση	Πολύ	Λίγο	Ελάχιστα	Καθόλου
2.	Στην επιλογή του επαγγέλματος μου ήταν καθοριστική				
Α.	Η παρότρυνση της οικογένειάς μου.				
Β.	Οι επιδόσεις μου σε ένα μάθημα .				
Γ.	Η τελική μου βαθμολογία				
Δ.	Η παρότρυνση ενός άλλου προσώπου				
Ε.	Συγκυρίες				
Στ.					

	Επάγγελμα - εκπλήρωση στόχων	Πολύ	Λίγο	Ελάχιστα	Καθόλου
3.	Στο επάγγελμά μου				
	Νιώθω δημιουργικός				
	Νιώθω χαρούμενος				
	Μου προκαλεί άγχος				
	Μου προκαλεί ανασφάλεια				
	Νιώθω ότι κατέκτησα τους στόχους μου				
	Είμαι καλός σε αυτό που κάνω				
	Εκμεταλεύομαι τις ικανότητές που έχω				



	Έχω ικανότητες που μένουν ανεκμετάλλευτες				
--	---	--	--	--	--

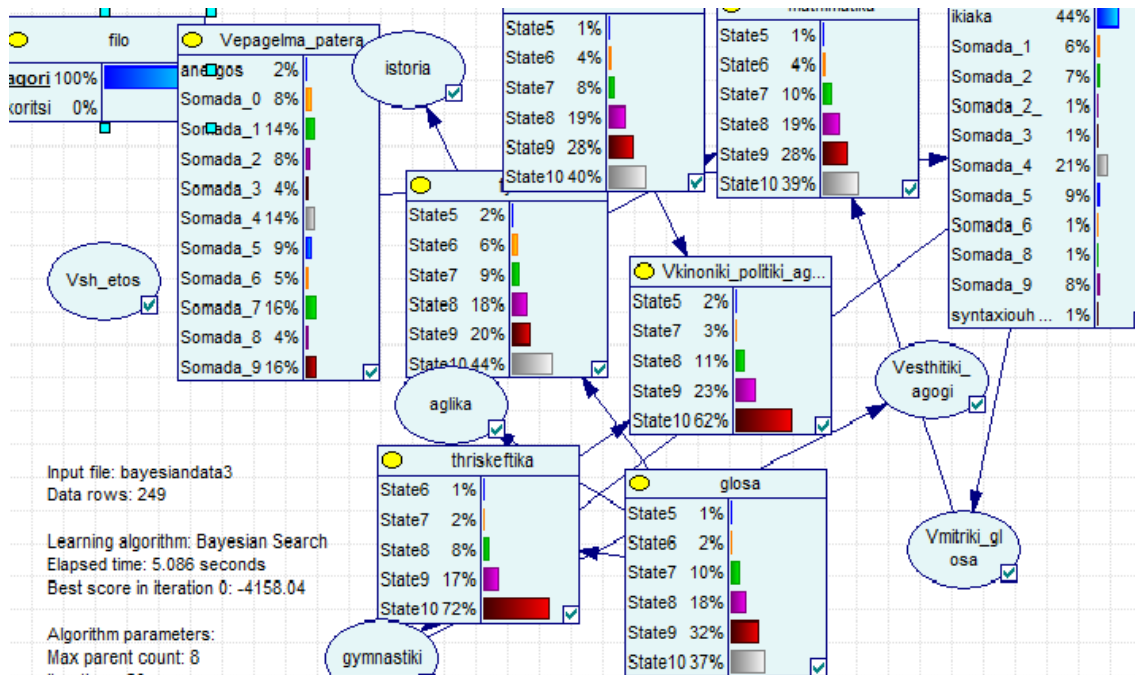
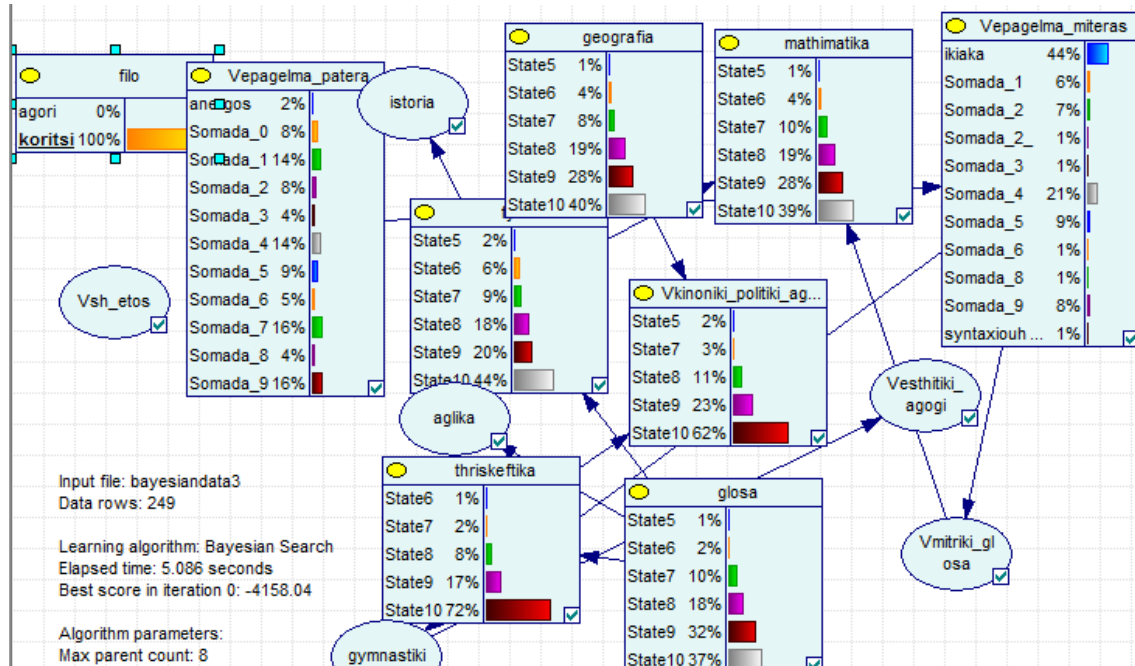
	Μαθητικό προφίλ	Πολύ	Λίγο	Ελάχιστα	Καθόλου
4.	Σαν μαθητής				
A.	Κατέβαλα μεγάλη προσπάθεια στην προετοιμασία				
B.	Κατέβαλα μικρή προσπάθεια στην προετοιμασία				
Γ.	Χρειάζομαι βοήθεια				
Δ.	Είχα βοήθεια				
Ε.	Δεν είχα καμιά βοήθεια				
Στ.	Είχα μαθησιακές δυσκολίες				
Z.	Μου άρεσε η Γλώσσα				
Η.	Μου άρεσαν τα μαθηματικά				
Θ.	Υστερούσα στη Γλώσσα				
Ι.	Υστερούσα στα Μαθηματικά				
Κ.	Ήμουν καλός στη ζωγραφική				



Τίτλος Μελέτης: Επάγγελμα γονιών, σχολική επίδοση, επιλογή επαγγέλματος

< Μαλιάριτση Διονυσία Μαρία >

Οι επισυναπτόμενοι πίνακες που δείχνουν τις πιθανότητες της βαθμολογίας όταν απομονώνεται η μεταβλητή φύλο. (Δίκτυα Bayes) Μοντέλο B





Μοντέλο Α Γραφική απεικόνιση του ερωτηματολογίου στα Δίκτυα Bayes

