



ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ (Τ.Ε.Ι.) ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

Τμήμα Διοίκησης Επιχειρήσεων (Πάτρα)

---

**Πτυχιακή Εργασία**

# **Εφαρμογές Υπολογιστικής Νοημοσύνης στις Γνωστικές Επιστήμες**

**(Μελέτη περίπτωσης: Συσχέτιση Χαρακτηριστικών Προσωπικότητας και  
Συναισθηματικής Νοημοσύνης)**

**Association Rules into social Networks users. Basic concepts in  
Emotional Intelligence and Personality Disorders' Occurrence**

**Μαργαρίτης Σωτήριος**

**Θεολόγος Κωνσταντίνος**

**Σπύρου Νικολέτα**

Υπό την επίβλεψη

**Κωνσταντίνος Χαλκιάπουλος**

**Πάτρα 2016**

## *Ευχαριστίες*

*Πρώτα, θέλουμε να ευχαριστήσουμε τον καθηγητή και επιβλέπον κ. Χαλκιάπουλο Κωνσταντίνο για την εμπιστοσύνη που μας έδειξε καθώς και για την καθοδήγηση καθ' όλη την περίοδο της ετοιμασίας αυτής της πτυχιακής εργασίας.*

*Θερμά ευχαριστώ στο καθηγητή κ. Αντζουλάτο Γεράσιμο για τις πολύτιμες συμβουλές-καθοδηγήσεις τόσο στο κείμενο, όσο και στο τεχνικό κομμάτι των αλγορίθμων αλλά και για τις επισημάνσεις σε διάφορα θέματα ακαδημαϊκής φύσεως. Καθώς επίσης και στον σπουδαστή του τμήματος ΕΠΔΟ Παναγιώτη Τόγια που μας πρόσφερε πολύτιμη βοήθεια στον εκτελεστικό τομέα της διατριβής μας.*

*Τέλος, ευχαριστούμε ιδιαίτερα τους φίλους και την οικογένειά μας για την στήριξη και την κατανόηση που επέδειξαν όλο αυτό τον καιρό.*

## ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας πτυχιακής είναι η μελέτη Γνωστικών Διεργασιών της Συναισθηματικής Νοημοσύνης Συσχετισμένες με Πτυχές της Προσωπικότητας και Συμπεριφοράς Φοιτητών και Φοιτητριών των Ελληνικών Πανεπιστημίων και Τεχνολογικών Εκπαιδευτικών Ιδρυμάτων της Χώρας. Έχει ως στόχο την εξόρυξη γνώσης, που προκύπτει βάσει των αποτελεσμάτων, μέσα από τη συμπλήρωση σταθμισμένων ψυχομετρικών κλιμάκων. Εφαρμόζοντας κατάλληλες τεχνικές εξόρυξης δεδομένων (data mining) πάνω σε ψυχομετρικές κλίμακες μέσω του λογισμικού πακέτου ανοιχτού στατιστικού κώδικα R, χρησιμοποιώντας αλγορίθμους Ανάλυσης Συσχέτισης, Κατηγοριοποίησης και Απομονωμένων Τιμών παράγονται κανόνες γνώσης που με την βοήθεια κατάλληλης ερμηνείας προκύπτουν ασφαλή συμπεράσματα. Τα αποτελέσματα που προέκυψαν, μετά από στάθμιση και βάση κριτηρίων, παρουσιάζουν τον βαθμό συσχέτισης των δύο ερωτηματολογίων της έρευνας μας καθώς και την συμπεριφορά των φοιτητών στο περιβάλλον τους.

Στο πρώτο κεφάλαιο αναλύονται εισαγωγικές έννοιες των Γνωστικών Επιστημών όπως η Υπολογιστική Νοημοσύνη και η Εξόρυξη γνώσης. Στο δεύτερο κεφάλαιο γίνεται αναφορά στην Ψυχομετρία αλλά και στα είδη των ψυχομετρικών εργαλείων. Στο τρίτο κεφάλαιο αναλύουμε τα ερωτηματολόγια Διαταραχών Προσωπικότητας & Συναισθηματικής Νοημοσύνης. Στην συνέχεια και συγκεκριμένα στο τέταρτο κεφάλαιο γίνεται αναφορά στους κανόνες Συσχέτισης, Κατηγοριοποίησης καθώς και στην Εύρεση των απομονωμένων τιμών αλλά και στον τρόπο λειτουργίας τους. Τέλος, στο πέμπτο κεφάλαιο αναλύονται με χρήση των προαναφερθέντων τεχνικών εξόρυξης γνώσης και εξάγονται αποτελέσματα και συμπεράσματα από τα δεδομένα που αντλήθηκαν ύστερα από την συμπλήρωση σταθμισμένων ερωτηματολογίων.

**Λέξεις-Κλειδιά:** Υπολογιστική Νοημοσύνη, Εξόρυξη Γνώσης, Συσχέτιση, Απομονωμένες Τιμές, Κατηγοριοποίηση, Ψυχομετρία.

## **ABSTRACT**

The main purpose of this thesis is the study of Cognitive Processes associated with Emotional Intelligence (EI) aspects of personality and students behavior. Students from all the Greek Universities and Institutes of the country took part in. This study aims at extracting knowledge which derived based on results through the completion of weighted psychometric scales. Applying these appropriate data mining techniques on psychometric scales through the open source R statistical software package, by using classification, association and Outlier Detection algorithms. They produced knowledge rules with the appropriate interpretation, resulting more secure conclusions. The results after weighing and criteria basis showing the degree of correlation between these questionnaires of our research and the behavior of the students in their environment.

In the first chapter analyzed the concepts and principles of Cognitive Sciences as Computational Intelligence and knowledge extraction. Furthermore the second chapter refers to psychometrics questionnaires of Disorders Personality & Emotional Intelligence. As follows the finding of outliers and the way they operated. In the end the fifth chapter analyzes by using the aforementioned data mining techniques and exported results and conclusions from the data which learned after the completion of weighted questionnaires.

**Keywords:** Computational Intelligence, Data Mining, Association, Outliers, Classification, Psychometrics

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

<b>1<sup>ο</sup> ΚΕΦΑΛΑΙΟ</b> .....	8
<b>1 Εισαγωγή στη Γνωστική Επιστήμη (Introduction in Cognitive Science)</b> .....	8
1.1 Ορισμός.....	8
1.2 Γνωστική Νευροεπιστήμη .....	8
1.3 Νευρωνικά Σύστοιχα Συνείδησης.....	9
1.3.1 Νοητικά Μοντέλα .....	10
1.4 Δεδομένα, Πληροφορία & Γνώση .....	11
1.5 Υπολογιστική Νοημοσύνη.....	13
1.5.1 Ορισμός.....	13
1.5.2 Κατηγορίες και Εφαρμογές Υπολογιστικής Νοημοσύνης.....	13
1.6 Ανάλυση Εξόρυξη Γνώσης (Data Mining).....	14
1.7 Ανακάλυψη Γνώσης Στις Βάσεις Δεδομένων - KDD (Knowledge Discovery Databases).....	15
1.7.1 Χρησιμότητα KDD.....	17
<b>2<sup>ο</sup> ΚΕΦΑΛΑΙΟ</b> .....	19
<b>2 Εισαγωγή στην Ψυχομετρία (Introduction in Psychometrics)</b> .....	19
2.1 Ορισμός.....	19
2.2 Μέτρηση και Ψυχολογία.....	19
2.2.1 Μέτρηση Ψυχολογικών Εννοιών .....	20
2.2.2 Στάδια στην Υλοποίηση μιας Ψυχομετρικής Δοκιμασίας .....	20
2.2.3 Ακρίβεια της Μέτρησης.....	21
2.2.4 Πηγές Σφαλμάτων Μέτρησης.....	22
2.3 Είδη Ψυχομετρικών Εργαλείων και Αξιολόγησης.....	23
2.3.1 Αξιοπιστία Εργαλείων .....	26
2.3.2 Παράγοντες που Επηρεάζουν την Αξιοπιστία ενός Εργαλείου .....	28
2.3.3 Εγκυρότητα Εργαλείων.....	29
<b>3<sup>ο</sup> ΚΕΦΑΛΑΙΟ</b> .....	32
<b>3 Ερωτηματολόγια Διαταραχών Προσωπικότητας &amp; Συναισθηματικής Νοημοσύνης (IPDE &amp; Emotional Intelligence Questionnaires)</b> .....	32
3.1 Εισαγωγή στη Συναισθηματική Νοημοσύνη .....	32
3.1.1 Κατηγορίες Συναισθηματικής Νοημοσύνης .....	34
3.1.2 Αυτεπίγνωση.....	35

3.1.3	Αυτοδιαχείριση .....	35
3.1.4	Κοινωνική Επίγνωση .....	36
3.1.5	Διαχείριση Σχέσεων .....	36
3.1.6	Διαστάσεις Συναισθηματικής Νοημοσύνης.....	37
3.1.7	Παράγοντες Μέτρησης Ερωτηματολογίου Συναισθηματικής Νοημοσύνης ...	38
3.2	Ερωτηματολόγιο Διαταραχών Προσωπικότητας (IPDE) .....	39
3.2.1	Ιστορικά Στοιχεία.....	40
3.2.2	Δομή.....	41
3.2.3	Παράγοντες Μέτρησης Ερωτηματολογίου Διαταραχών Προσωπικότητας.....	42
<b>4° ΚΕΦΑΛΑΙΟ</b>	.....	<b>44</b>
<b>4</b>	<b>Εισαγωγή στις Τεχνικές Εξόρυξης Γνώσης (Introduction to Data Mining Techniques).....</b>	<b>44</b>
4.1	Ορισμός –Έννοιες Data Mining.....	44
4.2	Εφαρμογές Data Mining .....	45
4.3	Βασικά Στάδια Της Εξόρυξης Δεδομένων .....	47
4.3.1	Περιγραφή Μοντέλου .....	47
4.3.2	Αξιολόγηση Μοντέλου .....	48
4.3.3	Αλγόριθμος Αναζήτησης .....	48
4.3.4	Διαχωρισμός Μεθόδων Εξόρυξης Δεδομένων .....	48
4.4	Περιγραφική Μοντελοποίηση (Descriptive Modeling) .....	49
4.4.1	Ορισμός.....	49
4.4.2	Κανόνες Συσχέτισης (Association Rules).....	49
4.4.2.1	Αλγόριθμος Apriori.....	51
4.4.2.2	Μέτρα Σημαντικότητας.....	52
4.5	Μοντελοποίηση Πρόβλεψης (Predictive Modeling).....	53
4.5.1	Κατηγοριοποίηση (Classification).....	54
4.6	Εύρεση Απομονωμένων τιμών (Outlier Detection).....	55
4.6.1	Εφαρμογές Εύρεσης Απομονωμένων Τιμών .....	57
4.7	Αλγόριθμος LOF (Local Outlier Factor) .....	59
<b>5° ΚΕΦΑΛΑΙΟ</b>	.....	<b>60</b>
<b>5</b>	<b>Αποτελέσματα.....</b>	<b>60</b>
5.1	Προεπεξεργασία δεδομένων .....	60
5.2	Apriori.....	61
5.3	Δέντρα Αποφάσεων .....	71

5.4	Εύρεση Απομονωμένων Τιμών (Outlier Detection) .....	77
5.4.1	Περιγραφική Στατιστική των Απομονωμένων Περιπτώσεων .....	78
6	Συμπεράσματα.....	81
	Βιβλιογραφία .....	83
	Παράρτημα .....	85

## ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

Εικόνα 1	Βήματα διαδικασίας KDD .....	16
Εικόνα 2	Μέθοδοι τους οποίους διαχειρίζεται η Συναισθηματική Νοημοσύνη .....	33
Εικόνα 3	Συναισθηματικές Δεξιότητες .....	35

## ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1	Διάγραμμα Διασποράς Κανόνων Δείγματος Ανδρών.....	63
Σχήμα 2	Διάγραμμα Συσχέτισης Κανόνων Δείγματος Ανδρών .....	63
Σχήμα 3	Διάγραμμα Διασποράς Κανόνων Δείγματος Γυναικών .....	65
Σχήμα 4	Διάγραμμα Συσχέτισης Κανόνων Δείγματος Γυναικών.....	66
Σχήμα 5	Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Δυτικής Ελλάδας .....	67
Σχήμα 6	Διάγραμμα Συσχέτισης Κανόνων Δείγματος Τει Δυτικής Ελλάδας.....	68
Σχήμα 7	3D Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Δυτικής Ελλάδας.....	68
Σχήμα 8	Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Ιονίων Νήσων.....	70
Σχήμα 9	Διάγραμμα Συσχέτισης Κανόνων Δείγματος Τει Ιονίων Νήσων.....	70
Σχήμα 10	Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Ιονίων Νήσων.....	71
Σχήμα 11	Διάγραμμα Πυκνότητας Πιθανοτήτων .....	77
Σχήμα 12	Ανάλυση Κύριων Συνιστωσών(PCA) .....	79
Σχήμα 13	Διάγραμμα Διασποράς Απομονωμένων Τιμών Emotionality-Avoidant.....	80
Σχήμα 14	Διάγραμμα Διασποράς Απομονωμένων Τιμών Sociability-Emotionalit	80

# 1<sup>ο</sup> ΚΕΦΑΛΑΙΟ

---

## 1 Εισαγωγή στη Γνωστική Επιστήμη (Introduction in Cognitive Science)

### 1.1 Ορισμός

Γνωστική επιστήμη (cognitive science) ονομάζουμε τη διεπιστημονική μελέτη του νου. Στην πραγματικότητα δεν υπάρχει μια ενιαία «γνωστική επιστήμη» υπάρχουν περισσότερες από μία. Δηλαδή, διαφορετικοί κλάδοι που συνεισφέρουν στην προσπάθεια περιγραφής και κατανόησης των νοητικών λειτουργιών. Στις γνωστικές επιστήμες συγκαταλέγονται και συνυπάρχουν πολλοί επιστημονικοί κλάδοι όπως γνωστική ψυχολογία, οι νευροεπιστήμες, η γλωσσολογία, η φιλοσοφία και η τεχνητή νοημοσύνη.

Υποστηρίζουν ότι ο νους είναι ένα είδος μηχανής που υλοποιείται από τον εγκέφαλο. Ο νους ταυτίζεται με την εγκεφαλική λειτουργία, αυτό σημαίνει πως όλες οι όψεις του νου πρέπει να ταυτίζονται με κάποιες όψεις της εγκεφαλικής λειτουργίας. Αυτό που θέλουμε να τονίσουμε σήμερα δεν είναι το «ταυτίζονται» αλλά το «όλες». Στο χώρο της γνωστικής επιστήμης έχει δοθεί έμφαση σε ένα πολύ μικρό υποσύνολο των νοητικών λειτουργιών, εστιάζοντας σε ορισμένα ζητήματα από το χώρο της γνωστικής ψυχολογίας και της ψυχολογίας: Αντίληψη, μνήμη και μάθηση, ομιλία και ανάγνωση, επίλυση προβλημάτων, λήψη αποφάσεων. (Πρωτόπαπας, 2008)

### 1.2 Γνωστική Νευροεπιστήμη

Η γνωστική νευροεπιστήμη αποτελεί την πιο πρόσφατη εξέλιξη στο χώρο των γνωσιακών επιστημών. Παράγει δεδομένα χωρικού και χρονικού εντοπισμού των εγκεφαλικών λειτουργιών, τα οποία συνδυάζονται με τα αποτελέσματα των συμπεριφορικών μελετών.



Η νευροαπεικόνιση δίνει σήμερα στοιχεία τα οποία συνδέουν εγκεφαλικά παρατηρήσιμα με συμπεριφορικά παρατηρήσιμα και με φαινομενολογικές αναφορές. Μπορούμε επίσης να καταγράψουμε τα στοιχεία αυτά. Για παράδειγμα, πότε έγινε αντιληπτή η λέξη μέσα από ένα θορυβώδες οπτικό ή ακουστικό περιβάλλον, ή πότε άλλαξε χρώμα ή ύψος φωνής. Η διερεύνηση των σχέσεων μεταξύ συμπεριφοράς, υποκειμενικής εμπειρίας, και εγκεφαλικής δραστηριότητας, αποτελεί το αντικείμενο της γνωστικής νευροεπιστήμης. Όταν πρόκειται για σχέσεις μεταξύ υποκειμενικής εμπειρίας και εγκεφαλικών καταγραφών, τότε μιλάμε για τα «*νευρωνικά σύστοιχα της συνείδησης*». Με τον τρόπο αυτό επιχειρούμε να αποκαλύψουμε τους εγκεφαλικούς μηχανισμούς που παράγουν την επίγνωση ή συνειδητοποίηση των περιεχομένων της υποκειμενικής ατομικής εμπειρίας.

### 1.3 Νευρωνικά Σύστοιχα Συνείδησης

Από το πρόγραμμα των νευρωνικών σύστοιχων της συνείδησης απουσιάζει συνήθως το βασικότερο συστατικό που είναι ο συνειδητός εαυτός, με τις ιδιότητες που του αποδίδονται από τους φιλοσόφους όπως είναι η ιδιοκτησία του συνειδητού περιεχομένου, η υποκειμενικότητα και η προοπτικότητα. Η απουσία αυτή ίσως δεν είναι τόσο εμφανής όταν το αντικείμενο της μελέτης είναι η οπτική αντίληψη. Εκεί θα μπορούσε κανείς να υποθέσει ότι αρκεί, καταρχήν να εντοπίσει τις επιμέρους συσχετίσεις. Για παράδειγμα τις συσχετίσεις μεταξύ της συνειδητοποίησης του χρώματος του ερεθίσματος και της δραστηριότητας των εγκεφαλικών δικτύων που είναι υπεύθυνες για την επεξεργασία, τον υπολογισμό και τη συνειδητοποίηση του χρώματος. Άρα, φαινόμενα όπως η βούληση, η εικόνα του σώματος και οι συναισθηματικές καταστάσεις θα πρέπει να ενταχθούν σε κοινό επεξηγηματικό πλαίσιο με τα πιο συνηθισμένα αντικείμενα μελέτης της γνωστικής ψυχολογίας όπως είναι: η αντίληψη, η προσοχή, η μνήμη και η γλώσσα. Η επιδιωκόμενη συνέπεια στην αντιμετώπιση των φαινομένων επιβάλλει τη χρήση των ενδοσκοπικών παρατηρήσεων στη μελέτη του νου.

Η **ενδοσκόπηση** είναι προφανώς νοητική λειτουργία, η ύπαρξή της δεν αμφισβητείται από κανέναν. Ως υποκειμενικές και ατομικές, δεν είναι δυνατό να επιβεβαιωθούν. Υπάρχουν όμως δύο σημαντικοί λόγοι να αναθεωρήσουμε αυτή τη

θέση και να συμπεριλάβουμε, με την αρμόζουσα μεθοδολογική αυστηρότητα, μια φαινομενολογική πηγή δεδομένων στη σύγχρονη μελέτη του νου.

- Ο πρώτος λόγος έχει να κάνει με τη διατομική συστηματικότητα:

Αν κάποιες ενδοσκοπικές παρατηρήσεις οδηγούν σε όμοια δεδομένα από πολλούς ανθρώπους, τότε έχουμε αξιοπιστία των παρατηρήσεων. Μάλιστα πρόκειται για ίδιου είδους αξιοπιστία με εκείνη οποιουδήποτε ψυχολογικού ή ψυχοφυσικού πειράματος.

- Ο δεύτερος λόγος αφορά στην ουσία της ενδοσκόπησης.

Εφόσον πρόκειται για μια αναμφισβήτητα υπαρκτή νοητική λειτουργία, κι εφόσον επιδιώκουμε την περιγραφή όλων των νοητικών λειτουργιών, είμαστε υποχρεωμένοι να συμπεριλάβουμε ως αντικείμενο μελέτης και την περιγραφή της ενδοσκόπησης, λαμβάνοντας τα φαινομενολογικά δεδομένα ως στοιχεία προς εξήγηση. Προς την κατεύθυνση αυτή υπάρχουν ήδη κάποιες μεθοδολογικές προτάσεις, όπως η νευροφαινομενολογία, η ετεροφαινομενολογία, με στόχο τη συστηματοποίηση και αξιοποίηση των υποκειμενικών παρατηρήσεων πρώτου προσώπου στη μελέτη του νου.

### 1.3.1 Νοητικά Μοντέλα

Τα τελευταία χρόνια παρατηρούμε ότι υπάρχει μεγάλο ενδιαφέρον για τη μελέτη των συναισθημάτων και των συγκινησιακών καταστάσεων γενικότερα, για την πρόθεση και τη βούληση, καθώς και για το ρόλο του σώματος στη νόηση. Τα φαινόμενα και ζητήματα αυτά που προσελκύουν το ενδιαφέρον παλαιών και νέων ερευνητών γίνονται εύκολα αποδεκτά. Πέρα από την κοινή αντίληψη ότι ο νους είναι μια μηχανή, οι ποικίλες γνωσιακές επιστήμες ενώνονται από την κοινή αντίληψη ότι για να μελετηθεί η μηχανή αυτή είναι χρήσιμο να κατασκευάζουμε τεχνητές μηχανές επιχειρώντας να προσομοιώσουμε τις νοητικές λειτουργίες. Με δυο λόγια, είναι κεντρική μεθοδολογική προσέγγιση των γνωσιακών επιστημών η κατασκευή και μελέτη νοητικών μοντέλων.

Τα νοητικά μοντέλα είναι υπολογιστικά συστήματα, συνήθως υλοποιημένα σε λογισμικό, με τα οποία εξετάζουμε λεπτομερειακά τις συνέπειες των θεωρητικών μας υποθέσεων. Χάρη στα μοντέλα υποχρεώνονται οι ερευνητές να κάνουν ρητές και πλήρεις παραδοχές ώστε να είναι εφικτή η υλοποίηση πραγματικών συστημάτων, λαμβάνοντας υπόψη τεχνικούς περιορισμούς. Βασικό πεδίο γόνιμης και εποικοδομητικής διαμάχης στις γνωσιακές επιστήμες αποτελεί το είδος της υπολογιστικής μηχανής που είναι κατάλληλο για την προσομοίωση των νοητικών λειτουργιών. Οι δυο βασικές αντίπαλες προσεγγίσεις είναι η συμβολική, που εμπνέεται από την αρχιτεκτονική και τη λειτουργική δομή των ψηφιακών υπολογιστών, και η συνδεδιστική, που εμπνέεται αντίστοιχα από τα βιολογικά νευρωνικά δίκτυα και δίνει έμφαση στη στατιστική. Τόσο η συμβολική όσο και η συνδεδιστική πρόταση είναι αναπαραστασιακές προσεγγίσεις.

Οι νευροεπιστήμονες, που δεν περιορίζονται ιδιαίτερα από θεωρητικές καταβολές, έχουν ήδη ξεκινήσει να αναζητούν και να χαρτογραφούν υποκειμενικές καταστάσεις και ιδιότητες στον εγκέφαλο μέσω της εικόνας του σώματος, της κοινωνικής νόησης και των συναισθημάτων. Οι θεωρητικοί των νοητικών λειτουργιών οφείλουν να ακολουθήσουν τις εξελίξεις, όχι μόνο για να διασώσουν τα επαγγελματικά τους δικαιώματα, αλλά επειδή είναι απαραίτητη η μεθοδολογική και θεωρητική συνεισφορά τους.

Στο πλαίσιο αυτό, τα μελλοντικά μοντέλα του νου θα πρέπει να συμπεριλάβουν υποκειμενικές καταστάσεις (βουλητικές, συγκινησιακές, ενδοσκοπικές) έτσι ώστε να είναι ευδιάκριτη σε αυτές η συγκρότηση του υποκειμένου – διαρκούς ή συγκυριακού. (Πρωτόπαππας, 2008)

### 1.4 Δεδομένα, Πληροφορία & Γνώση

Τα δεδομένα αφορούν συγκεκριμένες καταστάσεις ή οντότητες, συγκεκριμένα γεγονότα. Τα δεδομένα αποθηκεύονται σε βάσεις δεδομένων οι οποίες περιέχουν μεγάλο όγκο στοιχείων, όπως για παράδειγμα μία σχεσιακή βάση δεδομένων. Τα δεδομένα είναι στατικά, ενώ η πληροφορία είναι δυναμική. Παράγεται με δυναμικό τρόπο από το τρέχον σύνολο δεδομένων.

Η πληροφορία, από την άλλη πλευρά, δεν παύει να είναι συγκεκριμένη, έστω και αν είναι σε πιο υψηλό επίπεδο αφαιρετικότητας από τα δεδομένα. Η πληροφορία απορρέει από ένα συγκεκριμένο σύνολο δεδομένων, το οποίο αποτελεί την εμβέλεια της αλήθειας της (truth scope).

Η γνώση αποτελείται από γενικεύσεις, που καλύπτουν τα υπάρχοντα σχετικά δεδομένα, καθώς επίσης και μελλοντικά δεδομένα. Επομένως, η γνώση βρίσκεται σε υψηλότερο επίπεδο αφαιρετικότητας από την πληροφορία. Η εμβέλεια της αλήθειας της γνώσης πρέπει να είναι καθολική και όχι απλώς ένα συγκεκριμένο σύνολο δεδομένων.

Η γνώση είναι υπεράνω των δεδομένων, αφού η γνώση εξηγεί τα δεδομένα, δεν απορρέει από αυτά (όπως η πληροφορία η οποία τα συνοψίζει). Η γνώση, κυρίως η εμπειρική, είναι δυναμική όπως και η πληροφορία και έγκειται σε αναθεωρήσεις. Οι «καθολικές αλήθειες» (universal truths), που αποτελούν τη γνώση, δεν είναι στατικές, αλλά υπόκεινται σε μεταβολές, για να έλθουν πιο κοντά στην πραγματική αλήθεια. Ουσιαστικά το τι εννοούμε με τον όρο «γνώση» είναι «πεποιθήσεις» (beliefs) σε γενικό επίπεδο. Οι πεποιθήσεις είναι μεταβαλλόμενες και επομένως έχουν πεπερασμένη χρονική υπόσταση, ενώ θεωρητικά η πραγματική γνώση είναι διαχρονική. Από πρακτικής απόψεως, εάν συγκρίνουμε βάσεις δεδομένων με «βάσεις γνώσης» (knowledge bases) συστημάτων **Τεχνητής Νοημοσύνης** θα παρατηρήσουμε ότι οι βάσεις δεδομένων είναι πολύ μεγαλύτερες σε μέγεθος και αυτό είναι αυτονόητο αφού η γνώση είναι σε πολύ υψηλότερο επίπεδο αφαιρετικότητας.

Όμως οι βάσεις γνώσης είναι πλουσιότερες σε δομή. Συνήθως, οι διακρίσεις που χρειάζεται να γίνουν σε μία βάση δεδομένων, είναι απλές σε σχέση με μία βάση γνώσης όπου υπάρχουν πολλές και περίπλοκες σημασιολογικές διακρίσεις. Συνοψίζοντας, από τα δεδομένα απορρέουν πληροφορίες, από τις οποίες μπορεί να γεννηθεί γνώση. Πρόσφατα έχει αναπτυχθεί το πεδίο της εξόρυξης δεδομένων (data mining), στόχος του οποίου είναι η ανάλυση μεγάλων ποσοτήτων δεδομένων με σκοπό τον εντοπισμό προτύπων με τελικό σκοπό την ανακάλυψη γνώσης (Κεραυνού, 2000).

## 1.5 Υπολογιστική Νοημοσύνη

### 1.5.1 Ορισμός

Παρότι δεν υπάρχει κοινά αποδεκτός ορισμός, η Υπολογιστική Νοημοσύνη μπορεί να θεωρηθεί ως η μελέτη των μηχανισμών που κάνουν δυνατή την ευφυή συμπεριφορά ενός συστήματος σε ένα πολύπλοκο και μεταβαλλόμενο περιβάλλον. Υπολογιστική Νοημοσύνη είναι ένα σύνολο εμπνευσμένων από την φύση υπολογιστικών μεθοδολογιών και προσεγγίσεων για την αντιμετώπιση πολύπλοκων προβλημάτων του πραγματικού κόσμου, στον οποίο παραδοσιακές μεθοδολογίες και προσεγγίσεις είναι αναποτελεσματικές ή ανέφικτες (Παντελή, 2012).

Ο ορισμός μπορεί να επεκταθεί ακόμα περισσότερο στην μελέτη σχεδιασμού ευφυών πρακτόρων. Ένας πράκτορας μπορεί να είναι οτιδήποτε το οποίο δρα σε ένα περιβάλλον. Στην κατηγορία των πρακτόρων μπορεί να ανήκει οτιδήποτε από αεροπλάνα, ανθρώπους και οργανισμούς μέχρι και ολόκληρη την κοινωνία. Ένας ευφυής πράκτορας είναι ένα σύστημα το οποίο δρα με ευφυή τρόπο, δηλαδή αυτό που κάνει είναι το ιδανικό για κάθε περίπτωση και τους στόχους αυτής, είναι ικανό να αλλάζει εύκολα περιβάλλον και στόχους, μαθαίνει από προηγούμενες εμπειρίες και τέλος κάνει τις απαραίτητες αλλαγές με βάση δοσμένα υπάρχοντα όρια και πεπερασμένους υπολογισμούς.

Ο βασικός επιστημονικός στόχος της υπολογιστικής νοημοσύνης είναι να καταλάβουμε τις βασικές αρχές που κάνουν την ευφυή συμπεριφορά δυνατή σε φυσικά και τεχνητά συστήματα. Η κύρια υπόθεση, εδώ, είναι ότι η λογική ισούται του υπολογισμού. Όσο αναφορά το επίπεδο προγραμματισμού, είναι ο προσδιορισμός μεθόδων για τον σχεδιασμό χρήσιμων ευφυών αντικειμένων (Poole, Mcworth, & Goebel, 1998).

### 1.5.2 Κατηγορίες και Εφαρμογές Υπολογιστικής Νοημοσύνης

Η υπολογιστική νοημοσύνη μπορεί να διακριθεί στις παρακάτω κατηγορίες:

- Τα Νευρωνικά Δίκτυα

- Οι Γενετικοί Αλγόριθμοι<sup>1</sup>
- Ο Εξελικτικός Υπολογισμός
- Ασαφής Λογική

Έχει παρουσιάσει μεγάλη εξέλιξη και οι μέθοδοι της έχουν εφαρμοστεί με μεγάλη επιτυχία σε προβλήματα προερχόμενα από διαφορετικούς επιστημονικούς τομείς όπως η ιατρική, η μηχανική, η οικονομική κ.ά. Στα πλεονεκτήματα των μεθόδων υπολογιστικής νοημοσύνης, συγκαταλέγεται το ότι είναι σχετικά απλές στην εφαρμογή τους, ότι είναι αυτοματοποιημένες σε μεγάλο βαθμό και ότι δεν απαιτούν σύνθετες μαθηματικές ιδιότητες για την λύση κάποιου προβλήματος, γεγονός που τις καθιστά κατάλληλες για επίλυση προβλημάτων που προέρχονται από πραγματικές εφαρμογές.

Μερικές εφαρμογές της υπολογιστικής νοημοσύνης είναι οι εξής: (Παντελή, 2012)

- Λήψη αποφάσεων σε περιβάλλον αβεβαιότητας
- Κόστος κύκλου ζωής
- Προγραμματισμός παραγωγής
- Αλγόριθμοι δρομολόγησης (fleet routing)

### 1.6 Ανάλυση Εξόρυξη Γνώσης (Data Mining)

Το πεδίο *εξόρυξης γνώσης* (Data Mining) μπορεί να αναλυθεί με δύο έννοιες. Η πρώτη υφίσταται ως “Η σύνθετη διαδικασία εξαγωγής συγκεκριμένης, προηγουμένως άγνωστης και δυνητικά ωφέλιμης γνώσης από δεδομένα” και ως “Η επιστήμη της εξόρυξης χρήσιμης πληροφορίας από σύνολα ή βάσεις δεδομένων μεγάλου μεγέθους. Η σύγκληση της προόδου υπολογιστικών συστημάτων και τις εξέλιξης στην επικοινωνία έχει οδηγήσει στην δημιουργία μιας κοινωνίας ικανής να παρέχει διαρκώς νέες πληροφορίες. Το υλικό που συγκεντρώνεται καταγράφεται διαρκώς με αποτέλεσμα την δημιουργία τεραστίων βάσεων δεδομένων.

---

<sup>1</sup> Συμπεριλαμβάνουν συστήματα επίλυσης προβλημάτων βασισμένων στις αρχές της Γενετικής Εξέλιξης και της Κληρονομικότητας. Τα συστήματα αυτά διατηρούν ένα πληθυσμό κωδικοποιημένων πιθανών λύσεων και εφαρμόζουν σε αυτό διαδικασίες επιλογής του καλύτερου και γενετικού τελεστής. Αντιγράφουν τον τρόπο με τον οποίο αναπαράγονται και μεταλλάσσονται τα χρωμοσώματα των κυττάρων των ζωντανών οργανισμών και δημιουργούν συνεχώς νέους πληθυσμούς πιθανών λύσεων χρησιμοποιώντας, κομμάτια από την προηγούμενη γενιά και την καινούρια.

Για παράδειγμα σε ένα σύστημα βάσεων δεδομένων μπορούν να καταγραφούν οι συναλλαγές που γίνονται σε ένα κατάστημα ή η χρήση πιστωτικών καρτών και δανείων από τους πελάτες μιας τράπεζας. Από την άλλη πλευρά υπάρχουν και πολυπλοκότερα ζητήματα που χρειάζεται να οργανωθούν μέσω μιας βάσης δεδομένων όπως θέματα ιατρικής φύσεως, φωτογραφίες από δορυφόρους και πειραματικά δεδομένα. Το ζήτημα λοιπόν που προκύπτει είναι αν υπάρχει τρόπος να διαχειριστούμε τις πολύ μεγάλες αυτές βάσεις δεδομένων που ανανεώνονται διαρκώς από τους χρήστες. Επίσης θεωρείται αρκετά δύσκολη η άντληση του απαραίτητου υλικού από αυτές.

Όλα αυτά τα θέματα προκάλεσαν το ενδιαφέρον και οδήγησαν στη διαδικασία της Εξόρυξης Δεδομένων (Data Mining). Πρόκειται για μια σειρά από τεχνικές που βασίζονται σε ανάπτυξη αλγορίθμων και είναι χρήσιμες σε πολλούς και ετερόκλητους κλάδους όπως οι : οικονομία, γεωλογία, βιοιατρική, μετεωρολογία κ.α. (Παπακωνσταντίνου, 2009)

### 1.7 Ανακάλυψη Γνώσης Στις Βάσεις Δεδομένων - KDD (Knowledge Discovery Databases)

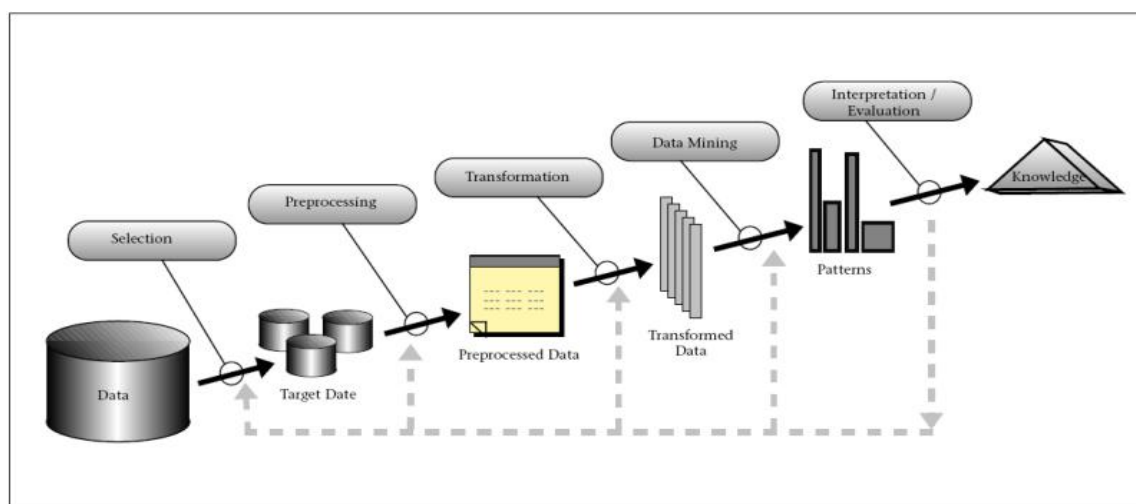
Σε μια ευρεία ποικιλία πεδίων, τα δεδομένα μπορούν να συλλέγονται και να συσσωρεύονται με δραματικό ρυθμό. Υπάρχει επείγουσα ανάγκη για μια νέα γενιά υπολογιστικών θεωριών και εργαλείων που θα βοηθήσει τους ανθρώπους στην εξόρυξη χρήσιμων πληροφοριών (γνώση) από τους ταχέως αναπτυσσόμενους όγκους ψηφιακών δεδομένων. Αυτές οι θεωρίες και τα εργαλεία είναι το αντικείμενο του αναδυόμενου αυτού πεδίου της ανακάλυψης γνώσης στις βάσεις δεδομένων (KDD).

Σε ένα θεωρητικό επίπεδο, το πεδίο των KDD ανησυχεί για την ανάπτυξη των μεθόδων και τις τεχνικές για την ερμηνεία των δεδομένων. Το βασικό πρόβλημα που αντιμετωπίζει η διαδικασία των KDD είναι ότι τα δεδομένα χαρτογράφησης χαμηλού επιπέδου (που είναι τυπικά πάρα πολύ ογκώδη για να κατανοηθούν και να αφομοιωθούν εύκολα) σε άλλες μορφές θα μπορούσαν να είναι πιο συμπαγή, πιο αφηρημένα (για παράδειγμα, μια περιγραφική προσέγγιση ή το μοντέλο της διεργασίας που δημιουργούνται τα δεδομένα), ή πιο χρήσιμα. Στον πυρήνα της

διαδικασίας είναι η εφαρμογή των συγκεκριμένων μεθόδων εξόρυξης δεδομένων για την ανακάλυψη ενός μοτίβου εξόρυξης. (Han & Kamber, 2001)

Τα βήματα της διαδικασίας των KDD είναι: (Han & Kamber, 2001)

1. Data cleaning: Καθαρισμός Δεδομένων (αφαίρεση θορύβου σε ασυνεπή δεδομένα).
2. Data integration: Ενοποίηση Δεδομένων (εκεί που οι πηγές πολλαπλών δεδομένων μπορούν να συνδυαστούν).
3. Data selection: Επιλογή Δεδομένων (εκεί που τα κοινά δεδομένα της ανάλυσης ανακτώνται από την βάση δεδομένων).
4. Data transformation: Μεταμόρφωση Δεδομένων (εκεί που τα δεδομένα μεταμορφώνονται ή παγιώνονται σε φόρμες επιτρεπτές για εξόρυξη εκτελώντας περίληψη ή συνάθροιση πράξεων).
5. Data mining: Εξόρυξη Δεδομένων (μια ουσιώδες επεξεργασία όπου έξυπνοι μέθοδοι χρησιμοποιούνται για να εξάγουν δεδομένα προτύπων).
6. Pattern evaluation: Αξιολόγηση Προτύπων (για να αναγνωρίσουμε τα πραγματικά ενδιαφέροντα πρότυπα γνώσης που παρουσιάζονται βασισμένα σε κάποιες ενδιαφέροντες μετρήσεις).
7. Knowledge presentation: Παρουσίαση Γνώσης (εκεί που η απεικόνιση και η γνώση αντιπροσωπεύουν τεχνικές που χρησιμοποιούνται για την παρουσίαση εξορυγμένης γνώσης στο χρήστη).



Εικόνα 1 Βήματα διαδικασίας KDD  
[Πηγή: Fayyad, Piatetsky-Shapiro, & Smyth ]



## 1.7.1 Χρησιμότητα KDD

Η παραδοσιακή μέθοδος της μετατροπής των δεδομένων σε γνώση στηρίζεται στη χειροκίνητη ανάλυση και ερμηνεία. Για παράδειγμα, στην υγειονομική περίθαλψη είναι σύνηθες για τους ειδικούς σε περιοδικά να αναλυθούν οι τρέχουσες τάσεις και αλλαγές στα δεδομένα του τομέα της υγείας σε τριμηνιαία βάση. Οι ειδικοί στη συνέχεια θα παρέχουν μια έκθεση σχετικά με την ανάλυση για την οργάνωση σε χορηγίες για την υγεία. Η έκθεση αυτή γίνεται η βάση για την μελλοντική λήψη αποφάσεων και τον σχεδιασμό-διαχείριση της υγειονομικής περίθαλψης. Σε ένα εντελώς διαφορετικό τύπο της εφαρμογής, πλανητικοί γεωλόγοι έχουν κοσκινίσει μέσω τηλεσκοπικών εικόνων τους πλανήτες και τους αστεροειδείς, προσεκτικά για τον εντοπισμό και την καταλογογράφηση όπως ενδιαφέροντα γεωλογικά αντικείμενα και των επιπτώσεων τους στους κρατήρες. Είτε πρόκειται για την επιστήμη, το μάρκετινγκ, τη χρηματοδότηση υγειονομικής περίθαλψης, λιανικό εμπόριο, ή οποιοδήποτε άλλο τομέα, η κλασική προσέγγιση για την ανάλυση των δεδομένων βασίζεται πρωταρχικά σε έναν ή περισσότερους αναλυτές οι οποίοι είναι εξοικειωμένοι στενά με τα στοιχεία και το σερβίρισμα αυτών ως διεπαφή μεταξύ των δεδομένων, των χρηστών και των προϊόντων.

Γι' αυτές και πολλές άλλες εφαρμογές, αυτή η μορφή ανίχνευσης του εγχειριδίου ενός συνόλου δεδομένων είναι αργή, δαπανηρή, και άκρως υποκειμενική. Πράγματι, καθώς ο όγκος των δεδομένων αυξάνεται δραματικά, η χρήση αυτού του είδους της ανάλυσης των δεδομένων γίνεται εντελώς ανέφικτη σε πολλούς τομείς. Οι Βάσεις δεδομένων αυξάνονται σε μέγεθος με δύο τρόπους:

1. αριθμός  $N$  των εγγραφών ή τα αντικείμενα στις βάσεις δεδομένων
2. Ο αριθμός των πεδίων  $d$  ή οι ιδιότητες σε ένα αντικείμενο.

Βάσεις δεδομένων που περιέχουν σε σειρά  $N = 10$  εις την 9 αντικείμενα γίνονται όλο και πιο συχνά, για παράδειγμα, στις αστρονομικές επιστήμες. Ομοίως, ο αριθμός των πεδίων  $d$  μπορεί εύκολα να είναι της τάξης των 10 εις την 2 ή ακόμη και 10 εις την 3, για παράδειγμα, εφαρμογές στην ιατρική διαγνωστική. Ποιος θα περίμενε να χωνεύουν εκατομμύρια εγγραφών, το καθένα με δεκάδες ή εκατοντάδες πεδία; Πιστεύουμε ότι αυτή η δουλειά δεν είναι σίγουρα για τον άνθρωπο, ως εκ τούτου, η ανάλυση λειτουργεί όταν είναι αυτοματοποιημένη, εν μέρει τουλάχιστον. Η ανάγκη

για αναβάθμιση των δυνατοτήτων της ανθρώπινης ανάλυσης στο χειρισμό του μεγάλου αριθμού των bytes, έτσι ώστε να μπορούμε να συλλέξουμε τα στοιχεία τόσο οικονομικά όσο και επιστημονικά. Οι επιχειρήσεις που χρησιμοποιούν τα δεδομένα για να αποκτήσουν ανταγωνιστικό πλεονέκτημα, παρουσιάζουν αύξηση της αποτελεσματικότητας, και μπορούν να παρέχουν πιο πολύτιμες υπηρεσίες στους πελάτες.

Τα δεδομένα που έχουμε συλλάβει για το περιβάλλον μας είναι η βασική μαρτυρία που χρησιμοποιούμε για την κατασκευή των θεωριών και των μοντέλων του σύμπαντος που ζούμε. Επειδή οι υπολογιστές επέτρεψαν στους ανθρώπους να συγκεντρώνουν περισσότερα δεδομένα από όσα μπορεί να αφομοιώσει, δεν είναι μόνο φυσικό να στραφούν σε υπολογιστικές τεχνικές για να βοηθηθούν να ξεθάψουν νόημα από τα μοτίβα και τις δομές από τη μαζική αποθήκευση μεγάλων σε όγκων δεδομένων.

Στο πρώτο εργαστήριο KDD το 1989 (Piatetsky-Shapiro 1991), η φράση ανακάλυψη γνώσης σε βάσεις δεδομένων επινοήθηκε για να τονιστεί ότι η γνώση είναι το τελικό προϊόν μια διαπίστωση που βασίζεται σε δεδομένα.

Οι KDD αναφέρονται στη συνολική διαδικασία της ανακάλυψης γνώσης από χρήσιμα δεδομένα, και η εξόρυξη δεδομένων αναφέρεται σε μια συγκεκριμένη βαθμίδα σε αυτή τη διαδικασία καθώς εκφράζεται ως η εφαρμογή από ειδικούς αλγόριθμους για την εξαγωγή μοντέλων από τα δεδομένα.

Τα επιπλέον βήματα στη διαδικασία KDD, όπως η προετοιμασία των δεδομένων, η επιλογή των δεδομένων, ο καθαρισμός των δεδομένων, η ενσωμάτωση των κατάλληλων και προηγμένων γνώσεων και η σωστή ερμηνεία των αποτελεσμάτων της εξόρυξης είναι απαραίτητο να διασφαλίζουν τη χρήσιμη γνώση που προέρχεται από τα δεδομένα (Fayyad, Piatetsky-Shapiro, & Smyth).

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ

---

### 2 Εισαγωγή στην Ψυχομετρία (Introduction in Psychometrics) <sup>2</sup>

#### 2.1 Ορισμός

Ψυχομετρία είναι η εφαρμοσμένη επιστήμη που μελετά τη μέτρηση ψυχολογικών χαρακτηριστικών (αξιολόγηση νοημοσύνης, χαρακτηριστικών προσωπικότητας στάσεων και συμπεριφορών) μέσω της χρήσης σταθμισμένων ερωτηματολογίων και κλιμάκων.

Εναλλακτικά μπορεί κανείς να πει ότι η ψυχομετρία είναι :

- Ο κλάδος της ψυχολογίας που ασχολείται με τον σχεδιασμό και την αξιολόγηση ποσοτικών δοκιμασιών που μετρούν ψυχολογικές μεταβλητές, όπως νοημοσύνη, ικανότητα και προσωπικότητα
- Η θεωρία και η τεχνική νοητικών μετρήσεων
- Η ποσοτικοποίηση και ανάλυση των ατομικών διαφορών, που περιλαμβάνει την ανάπτυξη και την αξιολόγηση ψυχολογικών δοκιμασιών
- Η συστηματική χρήση δοκιμασιών (τεστ) με σκοπό την ποσοτική μέτρηση των ψυχικών φαινομένων και της ανθρώπινης συμπεριφοράς.

#### 2.2 Μέτρηση και Ψυχολογία

Η ψυχομετρία αναφέρεται στη μέτρηση ψυχολογικών, δηλαδή αφηρημένων χαρακτηριστικών. Σύμφωνα με τον Smith Stevens, *μέτρηση είναι η απόδοση αριθμών σε αντικείμενα ή γεγονότα με βάση κάποιον κανόνα*. Ο καλύτερος κανόνας που θα μπορούσε να βρεθεί σε αυτήν την περίπτωση είναι η σύγκριση με κάτι άλλο, παρόμοιο. Έτσι με βάση τον Michell (1997) *μέτρηση είναι η εκτίμηση του μεγέθους*

---

<sup>2</sup> Πηγή Κεφαλαίου: (Γιαγλής, 2004)

*μιας ποσότητας σε σύγκριση με κάποια άλλη. Ο ορισμός μιας ποσότητας συγκριτικά με κάποιαν άλλη χρησιμοποιείται σε όλες τις επιστήμες.*

### 2.2.1 Μέτρηση Ψυχολογικών Εννοιών

Οι ψυχολογικές έννοιες (constructs) που μετρώνται από την ψυχομετρία αφορούν συνήθως κοινά χαρακτηριστικά των ατόμων (π.χ. γνώσεις, στάσεις, προσωπικότητα). Για παράδειγμα μπορεί κανείς να πει ότι ένα από τα στοιχεία που χαρακτηρίζουν ένα άτομο είναι η αισιοδοξία ή απαισιοδοξία του. Εάν γνωρίζουμε ότι το άτομο είναι πολύ αισιόδοξο, τότε μπορούμε να προβλέψουμε μέχρι ενός σημείου την αντίδρασή του απέναντι σε μια δυσκολία.

### 2.2.2 Στάδια στην Υλοποίηση μιας Ψυχομετρικής Δοκιμασίας

Το πρώτο βήμα στην κατασκευή μιας δοκιμασίας που μετρά μια ψυχολογική έννοια είναι η διερεύνηση και συλλογή ενδεικτικών λειτουργιών, δηλαδή διαδικασιών, που θα χρησιμεύσουν στη μέτρηση της ψυχολογικής έννοιας. Για παράδειγμα, στην προσπάθεια κατασκευής μιας κλίμακας μέτρησης του «καλού μαθητή» θα πρέπει κανείς να διερωτηθεί τι ορίζει τον «καλό» μαθητή, τι κάνει ο «καλός» μαθητής που δεν κάνει ο μη «καλός». Στη συνέχεια, ο ερευνητής θα διαλέξει ορισμένες από αυτές τις διαδικασίες και πιο συγκεκριμένα τις πιο χρήσιμες, τις πιο σαφείς και τις πιο εύκολα μετρήσιμες, ώστε να σχηματίσει έναν λειτουργικό ορισμό της έννοιας. Με βάση αυτόν, θα κατασκευαστεί ένα αδρό εργαλείο, που θα μετρά τα στοιχεία του ορισμού αυτού.

Στο παράδειγμά μας, ο «καλός μαθητής» μπορεί να οριστεί λειτουργικά ως αυτός που παίρνει καλούς βαθμούς, συμμετέχει στο μάθημα, τηρεί τις προθεσμίες στις εργασίες του, έρχεται εγκαίρως στις παραδόσεις, και επιδεικνύει σεβασμό στους διδάσκοντες. Επομένως, το ψυχομετρικό εργαλείο θα μετρά με συγκεκριμένο τρόπο τα πιο πάνω έξι χαρακτηριστικά που επιλέχθηκαν από τον ερευνητή.

Το βασικότερο βήμα στη διαδικασία κατασκευής ενός ψυχομετρικού εργαλείου είναι ο πιλοτικός έλεγχός του. Ο ερευνητής θα πρέπει να το χρησιμοποιήσει για την αξιολόγηση ενός επαρκούς αριθμού μαθητών, ώστε να διαπιστώσει το πώς λειτουργεί στην πράξη. Τα εμπειρικά δεδομένα (μετρήσεις) που θα προκύψουν από την πιλοτική έρευνα, ο ερευνητής θα τα χρησιμοποιήσει για να ελέγξει την αξιοπιστία και την εγκυρότητα του εργαλείου. Εάν τα αποτελέσματα δεν είναι ικανοποιητικά, ο ερευνητής θα προβεί σε διόρθωση του εργαλείου και θα επαναλάβει τον πιλοτικό του έλεγχο, μέχρι να μείνει ικανοποιημένος.

Στην πράξη, ποτέ ένα εργαλείο δεν είναι τέλειο αλλά και με την πάροδο των ετών η κουλτούρα των ανθρώπων (γλώσσα, γνώσεις, στάσεις, συμπεριφορές) αλλάζει. Αυτό έχει ως αποτέλεσμα όλα τα ψυχομετρικά εργαλεία να βρίσκονται σε μια διαδικασία διαρκούς εξέλιξης και αναθεώρησης (έρευνα → ψυχομετρική αξιολόγηση → διόρθωση → έρευνα). Παράδειγμα: η πιο διαδεδομένη κλίμακα μέτρησης της νοημοσύνης, η κλίμακα WAIS (Wechsler Adult Intelligence Scale), που ξεκίνησε το 1939 (ως Wechsler-Bellevue), αναθεωρήθηκε σε WAIS (1955), WAIS-R (1981), WAISIII (1997) και, προς το παρόν, WAIS-IV (2008).

### 2.2.3 Ακρίβεια της Μέτρησης

Καμία μέτρηση δεν είναι απόλυτα ακριβής. Για παράδειγμα, όταν λέμε «το ύψος μου είναι 1 μέτρο και 73 εκατοστά» δεν είμαστε σίγουροι αν το ύψος είναι ακριβώς 173 εκατοστά ή 173,2... Η διαφορά της πραγματικής τιμής από τη μετρούμενη τιμή ονομάζεται λάθος ή σφάλμα μέτρησης (Measurement Error). Δηλαδή:

$$\text{Λάθος μέτρησης} = \text{Πραγματική τιμή} - \text{Μετρούμενη τιμή}$$

Με άλλα λόγια, λάθος μέτρησης είναι κάθε απόκλιση από την πραγματική τιμή που οφείλεται στη διαδικασία μέτρησης.

Σε γενικές γραμμές τα σφάλματα μέτρησης χωρίζονται σε δυο κατηγορίες:

1. Τυχαία σφάλματα (Random Errors), που οφείλονται σε άγνωστες, αστάθμητες αιτίες. Τις περισσότερες φορές κατανέμονται ομοιόμορφα, δηλαδή δεν επηρεάζουν τη μέτρηση προς μια ορισμένη κατεύθυνση. Παράδειγμα τυχαίου σφάλματος παρουσιάζει μια παλιά, αναλογική ζυγαριά

που ο δείκτης της ταλαντεύεται προτού σταματήσει και μερικές φορές δείχνει λίγο περισσότερο, ενώ άλλοτε δείχνει λίγο λιγότερο από την πραγματική τιμή.

2. Συστηματικά σφάλματα (Systematic Errors), που οφείλονται σε μια συγκεκριμένη αιτία και επηρεάζουν με συστηματικό και επαναλαμβανόμενο τρόπο τη μέτρηση. Παράδειγμα συστηματικού σφάλματος παρουσιάζει μια ζυγαριά που μονίμως μας δείχνει βαρύτερους από ό,τι είμαστε ή ένας αυστηρός καθηγητής που, διορθώνοντας τα γραπτά, δίνει μικρότερες βαθμολογίες από ό,τι μας αξίζουν.

#### 2.2.4 Πηγές Σφαλμάτων Μέτρησης

Μερικές από τις αιτίες που μια έρευνα μπορεί να οδηγήσει σε εσφαλμένες μετρήσεις μιας ψυχολογικής έννοιας είναι οι εξής:

- Σφάλμα δειγματοληψίας: Αντί να χρησιμοποιήσει ένα αντιπροσωπευτικό δείγμα, ο ερευνητής έκανε μετρήσεις σε συγγενείς, φίλους, γνωστούς και επέλεξε βολικά και όχι τυχαία υποκείμενα. Επομένως, τα αποτελέσματα δεν μπορούν να γενικευθούν στον πληθυσμό.
- Ανακριβή εργαλεία: Το ψυχομετρικό εργαλείο που χρησιμοποιήθηκε δεν έχει την επιθυμητή ακρίβεια είτε επειδή δεν κατασκευάστηκε σωστά είτε επειδή έγινε κακή προσαρμογή του είτε επειδή χρησιμοποιήθηκε για να μετρήσει κάτι διαφορετικό από αυτό που κατασκευάστηκε να μετράει. Επομένως, οι μετρήσεις που προέκυψαν από αυτό είναι αναξιόπιστες και ανακριβείς.
- Μεροληψία των υποκειμένων: Τα υποκείμενα δεν δίνουν ειλικρινείς απαντήσεις είτε επειδή θέλουν να φανούν καλύτερα από ό,τι είναι στην πραγματικότητα, είτε επειδή ντρέπονται και θέλουν να αποκρύψουν ένα πρόβλημά τους, είτε επειδή απλά θέλουν να ευχαριστήσουν τον ερευνητή και απαντούν αυτό που νομίζουν ότι αυτός επιθυμεί.
- Μεροληψία του ερευνητή: Ο ερευνητής έχει μια ξεκάθαρη ιδέα για το τι περιμένει ότι θα προκύψει από την έρευνά του και πιθανότατα, επιθυμεί για προσωπικούς λόγους ένα συγκεκριμένο συμπέρασμα. Έτσι, επηρεάζει υποσυνείδητα ή και συνειδητά τα υποκείμενα της έρευνας και τους ωθεί να απαντήσουν προς την κατεύθυνση που εκείνος προσδοκά ή επιθυμεί.

- Εσφαλμένο/διαφορετικό επιστημολογικό πλαίσιο: Ο ερευνητής μελετά ένα φαινόμενο ξεκινώντας από εσφαλμένες ή απλά διαφορετικές αρχές από αυτές που χρειάζεται για να κατανοήσει πλήρως το φαινόμενο αυτό. Για παράδειγμα, προσπαθεί να μελετήσει τη σχολική επίδοση σε ένα σχολείο μιας διαφορετικής χώρας θεωρώντας ότι ο καλύτερος μαθητής είναι αυτός που απαντά σωστά στις περισσότερες ερωτήσεις που θέτει ο δάσκαλος στην τάξη. Αγνωρίζει όμως ότι στη συγκεκριμένη χώρα θεωρείται κοινωνικά μη αποδεκτό ένας μαθητής να απαντά σωστά μian ερώτηση, μπροστά σε όλους τους συμμαθητές του που δε γνωρίζουν την απάντηση.

### 2.3 Είδη Ψυχομετρικών Εργαλείων και Αξιολόγησης

Τα ψυχομετρικά εργαλεία μπορούν να κατηγοριοποιηθούν ανάλογα με το περιεχόμενο των ψυχολογικών εννοιών που μετρούν, ανάλογα με τον τρόπο χορήγησής τους και ανάλογα με τον τρόπο αξιολόγησης των αποτελεσμάτων.

Σε γενικές γραμμές ψυχομετρική ονομάζεται μια δοκιμασία όχι λόγω του περιεχομένου της (δηλαδή της ψυχολογικής έννοιας που μετρά) αλλά λόγω του τρόπου με τον οποίο έχει κατασκευαστεί ή/και βαθμολογείται.

Ανάλογα με το χρόνο που διατίθεται για την ολοκλήρωσή τους, οι ψυχομετρικές δοκιμασίες ικανοτήτων (Ability or Aptitude Tests) διακρίνονται σε:

- Δοκιμασίες ικανότητας (Power Tests) στις οποίες τα θέματα είναι συνήθως πιο δύσκολα και πολύπλοκα αλλά υπάρχει επαρκής χρόνος για να ολοκληρωθεί η δοκιμασία στο σύνολό της.
- Δοκιμασίες ταχύτητας (Speed Tests) στις οποίες τα θέματα ίσως να είναι πιο απλά αλλά υπάρχει περιορισμένος χρόνος για τη συμπλήρωση μεγάλου αριθμού θεμάτων.

Οι ψυχομετρικές ιδιότητες που έχει μια δοκιμασία μπορεί να διαφέρουν πολύ, ανάλογα με το αν χορηγείται υπό πίεση χρόνου (ταχύτητα) ή με άπλετο χρόνο (ικανότητα). Επομένως ο τρόπος χορήγησης είναι πολύ σημαντικός. Εκτός από τις δοκιμασίες ικανοτήτων (π.χ. νοημοσύνης, νευροψυχολογικές, εκπαιδευτικής ή

επαγγελματικής επάρκειας), η χορήγηση υπό πίεση χρόνου δε χρησιμοποιείται συνήθως σε άλλα είδη δοκιμασιών (π.χ. ψυχοπαθολογίας, προσωπικότητας κλπ).

Αξίζει να σημειωθεί ότι τις περισσότερες φορές κατά τη διάρκεια μιας δοκιμασίας παροτρύνουμε τον εξεταζόμενο και περιμένουμε από αυτόν να πετύχει την καλύτερη δυνατή επίδοση (μέγιστη προσπάθεια). Αυτός είναι και ο λόγος που είναι απαραίτητο, πριν από τέτοιου τύπου αξιολογήσεις να προηγείται ένα διάστημα εξοικείωσης του εξεταζομένου με τον χώρο εξέτασης, καθώς και εγκαθίδρυσης μιας καλής σχέσης μεταξύ εξεταστή και εξεταζόμενου (rapport).

Στις δοκιμασίες αυτές συνήθως αξιολογούμε το μέσο όρο των επιδόσεων του εξεταζόμενου (τα περισσότερα ψυχομετρικά εργαλεία), ενώ σπανίως μπορεί να λαμβάνεται υπόψη μόνον η καλύτερη επίδοση (π.χ. Longest Digit Span Forward, στη μνήμη αριθμών, ή στις αθλητικές δοκιμασίες) ή μόνον η χειρότερη επίδοση (π.χ. στις εξετάσεις οδήγησης). Σε αντιδιαστολή με τη μέγιστη προσπάθεια, η συνήθης προσπάθεια λαμβάνεται από πληροφορίες είτε αυτοαξιολόγησης (π.χ. ερωτηματολόγια) είτε από συγγενείς είτε από άμεση παρατήρηση του εξεταζόμενου στο περιβάλλον του.

Τέλος, ανάλογα με τον τρόπο αξιολόγησης της βαθμολογίας τα ψυχομετρικά εργαλεία μπορούν να διαχωριστούν (Glaser, 1963) σε:

- Αξιολόγηση με κριτήριο (Criterion - Referenced).

Σε αυτόν τον τρόπο αξιολόγησης υπάρχει συνήθως ένα όριο διαχωρισμού (cut-off score). Τα άτομα που έχουν βαθμολογία πάνω από το διαχωριστικό όριο, θεωρείται ότι έχουν πετύχει το κριτήριο, δηλαδή ξέρουν όσα χρειάζεται να ξέρουν ή μπορούν να κάνουν όσα χρειάζεται να μπορούν να κάνουν.

Συνήθως τα εργαλεία αυτά χρησιμοποιούνται στην εκπαιδευτική αξιολόγηση και για να κρίνουν εάν π.χ. ένας μαθητής θα περάσει την τάξη, ένας φοιτητής θα πάρει πτυχίο ή ένας ιατρός θα κριθεί ικανός να ασκήσει την ιατρική (mastery test). Το διαχωριστικό όριο μπορεί να είναι αυθαίρετο ή να έχει επιλεγεί με κάποιον ειδικό τρόπο.

- Αξιολόγηση με νόρμες (Norm - Referenced).



Σε αυτόν τον τρόπο αξιολόγησης η βαθμολογία ενός υποκειμένου δεν έχει σημασία αυτή καθαυτή αλλά μόνο συγκριτικά με τον πληθυσμό. Για παράδειγμα μια βαθμολογία σε δοκιμασία νοημοσύνης που αντιστοιχεί σε IQ 100 δε μας λέει σε πόσες ερωτήσεις ή δοκιμασίες απάντησε επιτυχώς ο εξεταζόμενος αλλά ότι πέτυχε βαθμολογία καλύτερη από το 50% του πληθυσμού (ατόμων ίδιου φύλου και ηλικίας). Με τον τρόπο αυτό μπορεί κανείς να αξιολογήσει τη σχετική σειρά κατάταξης ενός ατόμου στον πληθυσμό και να αξιολογήσει πόσο «σπάνια» είναι μια βαθμολογία.

Ένα χαρακτηριστικό παράδειγμα είναι οι εξετάσεις εισαγωγής στα ανώτερα και ανώτατα εκπαιδευτικά ιδρύματα, όπου δεν έχει σημασία ο συνολικός βαθμός ενός υποψηφίου αλλά η επίδοσή του συγκριτικά με τους υπόλοιπους υποψηφίους. Ο όρος νόρμες αναφέρεται στο πόσο καλά τα πήγε ο πληθυσμός σε μια δοκιμασία και συνήθως απεικονίζονται σε ειδικούς πίνακες.

- Ιδιοθετική αξιολόγηση (Ipsative).

Αυτός ο τρόπος αξιολόγησης συγκρίνει τη βαθμολογία ενός ατόμου σε σχέση με τον εαυτό του και όχι με άλλα άτομα. Αυτό μπορεί να πάρει τις εξής δύο μορφές:

1. Σύγκριση με παλαιότερες βαθμολογίες του ίδιου ατόμου στην ίδια δοκιμασία. Εφαρμόζεται κυρίως σε εκπαιδευτικό ή σε κλινικό περιβάλλον και αξιολογεί π.χ. τη βελτίωση του μαθητή μετά από μια συγκεκριμένη μέθοδο εκπαίδευσης ή τη βελτίωση ενός ασθενή μετά τη χορήγηση ενός συγκεκριμένου φαρμάκου.
2. Σύγκριση των απαντήσεων του ατόμου μεταξύ τους ή σύγκριση της βαθμολογίας του σε μια δοκιμασία με τις βαθμολογίες του σε άλλες δοκιμασίες. Με τον τρόπο αυτό μπορούν να αναγνωριστούν τα δυνατά ή αδύναμα σημεία ενός ατόμου ή να ανιχνευθούν οι προτιμήσεις του. Ως παραδείγματα αυτής της μεθόδου αναφέρονται
  - α) η αξιολόγηση των δοκιμασιών επαγγελματικού προσανατολισμού, που κρίνουν εάν π.χ. ένας μαθητής προτιμά τα χειρωνακτικά από τα πνευματικά επαγγέλματα, και β) η αξιολόγηση εάν ένα παιδί έχει π.χ. μεγαλύτερη λεκτική (verbal) ή μη-λεκτική, πρακτική (performance) νοημοσύνη.

### 2.3.1 Αξιοπιστία Εργαλείων

Όπως αναφέρθηκε πιο πάνω, ένα «καλό» εργαλείο είναι ένα εργαλείο που έχει μικρό σφάλμα μέτρησης. Αξιοπιστία (reliability) είναι η ιδιότητα ενός εργαλείου μέτρησης να δίνει την ίδια τιμή, όταν μετράει το ίδιο αντικείμενο. Επομένως, το αξιόπιστο εργαλείο μετρά πάντοτε με τον ίδιο (σχεδόν) τρόπο.

Για παράδειγμα, μια ζυγαριά είναι αξιόπιστη, εάν κάθε φορά που ανεβαίνουμε μας δείχνει την ίδια τιμή βάρους.

Στην ψυχομετρία υπάρχουν τέσσερις διαφορετικές μέθοδοι:

#### 1. Αξιοπιστία επαναληπτικών μετρήσεων (test retest reliability).

Σε αυτή τη μέθοδο χορηγούμε την ίδια δοκιμασία δύο (ή περισσότερες) φορές στα ίδια άτομα και στη συνέχεια συγκρίνουμε τις τιμές που βρήκαμε. Αυτή η μορφή αξιοπιστίας ονομάζεται και σταθερότητα (stability). Εάν μια δοκιμασία έχει αυτή τη μορφή αξιοπιστίας, τότε οι διαφορετικές μετρήσεις στα ίδια άτομα θα διαφέρουν μόνον λίγο. Ο χρόνος που μεσολαβεί ανάμεσα στις διαδοχικές χορηγήσεις της δοκιμασίας αποφασίζεται από τον ερευνητή και έχει σχέση και με τη μετρούμενη έννοια.

Από τη μια μεριά, το χρονικό διάστημα θα πρέπει να είναι αρκετά μικρό, ώστε οι συμμετέχοντες να μην έχουν προλάβει να «αλλάξουν» πολύ. Για παράδειγμα, εάν ένα εργαλείο που μετρά τα επίπεδα της κατάθλιψης το χορηγήσουμε δυο φορές με χρονική απόσταση δύο μηνών, είναι πιθανόν να δούμε διαφορετικές τιμές, όχι επειδή το εργαλείο δεν είναι αξιόπιστο αλλά επειδή μέσα σε αυτό το χρονικό διάστημα έχει μεταβληθεί η συναισθηματική κατάσταση των υποκειμένων.

Από την άλλη μεριά, το χρονικό διάστημα θα πρέπει να είναι αρκετά μεγάλο, ώστε οι συμμετέχοντες να μη θυμούνται τι απάντησαν την πρώτη φορά. Για παράδειγμα, εάν χορηγήσουμε μια δοκιμασία ξανά, πέντε λεπτά αργότερα από την πρώτη φορά, τότε είναι πιθανόν τα υποκείμενα να δώσουν απαντήσεις από μνήμης, παρά να ξανασκεφτούν τις απαντήσεις τους από την αρχή. Κι έτσι, θα καταλήξουμε να αξιολογήσουμε περισσότερο τη μνήμη των υποκειμένων, παρά την αξιοπιστία του εργαλείου μας.

## 2. Αξιοπιστία παράλληλων μορφών (alternate or parallel forms reliability).

Σε αυτή τη μέθοδο κατασκευάζουμε δύο ισοδύναμες μορφές της ίδιας δοκιμασίας που έχουν παρόμοιο περιεχόμενο. Για παράδειγμα, στη δοκιμασία «αριθμητικής ικανότητας» χρησιμοποιούμε πολλαπλασιασμό τριψηφίων αριθμών αλλά χρησιμοποιούμε διαφορετικούς αριθμούς στην πρώτη (Α) και διαφορετικούς στη δεύτερη (Β) μορφή. Χορηγούμε και τις δυο μορφές στα ίδια άτομα και συγκρίνουμε τα αποτελέσματα. Θεωρητικά θα πρέπει οι δυο μορφές να δίνουν παρόμοια αποτελέσματα.

Διαφορετικά θα πρέπει να καταλήξουμε στη χρήση πολύπλοκων στατιστικών μεθόδων που μπορούν να εξισώσουν τις βαθμολογίες των δυο δοκιμασιών.

3. Αξιοπιστία διχοτόμησης ή ημικλάστων (split-half reliability)<sup>3</sup>.

Αντί να κατασκευαστούν εναλλακτικές μορφές, είναι δυνατόν κάτω από κάποιες συνθήκες να χωρίσουμε τη δοκιμασία σε δυο ίσα τμήματα. Μπορούμε, δηλαδή, να θεωρήσουμε ως ένα τμήμα τα πρώτα μισά θέματα (ερωτήσεις) και ως δεύτερο τμήμα τα τελευταία μισά. Άλλος τρόπος είναι να θεωρήσουμε ως ένα τμήμα τα θέματα με περιττό (μονό) αριθμό και ως δεύτερο τμήμα τα θέματα με άρτιο (ζυγό) αριθμό. Σε αυτή τη μέθοδο χορηγούμε τη δοκιμασία στην ίδια ομάδα υποκειμένων και συγκρίνουμε τους βαθμούς τους στα δύο αυτά τμήματα. Μια βασική προϋπόθεση για την εφαρμογή αυτής της μεθόδου μελέτης της αξιοπιστίας είναι η δοκιμασία να είναι ομοιογενής, δηλαδή όλες οι ερωτήσεις να είναι παρόμοιες και ισοδύναμες ως προς τη σημασία τους. Διαφορετικά θα καταλήξει το ένα τμήμα να είναι διαφορετικό ή και σημαντικότερο από το άλλο.

## 4. Αξιοπιστία μεταξύ κριτών ή βαθμολογητών (Inter - Rater Reliability).

Όλες σχεδόν οι ψυχομετρικές δοκιμασίες, εκτός ίσως από τις ερωτήσεις πολλαπλής επιλογής, απαιτούν την κρίση ενός εξεταστή για να βαθμολογηθούν. Για παράδειγμα, οι δοκιμασίες που απαιτούν προφορικές απαντήσεις, αντιγραφή ενός σχήματος, παράθεση ενός ορισμού κλπ απαιτούν από το βαθμολογητή να λάβει κάποιες αποφάσεις. Αυτή είναι μια επιπρόσθετη πηγή λάθους μέτρησης, γιατί ένας

<sup>3</sup> Μια επέκταση αυτής της μορφής αξιοπιστίας είναι ο υπολογισμός του συντελεστή α του Cronbach. Ο συντελεστής αυτός αποτελεί ένα μέτρο της συσχέτισης όλων των θεμάτων μεταξύ τους και υψηλές τιμές του ερμηνεύονται συχνά ως ένδειξη ότι όλα τα θέματα μετρούν την ίδια ψυχολογική έννοια. Τότε η μορφή αυτή της αξιοπιστίας ονομάζεται και εσωτερική συνέπεια (internal consistency).

κριτής μπορεί π.χ. να είναι επιεικής, ενώ κάποιος άλλος αυστηρός. Για να μελετήσουμε κατά πόσο μια δοκιμασία επηρεάζεται ή όχι από τέτοιους παράγοντες, τη χορηγούμε μία φορά και στη συνέχεια ζητούμε να τη βαθμολογήσουν δύο (ή περισσότεροι) κριτές.

### 2.3.2 Παράγοντες που Επηρεάζουν την Αξιοπιστία ενός Εργαλείου

Ένα εργαλείο μπορεί να επηρεαστεί από τους παρακάτω παράγοντες.

1. Το μέγεθος μιας δοκιμασίας μπορεί να επηρεάσει αρνητικά την αξιοπιστία της, καθώς μακροσκελείς δοκιμασίες τείνουν να προκαλούν κόπωση και ανία στα υποκείμενα. Αυτό έχει ως αποτέλεσμα να απαντούν χωρίς προσοχή και άρα χωρίς ακρίβεια.
2. Η ομοιογένεια στη μορφή αλλά και στις ερωτήσεις είναι σημαντική. Για παράδειγμα, ένα εργαλείο που έχει ερωτήσεις “Ναι / Όχι”, ερωτήσεις με βαθμολόγηση σε 5-βαθμη και μετά σε 4-βαθμη κλίμακα Likert και μετά επιλογή της σημαντικότερης απάντησης από μια λίστα επιλογών είναι πιθανόν να δημιουργήσει κόπωση και σύγχυση στα υποκείμενα. Ένα ερωτηματολόγιο με ομοιόμορφες ερωτήσεις (π.χ. όλες “Ναι / Όχι”) είναι πιο εύκολο να απαντηθεί και άρα λιγότερο πιθανό να γίνουν λάθη ή απροσεξίες.
3. Συνεργασιμότητα των υποκειμένων. Είναι αυτονόητο ότι οι ερωτώμενοι θα πρέπει να έχουν ένα κίνητρο, για να συμμετέχουν στην αξιολόγηση με διάθεση και ειλικρίνεια αλλιώς θα δώσουν τυχαίες ή καθόλου απαντήσεις. Στα πιθανά κίνητρα συμπεριλαμβάνονται η αμοιβή (χρηματική ή συμβολική), η προσδοκία κέρδους (π.χ. λαχνός σε κλήρωση), το άμεσο θεραπευτικό όφελος (π.χ. δοκιμασίες που θα βοηθήσουν στη διάγνωση ή θα κατευθύνουν τη θεραπεία τους) ή το έμμεσο θεραπευτικό όφελος (π.χ. τυποποίηση δοκιμασιών που θα βοηθήσουν την επιστήμη και την υγεία).

4. Οι σταθερές ή παρόμοιες συνθήκες χορήγησης, π.χ. σε δωμάτιο ήσυχο, φωτεινό, όπου είναι παρόντες μόνο ο εξεταστής και ο εξεταζόμενος, θα οδηγήσουν σε αξιόπιστες βαθμολογίες. Αντίθετα, θορυβώδεις συνθήκες, διακοπή από εξωτερικούς παράγοντες ή χορήγηση π.χ. αργά το απόγευμα είναι δυνατόν να επηρεάσουν την επίδοση των συμμετεχόντων και να προκαλέσουν ανακρίβεια στις μετρήσεις.
5. Αντίστοιχα ή και περισσότερο σημαντική είναι η διαδικασία της βαθμολόγησης. Το πρωτόκολλο (διαδικασία) αξιολόγησης μιας δοκιμασίας είναι ένας αυστηρά καθορισμένος τρόπος βαθμολόγησης που συχνά απαιτεί εκπαίδευση και εμπειρία.
6. Πέραν των ερωτήσεων πολλαπλής επιλογής, οι ψυχομετρικές δοκιμασίες έχουν ένα βαθμό αξιοπιστίας μεταξύ κριτών που σχεδόν ποτέ δεν είναι τέλειος. Για το λόγο αυτό, όσο λιγότεροι είναι οι βαθμολογητές των δοκιμασιών, τόσο περισσότερο αξιόπιστα αναμένεται να είναι τα αποτελέσματα.

### 2.3.3 Εγκυρότητα Εργαλείων

Εγκυρότητα (Validity) είναι η ιδιότητα ενός εργαλείου με σκοπό την μέτρηση. Για παράδειγμα, μια δοκιμασία μνήμης θα πρέπει να μετράει τη μνημονική ικανότητα και όχι π.χ. το εύρος των γενικών γνώσεων ενός υποκειμένου. Εννοείται ότι ένα εργαλείο με χαμηλή αξιοπιστία δεν μπορεί να έχει υψηλή εγκυρότητα, αφού λόγω της χαμηλής αξιοπιστίας του δεν μπορούμε να είμαστε σίγουροι για το τι μετράει ακριβώς.

Οι κυριότερες μορφές εγκυρότητας ενός ψυχομετρικού εργαλείου είναι οι ακόλουθες:

#### 1. Εγκυρότητα περιεχομένου (Content Validity).

Αυτή η μορφή εγκυρότητας αφορά κυρίως τις δοκιμασίες εκπαιδευτικής αξιολόγησης και αναφέρεται στο κατά πόσο η δοκιμασία καλύπτει επαρκώς το εξεταζόμενο αντικείμενο. Για παράδειγμα, μια δοκιμασία «αριθμητικής ικανότητας» θα πρέπει να καλύπτει επαρκώς τις τέσσερις βασικές αριθμητικές

πράξεις (πρόσθεση, αφαίρεση, πολλαπλασιασμός, διαίρεση). Επομένως μια δοκιμασία «αριθμητικής ικανότητας» που αποτελείται από 20 αριθμητικές πράξεις, 5 από κάθε είδος, μπορούμε να υποστηρίξουμε ότι έχει εγκυρότητα περιεχομένου. Βασική προϋπόθεση για την αξιολόγηση της εγκυρότητας περιεχομένου μιας δοκιμασίας είναι ότι ένας ή περισσότεροι ειδικοί θα συμφωνήσουν ως προς το περιεχόμενο, τη δομή και τη σημασία των στοιχείων που συναποτελούν την έννοια.

### 2. Φαινομενική εγκυρότητα (Face Validity).

Παρότι θεωρητικά αυτή είναι η λιγότερο σημαντική μορφή εγκυρότητας, πρακτικά έχει σημασία στο βαθμό ότι βοηθάει τα υποκείμενα να δεχτούν καλύτερα την αξιολόγηση, να πειστούν ότι κάνουμε όντως αυτό που τους είπαμε ότι κάνουμε.

### 3. Εγκυρότητα εξωτερικού κριτηρίου (Criterion Validity).

Αυτή η μέθοδος αξιολόγησης της εγκυρότητας στηρίζεται σε άλλα εργαλεία ή δοκιμασίες που συμφωνούμε ότι μετρούν την ίδια έννοια. Για παράδειγμα, ένα νεοσύστατο ερωτηματολόγιο κατάθλιψης δεχόμαστε ότι μετράει όντως την κατάθλιψη, εάν παρουσιάζει υψηλή συσχέτιση στις βαθμολογίες του με ένα άλλο ερωτηματολόγιο (ή γενικότερα μια άλλη αξιολόγηση) κατάθλιψης, που είναι γνωστό και αποδεδειγμένο ότι μετράει πραγματικά την κατάθλιψη. Αυτή η δεύτερη αξιολόγηση είναι το εξωτερικό κριτήριο και θεωρείται το βασικό πρότυπο (Gold Standard) μέτρησης της κατάθλιψης.

Η εγκυρότητα εξωτερικού κριτηρίου διακρίνεται σε:

- Συγχρονική εγκυρότητα (concurrent validity), όταν η συσχέτιση υπολογίζεται με άλλα εργαλεία που χορηγήθηκαν στους ίδιους συμμετέχοντες την ίδια περίπου χρονική περίοδο, και
- Προβλεπτική εγκυρότητα (predictive validity), όταν το ψυχομετρικό εργαλείο συσχετίζεται και άρα προβλέπει τη βαθμολογία άλλων εργαλείων που χορηγούνται στους ίδιους συμμετέχοντες μετά από κάποιο ικανό χρονικό διάστημα.

4. Εγκυρότητα εννοιολογικής κατασκευής (Construct Validity).

Αυτή η μέθοδος αξιολόγησης της εγκυρότητας εξετάζει κατά πόσο ο λειτουργικός ορισμός που χρησιμοποιείται από μια ψυχομετρική δοκιμασία για τη μέτρηση μιας ψυχολογικής έννοιας συμφωνεί με την τρέχουσα θεωρία για την έννοια αυτή.

Από πρακτικής άποψης, η βαθμολογία στη δοκιμασία που μετράει την έννοια αυτή θα πρέπει:

- Να παρουσιάζει υψηλή συσχέτιση με δοκιμασίες που μετρούν παρόμοιες ή συγγενικές έννοιες, δηλαδή να παρουσιάζει συγκλίνουσα εγκυρότητα (Convergent Validity)
- Να παρουσιάζει χαμηλή ή καθόλου συσχέτιση με δοκιμασίες που μετρούν άσχετες έννοιες, δηλαδή να παρουσιάζει αποκλίνουσα εγκυρότητα (Divergent Validity).

## 3<sup>ο</sup> ΚΕΦΑΛΑΙΟ

---

### 3 Ερωτηματολόγια Διαταραχών Προσωπικότητας & Συναισθηματικής Νοημοσύνης (IPDE & Emotional Intelligence Questionnaires)

#### 3.1 Εισαγωγή στη Συναισθηματική Νοημοσύνη

Η **συναισθηματική νοημοσύνη** αποτελεί ταυτόχρονα χαρακτηριστικό της προσωπικότητας και νοητική ικανότητα, την οποία το άτομο κληρονομεί από το γενετικό υλικό των γονέων του και την εξελίσσει - αναπτύσσει καθ' όλη την διάρκεια της ζωής του. Αναφέρεται στην ικανότητα επεξεργασίας πληροφοριών προερχόμενων από τα συναισθήματα και η χρησιμοποίησή τους για την καθοδήγηση της δράσης σε περιστάσεις που απαιτούν ενεργοποίηση του γνωστικού συστήματος. Δεν βάζει σε πρώτη γραμμή τις πνευματικές ικανότητες των ανθρώπων αλλά τη νοημοσύνη των συναισθημάτων, καθώς θεωρείται πολύ πιο σημαντική από τη ψυχρή νοημοσύνη της σκέψης προκειμένου να οικοδομηθεί μια πιο ολοκληρωμένη ζωή.

Θεωρείται μια ανώτερη ικανότητα από την οποία εξαρτάται στην ουσία το πόσο καλά μπορεί κανείς να χρησιμοποιήσει τις ικανότητες του για να έχει επιτυχία στο επάγγελμα του, μια ευτυχισμένη συντροφική σχέση, ενδιαφέρουσες και στέριαις κοινωνικές επαφές, γενική ευεξία και σε ένα βαθμό ευχαρίστησης στην ζωή.

Ο Daniel Goleman, που έκανε ευρέως γνωστή τη Συναισθηματική Νοημοσύνη, μιλά για τη "νοημοσύνη της καρδιάς" και την ορίζει ως "την ικανότητα ενός ατόμου να αναγνωρίζει τα δικά του συναισθήματα και τα συναισθήματα των άλλων, να τα χειρίζεται αποτελεσματικά και να δημιουργεί κίνητρα για τον εαυτό του".

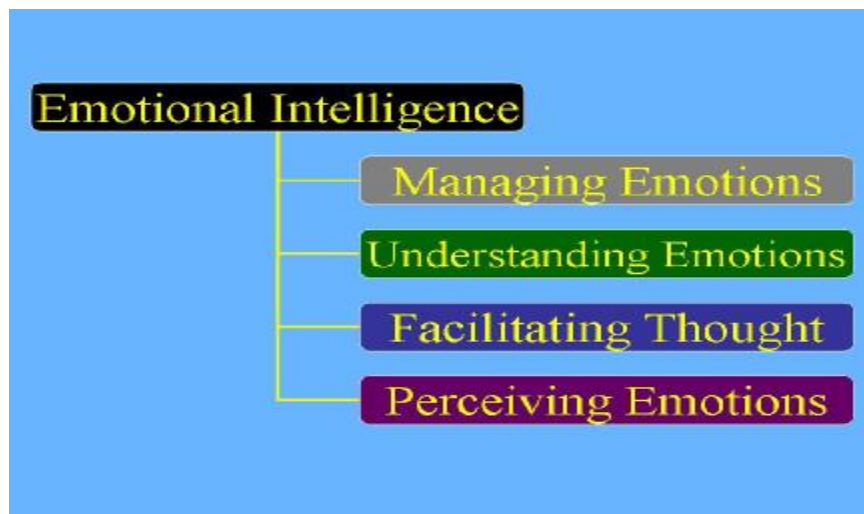
Επιπλέον οι Bradberry & Greaves κάνουν μια αξιόλογη προσπάθεια να παρουσιάσουν την έννοια της Συναισθηματικής Νοημοσύνης με έναν πιο κατανοητό και πρακτικό τρόπο, σε σχέση με τον Goleman. Κάνουν λόγο για τη "συναισθηματική διάσταση της λογικής" και ορίζουν τη Συναισθηματική Νοημοσύνη ως την ικανότητα ενός ατόμου να αναγνωρίζει και να συνειδητοποιεί τα συναισθήματα του και να



χρησιμοποιεί την επίγνωση αυτή για να χειρίζεται τη συμπεριφορά και τις σχέσεις του. (Δαλακούρα, 2011)

Τονίζεται ότι η **Συναισθηματική Νοημοσύνη** καθορίζει τον τρόπο με τον οποίο :

1. Συμπεριφερόμαστε
2. Χειριζόμαστε την πολυπλοκότητα των κοινωνικών μας σχέσεων
3. Παίρνουμε προσωπικές αποφάσεις που καταλήγουν σε θετικά αποτελέσματα



Εικόνα 2 Μέθοδοι τους οποίους διαχειρίζεται η Συναισθηματική Νοημοσύνη

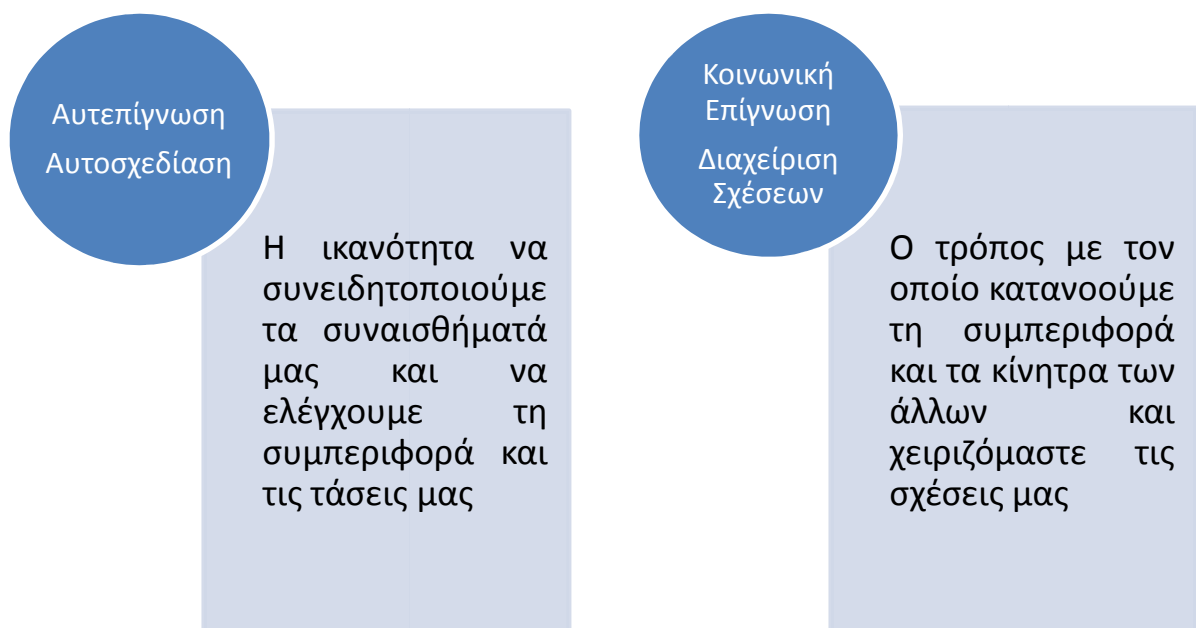
Τέσσερις συναισθηματικές δεξιότητες οι οποίες συγκροτούν την έννοια της Συναισθηματικής στην πράξη:

- **Αυτεπίγνωση:** Αναφέρεται στην ικανότητα να αντιλαμβανόμαστε με ακρίβεια τα συναισθήματά μας, τις τάσεις μας και την προδιάθεση που έχουμε προς μία συγκεκριμένη συμπεριφορά.
- **Αυτοδιαχείριση:** Σημαίνει να χρησιμοποιούμε την επίγνωση των συναισθημάτων μας για να είμαστε ευέλικτοι και να κατευθύνουμε προς την επιθυμητή κατεύθυνση τη συμπεριφορά μας.
- **Κοινωνική επίγνωση:** Αναφέρεται στην ικανότητα να αντιλαμβανόμαστε τι σκέφτονται και τι αισθάνονται οι άλλοι, ακόμη και αν εμείς δεν αισθανόμαστε το ίδιο.

- **Διαχείριση σχέσεων:** Αναφέρεται στην ικανότητα να χειριζόμαστε με επιτυχία και σε βάθος χρόνου τις αλληλεπιδράσεις μας με τους άλλους ανθρώπους.

### 3.1.1 Κατηγορίες Συναισθηματικής Νοημοσύνης

Ένας άνθρωπος για να αυξήσει τη Συναισθηματική του Νοημοσύνη πρέπει να αυξήσει τις επιμέρους δεξιότητές του και στους τέσσερις παρακάτω τομείς:



Οι δύο πρώτες αφορούν εσωτερικές διεργασίες που καθορίζουν τον τρόπο με τον οποίο το άτομο χειρίζεται τα θέματα του εαυτού του, ενώ οι δύο τελευταίες αφορούν εξωτερικές διεργασίες που καθορίζουν πόσο καλά το άτομο χειρίζεται τις σχέσεις του με τους άλλους.



Εικόνα 3 Συναισθηματικές Δεξιότητες

Πρέπει να τονιστεί ότι οι παραπάνω κατηγορίες της Συναισθηματικής Νοημοσύνης είναι ιεραρχικά αλληλοεξαρτώμενες, δηλαδή η μια επηρεάζει την άλλη σε κάποιο βαθμό και η μια οικοδομείται πάνω στην άλλη. Για παράδειγμα, η αυτοδιαχείριση προϋποθέτει την αυτεπίγνωση, ενώ οι κοινωνικές δεξιότητες βασίζονται σημαντικά στην ενσυναίσθηση. (Δαλακούρα, 2011)

### 3.1.2 Αυτεπίγνωση

Η αυτοεπίγνωση, περιλαμβάνει ως υποκατηγορίες την συναισθηματική μας επίγνωση, την ακριβή μας αυτοαξιολόγηση και τέλος την αυτοπεποίθηση μας. Είναι σημαντικό να ξέρουμε καλά ποιες είναι οι τάσεις μας, ώστε να κατανοήσουμε πιο γρήγορα τα συναισθήματά μας. Ο μοναδικός τρόπος για να το πετύχουμε είναι να αφιερώσουμε χρόνο ώστε να αντιληφθούμε από πού πηγάζουν και κυρίως για ποιό λόγο αισθανόμαστε έτσι. Πολλές φορές μοιάζουν να εμφανίζονται από το πουθενά, αυτό όμως δεν ισχύει, και έχει μεγάλη σημασία να καταλάβουμε γιατί οι εκάστοτε συνθήκες έχουν τη δύναμη να πυροδοτήσουν σε εμάς μια συγκεκριμένη αντίδραση. Οι καταστάσεις που δημιουργούν έντονα συναισθήματα απαιτούν πάντα περισσότερη σκέψη. (Λιάκου, 2008)

### 3.1.3 Αυτοδιαχείριση

Η ικανότητα να διαχειριζόμαστε την εσωτερική μας κατάσταση, τις παρορμήσεις και τις διαθέσεις μας. Τα άτομα που διαθέτουν την ικανότητα αυτή μπορούν να

αποφύγουν συναισθηματικές εκρήξεις, οι οποίες μπορεί να έχουν αρνητικές συνέπειες τόσο για τους ίδιους όσο και για τους άλλους. (Δαλακούρα, 2011) Εξαρτάται από την αυτοεπίγνωση και αποτελεί ένα μεγάλο κομμάτι της ατομικής ικανότητας. Αυτοδιαχείριση σημαίνει να χρησιμοποιούμε την επίγνωση των συναισθημάτων μας για να είμαστε ευέλικτοι και να κατευθύνουμε θετικά την συμπεριφορά μας. Χωρίς την κατανόηση των συναισθημάτων μας δεν μπορούμε να τα διαχειριστούμε με συνέπεια να μας ελέγχουν αυτά. Η ξαφνική εφόρμηση αρνητικών συναισθημάτων αποδυναμώνει την ικανότητα του σκεπτόμενου εγκεφάλου να εστιάζει σε κάποιο συγκεκριμένο έργο π.χ. (στρατηγικός σχεδιασμός), είτε στη διαχείριση του συναισθήματος του θυμού, έτσι ώστε στην συνέχεια να επιλέξει ψύχραιμα την αντίδραση που θεωρεί πιο αποτελεσματική. (Αμοργιανός, 2010)

#### 3.1.4 Κοινωνική Επίγνωση

Η κοινωνική επίγνωση είναι η ικανότητα μας να συλλαμβάνουμε με ακρίβεια τα συναισθήματα των άλλων και να καταλαβαίνουμε τι συμβαίνει. Αυτό σημαίνει να αντιλαμβανόμαστε τι σκέφτονται οι άλλοι και τι αισθάνονται ακόμα και όταν εμείς δεν αισθανόμαστε το ίδιο. (Αμοργιανός, 2010) Ο ηγέτης που κατέχει την ικανότητα της ενσυναίσθησης, αντιλαμβάνεται ότι το συναισθηματικό μίγμα που δημιουργείται από μέλη μίας ομάδας και μπορεί να κατευθύνει προς όφελος του έργου αλλά και της ομάδας, αποφεύγοντας εντάσεις που καταστρέφουν το κλίμα και δυσκολεύουν την επίτευξη του στόχου τους. (Δαλακούρα, 2011) Η κοινωνική επίγνωση είναι εξαιρετικά σημαντική για την επίτευξη αρμονίας. Όταν ο ηγέτης είναι συντονισμένος με αυτό που νιώθουν οι άλλοι τη συγκεκριμένη στιγμή, μπορεί να πει και να πράξει ότι είναι κατάλληλο. (Αμοργιανός, 2010)

#### 3.1.5 Διαχείριση Σχέσεων

Η διαχείριση των σχέσεων συνίσταται από τις ακόλουθες υποκατηγορίες, την ηγεσία, την παρακίνηση, τη διαχείριση των συγκρούσεων, τη σύσφιξη των σχέσεων και τέλος την ομαδικότητα και τη συνεργασία. (Λιάκου, 2008) Η ικανότητα ενός ατόμου να

επικοινωνεί με τους άλλους, να δημιουργεί σχέσεις και να προκαλεί στους άλλους ανθρώπους τις αντιδράσεις που θέλει. Τα άτομα που κατέχουν τις ικανότητες αυτές δεν είναι απλώς φιλικοί με τους άλλους, αλλά είναι φιλικοί έχοντας ένα στόχο. Ο στόχος αυτός μπορεί να είναι συμφωνία σε ένα επιχείρημα, ο ενθουσιασμός για ένα καινούργιο προϊόν ή η αναγκαιότητα μιας αλλαγής στην επιχείρηση.

Τα άτομα αυτά δημιουργούν εύκολα ένα πλέγμα σχέσεων με τους ανθρώπους, το οποίο μπορούν να αξιοποιήσουν όποια στιγμή το χρειαστούν. Πιστεύουν στην αναγκαιότητα των άλλων και στην αξία της ομαδικής εργασίας. Αναγνωρίζουν έγκαιρα την ανάγκη για αλλαγές, δημιουργούν συμμάχους για να επιτύχουν τις αλλαγές που επιθυμούν και αποτελούν οι ίδιοι παράδειγμα προς μίμηση για τους άλλους (Δαλακούρα, 2011). Η επιδέξια διαχείριση των σχέσεων παραπέμπει στην ουσία του χειρισμού των συναισθημάτων των άλλων. Αυτό απαιτεί από τον ηγέτη να έχει επίγνωση των δικών του συναισθημάτων και να συντονίζεται με τους ανθρώπους που καθοδηγεί μέσω της ενσυναίσθησης. Η τέχνη της καλής διαχείρισης σχέσεων προαπαιτεί αυθεντικότητα και ειλικρίνεια.

Όταν ο ηγέτης είναι εναρμονισμένος με τα προσωπικά του οράματα και αξίες, πάντοτε στο θετικό συναισθηματικό φάσμα, και συντονίζεται με τα συναισθήματα της ομάδας, τότε είναι βέβαιο ότι οι δεξιότητες του στη διαχείριση των σχέσεων θα λειτουργήσουν καταλυτικά επιτυγχάνοντας την αρμονία (Αμοργιανός, 2010).

### 3.1.6 Διαστάσεις Συναισθηματικής Νοημοσύνης

Οι διαστάσεις της συναισθηματικής νοημοσύνης μπορεί να ταξινομηθούν σε:

- 1. Ανεξάρτητες:** Η καθεμιά συνεισφέρει με μοναδικό τρόπο στην εργασιακή και επικοινωνιακή αποτελεσματικότητα.
- 2. Αλληλεξαρτώμενες:** Η καθεμιά από αυτές επηρεάζει ως ένα βαθμό κάποιες άλλες, ενώ υπάρχουν πολλές ισχυρές αλληλεπιδράσεις.
- 3. Ιεραρχικές:** Οι διαστάσεις της συναισθηματικής νοημοσύνης “οικοδομούνται” η μία πάνω στην άλλη. Για παράδειγμα, η αυτοεπίγνωση είναι σημαντική για την αυτορύθμιση και την ενσυναίσθηση. Η αυτορύθμιση και η αυτοεπίγνωση συμβάλουν στην απόκτηση κινήτρων και όλες μαζί συμμετέχουν στις κοινωνικές δεξιότητες.

- 4. Αναγκαίες αλλά όχι ικανές:** Το να έχει κανείς μια βασική ικανότητα συναισθηματικής νοημοσύνης δεν του εγγυάται ότι θα αναπτύξει ή θα επιδείξει και τις σχετικές δεξιότητες, όπως είναι η συνεργασία ή η ηγεσία. Παράγοντες όπως το κλίμα ενός εκπαιδευτικού οργανισμού ή το ενδιαφέρον ενός ατόμου για την εργασία του θα καθορίσουν επίσης αν θα εκδηλωθεί η συγκεκριμένη ικανότητα.
- 5. Γενικού τύπου:** Ο γενικός κατάλογος ισχύει, σε κάποιο βαθμό, για όλες τις εργασίες. Ωστόσο, διαφορετικές εργασίες έχουν και διαφορετικές απαιτήσεις σε ότι αφορά τις απαραίτητες δεξιότητες. (Λιάκου, 2008)

### 3.1.7 Παράγοντες Μέτρησης Ερωτηματολογίου Συναισθηματικής Νοημοσύνης

Το ερωτηματολόγιο που χρησιμοποιήθηκε για την αξιολόγηση της συναισθηματικής νοημοσύνης στην παρούσα εργασία ήταν το Trait Emotional Intelligence (TEIQue) του K.V Petrides η συντομευμένη μορφή (short form) που περιλαμβάνει 30 ερωτήσεις οι οποίες ομαδοποιούνται και διαχωρίζονται σε τέσσερις βασικές κατηγορίες που αναλύονται ως εξής:

**Ευημερία (Well-being):** Υψηλές τιμές σε αυτόν τον παράγοντα αντικατοπτρίζουν μια γενική εικόνα της ευημερίας του ερωτηθέντος, βάση τόσο παρελθοντικών όσο και μελλοντικών προσδοκιών. Σε γενικές γραμμές, άτομα με υψηλά αποτελέσματα αισθάνονται ευτυχισμένοι και ολοκληρωμένοι. Ενώ, από την άλλη μεριά, άτομα με χαμηλά αποτελέσματα τείνουν να είναι δυσαρεστημένοι από την ζωή, όπως αυτή είναι στο παρόν.

**Αυτοέλεγχος (Self-control):** Υψηλά αποτελέσματα ερωτηθέντων δείχνουν υγιή βαθμό ελέγχου πάνω στις παρορμήσεις και επιθυμίες τους. Επιπρόσθετα, εκτός από την παράκρουση παρορμήσεων, τα άτομα με υψηλό αυτοέλεγχο είναι καλά στο να διαχειρίζονται εξωτερικές πιέσεις και άγχος. Ακόμα, δεν είναι ούτε καταπιεσμένοι, ούτε εκφραστικοί. Σε αντίθεση, τα άτομα με χαμηλά αποτελέσματα, τείνουν σε παρορμητικές συμπεριφορές και δείχνουν να είναι ανίκανοι να διαχειριστούν το άγχος.

**Συναισθηματικότητα (Emotionality):** Άτομα με υψηλά αποτελέσματα σε αυτόν τον παράγοντα πιστεύουν ότι έχουν ένα μεγαλύτερο εύρος ικανοτήτων που αφορά

διαχείριση συναισθημάτων. Μπορούν να δεχθούν και να εκφράσουν συναισθήματα και να χρησιμοποιήσουν τις ικανότητες αυτές για να αναπτύξουν και να διατηρήσουν στενές σχέσεις με σημαντικά, για αυτούς, άτομα. Στην περίπτωση που υπάρχουν χαμηλά αποτελέσματα, οι ερωτηθέντες, βρίσκουν δύσκολο το να αναγνωρίζουν τις εσωτερικές συναισθηματικές τους καταστάσεις και να τις εκφράσουν σε άλλους, πράγμα το οποίο οδηγεί σε λιγότερο καρποφόρες προσωπικές σχέσεις.

**Κοινωνικότητα (Sociability):** Ο παράγοντας της κοινωνικότητας διαφέρει κυρίως από αυτόν της συναισθηματικότητας λόγω ότι ο πρώτος δίνει έμφαση στις κοινωνικές σχέσεις και επιρροή. Εστιάζει στην περίπτωση στην οποία, το άτομο, λειτουργεί κάτω από διαφορετικά κοινωνικά πλαίσια και όχι στις προσωπικές σχέσεις με την οικογένεια και στενούς φίλους. Υψηλά αποτελέσματα παρουσιάζουν καλύτερη δυνατότητα κοινωνικής συναναστροφής. Τα άτομα πιστεύουν ότι διαθέτουν καλές δυνατότητες ως προς το να ακούν και να καταλαβαίνουν, να μπορούν να επικοινωνήσουν καθαρά και με αυτοπεποίθηση με ανθρώπους από διαφορετικά υπόβαθρα. Αυτοί με χαμηλά αποτελέσματα, πιστεύουν πως είναι ανίκανοι να επηρεάσουν συναισθήματα άλλων και είναι λιγότερο πιθανόν να είναι καλοί διαπραγματευτές. Είναι αμφίβολοι για το τι μπορεί να πουν ή να κάνουν σε κοινωνικές περιστάσεις και ως αποτέλεσμα εμφανίζονται ντροπαλοί και απομονωμένοι. (Petrides, 2001)

### 3.2 Ερωτηματολόγιο Διαταραχών Προσωπικότητας (IPDE)

Το ερωτηματολόγιο Διεθνούς Διερεύνησης Διαταραχών Προσωπικότητας (IPDE) είναι μια ημιδομημένη κλινική συνέντευξη που έχει σχεδιαστεί με σκοπό την αξιολόγηση των διαταραχών προσωπικότητας σύμφωνα με τα ταξινομικά συστήματα διάγνωσης DSM-IV και ICD-10. Το ερωτηματολόγιο δεν είναι σχεδιασμένο για να διερευνήσει ολόκληρο το πεδίο και τις πτυχές της προσωπικότητας ενός ατόμου, καθώς σκοπός του επικεντρώνεται στο να διερευνήσει εκείνα τα χαρακτηριστικά και τις συμπεριφορές που είναι σχετικά και σύμφωνα με μια αξιολόγηση των κριτηρίων βάση των ταξινομικών συστημάτων διάγνωσης. Η διερεύνηση, γίνεται εξετάζοντας έξι ενότητες: Εργασία, εαυτός, διαπροσωπικές σχέσεις, συναισθήματα, έλεγχος της

πραγματικότητας και έλεγχος παρορμήσεων. (Loranger) (Vegue-Gonzalez & Alvaro-Brun, 2008)

*Ως διαταραχή προσωπικότητας στο DSM-V ορίζεται ένα σταθερό πρότυπο εσωτερικής εμπειρίας και συμπεριφοράς το οποίο αποκλίνει σημαντικά από τις προσδοκίες που καθορίζονται από την κουλτούρα του ατόμου και εκδηλώνεται τουλάχιστον σε δύο από τις παρακάτω περιοχές: Σκέψη, συναίσθημα, διαπροσωπική λειτουργικότητα ή έλεγχο των παρορμήσεων. Αυτό το σταθερό πρότυπο είναι άκαμπτο και εμμένον σε ένα μεγάλο εύρος κοινωνικών και προσωπικών καταστάσεων. Οδηγεί σε κλινικά σημαντική περιοχή της λειτουργικότητας. Το πρότυπο είναι σταθερό και χαρακτηρίζεται από μακρά διάρκεια και η έναρξή του ανιχνεύεται τουλάχιστον στην εφηβεία ή στην πρώτη νεότητα. (Ευσταθίου, 2005)*

### 3.2.1 Ιστορικά Στοιχεία

Το IPDE είναι μία εξέλιξη με τροποποιήσεις κατάλληλες για διεθνή χρήση της Διερεύνησης Διαταραχών Προσωπικότητας (PDE). Για να γίνει με επιτυχία αυτή η εξέλιξή του, χρειάστηκαν πολλαπλές συναντήσεις εκπροσώπων της ψυχιατρικής κοινότητας από όλες τις χώρες, με αφετηρία αυτών, το 1985. Μερικά από τα κύρια θέματα συζήτησης των συναντήσεων ήταν η δομή του ερωτηματολογίου, η αλλαγή ορολογίας και η δημιουργία κλιμάκων αξιολόγησης. Μετά από αρκετές δοκιμές και μεταφράσεις του, την περίοδο 1988-1989, έλαβε μέρος η γενική δοκιμή στον γενικό πληθυσμό δεκατεσσάρων συμμετεχόντων κέντρων σε έντεκα χώρες στην Βόρεια Αμερική, Ευρώπη, Αφρική και Ασία.

Περαιτέρω τροποποίηση στο ερωτηματολόγιο έγινε τον Αύγουστο του 1991, όπου οι διευθυντές ερευνητές της γενικής δοκιμής συγκεντρώθηκαν στα κεντρικά του Παγκόσμιου Οργανισμού Υγείας με σκοπό να συζητήσουν τα αποτελέσματα της έρευνας και την εμπειρία των ερωτηθέντων με το IPDE, όπως καθώς έγιναν συμπληρωματικές τροποποιήσεις για την σωστή μετάβαση από το ταξινομικό σύστημα DSM-III-R στο DSM-IV. Για να εξαλειφθούν οι ανησυχίες που αφορούσαν την διάρκεια της εξέτασης και για να γίνει κοινώς αποδεκτό από κλινικό προσωπικό και ερευνητές, αποφασίστηκε ο ορισμός ενοτήτων στο IPDE.



Μεταξύ άλλων, το IPDE, έχει μεταφραστεί στις εξής γλώσσες: Γαλλικά, Γερμανικά, Ιαπωνικά, Ελληνικά, Ισπανικά κ.α. Για την αποφυγή λαθών, αφού τελείωνε κάποια μετάφραση, η διαδικασία άρχιζε ξανά προς τα Αγγλικά, και με αυτόν τον τρόπο το ερωτηματολόγιο αποδιδόταν ουσιαστικά προς άλλες γλώσσες.

Προβλήματα μπορεί να παρουσιαστούν όταν οι ερωτηθέντες μιλάνε κάποια από τις μη διαδεδομένες διαλέκτους ή κάποια που να μην έχει ακόμα μεταφραστεί το IPDE. Σε αυτή την περίπτωση, ο ερευνητής, πρέπει να αυτοσχεδιάσει τις ερωτήσεις του ερωτηματολογίου με τέτοιο τρόπο όπου θα είναι και αντιληπτές από τον ερωτηθέντα και οι απαντήσεις αυτού να ανήκουν σε ένα ανεκτό όριο λάθους, λόγω ελλιπής αντιστοίχισης του κειμένου με αυτό κάποιας μεταφρασμένης γλώσσας. (Janca, Sartorius, & Loranger, 2007)

### 3.2.2 Δομή

Το IPDE διατάσσεται σε μορφή που επιχειρεί να παρέχει τη βέλτιστη ισορροπία μεταξύ μιας αυθόρμητης στην απάντηση, φυσικής κλινικής συνέντευξης.

Οι ερωτήσεις, στην συνέντευξη IPDE, είναι 99 και ακολουθούν η μία την άλλη με τρόπο που είναι αποδεκτές για τον κλινικό ειδικό. Είναι τοποθετημένες σε έξι κατηγορίες : Εργασία, εαυτός, διαπροσωπικές σχέσεις, συναισθήματα, έλεγχος της πραγματικότητας και έλεγχος παρορμήσεων. Οι τίτλοι δεν είναι μόνο απλές ονομασίες της αρεσκείας, αλλά διαδραματίζουν οργανωτικούς ή θεματικούς ρόλους. Κατά καιρούς η διπλή φύση των έξι τομέων απαιτούσε την κάπως αυθαίρετη κατανομή των ερωτήσεων. Για καλύτερη αποτελεσματικότητα και ευκολία, μερικές φορές ένα ζήτημα εκτείνεται πέρα από το πεδίο του τμήματος όπου εμφανίζεται. Για παράδειγμα, πολλά καταναγκαστικά κριτήρια είναι καλύτερα αξιολογούμενα στο πλαίσιο της λειτουργίας στην εργασία, αλλά λαμβάνεται υπόψιν, επίσης, και η συμπεριφορά έξω από την εργασία, ακόμη και αν οι ερωτήσεις εμφανίζονται στην ενότητα «Εργασία» της συνέντευξης.

Οι ενότητες αρχίζουν με ερωτήσεις ανοιχτού τύπου προσφέροντας στον ερωτώμενο την ευκαιρία να αναλύσει το θέμα αυτής. Αυτό βοηθά στην ανάπτυξη μιας εικόνας για τις ερωτήσεις που ακολουθούν, και παρέχει μια περίληψη της προηγούμενης ενότητας. Παρόλο που δεν βαθμολογούνται, ως εκ τούτου, αυτές οι αναλύσεις του

θέματος παρέχουν ένα υπόβαθρο βάσει του οποίου θα κρίνουν την επιστημονική σημασία ορισμένων από τις απαντήσεις σε ερωτήματα που ακολουθούν.

Για την εξοικονόμηση χρόνου, χρησιμοποιείται το ερωτηματολόγιο IPDE. Είναι ένας τρόπος ανίχνευσης διαταραχών προσωπικότητας με στόχο να αποφευχθεί η συνέντευξη σε εκείνους που δεν έχουν λάβει διάγνωση διαταραχής προσωπικότητας. Περιλαμβάνει 77 ερωτήσεις με δύο πιθανές απαντήσεις, σωστό ή λάθος, και μπορεί να εφαρμοστεί και να συμπληρωθεί με τη βοήθεια ενός λογισμικού υπολογιστή. Το αποτέλεσμα της είναι μόνο κατηγορικά, έτσι δεν μπορούν να χρησιμοποιηθούν για προσδιορισμό βαθμολογικών διαστάσεων. (Janca, Sartorius, & Loranger, 2007)

### 3.2.3 Παράγοντες Μέτρησης Ερωτηματολογίου Διαταραχών Προσωπικότητας

Οι Διαταραχές Προσωπικότητας και παράγοντες του ερωτηματολογίου IPDE, μαζί με την ερμηνεία τους (Ευσταθίου, 2005):

#### **Παρανοειδής (Paranoid):**

- Έλλειψη εμπιστοσύνης και καχυποψία.
- Απόδοση αρνητικών κινήτρων στη συμπεριφορά των άλλων.

#### **Σχιζοειδής (Schizoid):**

- Απομάκρυνση από τις κοινωνικές σχέσεις.
- Περιορισμένο εύρος συναισθηματικής έκφρασης.

#### **Σχιζοτυπική (Schizotypal):**

- Έντονη δυσφορία στις στενές διαπροσωπικές σχέσεις.
- Γνωστικές ή αντιληπτικές διαστρεβλώσεις.
- Εκκεντρική συμπεριφορά.

#### **Αντικοινωνική (Antisocial):**

- Ανεύθυνη συμπεριφορά.
- Απειθαρχία.
- Ασέβεια προς τα κοινωνικά πρότυπα και τους κανόνες.

#### **Οριακή (Borderline):**

- Αστάθεια στις διαπροσωπικές σχέσεις, στην εικόνα του εαυτού και στο συναίσθημα.

- Έντονη παρορμητικότητα.

**Οιστριονική (Histrionic):**

- Υπερβολική έκφραση συναισθήματος και αναζήτηση προσοχής.

**Ναρκισσιστική (Narcissistic):**

- Αίσθηση προσωπικού μεγαλείου.
- Ανάγκη για θαυμασμό.
- Έλλειψη ενσυναίσθησης.
- Αποφευκτή Κοινωνική αναστολή.
- Αισθήματα αναξιότητας.
- Υπερβολική ευαισθησία στην αρνητική κριτική.

**Εξαρτητική (Dependent):**

- Συμπεριφορές υποταγής και προσκόλλησης.

**Ιδεοψυχαναγκαστική (Obsessive – Compulsive):**

- Εμμονή με την τάξη.
- Τελειοθηρία.
- Ανάγκη για έλεγχο

**Αποφευκτική (Avoidant):**

- Διστακτικότητα
- Αμηχανία
- Άγχος

**Δημογραφικά Στοιχεία**

- Birthyear: Γενέθλια
- Sex: Φύλλο
- EducationMe: Επίπεδο Μόρφωσης
- cityBirthMe: Πόλη Καταγωγής
- citySameMe: Πόλη Διαμονής & Καταγωγής
- cityStayMe: Πόλη Διαμονής
- AgeFather: Ηλικία Πατέρα
- ageMother: Ηλικία Μητέρας
- educationFather: Μόρφωση Πατέρα
- educationMother: Μόρφωση Μητέρας
- JobFather: Εργασία Πατέρα
- JobMother: Εργασία Μητέρας

## 4<sup>ο</sup> ΚΕΦΑΛΑΙΟ

---

### 4 Εισαγωγή στις Τεχνικές Εξόρυξης Γνώσης (Introduction to Data Mining Techniques)

#### 4.1 Ορισμός – Έννοιες Data Mining

Η εξόρυξη δεδομένων έχει προσελκύσει την προσοχή των εταιριών πληροφόρησης και της κοινωνίας ευρύτερα. Η τεράστια διαθεσιμότητα ποσοτήτων δεδομένων έχει κάνει την εξόρυξη επιτακτική ανάγκη για την μετατροπή όλων αυτών των δεδομένων σε χρήσιμες πληροφορίες που μας παρέχουν γνώση πάνω στο αντικείμενο που εξετάζουμε. Επιπροσθέτως χρησιμεύει στην ανάλυση της αγοράς ,την ανίχνευση απάτης καθώς επίσης και τον έλεγχο της παραγωγής και την εξερεύνηση της επιστήμης. Κάποιες από τις βασικές τεχνικές που χρησιμοποιούμε είναι η κατηγοριοποίηση και οι κανόνες συσχέτισης. (Μ.Βαζιργιάννης, 2003) Σε αυτό θα μας βοηθήσουν οι διαφορές εφαρμογές αλγορίθμων που έχουμε στην κατοχή μας.

Η δήλωση των σχέσεων και η σύναψη των στοιχείων συχνά αναφέρονται ως μοντέλο ή πρότυπο. Βασικοί στόχοι της Εξόρυξης Δεδομένων είναι η περιγραφή και η πρόβλεψη. Δηλαδή η αναγνώριση των προτύπων που επικρατούν σε ένα μεγάλο σύνολο δεδομένων και η δημιουργία προβλέψεων όσον αφορά τη μελλοντική αξία ή συμπεριφορά κάποιων μεταβλητών. Η αναγνώριση των προτύπων γίνεται μέσω γραμμικών εξισώσεων, κανόνων, απόδοσης γραφημάτων και δομών σε μορφή δέντρου καθώς και επαναλαμβανόμενων προτύπων σε μορφή χρονοσειρών.

Ο σκοπός της Εξόρυξης Δεδομένων δεν είναι η ανάπτυξη στρατηγικής ως προς την συλλογή δεδομένων. Αυτός είναι ένας λόγος που η Εξόρυξη Δεδομένων διαφέρει εν μέρει από τη Στατιστική. Στην ουσία η διάφορα της Στατιστικής είναι ότι εκεί συλλέγονται συχνά δεδομένα χρησιμοποιώντας αποτελεσματικές στρατηγικές με σκοπό να απαντηθούν συγκεκριμένα ζητήματα. Για τον λόγο αυτό, η Εξόρυξη Δεδομένων αναφέρεται και ως «δευτερογενής» ανάλυση δεδομένων.

Η Εξόρυξη Δεδομένων χρησιμοποιεί έννοιες όπως δειγματοληψία, εκτίμηση και έλεγχος υποθέσεων από τη Στατιστική, καθώς και εφαρμογές όπως αναζήτηση αλγορίθμων, τεχνικές δημιουργίας υποδειγμάτων (*modeling techniques*), θεωρίες

τεχνητής νοημοσύνης, αναγνώρισης προτύπων και μηχανικές εκμάθησης. Άλλες έννοιες στις οποίες βασίζεται η Εξόρυξη Δεδομένων είναι η θεωρία της βελτιστοποίησης, εξελικτικός υπολογισμός, θεωρία της πληροφορίας, έλεγχος σημάτων και ανάκτηση πληροφορίας. Επιπλέον υπάρχουν αρκετοί άλλοι τομείς των επιστήμων που στήριξαν την πρόοδο της Εξόρυξης Δεδομένων. Για παράδειγμα, η τεχνολογία βάσεων δεδομένων δύναται να παρέχει βοήθεια μέσω τεχνικών αποτελεσματικής αποθήκευσης, διεξαγωγής ερωτημάτων και ευρετηριοποίησης. Ακόμη τεχνικές υψηλής απόδοσης από υπολογιστικής πλευράς και σχετικές με την ταξινόμηση παρέχουν βοήθεια σε σχέση με την διαχείριση του μεγέθους και της συλλογής των τεραστίων συνόλων δεδομένων. (Hand, Mannila, & Smyth, 2001)

## 4.2 Εφαρμογές Data Mining

Στο σημείο αυτό θα αναφέρουμε συνοπτικά ορισμένες εφαρμογές της μεθόδου Data Mining οι οποίες έχουν συλλεχθεί από διάφορα επιστημονικά πεδία. ([www.datamining.gr](http://www.datamining.gr), n.d.)

- Direct mail marketing – Μάρκετινγκ μέσω mail

Το Body Shop International δοκιμάζει τεχνικές Data Mining ώστε να καταφέρει να αυξήσει την αποτελεσματικότητα των παραγγελιών μέσω mail. Τα διοικητικά στελέχη ενδιαφέρονται στο να μειώσουν το κόστος αποστολής διαφημιστικών καταλόγων, εστιάζοντας μόνο σε πελάτες οι οποίοι θεωρούνται ως «κερδοφόροι».

- Category management and inventory control – Κατηγοριοποίηση μέσω μάνατζμεντ, και έλεγχος για κάτι νέο

Η εταιρεία Rubbermaid χρησιμοποιεί μεθόδους Data Mining για να κατηγοριοποιεί το στυλ της στρατηγικής που χρησιμοποιεί ανάλογα με την αγορά που εστιάζεται, στο να αγοράζει και να πουλά τα οικονομικά αγαθά που παράγει.

- Market Basket Analysis (MBA) – Ανάλυση καλαθιού της νοικοκυράς

Τα στελέχη της J.Crew Group συνδυάζουν click system analysis μέσα από το επίσημο web-site τους σε συνδυασμό με την μέθοδο point-of-sale (POS) στις λιανικές τους πωλήσεις. Με αυτό τον τρόπο θέλουν να δούνε τι ρούχα, τι υποδήματα και άλλα

αξεσουάρ αγοράζονται μαζί. Τα δεδομένα μετά θα αναλυθούν και έπειτα θα σταλθούν κατάλογοι με οικονομικά αγαθά και προσφορές των σε on-line αγοραστές.

- Customer relationship management (CRM) and customer profiles – Καταλαβαίνοντας το προφίλ του κάθε πελάτη

Σύμφωνα με διάφορες εταιρείες λιανικής, άλλες από τον τραπεζικό τομέα είναι δυνατό χρησιμοποιώντας τεχνικές Data Mining, όπως κανόνες ταξινόμησης ή και ομαδοποίησης να κατατάσσουμε τους πελάτες ανάλογα με προσωπικά τους στοιχεία και την προηγούμενη τους συμπεριφορά προς την εταιρεία που μας ενδιαφέρει σε «καλούς», «μέτριους» και «κακούς». Έτσι δίνεται η δυνατότητα να επιλέγουν το αγοραστικό κοινό με το οποίο συναλλάσσονται και στο οποίο εστιάζουν την πολιτική της εταιρείας.

- Astronomy - Αστρονομία

Έχει κατασκευαστεί ένα σύστημα με την ονομασία SKICAT από το JPL/Caltech και χρησιμοποιείται από τους αστρονόμους στο να αναγνωρίζουν αυτόματα τους διάφορους γαλαξίες και αστεροειδείς σε μία μεγάλη κλίμακα η οποία περιέχει διάφορα αστρονομικά μεγέθη.

- Biology – Βιολογία

Διάφορα συστήματα έχουν κατασκευαστεί στο να εξάγουν κανόνες που αφορούν την δομή των οργανισμών, την ανάλυση του DNA, καθώς και την δυνατότητα για εύρεση φαρμάκων για την καταπολέμηση ασθενειών.

- Global Climate Modeling – Παγκόσμιο μοντέλο κλιματολογικών συνθηκών

Διάφορα συστήματα έχουν υιοθετηθεί τα οποία επιτρέπουν την ανάλυση κλιματολογικών συνθηκών και με αυτό τον τρόπο δίνουν την δυνατότητα προβλέψεων κλιματολογικών φαινομένων όπως οι κυκλώνες, οι καταιγίδες, οι καύσωνες και άλλα πολλά.

- Data Mining for financial applications - Η μέθοδος Εξόρυξης από Δεδομένα σε οικονομικές εφαρμογές

Πολλές φορές τεχνικές της μεθόδου Εξόρυξης από Δεδομένα όπως τα νευρωνικά δίκτυα και τα δένδρα αποφάσεων μπορούν να χρησιμοποιηθούν από οικονομικούς

αναλυτές για την λήψη στρατηγικών αποφάσεων στο ανάλογο οικονομικό πεδίο που ενδιαφέρει κάθε φορά. Φυσικά χρειάζεται και το ανάλογο υπόβαθρο από ιστορικά δεδομένα ώστε να είναι δυνατή η ανάλυση που θα γίνει. (Παγουρόπουλος , 2006)

### 4.3 Βασικά Στάδια Της Εξόρυξης Δεδομένων

Ο διαχωρισμός αυτός σε συγκεκριμένα στάδια θα μας βοηθήσει να κάνουμε μια αρχική προσέγγιση των διαδικασιών της Εξόρυξης Δεδομένων, αλλά και να εμπεδώσουμε στην συνέχεια τις απατήσεις της. (Frawley, Piatetsky, & Matheus, 1992)

Ως βασικά στάδια εξόρυξης δεδομένων ορίζονται:

1. Περιγραφή Μοντέλου
2. Αξιολόγηση Μοντέλου
3. Αλγόριθμος Αναζήτησης
4. Διαχωρισμός Μεθόδων Εξόρυξης Δεδομένων

#### 4.3.1 Περιγραφή Μοντέλου

Στο πρώτο στάδιο της Εξόρυξης Δεδομένων, επιχειρούμε να δηλώσουμε τον στόχο μας, όπως για παράδειγμα την ταξινόμηση, την παλινδρόμηση την συσταδοποίηση ή τους κανόνες συσχέτισης. Αυτοί οι εναλλακτικοί τύποι Εξόρυξης Δεδομένων θα σχολιαστούν εκτενέστερα παρακάτω καθώς καθένας από αυτούς εξυπηρετεί διαφορετικό σκοπό. Επίσης μας ενδιαφέρει η παραστατική μορφή του μοντέλου δηλαδή η απεικόνιση των δεδομένων και να είναι δυνατόν να ερμηνευθεί. Χαρακτηριστικά παραδείγματα μοντέλων αποτελούν τα δέντρα απόφασης, τα γραφικά μοντέλα, τα νευρωνικά δίκτυα και τα μοντέλα-σύστημα που βασίζονται σε παραδείγματα ή πιθανότητες.

#### 4.3.2 Αξιολόγηση Μοντέλου

Ύστερα από την δημιουργία του μοντέλου οφείλουμε να εξετάσουμε κατά ποσό ταιριάζει με τις συνθήκες της KDD (Knowledge Discovery Database). Προχωρούμε δηλαδή στην αξιολόγηση του μοντέλου έτσι ώστε να κρίνουμε την εγκυρότητα των προτύπων και την ακρίβεια και χρησιμότητα του μοντέλου. Ένα κριτήριο αξιολόγησης είναι αυτό της μεγίστης πιθανότητας (Frawley, Piatetsky, & Matheus, 1992).

#### 4.3.3 Αλγόριθμος Αναζήτησης

Στόχος του αλγόριθμου αναζήτησης είναι η σύγκριση μοντέλων και παραμέτρων, του συνόλου δεδομένων, τους οικογένειας μοντέλων και του κριτηρίου αξιολόγησης. Οι βασικότεροι τύποι αλγορίθμων αναζήτησης είναι αυτοί που αναζητούν παραμέτρους βελτιστοποίησης τους κριτηρίου αξιολόγησης και αυτοί που αναζητούν μοντέλα αντιπροσώπευσης των δεδομένων.

#### 4.3.4 Διαχωρισμός Μεθόδων Εξόρυξης Δεδομένων

Έχει γίνει κατανοητό ότι υπάρχουν αρκετές μέθοδοι Εξόρυξης Δεδομένων, οι οποίες χρησιμοποιούνται για διαφορετικούς σκοπούς και μπορούν να καλύψουν τους στόχους. Η ποικιλία των μεθόδων είναι τόσο μεγάλη που καθιστά αναγκαία την ταξινόμηση τους σε σχετικές ομάδες ανάλογα με τους στόχους που μπορεί να καλύψει κάθε ομάδα. Οι μέθοδοι ανακάλυψης είναι αυτές που εντοπίζουν αυτόματα πρότυπα στα δεδομένα. Το στάδιο αυτό χωρίζεται στην *περιγραφή* και την *πρόβλεψη*. (Μ.Βαζιργιάννης, 2003)



## 4.4 Περιγραφική Μοντελοποίηση (Descriptive Modeling)

### 4.4.1 Ορισμός

Ο στόχος ενός μοντέλου περιγραφής είναι να γίνει περιγραφή όλου του συνόλου δεδομένων ή της διαδικασίας που παρέχει τα δεδομένα. Ας σκεφτούμε ως εφαρμογή περιγραφές που περιλαμβάνουν μοντέλα για την κατανομή πιθανότητας των δεδομένων, τη διαμέριση ενός χώρου  $n$  διαστάσεων σε ομάδες ή την περιγραφή των σχέσεων μεταξύ των μεταβλητών (εξαρτημένα μοντέλα).

Οι μέθοδοι περιγραφής έχουν ως στόχο την ερμηνεία των δεδομένων και επικεντρώνονται στην κατανόηση του τρόπου που σχετίζονται τα δεδομένα. Αυτό για παράδειγμα γίνεται μέσω την νοερής απεικόνισης (visualization) ή της σύνοψης (summarization) οι οποίες μπορούμε να πούμε ότι αποτελούν μέρος της Διερευνητικής Ανάλυσης Δεδομένων. Η σημαντικότερη εφαρμογή των περιγραφικών μοντέλων είναι η ανάλυση συστάδων που αναφέραμε και προηγουμένως.

### 4.4.2 Κανόνες Συσχέτισης (Association Rules)

Η εξαγωγή κανόνων συσχέτισης (association rules) θεωρείται μια από τις σημαντικότερες διεργασίες εξόρυξης δεδομένων. Έχει προσελκύσει ιδιαίτερο ενδιαφέρον καθώς οι κανόνες συσχέτισης παρέχουν έναν συνοπτικό τρόπο για να εκφραστούν οι ενδεχομένως χρήσιμες πληροφορίες που γίνονται εύκολα κατανοητές από τους τελικούς χρήστες. Οι κανόνες συσχέτισης ανακαλύπτουν κρυμμένες «συσχετίσεις» μεταξύ των γνωρισμάτων ενός συνόλου των δεδομένων. Αυτοί οι συσχετισμοί παρουσιάζονται στην ακόλουθη μορφή:  $A \rightarrow B$  όπου το  $A$  και το  $B$  αναφέρονται στα σύνολα γνωρισμάτων που υπάρχουν στα υπό ανάλυση δεδομένα. Υποθέστε ότι μας δίνεται ένα σύνολο συναλλαγών  $S = \{S_1, \dots, S_n\}$ , όπου κάθε συναλλαγή είναι ένα υποσύνολο του  $A = \{A_1, \dots, A_k\}$  ( $A_i, i = 1, 2, \dots, K$ , είναι οι ιδιότητες του συνόλου δεδομένων). Για ένα δεδομένο σύνολο  $A \subseteq A$ , η υποστήριξη του  $A$ ,  $\text{sup}(A)$ , καθορίζεται ώστε να είναι ο αριθμός συναλλαγών στο  $S$  που είναι υπερσύνολα του  $A$  (δηλαδή το  $A$  εμφανίζεται σε αυτές τις συναλλαγές). Εάν η υποστήριξη ενός συνόλου αντικειμένων  $A$  είναι μεγαλύτερη από ένα καθορισμένο από τον χρήστη κατώτατο όριο υποστήριξης  $T$ , τότε ονομάζουμε το  $A$  ως συχνό σύνολο.

Η σημασία ενός τέτοιου κανόνα είναι ότι οι συναλλαγές στο σύνολο δεδομένων, που περιέχουν τις ιδιότητες του A, τείνουν επίσης να περιέχουν τις ιδιότητες του B. Σημειώνουμε επίσης ότι οι κανόνες συσχέτισης που εξάγονται πρέπει να μπορούν επίσης να ικανοποιούν κι άλλους περιορισμούς που καθορίζονται από το χρήστη, σχετικούς με τα μέτρα των κανόνων συσχέτισης. Λαμβάνοντας υπόψη την ανωτέρω περιγραφή, μια σημαντική δευτερεύουσα λειτουργία που συνήθως λύνεται πρώτη είναι αυτή του υπολογισμού των συχνών συνόλων. Δηλαδή, λαμβάνοντας υπόψη ένα σύνολο συναλλαγών S, υπολογίζονται όλα τα συχνά υποσύνολα του A (για το δεδομένο κατώτατο όριο υποστήριξης T). Μόλις βρεθούν τα συχνά σύνολα, το πρόβλημα του υπολογισμού των κανόνων συσχέτισης από αυτά γίνεται πολύ απλό. Για κάθε συχνό σύνολο A, και για κάθε  $B \subset A$  μπορεί να εξεταστεί η εμπιστοσύνη του κανόνα  $A/B \rightarrow B$ . (Μπαμπαλιάρης, 2011)

Γενικότερα το πρόβλημα εύρεσης των κανόνων συσχέτισης που πληρούν τις επιθυμητές τιμές επιβεβαίωσης και αξιοπιστίας μπορούν να διαιρεθούν σε δύο υποπροβλήματα:

- ∅ **Εύρεση** όλων των συνδυασμών των προϊόντων που έχουν επιβεβαίωση πάνω από την ελάχιστη επιβεβαίωση. Αυτοί οι συνδυασμοί ονομάζονται μεγάλες λίστες από προϊόντα (large itemsets) και όλοι οι υπόλοιποι συνδυασμοί μικρές λίστες από προϊόντα (small itemsets).
- ∅ **Χρήση** όλων των μεγάλων λιστών από προϊόντα για εξόρυξη των κανόνων Συσχέτισης που ικανοποιούν την ελάχιστη αξιοπιστία. Για παράδειγμα, έστω ότι τα ABCD και AB είναι μεγάλες λίστες από προϊόντα. Μπορούμε να καθορίσουμε αν ο κανόνας Συσχέτισης  $AB \Rightarrow CD$  ξεπερνά την ελάχιστη αξιοπιστία, υπολογίζοντας το λόγο  $r$ , που δίνει  $r = \text{επιβεβαίωση (ABCD)} / \text{επιβεβαίωση (AB)}$ . Στην περίπτωση που  $r \geq \text{ελάχιστη αξιοπιστία}$ , ο κανόνας γίνεται αποδεκτός. Η εύρεση μεγάλων λιστών από προϊόντα, για να αποφεύγει ένα εξαντλητικό ψάξιμο όλων των συνδυασμών βασίζεται στο γεγονός ότι μία λίστα είναι μεγάλη όταν κάθε υποσύνολό της είναι μεγάλη λίστα από προϊόντα.

#### 4.4.2.1 Αλγόριθμος Apriori

Ο αλγόριθμος Apriori έχει προταθεί από τους R. Agrawal R. Srikant το 1994. Ο αλγόριθμος χρησιμοποιείται για ανόρυξη συχνών συνόλων αντικειμένων (itemsets) για εξόρυξη κανόνων συσχέτισης. Ο αλγόριθμος έχει πάρει το όνομα του από την προγενέστερη γνώση (prior knowledge) των χαρακτηριστικών των συχνών συνόλων αντικειμένων, που χρησιμοποιεί. Ο Apriori υιοθετεί την τεχνική αναζήτηση, level-wise, η οποία είναι μια επαναλαμβανόμενη τεχνική που χρησιμοποιεί τα k-itemsets για να κτίσει τα (k+1)-itemsets. Στην αρχή, ο αλγόριθμος βρίσκει τα συχνά εμφανιζόμενα 1-itemsets (το σύνολο αντικειμένων με 1 μόνο χαρακτηριστικό). Ο αλγόριθμος αναζητά και συναθροίζει τον αριθμό που εμφανίζεται κάθε αντικείμενο – χαρακτηριστικό στη βάση δεδομένων, και μετά συλλέγει τα αντικείμενα που ικανοποιούν το ελάχιστο support, στο σύνολο L1. Κατόπιν, χρησιμοποιώντας το σύνολο L1, χτίζεται το σύνολο L2 το οποίο περιλαμβάνει όλα τα συχνά σύνολα αντικειμένων με 2 χαρακτηριστικά (2-itemsets), το οποίο κι αυτό χρησιμοποιείται για να χτιστεί το L3, και ούτω κάθε εξής, μέχρι που να μην μπορεί βρεθεί άλλο σύνολο με k-itemsets, δηλαδή το Lk να είναι κενό. Για να βρεθεί κάθε Lk απαιτείται μία αναζήτηση της βάσης δεδομένων. (Παπακωνσταντίνου, 2009)

##### **Βήματα Αλγορίθμου Apriori**

1. Βρίσκουμε τα αγαθά που εμφανίζονται περισσότερο από την ελάχιστη επιβεβαίωση (minimum support), δηλαδή το σύνολο L1= μεγάλες λίστες από 1 αγαθό (large 1-item sets)
2. Από k=2 και όσο Lk-1 δεν είναι κενό κάνε: A) Βρες το σύνολο Ck των υποψήφιων μεγάλων λιστών από k αγαθά (candidate large k-item sets) με βάση το Lk-1 B) Βρες ποία από αυτά εμφανίζονται περισσότερο από την ελάχιστη επιβεβαίωση και φτιάξε το σύνολο Lk= μεγάλες λίστες από k αγαθά.
3. Για κάθε στοιχείο των L1, ... , Ln βρες ποία ικανοποιούν την ελάχιστη αξιοπιστία (minimum confidence).

Στο πρώτο βήμα ο αλγόριθμος μετρά τις εμφανίσεις του κάθε προϊόντος ξεχωριστά για να καθοριστούν οι μεγάλες λίστες μεγέθους ενός προϊόντος. Το δεύτερο βήμα είναι μία επαναλαμβανόμενη σειρά από υποβήματα. Κάθε επανάληψη, έστω η k, περιλαμβάνει δύο υποβήματα. Αρχικά, μεγάλες λίστες από k-1 προϊόντα Lk-1, που βρέθηκαν στην προηγούμενη k-1 επανάληψη χρησιμοποιούνται για να

δημιουργηθούν οι υποψήφιες μεγάλες λίστες από  $k$  προϊόντα ( $C_k$ ), με βάση την αρχή ότι μια λίστα από προϊόντα είναι μεγάλη, αν κάθε υποσύνολό της είναι μεγάλη λίστα από προϊόντα.

Στην συνέχεια, σαρώνουμε το σύνολο εκπαίδευσης για να βρούμε την επιβεβαίωση των υποψήφιων μεγάλων λιστών από  $k$  προϊόντα. Οι επαναλήψεις σταματούν όταν δεν υπάρχουν υποψήφιες μεγάλες λίστες από προϊόντα. Τότε, στο επόμενο βήμα (τρίτο), από κάθε μία μεγάλη λίστα από προϊόντα προκύπτουν κανόνες από τους οποίους γίνονται τελικά αποδεκτοί όσοι έχουν μεγαλύτερη από την ελάχιστη αξιοπιστία. Στο σημείο αυτό θα αναφέρουμε ένα παράδειγμα για το πως μπορεί να εφαρμόσουμε την μέθοδο της Συσχέτισης κυρίως σε επιχειρήσεις εμπορίας τροφίμων αλλά και άλλων όπως οι υπεραγορές. (Παγουρόπουλος , 2006)

#### 4.4.2.2 Μέτρα Σημαντικότητας

Ο σκοπός των κανόνων συσχέτισης είναι να αποκαλύπτουν ενδιαφέρουσες σχέσεις μεταξύ δεδομένων. Γι' αυτό το λόγο χρησιμοποιούνται ορισμένα μέτρα τα οποία αξιολογούν το επίπεδο σημαντικότητας του κάθε κανόνα συσχέτισης. Αυτά είναι:

- **Confidence (Strength, Εμπιστοσύνη):** Η εμπιστοσύνη ενός κανόνα συσχέτισης είναι το ποσοστό των περιπτώσεων που καλύπτονται από το LHS του κανόνα και οι οποίες καλύπτονται επίσης από το RHS. Μια τιμή της εμπιστοσύνης κοντά στο 1 είναι ένδειξη ενός σημαντικού κανόνα συσχέτισης.
- **Support (Υποστήριξη):** Η υποστήριξη ενός κανόνα συσχέτισης είναι το ποσοστό όλων των περιπτώσεων στο σύνολο δεδομένων που ικανοποιούν έναν κανόνα, δηλαδή ικανοποιούν το LHS και το RHS του κανόνα. Η υποστήριξη μπορεί να θεωρηθεί ως ένδειξη του πόσο συχνά ένας κανόνας εμφανίζεται σε ένα σύνολο στοιχείων και κατά συνέπεια πόσο σημαντικός είναι ο κανόνας.
- **Coverage (Κάλυψη):** Η κάλυψη ενός κανόνα συσχέτισης είναι το ποσοστό των περιπτώσεων των δεδομένων που έχουν τις τιμές των γνωρισμάτων ή

των αντικειμένων που ορίζονται στο αριστερό μέλος του κανόνα. Ένας κανόνας συσχέτισης με τιμή κάλυψης κοντά σε 1, μπορεί να θεωρηθεί ως κανόνας με ενδιαφέρον.

- **Leverage (Μόχλευση):** Το leverage ενός κανόνα είναι το ποσοστό των πρόσθετων περιπτώσεων που καλύπτονται και από το LHS και από το RHS πάνω από εκείνο που αναμένονται εάν τα LHS και RHS ήταν ανεξάρτητα. Το leverage παίρνει τιμές στο διάστημα  $[-1, 1]$ . Τιμές κάτω από το 0 δείχνουν ισχυρή ανεξαρτησία μεταξύ LHS και RHS, ενώ τιμές κοντά στο 1 είναι ένδειξη ενός σημαντικού κανόνα συσχέτισης.
- **Lift (Εμπιστοσύνη):** Το lift ορίζεται ως η εμπιστοσύνη διαιρούμενη με το ποσοστό όλων των περιπτώσεων που καλύπτονται από το RHS. Είναι ένα μέτρο της σπουδαιότητας της συσχέτισης και είναι ανεξάρτητο από την κάλυψη.

#### 4.5 Μοντελοποίηση Πρόβλεψης (Predictive Modeling)

Η κατασκευή ενός μοντέλου πρόβλεψης στοχεύει στη δυνατότητα πρόγνωσης της τιμής μιας μεταβλητής(απόκριση), μέσα από τιμές άλλων μεταβλητών(επεξηγηματικές) που είναι γνώστες. Εάν η μεταβλητή απόκρισης είναι κατηγορική τότε είμαστε σε θέση να εφαρμόσουμε μια μέθοδο ταξινόμησης(classification). Ένα παράδειγμα είναι η πρόβλεψη αγοράς ενός προϊόντος : ναι ή όχι(δίτιμη μεταβλητή). Όμως αν έχουμε συνεχή απόκριση, τότε προχωράμε σε παλινδρόμηση(regression). Μια ενδεικτική εφαρμογή είναι η πρόγνωση της μελλοντικής τιμής ενός αποθέματος.

Όπως και αν δράσουμε όμως κοινός στόχος των δυο εφαρμογών είναι η δημιουργία ενός μοντέλου που ελαχιστοποιεί το σφάλμα στις πραγματικές και προβλεφθείσες τιμές. Ο όρος «πρόβλεψη» χρησιμοποιείται κυρίως ως έννοια και δεν θεωρείται συνέχεια στον χρόνο. Για παράδειγμα μπορεί να θέλουμε να προβλέψουμε την αποπληρωμή ενός δανείου σε μια τράπεζα σε συγκεκριμένο μελλοντικό χρονικό διάστημα αλλά μπορεί και να μας ενδιαφέρει ο προσδιορισμός της διάγνωσης για έναν ασθενή.

Το στοιχείο για να διακρίνουμε τις διαδικασίες πρόβλεψης από αυτές της περιγραφής είναι ότι αντικειμενικός σκοπός της πρόβλεψης είναι μια συγκεκριμένη μεταβλητή, πράγμα που δεν συμβαίνει στην περιγραφή. (Frawley, Piatetsky, & Matheus, 1992)

#### 4.5.1 Κατηγοριοποίηση (Classification)

Η διαδικασία της κατηγοριοποίησης, ή αλλιώς ταξινόμησης (classification) περιλαμβάνει την οργάνωση ενός συνόλου αντικειμένων (objects) που περιγράφονται από ένα σύνολο χαρακτηριστικών (attributes), σε μια σειρά από προκαθορισμένες κλάσεις (classes), χρησιμοποιώντας μεθόδους μάθησης με επίβλεψη (supervised learning methods). Οι τεχνικές της ταξινόμησης ή αλλιώς κατηγοριοποίησης χρησιμοποιούν κατά κανόνα ένα σύνολο εκπαίδευσης (training set), όπου όλα τα αντικείμενα είναι ήδη συνδεδεμένα με γνωστές κλάσεις. Ο αλγόριθμος ταξινόμησης μαθαίνει από αυτό το σύνολο, χρησιμοποιώντας την μάθηση αυτή για την κατασκευή ενός μοντέλου και το μοντέλο αυτό στην συνέχεια ταξινομεί νέα αντικείμενα στις κατάλληλες κλάσεις (Fayyad, Piatetsky-Shapiro, & Smyth).

Άρα μπορούμε να πούμε ότι η κατηγοριοποίηση μαθαίνει σε μία λειτουργία να χαρτογραφεί ή πιο απλά να ταξινομεί ένα στοιχείο δεδομένων σε μία από τις διάφορες προκαθορισμένες κατηγορίες. Η κατηγοριοποίηση πρόκειται ίσως για την πιο δημοφιλή τεχνική με πλήθος εφαρμογών στην αναγνώριση προτύπων και εικόνας σε διάφορους κλάδους.

Στην πράξη μια διαδικασία κατηγοριοποίησης μπορεί να οριστεί ως η εκτέλεση δύο συγκεκριμένων βημάτων: ([www.ecourse.uoi.gr](http://www.ecourse.uoi.gr)) ([evdoxos.ds.uni.gr](http://evdoxos.ds.uni.gr)) ([www-users.cs.umn.edu](http://www-users.cs.umn.edu))

1. Δημιουργία μοντέλου βασισμένου σε δεδομένα εκπαίδευσης
2. Εφαρμογή του μοντέλου στο σύνολο των δεδομένων.

#### 4.6 Εύρεση Απομονωμένων τιμών (Outlier Detection)

Η Εύρεση *Απομονωμένων Τιμών* ή αλλιώς *Ακραίων Τιμών* είναι η πλέον συγγενής εφαρμογή, από τη γενική διαδικασία εξόρυξης γνώσης, ως προς την έννοια της ανακάλυψης χρήσιμων και αξιοποιήσιμων πληροφοριών σε βάσεις δεδομένων. Σύμφωνα με τον Hawkins (1980), που πρώτος έδωσε έστω ένα διαισθητικό ορισμό (Hawkin, 1980) (Kriegel, Kroger, & Zimek, 2010).

*«Απομονωμένη είναι μία παρατήρηση που αποκλίνει τόσο πολύ από τις άλλες παρατηρήσεις, ώστε να μας εξεγείρει τις υποψίες ότι δημιουργήθηκε από διαφορετικό μηχανισμό.»*

Η έννοια μίας παρατήρησης ως απομονωμένη πρέπει να θεωρηθεί τέτοια μονάχα σε ένα πλαίσιο ενός συγκεκριμένου συνόλου δεδομένων. Ένα απτό παράδειγμα είναι η παρουσία ενός σκύλου μέσα σε ένα κοπάδι από πρόβατα. Η ένα διώροφο οίκημα σε ένα δρόμο της Νέας Υόρκης γεμάτο ουρανοξύστες. Αντίθετα η ύπαρξη ενός προβάτου σε μία αγέλη σκύλων θα ήταν επίσης έκτοπο. Η διαφορά τόσο στα χαρακτηριστικά όσο και στις ποιότητες από την πλειονότητα του δείγματος σηματοδοτεί την διαφορετικότητα που εδώ ονομάζουμε απομονωμένες παρατηρήσεις. Λόγω του ότι κάποιος θα θεωρούσε την εύρεση *ακραίων τιμών* ως υποδιαδικασία της εφαρμογής ομαδοποίησης ενός δείγματος πρέπει να αποσαφηνίσουμε περαιτέρω την έννοια της εύρεσης *απομονωμένων τιμών*.

Στο παράδειγμα που αναφέραμε πριν, η εύρεση απομονωμένων τιμών είναι υπεύθυνη για τον εντοπισμό του σκύλου σε ένα κοπάδι πρόβατα. Αντίθετα η διαδικασία της ομαδοποίησης στο ίδιο δείγμα θα μας ομαδοποιούσε τα λευκά πρόβατα, τα μαύρα και ίσως δημιουργούσε μία νέα ομάδα μόνο για το σκύλο. Ενδεχομένως, να τον χαρακτήριζε ως μία παρατήρηση που δεν ανήκει σε καμία από τις δύο τάξεις. Επίσης εάν ένα πρόβατο είχε τρία πόδια η ομαδοποίηση ίσως το αγνοούσε και απλά ομαδοποιούσε το δείγμα βάση χρώματος.

Αντιθέτως η εύρεση απομονωμένων τιμών δεν γίνεται βάση κάποιου χαρακτηριστικού αλλά ο χώρος αναζήτησης (search space) εκτείνεται σε όλα τα χαρακτηριστικά. Επομένως, παρατηρώντας την ομοιότητα και τις διαφορές ανάμεσα σε αυτές τις δύο τεχνικές, μπορούμε να θεωρήσουμε ότι η ομαδοποίηση έχει τη δυνατότητα εύρεσης *ακραίων τιμών* σε περιορισμένο πεδίο μιας και επικεντρώνεται

στην δημιουργία ομάδων, αλλά η εύρεση *ακραίων τιμών*, αναγνωρίζει, για όλα τα χαρακτηριστικά ακόμη και για ομάδες αυτών, τις διαφορές βάση μίας γενικότερης εικόνας που έχει διαμορφωθεί για το δείγμα. Έχοντας λοιπόν την ανεξαρτησία από κάποιο περιορισμό στις δυνατότητες αναζήτησης, ανακαλύπτεται αρκετή και ενδιαφέρουσα γνώση για τα δεδομένα μας, που δεν θα μπορούσε ενδεχομένως να ανακαλυφθεί διαφορετικά.

Τέλος υπάρχουν πολλοί τρόποι για να εφαρμόσουμε την εύρεση *απομονωμένων τιμών* και ένας τρόπος είναι η ομαδοποίηση. Τελευταίες έρευνες έχουν παράγει αλγόριθμους που ενσωματώνουν ομαδοποίηση με ταυτόχρονη σήμανση όπως ο αλγόριθμος DBSCAN.

Ο δειγματοχώρος με τα χαρακτηριστικά γνωρίσματα είναι η αναπαράσταση του παρατηρούμενου πραγματικού κόσμου και αυτή η περιοριστική αναπαράσταση είναι που μας επιτρέπει την Εύρεση *απομονωμένων τιμών* δίνοντας έμφαση σε συγκεκριμένα χαρακτηριστικά. Η εξαγωγή χαρακτηριστικών γνωρισμάτων είναι η αντιστοίχιση στο χώρο δειγματοποίησης του παρατηρούμενου δείγματος βάση των σημαντικότερων χαρακτηριστικών τους. Αυτή η διαδικασία μπορεί να καταστεί αρκετά πολύπλοκη ή αρκετά απλή αναλόγως το περιβάλλον αναζήτησης. Επί παραδείγματι στο παράδειγμα με το κοπάδι από πρόβατα θα πρέπει να αποφασίσουμε πώς θα αναπαραστήσουμε ένα πρόβατο καθώς και τον τρόπο συλλογής δεδομένων. Η πολυπλοκότητα του μετασχηματισμού αυτού εξαρτάται από τον ορισμό των *απομονωμένων τιμών*. Ο ορισμός των *απομονωμένων τιμών* σε ένα συγκεκριμένο χώρο αποτελείται από μία περιγραφή του τι καθιστά μία παρατήρηση ως ακραία. Μπορούν να υπάρξουν και επιπλέον της μίας περιγραφής *ακραίων τιμών* σε ένα δειγματοχώρο. Για παράδειγμα *απομονωμένη τιμή* θα μπορούσε να είναι όποιο αντικείμενο δεν ανήκει σε κάποια ομάδα, ή όποια παρατήρηση έχει χαρακτηριστικό  $q=153,6$ .

Μία άλλη ορολογία που επισυνάπτεται του αγγλικού outlier detection, είναι η *ανίχνευση ανωμαλιών*. Στην ελληνική βιβλιογραφία και επιστημονική κοινότητα δεν υπάρχει μία καθορισμένη μετάφραση. Συνεπώς, θα χρησιμοποιήσουμε το όρο *απομονωμένη τιμή* αντί *ανωμαλίας* μιας και η *ανωμαλία* προϋποθέτει γνώση και οριοθέτηση φυσιολογικής συμπεριφοράς, κάτι που ίσως απουσιάζει στα παραδείγματα εξόρυξης γνώσης από βάσεις δεδομένων. Απεναντίας ο όρος



απομονωμένη τιμή ίσως μας καθοδηγεί οπτικά καθώς και γλωσσικά να καταλάβουμε απ' ευθείας την έννοια του outlier στα ελληνικά.

*Η Εύρεση απομονωμένων τιμών είναι η αυτοματοποιημένη ή ημι-αυτοματοποιημένη διαδικασία ανεύρεσης ακραίων τιμών. Η διαδικασία αυτή είναι σημαντική διότι οι απομονωμένες τιμές υποδεικνύουν ότι κάποιου είδους ενέργεια ή δράση είναι απαραίτητη.*

Σκοπός της Εύρεσης απομονωμένων τιμών είναι η αποκάλυψη των μηχανισμών που γεννούν, καθορίζουν ή διέπουν την ύπαρξή τους. Οι μηχανισμοί αυτοί είναι εύλογο ότι δεν θα συμβαδίζουν πλήρως με τους μηχανισμούς γένεσης/ύπαρξης των παρατηρήσεων που δεν λογίζονται ως *ακραίες τιμες*. Το επόμενο βήμα εάν ανακαλύψουμε τους μηχανισμούς δημιουργίας απομονωμένων τιμών είναι να δημιουργήσουμε έναν ταξινομητή ή ομαδοποιητή ο οποίος θα είναι σε θέση να μάθει και να δημιουργήσει ομάδες από τα δείγματα ακραίων τιμών.

Στην πραγματικότητα όμως τέτοιοι μηχανισμοί είναι υπερβολικά πολύπλοκοι για να καθοριστούν σύμφωνα με την τρέχουσα επιστημονική γνώση. Τα αίτια παραγωγής απομονωμένων τιμών συνήθως είναι εγγενή σε συμπεριφορές φυσικών μηχανισμών ή ανθρώπινης συμπεριφοράς. Αυτό καθιστά σχεδόν αδύνατο το λογικό καθορισμό των μηχανισμών αυτών και συνήθως οι ειδικοί του χώρου για τον οποίον γίνεται η εύρεση ακραίων τιμών είναι εκείνοι που επιβαρύνονται με την εξήγηση τέτοιων μηχανισμών. (Unified Model for Outliers) (itl.nist.gov, n.d.) ([www-users.cs.umn.edu](http://www-users.cs.umn.edu)) (Kriegel, Kroger, & Zimek, 2010).

#### 4.6.1 Εφαρμογές Εύρεσης Απομονωμένων Τιμών

Αναλύοντας τις εφαρμογές του αλγόριθμου outlier (απομονωμένων τιμών) παρατηρούμε ότι, έτσι ονομάζουμε μια ακραία τιμή σε μια παρατήρηση, που βρίσκεται σε «ακραία» θέση σε σχέση με τις συνηθισμένες θέσεις των τιμών σε ένα τυχαίο δείγμα πληθυσμού. Κατά μια έννοια ο ορισμός αυτός αφήνει στον αναλυτή τη δυνατότητα να αποφασίσει τι θεωρεί ως μη φυσιολογική τιμή. Πριν από αυτή τη διαδικασία θα πρέπει πρώτα ο ίδιος να απομονώσει και να χαρακτηρίσει τις φυσιολογικές παρατηρήσεις. Ο προσδιορισμός των πιθανών ακραίων τιμών είναι σημαντικός για τους ακόλουθους λόγους :

1. Μια ακραία τιμή μπορεί να υποδεικνύει εσφαλμένα δεδομένα. Για παράδειγμα τα δεδομένα μπορεί να έχουν κωδικοποιηθεί λανθασμένα ή ένα πείραμα δεν έχει εκτελεστεί σωστά. Αν μπορεί δηλαδή να προσδιοριστεί ότι ένα απόκεντρο σημείο είναι στην πραγματικότητα λανθασμένο, τότε οι απομακρυσμένες τιμές θα πρέπει να διαγραφούν από την ανάλυση ή να διορθωθούν αν είναι δυνατόν.
2. Σε ορισμένες περιπτώσεις, μπορεί να μην είναι δυνατό να προσδιοριστεί αν ένα απόκεντρο σημείο εμπεριέχει εσφαλμένα δεδομένα. Οι ακραίες τιμές μπορεί να οφείλονται σε τυχαία διακύμανση ή μπορεί να υποδεικνύουν κάτι επιστημονικά ενδιαφέρον. Συνήθως δεν θέλουμε να διαγράψουμε απλώς μια απομακρυσμένη παρατήρηση. Ωστόσο, αν τα δεδομένα περιέχουν σημαντικές ακραίες τιμές μπορεί να χρειαστεί η χρήση ισχυρών στατιστικών τεχνικών.

Οι ακραίες τιμές θα πρέπει να εξετάζονται προσεκτικά. Συχνά περιέχουν πολύτιμες πληροφορίες σχετικά με τη διαδικασία της έρευνας ή της συλλογής δεδομένων και της διαδικασίας εγγραφής. Πριν εξεταστεί η πιθανή εξάλειψη των σημείων αυτών από τα δεδομένα, θα πρέπει να προσπαθήσουμε να κατανοήσουμε γιατί εμφανίστηκαν και κατά πόσον είναι πιθανό παρόμοιες τιμές να συνεχίζουν να εμφανίζονται. Φυσικά, οι ακραίες τιμές είναι συχνά *απαλείψιμα* σημεία δεδομένων.

Οι τεχνικές εύρεσης *απομονωμένων τιμών* βρίσκουν εφαρμογή όταν σπάνιες ή σημαντικές περιπτώσεις είναι πιο ενδιαφέρον από τις σύνηθες. Τέτοιες εφαρμογές περιλαμβάνουν από εντοπισμό επίθεσης σε δίκτυα, πλαστές συναλλαγές πιστωτικών καρτών μέχρι και εξόρυξη μη τυπικών ουρανίων σωμάτων σε αστρονομικές βάσεις δεδομένων. Μία πολύ διάσημη εφαρμογή είναι εκείνη που χρησιμοποιεί δεδομένα από κάμερες αστικής παρακολούθησης προκειμένου να ανιχνεύσει ύποπτη συμπεριφορά κίνησης. Επίσης δεδομένα από σπάνια γεγονότα όπως μηχανικά σφάλματα ή ακραίες αθλητικές επιδόσεις έχουν δώσει έναυσμα για ενασχόληση εύρεσης *απομονωμένων τιμών*. (itl.nist.gov, n.d.)

#### 4.7 Αλγόριθμος LOF (Local Outlier Factor)

Σε μια ανώμαλη ανίχνευση που κάνουμε ο παράγοντας LOF (*Local Outlier Factor*) είναι ένας αλγόριθμος που έχει προταθεί από τους (Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng and Jörg Sander), προκειμένου να βρίσκουν τα ανώμαλα σημεία δεδομένων μετρώντας την τυπική απόκλιση ενός συγκεκριμένου σημείου δεδομένων σε σχέση με τους «γείτονές» της. Ο Lof μοιράζεται στοιχεία και ορισμένες έννοιες από την DBSCAN όπως η «κεντρική απόσταση» (core distance) και την «εξ αποστάσεως προσβασιμότητα» (reachability distance), που χρησιμοποιείται για την τοπική εκτίμηση πυκνότητας.

Η κύρια ιδέα του αλγορίθμου LOF βασίζεται στην έννοια της τοπικής πυκνότητας, όπου η *τοποθεσία* δίνεται από τους πλησιέστερους γείτονες του  $\mathbf{k}$  (κέντρων), των οποίων η απόσταση χρησιμοποιείται για την εκτίμηση της πυκνότητας. Συγκρίνοντας την τοπική πυκνότητα ενός αντικειμένου με τις τοπικές πυκνότητες των γειτόνων της, μπορούμε να προσδιορίσουμε τις περιοχές παρόμοιας πυκνότητας και τα σημεία που έχουν σημαντικά χαμηλότερη πυκνότητα από τους «γείτονες» τους. Αυτά θεωρούνται **ακραίες τιμές**. Η τοπική πυκνότητα υπολογίζεται από την τυπική απόσταση στην οποία ένα σημείο μπορεί να «φτάσει» από τους «γείτονες» του. Ο ορισμός της «εξ αποστάσεως προσβασιμότητας» που χρησιμοποιείται στον αλγόριθμο LOF είναι ένα επιπλέον μέτρο για την παραγωγή σταθερών αποτελεσμάτων μέσα στις συστάδες (clusters).

Λόγω της τοπικής προσέγγισης η LOF είναι σε θέση να προσδιορίσει ακραίες τιμές σε ένα σύνολο δεδομένων που δεν θα ήταν ακραίες τιμές σε μια άλλη περιοχή του συνόλου δεδομένων. Οι τοπικές πυκνότητες προσβασιμότητας συγκρινόμενες με τον γειτόνων ορίζονται ως, **Μέση Τοπική Προσβασιμότητα Των Γειτόνων** ([www.wikipedia.org](http://www.wikipedia.org), n.d.).

$$LOF_k(A) := \frac{\sum_{B \in N_k(A)} lrd(B)}{|N_k(A)|} / lrd(A)^4$$

<sup>4</sup> Όπου τοπική πυκνότητα προσβασιμότητας ενός αντικειμένου  $\mathbf{A}$  ορίζεται:

$$lrd(A) := 1 / \left( \frac{\sum_{B \in N_k(A)} reachability - distance_k(A, B)}{|N_k(A)|} \right)$$

## 5<sup>ο</sup> ΚΕΦΑΛΑΙΟ

### 5 Αποτελέσματα<sup>5</sup>

#### 5.1 Προεπεξεργασία δεδομένων

Προκειμένου να πραγματοποιήσουμε τις τεχνικές εξόρυξης δεδομένων Apriori και Δέντρα αποφάσεων, είναι αναγκαίο να γίνει προεπεξεργασία των δεδομένων του ερωτηματολογίου ώστε να αποκτήσουν την κατάλληλη μορφή για τα «τρεξίματα» των αλγορίθμων με σκοπό την εξόρυξη αποτελεσμάτων. Ο τρόπος επίτευξης αυτών είναι η μετατροπή των αριθμητικών σε διακριτοποιημένες τιμές.

Παρακάτω παρατίθεται πίνακας με στατιστικά στοιχεία των ψυχολογικών δεικτών που περιλαμβάνουν την διάμεσο (Median ή το τεταρτημόριο Q2) , τα τεταρτημόρια Q1, Q3, μέγιστη και ελάχιστη τιμή πριν την διακριτοποίηση.

```
summary(dataset)
```

#### *Dataset*

	<i>Min</i>	<i>1<sup>st</sup> Qu.</i>	<i>Median</i>	<i>Mean</i>	<i>3<sup>rd</sup> Qu.</i>	<i>Max</i>
Well_being	1.830	3.670	4.500	4.383	5.170	7.000
Self_control	1.50	3.50	4.00	4.02	4.50	6.17
Emotionality	1.630	3.250	3.880	3.939	4.630	6.500
Sociability	1.830	3.830	4.500	4.415	5.000	6.670
Paranoid	0.0000	0.2900	0.4300	0.4563	0.5700	1.0000
Schizoid	0.0000	0.0000	0.1400	0.1749	0.2900	0.8600
Schizotypal	0.0000	0.1100	0.2200	0.2918	0.4400	0.8900
Histrionic	0.0000	0.3800	0.5000	0.4874	0.6300	1.0000
Antisocial	0.0000	0.1400	0.1400	0.2061	0.2900	0.7100
Narcissistic	0.1100	0.2200	0.3300	0.3822	0.5600	0.8900
Borderline	0.0000	0.2200	0.3300	0.3998	0.5600	0.8900
Obsessive_compulsive	0.000	0.250	0.380	0.404	0.500	0.880
Avoidant	0.000	0.380	0.500	0.516	0.750	1.000

Πίνακας 1 Στατιστικά Στοιχεία Ψυχολογικών Δεικτών

<sup>5</sup> Στο κεφάλαιο 3 παρατίθενται αναλυτικά τα δημογραφικά στοιχεία των ερωτηματολογίων

Πρέπει πρώτα να εξετάσουμε τις συσχετίσεις μεταξύ των κλιμάκων των δύο ερωτηματολογίων.

Ακολουθεί ο σχετικός πίνακας:

	WELL_BEING	SELF_CONTROL	EMOTIONALITY	SOCIABILITY
PARANOID	-0.07783308	0.038696252	0.019372126	0.036834051
SCHIZOID	-0.22728511	0.038702336	0.005391676	-0.120198677
SCHIZOTYPAL	-0.22762035	-0.008920092	-0.047685587	-0.166256772
HISTRIONIC	-0.10365655	-0.009511488	0.069630627	0.048569593
ANTISOCIAL	-0.06988480	0.022501479	0.017200888	0.002614922
NARCISSISTIC	-0.16133715	-0.026749823	-0.049528851	-0.054682007
BORDERLINE	-0.20111602	-0.007302467	0.143430240	-0.045176620
OBSESSIVE_COMP	-0.02748412	-0.022441925	0.080773566	-0.028212522
DEPENDENT	-0.25809074	-0.014484800	0.022931206	-0.072627819
AVOIDANT	-0.19542242	-0.058127604	-0.005988254	-0.072538316

Πίνακας 2 Θετικές και Αρνητικές Συσχετίσεις Μεταξύ E.I. και IPDE

Όπως παρατηρούμε από τον πίνακα, οι τιμές με πράσινο παρουσιάζουν «ισχυρές» συσχετίσεις μεταξύ κλιμάκων των δύο ερωτηματολογίων ενώ οι τιμές με κόκκινο παρουσιάζουν «ασθενείς» συσχετίσεις, αντίστοιχα, δεδομένου ότι θεωρούμε «ισχυρή» μία συσχέτιση όταν τείνει στο 1, ενώ «ασθενής» όταν τείνει στο -1.

Διαπιστώνουμε ότι η πλειοψηφία των ψυχολογικών δεικτών των δύο ερωτηματολογίων δεν συσχετίζεται ικανοποιητικά, με μόνο ένα μέρος του συνόλου να παρουσιάζει θετική συσχέτιση.

## 5.2 Apriori

Όπως αναφέραμε και παραπάνω ο σκοπός των κανόνων συσχέτισης είναι να αποκαλύπτουν ενδιαφέρουσες σχέσεις μεταξύ δεδομένων, στην προκειμένη περίπτωση μεταξύ των δύο ερωτηματολογίων με τον αλγόριθμο apriori.

Τα σενάρια τα οποία εξετάσαμε παρατίθενται ως εξής:

### Σενάριο 1<sup>ο</sup>

Εξέταση περίπτωσης όπου το φύλο (sex) είναι «Άνδρας / (Male)».

*Κανόνας 1<sup>ος</sup>*

```
{self_control=Normal,
cityStayMe=Urban,                support  confidence  lift
citySameMe=yes} => {Anti social=Normal} 0.4262295 0.9285714 1.110644
```

Με confidence 92% και lift 1,1 όταν:

- Η πόλη διαμονής είναι αστική περιοχή
- Η πόλη καταγωγής και διαμονής είναι ίδιες
- Ο δείκτης Αυτοελέγχου είναι «κανονικός» (E.I.)

Τότε, παρουσιάζονται στον δείκτη Αντικοινωνικότητας, του IPDE, , τιμές οι οποίες αντιστοιχούν στο κανονικό επίπεδο.

*Κανόνας 2<sup>ος</sup>*

```
{Paranoid=Normal,
self_control=Normal,                support  confidence  lift
cityBirthMe=Urban} => {Dependent=Normal} 0.3278689 0.9090909 1.38636
```

Με confidence 90% και lift 1,3 όταν:

- Η πόλη καταγωγής είναι αστική περιοχή
- Ο Παρανοειδής δείκτης είναι «κανονικός» (IPDE)
- Ο δείκτης Αυτοελέγχου είναι «κανονικός» (E.I.)

Τότε, παρουσιάζονται στον δείκτη Εξάρτησης, του IPDE, τιμές οι οποίες αντιστοιχούν στο κανονικό επίπεδο.

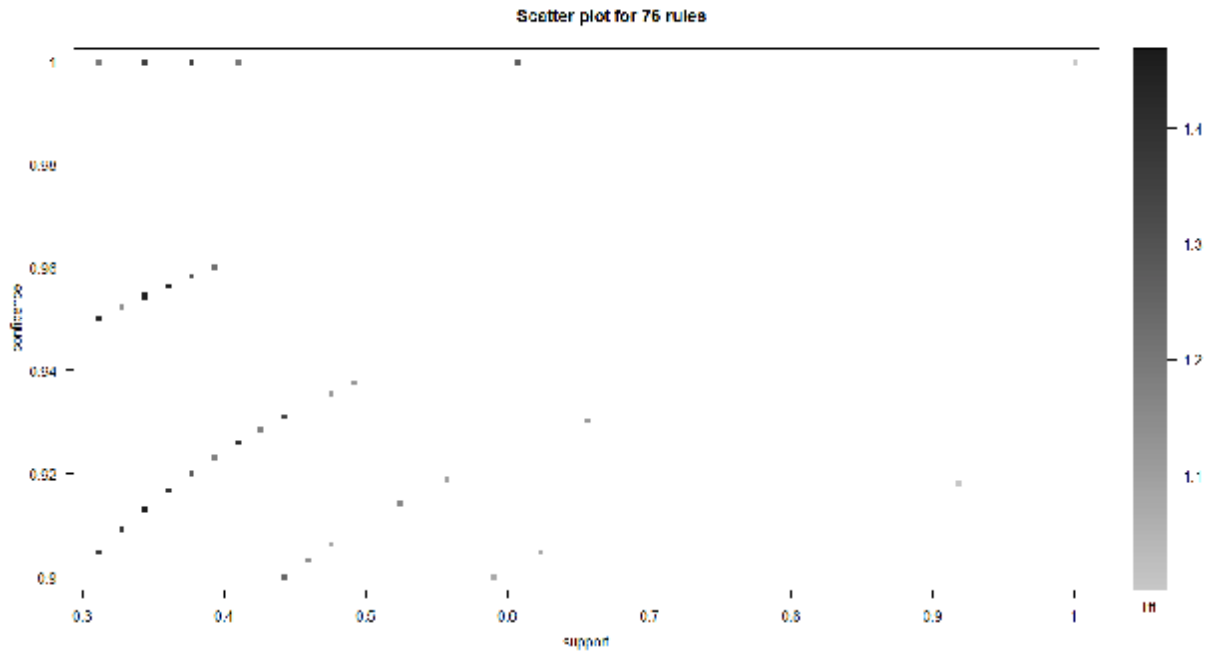
*Κανόνας 3<sup>ος</sup>*

```
{Obsessive-compulsive=Normal,
Borderline=Normal,                support  confidence  lift
sociality=Normal} =>{Dependent=Normal} 0.3114754 0.9047619 1.37976
```

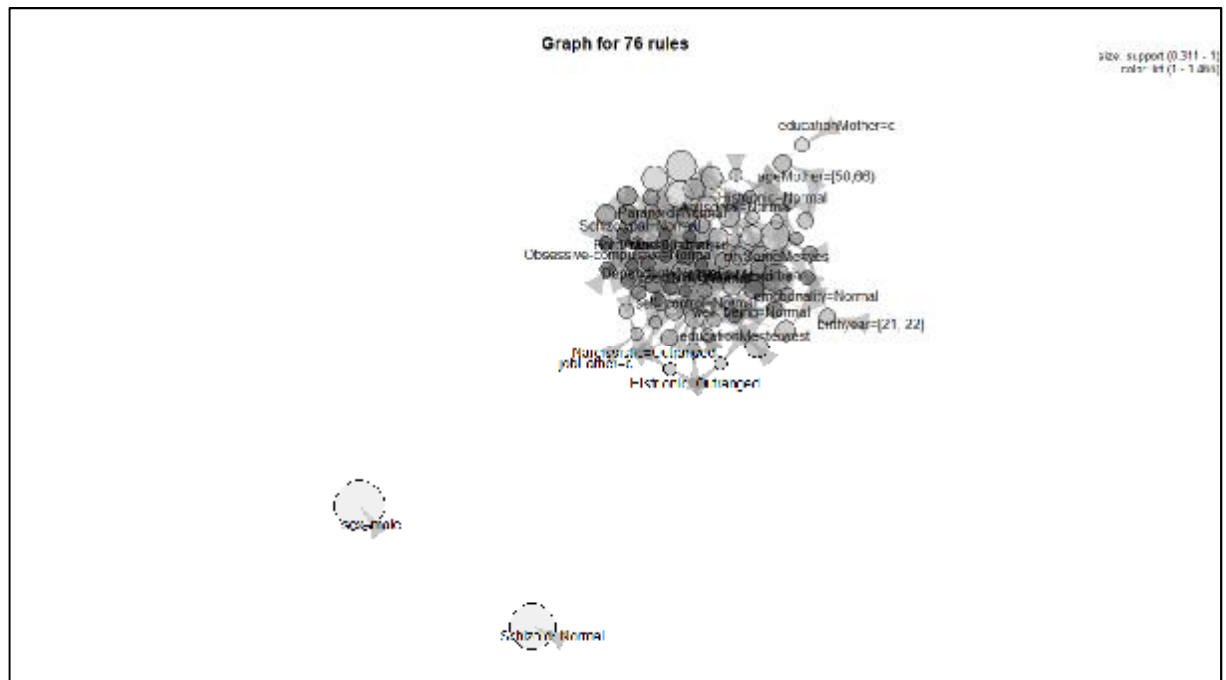
Με confidence 90% και lift 1,3 όταν:

- Ο Ιδιοψυχαναγκαστικός δείκτης είναι «κανονικός» (IPDE)
- Ο Οριακός δείκτης είναι «κανονικός» (IPDE)
- Ο δείκτης Κοινωνικότητας είναι «κανονικός» (E.I.)

Τότε, παρουσιάζονται στον δείκτη Εξάρτησης, του IPDE, τιμές οι οποίες αντιστοιχούν στο κανονικό επίπεδο.



Σχήμα 1 Διάγραμμα Διασποράς Κανόνων Δείγματος Ανδρών



Σχήμα 2 Διάγραμμα Συσχέτισης Κανόνων Δείγματος Ανδρών

**Σενάριο 2<sup>ο</sup>**

Εξέταση περίπτωσης όπου το αποτέλεσμα της συσχέτισης είναι το φύλο (sex)  
«Γυναίκα / (Female)».

Κανόνας 1<sup>ος</sup>

{Anti social =Normal , emotional ity=Normal , citySameMe=no}	=> {sex=female}	support 0.1560976	confidence 0.9696970	lift 1.380471
--	-----------------	----------------------	-------------------------	------------------

Με confidence 96% και lift 1,3 οι γυναίκες παρουσιάζουν:

- «Κανονικό» Δείκτη Αντικοινωνικότητας (IPDE)
- «Κανονικό» Δείκτη Συναισθηματικότητας (E.I.)
- Η πόλη καταγωγής και Διαμονής τους δεν είναι ο ίδιος

Κανόνας 2<sup>ος</sup>

{Anti social =Normal , well_being=Normal , cityBirthMe=NoUrban}	=> {sex=female}	support 0.1707317	confidence 0.9210526	lift 1.311221
---	-----------------	----------------------	-------------------------	------------------

Με confidence 92% και lift 1,3 οι γυναίκες παρουσιάζουν:

- «Κανονικό» Δείκτη Αντικοινωνικότητας (IPDE)
- «Κανονικό» Δείκτη Ευημερίας (E.I.)
- Η πόλη Καταγωγής τους να είναι μη αστική περιοχή

Κανόνας 3<sup>ος</sup>

{Anti social =Normal , Avoicant=Outranged , sociability=Normal , cityBirthMe=NoUrban}	=> {sex=female}	support 0.1609756	confidence 0.9166667	lift 1.304977
--	-----------------	----------------------	-------------------------	------------------

Με confidence 91% και lift 1,3 οι γυναίκες παρουσιάζουν:

- «Κανονικό» Δείκτη Αντικοινωνικότητας (E.I.)
- «Πάνω του κανονικού» Αποφευκτικό Δείκτη (IPDE)
- «Κανονικό» Δείκτη Κοινωνικότητας (E.I.)
- Η πόλη Καταγωγής τους να είναι μη αστική περιοχή

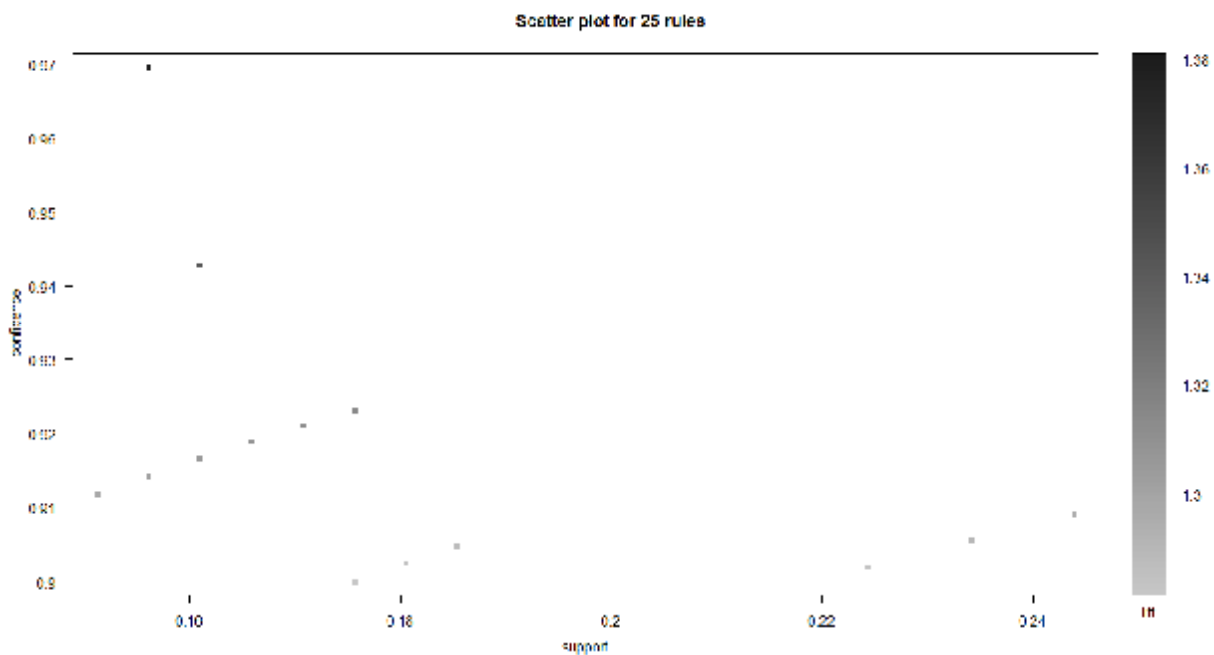


Κανόνας 4<sup>ος</sup>

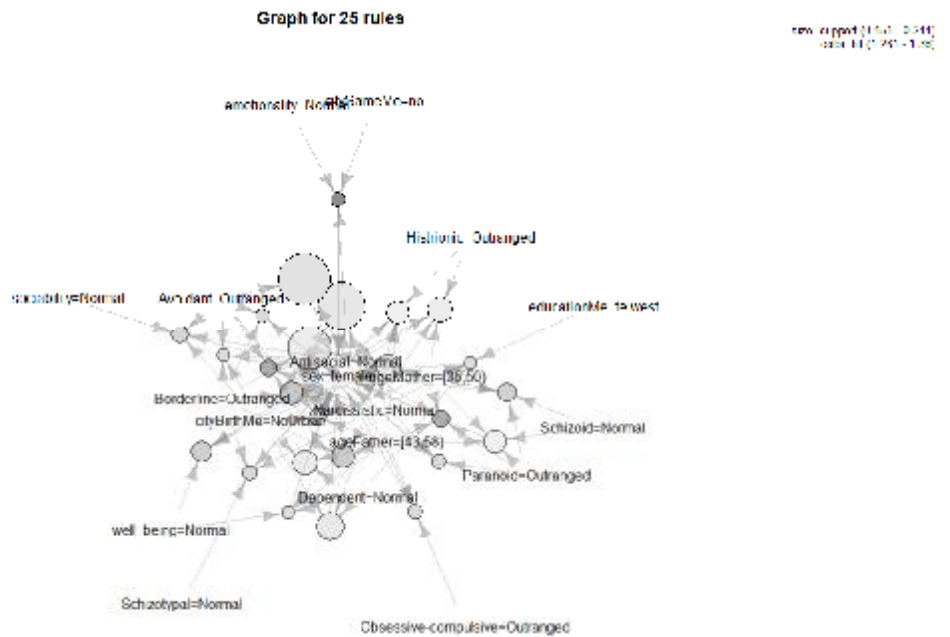
```
{Dependent=Normal ,
Narcissistic=Normal ,
well_being=Normal ,
ageFather=[ 43, 58) } => {sex=female} 0.1512195 0.9117647 1.297998
```

Με confidence 91% και lift 1,2 οι γυναίκες παρουσιάζουν:

- «Κανονικό» Δείκτη Εξάρτησης (IPDE)
- «Κανονικό» Δείκτη Ναρκισσισμού (IPDE)
- «Κανονικό» Δείκτη Ευημερίας (E.I.)
- Ηλικία Πατέρα από 43 έως 58 ετών



Σχήμα 3 Διάγραμμα Διασποράς Κανόνων Δείγματος Γυναικών



Σχήμα 4 Διάγραμμα Συσχέτισης Κανόνων Δείγματος Γυναίκων

### Σενάριο 3<sup>ο</sup>

Εξέταση περίπτωσης όπου το αποτέλεσμα της συσχέτισης είναι η εκπαίδευση (educationMe) «Τεί Δυτικής Ελλάδας / (teiwest)».

Κανόνας 1<sup>ος</sup>

{Hi stri oni c=Out ranged, soci abi li ty=Normal }=> {ci tySt ayMe=Urban}	support	confi dence	li ft
	0. 3214286	0. 9000000	1. 02

Με confidence 90% και lift 1 όταν:

- Ο Οιστριονικός Δείκτης είναι «πάνω του κανονικού» (IPDE)
- Ο Δείκτης Κοινωνικότητας είναι «κανονικός» (E.I.)

Τότε, παρουσιάζεται ότι η πόλη διαμονής των φοιτητών είναι αστική περιοχή.

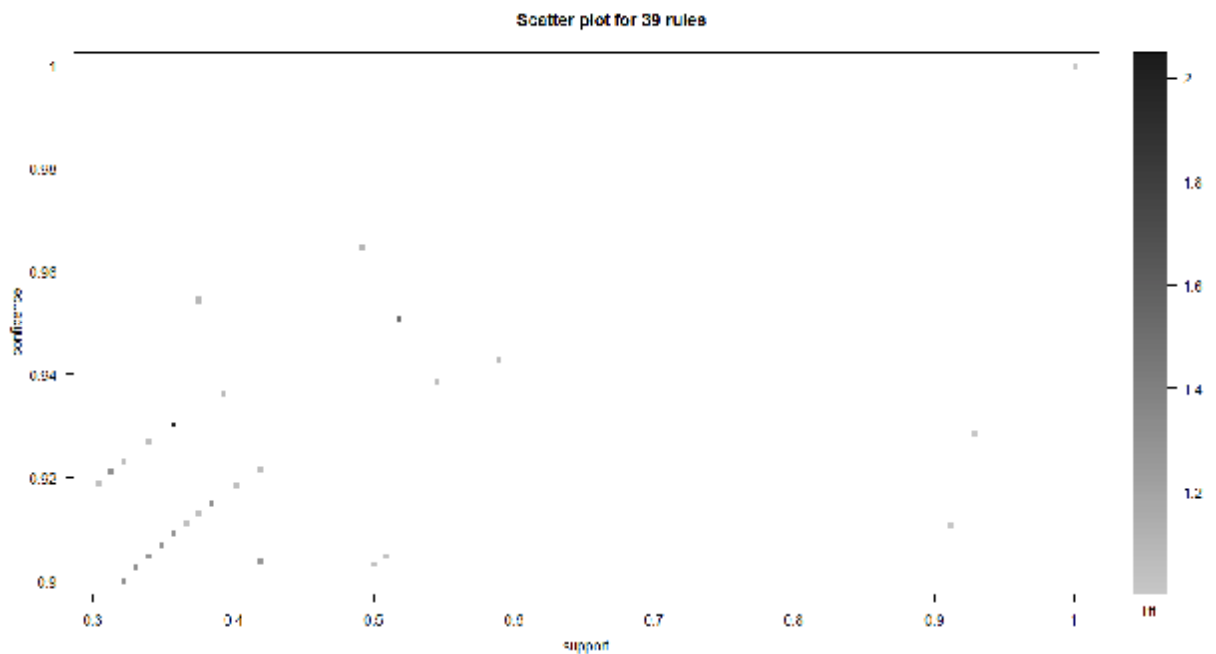
Κανόνας 2<sup>ος</sup>

	support	confidence	lift
{Dependent=Normal, emotionality=Normal} => {cityStayMe=Urban}	0.3214286	0.9230769	1.0549

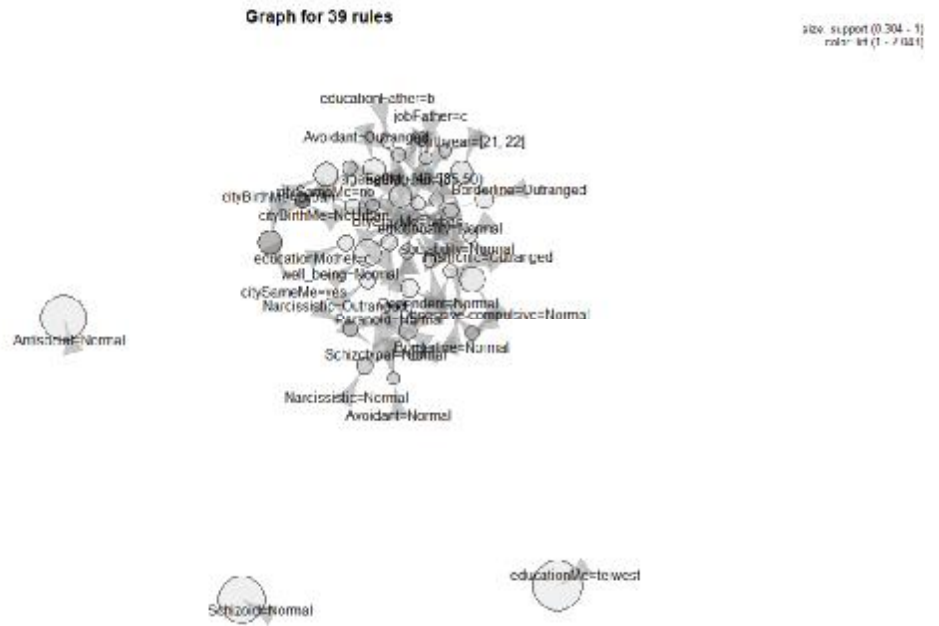
Με confidence 92% και lift 1 όταν:

- Ο Δείκτης Εξάρτησης είναι «κανονικός» (IPDE)
- Ο Δείκτης Συναισθηματικότητας είναι «κανονικός» (E.I.)

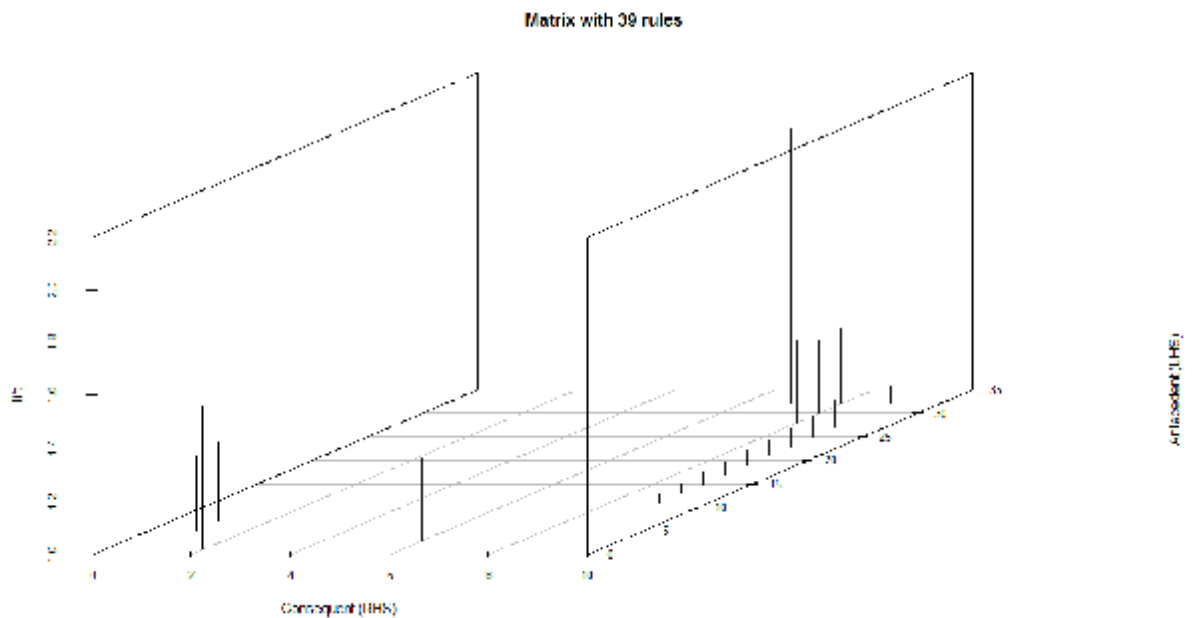
Τότε, παρουσιάζεται ότι η πόλη διαμονής των φοιτητών είναι αστική περιοχή.



Σχήμα 5 Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Δυτικής Ελλάδας



Σχήμα 6 Διάγραμμα Συσχέτισης Κανόνων Δείγματος Τει Δυτικής Ελλάδας



Σχήμα 7 3D Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Δυτικής Ελλάδας

**Σενάριο 4<sup>ο</sup>**

Εξέταση περίπτωσης όπου το αποτέλεσμα της συσχέτισης είναι η εκπαίδευση (educationMe) «Τεί Ιονίων Νήσων / (teiIwn)».

Κανόνας 1<sup>ος</sup>

{sociability=Normal, sex=female}	=> {Avoirdant=Outranged}	support	confidence	lift
		0.4285714	0.9000000	1.1812

Με confidence 90% και lift 1,1 όταν:

- Ο Δείκτης Κοινωνικότητας είναι «κανονικός» (E.I.)
- Το φύλλο είναι Γυναίκα

Τότε, παρουσιάζονται στον Αποφευκτικό δείκτη, του IPDE, τιμές οι οποίες αντιστοιχούν σε πάνω του κανονικού επίπεδο.

Κανόνας 2<sup>ος</sup>

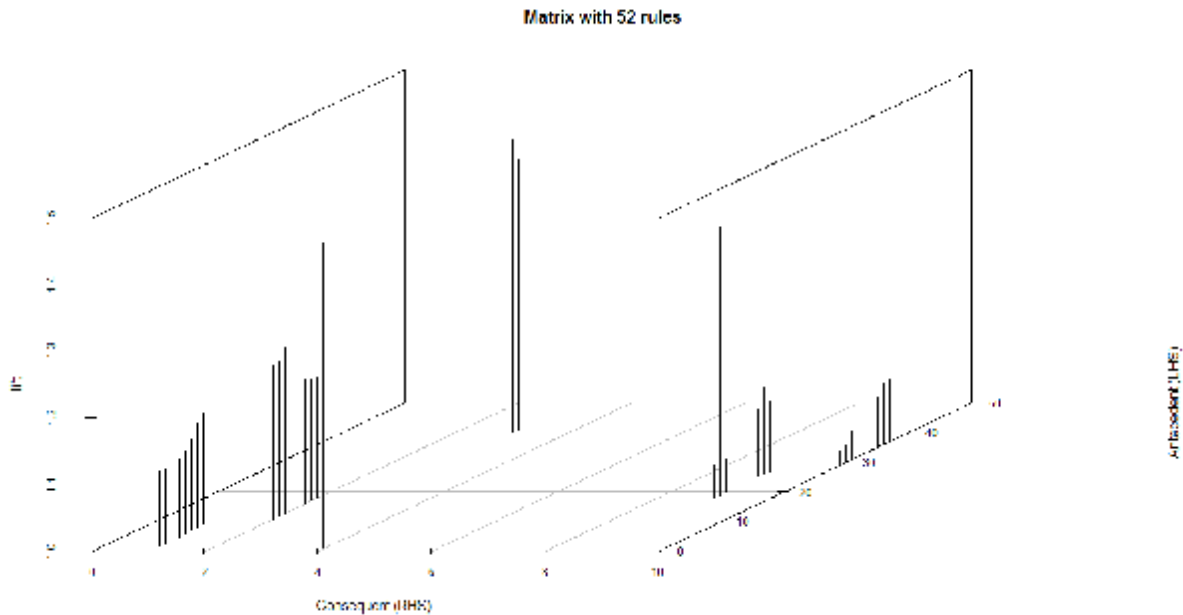
{Schizotypal=Normal, sociability=Normal}	=> {Avoirdant=Outranged}	support	confidence	lift
		0.3571429	0.9375000	1.2

Με confidence 93% και lift 1,2 όταν:

- Ο Δείκτης Σχιζοτυπικής Συμπεριφοράς είναι «κανονικός»
- Ο Δείκτης Κοινωνικότητας είναι «κανονικός»

Τότε, παρουσιάζονται στον Αποφευκτικό δείκτη, του IPDE, τιμές οι οποίες αντιστοιχούν σε πάνω του κανονικού επίπεδο.





Σχήμα 10 Διάγραμμα Διασποράς Κανόνων Δείγματος Τει Ιονίων Νήσων

Επομένως συμπεραίνουμε από τον πίνακα συσχέτισης των δύο ερωτηματολογίων, ότι με την περαιτέρω ανάλυση που βασίστηκαν στις τεχνικές εξόρυξης δεδομένων φαίνεται πώς υπάρχουν κανόνες οι οποίοι συσχετίζουν ικανοποιητικά τις κλίμακες όπως για παράδειγμα, *self\_control* με *Antisocial* => 0.22 «ισχυρή» και *Dependent*, *Narcissistic* με *well\_being* => *Dependent* = -0.25 , *Narcissistic*= -0.16 «ασθενείς» αντίστοιχα.

### 5.3 Δέντρα Αποφάσεων

Με την μέθοδο των δέντρων απόφασης έχουμε στην διάθεση μας το εργαλείο για την λήψη απόφασης με βάση συγκεκριμένα γνωρίσματα που χρειάζονται ανάλυση με σκοπό την εκμάθηση.

Η **Μήτρα Σύγκρισης** που χρησιμοποιείται παρακάτω είναι ένας πίνακας στον οποίο είναι αποθηκευμένο το πλήθος των εγγραφών ελέγχου που έχουν προβλεφθεί σωστά και λανθασμένα. Στην κύρια διαγώνιο της μήτρας παρουσιάζεται το πλήθος των σωστών προβλέψεων, ενώ και τα μη διαγώνια στοιχεία φανερώνουν το πλήθος των λανθασμένων προβλέψεων.

Τα στατιστικά στοιχεία που παρουσιάζονται είναι:

- **Ακρίβεια πρόβλεψης** (accuracy): ορίζεται ως η αναλογία των δειγμάτων θετικών και αρνητικών που έχουν προβλεφθεί σωστά.
- **Σφάλμα ταξινόμησης** (classification error): ορίζεται ως η ανάλυση των σφαλμάτων ταξινόμησης όπου ως στόχο έχει να μετρήσει τα ποσοστά των σφαλμάτων και να διορθώσει τις εκτιμήσεις των αναλογιών για ταξινόμηση.

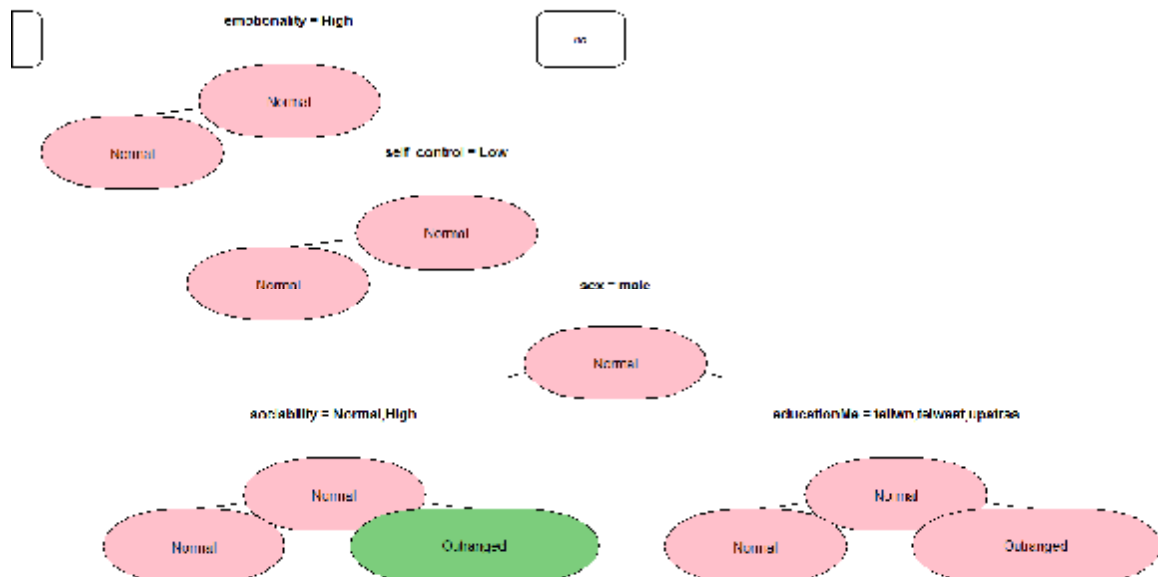
## Σενάριο 1<sup>ο</sup>

Εξέταση περίπτωσης όπου εξαρτημένη τιμή είναι ο Παρανοειδής Δείκτης (Paranoid) και ανεξάρτητες τιμές οι Δείκτες: Συναισθηματικότητας (Emotionality), Κοινωνικότητας (Sociability) και τα Δημογραφικά στοιχεία: Εκπαίδευση Φοιτητή (educationMe), Φύλλου (sex).

```
depend. var <- c("Paranoid")
ind. vars <- c("emotionality", "self_control", "educationMe",
"sociability", "sex")

> acc
[1] 43.90244
> cl.err
[1] 56.09756
> conf.mat
      pred
      Normal Outranged
Normal      18         5
Outranged   18         0
```

Οι φοιτητές με κανονικό Παρανοειδή δείκτη παρουσιάζουν χαμηλό ή κανονικό δείκτη συναισθηματικότητας, Κανονικό ή υψηλό δείκτη Αυτοελέγχου στους άντρες όταν έχουν κανονικό ή υψηλό δείκτη κοινωνικότητας και στις γυναίκες όταν σπουδάζουν στα Τει Ιονίων Νήσων, Δυτικής Ελλάδας ή Πανεπιστήμιο Πατρών.





## Σενάριο 2<sup>ο</sup>

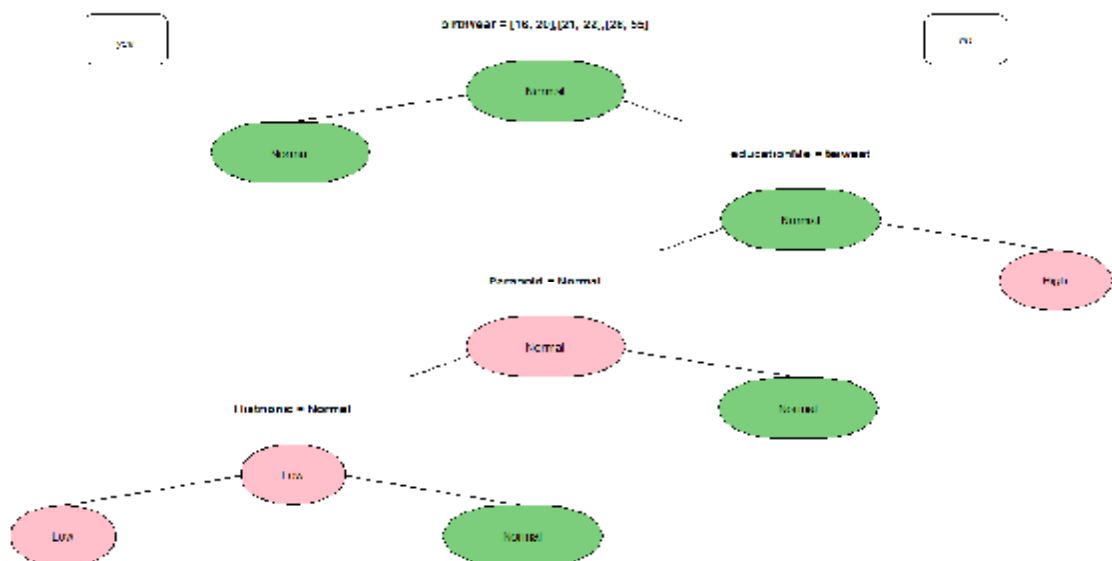
Εξέταση περίπτωσης όπου εξαρτημένη τιμή είναι ο Δείκτης Συναισθηματικότητας (Emotionality) και ανεξάρτητες τιμές οι Δείκτες: Παρανοειδής (Paranoid), Σχιζοειδής (Schizoid), Αντικοινωνική (Antisocial), Οιστριονική (Histrionic), Οριακή (Borderline) και τα Δημογραφικά στοιχεία: Εκπαίδευση Φοιτητή (educationMe), Ημερομηνία Γέννησης (birthyear).

```
depend.var <- c("emotionality")
ind.vars <- c("Paranoid", "Schizoid", "educationMe", "birthyear",
"Antisocial", "Histrionic", "Borderline")

> cl.err
[1] 0.4878049
> acc
[1] 0.5121951

> conf.mat
      pred
      Low Normal High
Low    0     10    0
Normal 6     15    4
High   1     3     2
```

Οι φοιτητές στις ηλικίες από 23 έως 24 και 25 έως 27 που σπουδάζουν στο Τει Δυτικής Ελλάδας και ο δείκτης Παρανοειδούς συμπεριφοράς είναι κανονικός, τότε παρουσιάζουν κανονικό δείκτη συναισθηματικότητας ενώ χαμηλό όταν ο Οιστριονικός δείκτης είναι κανονικός.



### Σενάριο 3<sup>ο</sup>

Εξέταση περίπτωσης όπου εξαρτημένη τιμή είναι ο Δείκτης Κοινωνικότητας (Sociability) και ανεξάρτητες τιμές οι Δείκτες: Παρανοειδής (Paranoid), Αντικοινωνική (Antisocial), Οιστριονική (Histrionic), Συναισθηματικότητα (Emotionality) και τα Δημογραφικά στοιχεία: Εκπαίδευση Φοιτητή (educationMe), Ημερομηνία Γέννησης (birthyear), Φύλλου (sex).

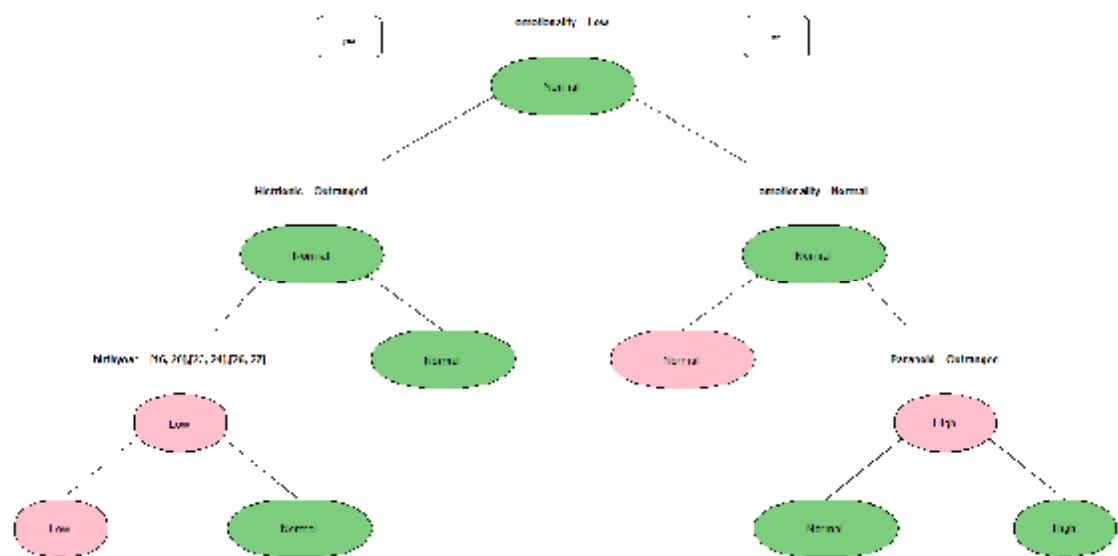
```
depend. var <- c("sociability")
ind. vars <- c("Paranoid", "educationMe", "birthyear", "Antisocial",
"Histrionic", "sex", "emotionality")
```

```
> cl.err
[1] 0.4878049
> acc
[1] 0.5121951

> conf.mat
      pred
      Low Normal High
Low    1      5    0
Normal 2     16    5
High   1     10    1
```

Οι φοιτητές στις ηλικίες από 16 έως 20 και 23 έως 27 με χαμηλό δείκτη Κοινωνικότητας παρουσιάζουν Οιστριονικό δείκτη πάνω του κανονικού και χαμηλή Συναισθηματικότητα όταν ο δείκτης κοινωνικότητας είναι κανονικός.

Οι φοιτητές με δείκτη Παρανοειδούς συμπεριφοράς όταν είναι πάνω του κανονικού παρουσιάζουν υψηλό δείκτη Κοινωνικότητας, ενώ όταν ο δείκτης Κοινωνικότητας είναι κανονικός τότε και ο δείκτης Συναισθηματικότητας κυμαίνεται σε κανονικό επίπεδο.



## Σενάριο 4<sup>ο</sup>

Εξέταση περίπτωσης όπου εξαρτημένη τιμή είναι ο Δείκτης Ευημερίας (Well\_Being) και ανεξάρτητες τιμές οι Δείκτες: Αυτοέλεγχου (Self\_Control), Οριακός (Borderline), Συναισθηματικότητα (Emotionality), Ναρκισσιστικός (Narcissistic) και τα Δημογραφικά στοιχεία: Εκπαίδευση Φοιτητή (educationMe), Ημερομηνία Γέννησης (birthyear), Φύλλου (sex).

```
depend. var <- c("well_being")
ind. vars <- c("emotionality", "self_control", "birthyear",
"educationMe", "Borderline", "sex", "Narcissistic")
```

```
> cl. err
[1] 0.4634146
> acc
[1] 0.5365854
```

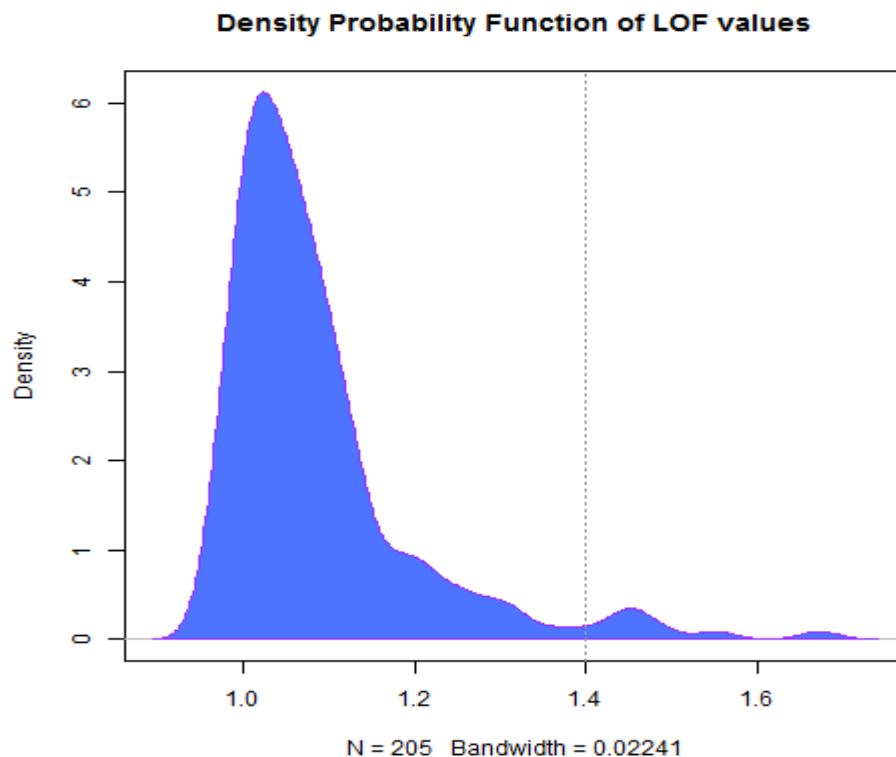
```
> conf. mat
      pred
      Low Normal High
Low    3     7    0
Normal 5    15    4
High   1     5    1
```

Οι φοιτητές με κανονικό δείκτη Ευημερίας παρουσιάζουν κανονικό ή υψηλό Αυτοέλεγχο με χαμηλό ή υψηλό δείκτη Συναισθηματικότητας όταν είναι άντρες και σπουδάζουν στο Πανεπιστήμιο Πατρών ή σε κάποιο άλλο Ίδρυμα με κανονικό δείκτη Ναρκισσισμού, ενώ όταν παρουσιάζουν υψηλό δείκτη Συναισθηματικότητας τότε



#### 5.4 Εύρεση Απομονωμένων Τιμών (Outlier Detection)

Ο σκοπός είναι να εντοπίσουμε ακραίες περιπτώσεις με βάση τα χαρακτηριστικά όλων των ψυχομετρικών κλιμάκων. Γιαυτό χρησιμοποιήθηκε ο αλγόριθμος Lof.



Σχήμα 11 Διάγραμμα Πυκνότητας Πιθανοτήτων

Outliers Position:

```
> pos.outliers
[1] 34 55 57 80 87 167 180
```

Outliers Values:

```
> val.outliers
[1] 1.487403 1.439551 1.455842 1.462654 1.441647 1.672956 1.551421
```

Στην κατανομή των **ακραίων τιμών** (Outlier) με βάση τον αλγόριθμο LOF, παρατηρούμε ότι οι περισσότερες παρατηρήσεις κυμαίνονται από 1.0 έως 1.2 σε δείγμα 205 παρατηρήσεων (Σχήμα 11). Επομένως θεωρούμε ότι οι παρατηρήσεις LOF που έχουν δείκτη 1.4 θεωρούνται **ακραίες τιμές** (Outlier). Στο δείγμα μας οι περιπτώσεις που είναι απομονωμένες είναι οι παρατηρήσεις: 34,55,57,80,87,167,180 με τιμές του δείκτη LOF 1.487, 1.439, 1.455, 1.462, 1.441, 1.672, 1.551.

Στην ενότητα που ακολουθεί πραγματοποιούμε μία ανάλυση των βασικών στατιστικών μεγεθών των απομονωμένων περιπτώσεων.

## 5.4.1 Περιγραφική Στατιστική των Απομονωμένων Περιπτώσεων

Με την βοήθεια του πακέτου *psych* θα εξετάσουμε έναν πίνακα περιγραφικών στατιστικών, έχοντας μετατρέψει το δείγμα μας σε πίνακα βασικών στατιστικών δεικτών των Κανονικών παρατηρήσεων και Outlier.

	<i>Mean</i>	<i>Median</i>	<i>Min</i>	<i>Max</i>	<i>Std.dev</i>
<i>Well_being</i>	4.42	4.5	1.83	7.00	0.95
<i>Self_control</i>	4.04	4.0	2.17	6.17	0.75
<i>emotionality</i>	3.92	3.88	1.75	6.38	0.97
<i>sociability</i>	4.43	4.42	2.33	6.67	0.91
<i>Paranoid</i>	0.45	0.43	0.0	1.00	0.17
<i>Schizoid</i>	0.18	0.14	0.0	0.86	0.18
<i>Schizotypal</i>	0.29	0.22	0.0	0.89	0.19
<i>Histrionic</i>	0.49	0.5	0.0	1.0	0.21
<i>Antisocial</i>	0.20	0.14	0.0	0.71	0.18
<i>Narcissistic</i>	0.38	0.33	0.11	0.89	0.21
<i>Borderline</i>	0.40	0.33	0.0	0.89	0.21
<i>Obsessive_comp</i>	0.40	0.38	0.0	0.88	0.22
<i>Dependent</i>	0.32	0.25	0.0	1.0	0.25
<i>Avoidant</i>	0.51	0.5	0.0	1.0	0.24

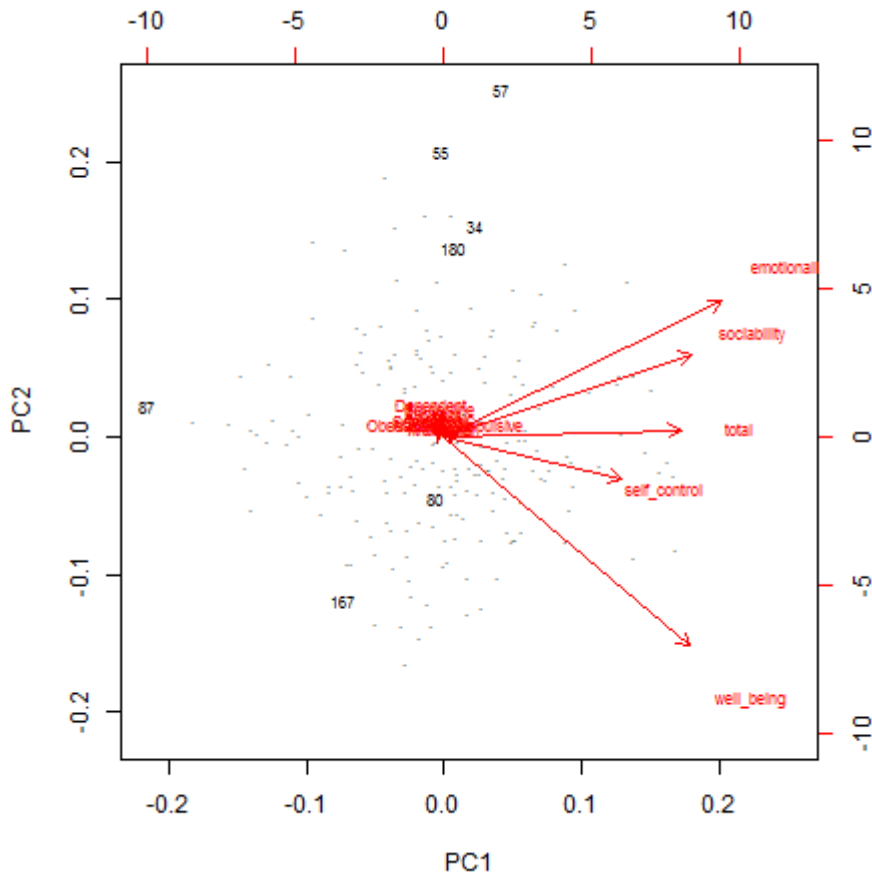
Πίνακας 3 Πίνακας Περιγραφικών Στατιστικών Κανονικών Παρατηρήσεων<sup>6</sup>

	<i>Mean</i>	<i>Median</i>	<i>Min</i>	<i>Max</i>	<i>Std.dev</i>
<i>Well_being</i>	3.33	3.0	2.0	5.17	1.19
<i>Self_control</i>	3.52	3.5	1.51	5.0	1.55
<i>emotionality</i>	4.38	5.0	1.63	6.5	1.68
<i>sociability</i>	4.0	4.5	1.83	5.83	1.63
<i>Paranoid</i>	0.57	0.57	0.43	0.71	0.14
<i>Schizoid</i>	0.14	0.14	0.00	0.29	0.12
<i>Schizotypal</i>	0.4	0.44	0.11	0.67	0.19
<i>Histrionic</i>	0.52	0.5	0.13	0.88	0.23
<i>Antisocial</i>	0.31	0.29	0.14	0.57	0.17
<i>Narcissistic</i>	0.33	0.22	0.22	0.67	0.17
<i>Borderline</i>	0.44	0.44	0.11	0.67	0.17
<i>Obsessive_comp</i>	0.5	0.5	0.25	0.88	0.21
<i>Dependent</i>	0.32	0.25	0.0	0.63	0.20
<i>Avoidant</i>	0.59	0.5	0.25	0.88	0.25

Πίνακας 4 Πίνακας Περιγραφικών Στατιστικών Παρατηρήσεων Outlier

<sup>6</sup>Όπου: mean η μέση τιμή, median η διάμεσος, std.dev η τυπική απόκλιση, min η ελάχιστη τιμή, max η μέγιστη τιμή και range το εύρος.

Παρατηρώντας τους δύο πίνακες των στατιστικών μεγεθών για κάθε μία ψυχομετρική κλίμακα ξεχωριστά παρατηρούμε ότι οι απομονωμένες παρατηρήσεις του *emotionality* είναι εκτός του εύρους τιμών (min,max) των κανονικών παρατηρήσεων. Εντούτοις χρησιμοποιώντας τον LOF ο οποίος εξετάζει όλα τα χαρακτηριστικά ταυτόχρονα, εντόπισε περιπτώσεις οι οποίες είναι απομονωμένες όπως φαίνετε και στο παρακάτω σχήμα.

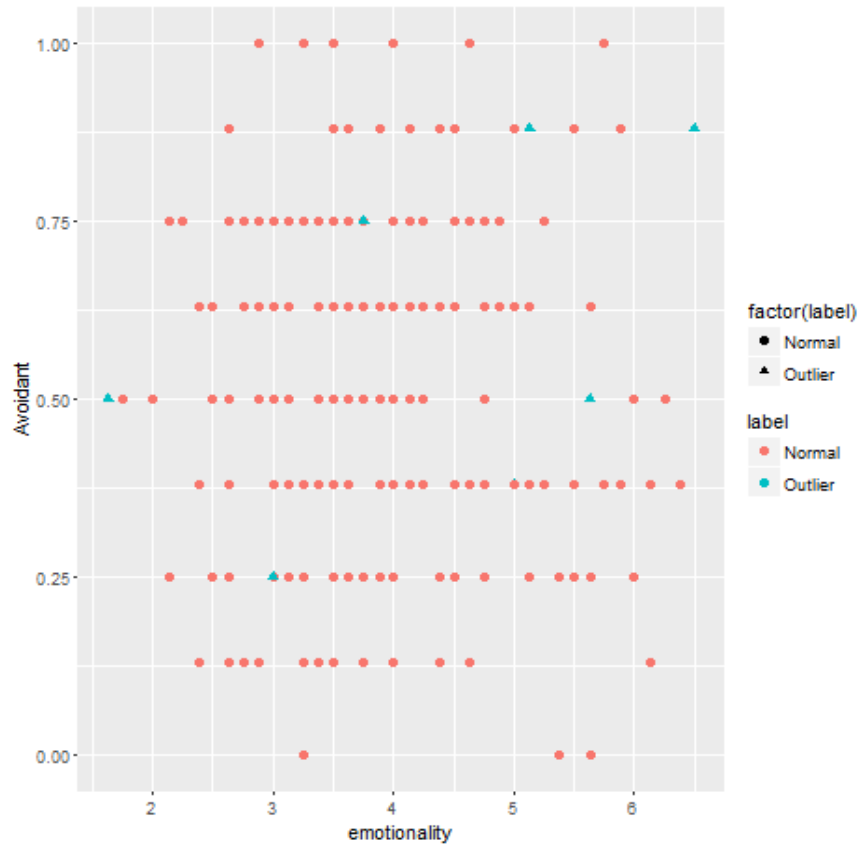


Σχήμα 12 Ανάλυση Κύριων Συνιστωσών(PCA)

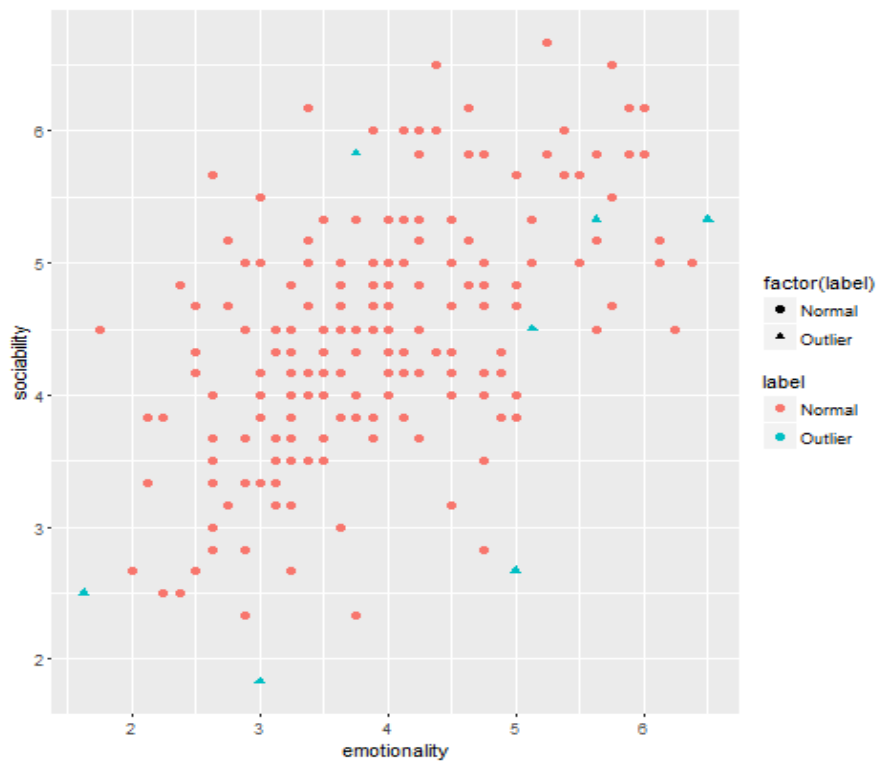
Απεικονίζοντας τις δύο κύριες συνιστώσες συμπεραίνουμε ότι οι παρατηρήσεις 34, 55, 57, 80, 87, 167, 180 είναι οι **ακραίες τιμές** (outlier) και με την μέθοδο της *Ανάλυσης Κύριων Συνιστωσών* (PCA)<sup>7</sup>.

Παρακάτω παρατίθενται ενδεικτικά *Διαγράμματα Διασποράς Απομονωμένων Τιμών* όπου οι *ακραίες τιμές* είναι απομακρυσμένες από τα κέντρα των κανονικών παρατηρήσεων.

<sup>7</sup> Αναζητά ένα k n-διάστατο ορθογώνιο παράγοντα, που αναπαριστά τα δεδομένα με αποτέλεσμα την εύρεση της μεγαλύτερης πληροφορίας στο πως κατανέμονται τα δεδομένα στο χώρο.



Σχήμα 13 Διάγραμμα Διασποράς Απομονωμένων Τιμών Emotionality-Avoidant



Σχήμα 14 Διάγραμμα Διασποράς Απομονωμένων Τιμών Sociability-Emotionality



## 6 Συμπεράσματα

Μέσα από τις ψυχομετρικές κλίμακες των ερωτηματολογίων Διαταραχών Προσωπικότητας και Συναισθηματικής Νοημοσύνης προσπαθήσαμε να αποκωδικοποιήσουμε τα χαρακτηριστικά της Προσωπικότητας τους αλλά και τον τρόπο με τον οποίο συμπεριφέρονται στους γύρω τους οι φοιτητές του ΤΕΙ Δυτικής Ελλάδας, Ιονίων Νήσων και άλλων εκπαιδευτικών ιδρυμάτων και πανεπιστημίων. Η τεχνική των Δέντρων αποφάσεων μας βοήθησε στο να περιγράψουμε το σύνολο δεδομένων βρίσκοντας κανόνες μεταξύ των χαρακτηριστικών που θέλαμε να εξετάσουμε. Όσο αναφορά τους κανόνες συσχέτισης χρησιμοποιήσαμε τον αλγόριθμο Apriori καθώς θέλαμε να εξάγουμε αποτελέσματα με υψηλή αξιοπιστία. Βασικές ρυθμιστικές παράμετροι του αλγορίθμου είναι η εμπιστοσύνη (**confidence**) που καθορίζει το πόσο ισχυρό είναι το πρότυπο που εντοπίστηκε και η υποστήριξη (**support**) το πόσο σημαντικό είναι για τον τελικό χρήστη. Τέλος για τις απομονωμένες τιμές χρησιμοποιήσαμε τον αλγόριθμο LOF ούτως ώστε να εντοπίσουμε περιπτώσεις που διαφέρουν από το σύνολο δεδομένων και να τις μελετήσουμε ξεχωριστά. Ορισμένοι αλγόριθμοι κατηγοριοποίησης είναι ευαίσθητοι στην παρουσία απομονωμένων περιπτώσεων οπότε καλό θα ήταν να εντοπίζονται και να απομακρύνονται από την ανάλυση. Συνεπώς οι αλγόριθμοι για απομονωμένες τιμές χρειάζονται ως ένα βήμα προτού εφαρμοστούν οι αλγόριθμοι για την Κατηγοριοποίηση και την Ανάλυση των Συσχετίσεων. Ωστόσο στην προκειμένη έρευνα θεωρήσαμε σκόπιμο να μην τις αφαιρέσουμε από το σύνολο δεδομένων καθώς από την σύγκριση των βασικών στατιστικών μεγεθών μεταξύ των απομονωμένων περιπτώσεων (κατά LOF) και των κανονικών περιπτώσεων δεν φαίνεται κάποια σημαντική απόκλιση σε όλα τα χαρακτηριστικά εκτός του *emotionality*.

Επομένως, εφαρμόζοντας τις κατάλληλες τεχνικές εξόρυξης δεδομένων χρησιμοποιώντας αλγορίθμους Κατηγοριοποίησης και Ανάλυσης Συσχετίσεων προέκυψαν τα εξής συμπεράσματα:

Οι φοιτητές του ΤΕΙ Δυτικής Ελλάδας στις κλίμακες ηλικιών από 23-24 και 25-27 με «κανονικό» δείκτη *παρanoiδούς* συμπεριφοράς (έλλειψη εμπιστοσύνης και καχυποψία στους γύρω τους) τότε παρουσιάζουν «κανονικό» δείκτη

*συναισθηματικότητας*, ενώ «χαμηλό» δείκτη *παρανοειδούς* συμπεριφοράς όταν ο *οιστριωνικός* δείκτης (αναζήτηση προσοχής) είναι «κανονικός».

Οι φοιτητές στις ηλικίες από 16-20, 23-24 & 25-27 με «χαμηλή» *κοινωνικότητα* παρουσιάζουν *οιστριωνικό* δείκτη (αναζήτηση προσοχής) «πάνω του κανονικού» και «χαμηλή» *συναισθηματικότητα* (δυσκολία αναγνώρισης εσωτερικών συναισθηματικών καταστάσεων, λιγότερο καρποφόρες προσωπικές σχέσεις) όταν ο δείκτης *κοινωνικότητας* είναι «κανονικός».

Συγκεκριμένα ο άνδρικός φοιτητικός πληθυσμός που σπουδάζει στο Πανεπιστήμιο Πατρών ή σε κάποιο άλλο Ίδρυμα με «κανονικό» δείκτη *ευημερίας* (ευτυχισμένοι-ολοκληρωμένοι) παρουσιάζει «κανονικό» ή «υψηλό δείκτη» *συναισθηματικότητας* και «κανονικό» δείκτη *ναρκισσισμού*. Δηλαδή οι άνδρες φοιτητές που είναι ευτυχισμένοι απ την ζωή τους σε κανονικό επίπεδο παρουσιάζουν αποτελέσματα ικανοποιητικού έως πολύ μεγαλύτερου εύρους ικανοτήτων όσο αναφορά την διαχείριση των συναισθημάτων τους όταν αυτο συνδιάζεται με την αίσθηση του προσωπικού τους μεγαλείου να κυμαίνεται σε κανονικά επίπεδα.

Τέλος παρατηρήσαμε στον δείκτη *αυτοελέγχου* όταν κυμαίνετε στο «κανονικό» αλλά και στο «υψηλό» επίπεδο ότι στους άντρες παρουσιάζεται «κανονικός» ή «υψηλός» δείκτης *κοινωνικότητας*, ενώ στις γυναίκες παρουσιάζεται όταν φοιτούν στα ΤΕΙ Ιονίων Νήσων, Δυτικής Ελλάδας ή Πανεπιστήμιο Πατρών. Συνεπώς οι άντρες φοιτητές με κανονικό και υψηλό επίπεδο διαχείρισης εξωτερικών πιέσεων και άγχους παρουσιάζουν υψηλή δυνατότητα κοινωνικής συναναστροφής στους γύρω τους.

## Βιβλιογραφία

- evdoxos.ds.uni.gr*. (n.d.). Ανάκτηση από <https://evdoxos.ds.unipi.gr/modules/document/document.php?course=DS154&openDir=/531995e8fihl>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (n.d.). *From Data Mining to Knowledge Discovery in Databases*.
- Frawley, W., Piatetsky, G., & Matheus, C. (1992). *Knowledge Discovery in Databases: An Overview*.
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*.
- Hawkin, D. (1980). *Identification Of Outliers*. Chapman and Hall.
- itl.nist.gov*. (n.d.). Ανάκτηση από <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>
- Janca, A., Sartorius, N., & Loranger, A. W. (2007). *Assessment and diagnosis of personality disorders The ICD-10 international personality disorder examination (IPDE)*.
- Kriegel, H.-P., Kroger, P., & Zimek, A. (2010). *Outlier Detection Techniques*. Munich, Germany.
- Loranger, A. (n.d.). *International Personality Disorder Examination According to DSM-IV*.
- Petrides, K. V. (2001). *TEIQue Interpretations*.
- Poole, D., Mcworth, A., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press .
- siam.org*. (n.d.). Ανάκτηση από <https://www.siam.org/meetings/sdm10/tutorial3.pdf>
- Vegue-Gonzalez, M., & Alvaro-Brun, E. (2008). *Validity of International Personality Disorder Examination (IPDE) questionnaire on a sample of prisoner inmates*.
- www.datamining.gr*. (n.d.). Ανάκτηση από [www.datamining.gr](http://www.datamining.gr)
- www.ecourse.uoi.gr*. (n.d.). Ανάκτηση από <http://ecourse.uoi.gr/course/view.php?id=1051>
- www.-users.cs.umn.edu*. (n.d.). Ανάκτηση από <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>
- www.whonet.aueb.gr*. (n.d.).
- www.wikipedia.org*. (n.d.). Ανάκτηση από <https://en.wikipedia.org/wiki/Outlier>
- www.wikipedia.org*. (n.d.). Ανάκτηση από [https://en.wikipedia.org/wiki/Local\\_outlier\\_factor](https://en.wikipedia.org/wiki/Local_outlier_factor)

- Αμοργιανός, Π. Γ. (2010). *Συναισθηματική Νοημοσύνη & Ηγεσία*. Πειραιάς.
- Γιαγλής, Γ. (2004). *Στοιχεία Ψυχομετρίας*.
- Δαλακούρα, Α. (2011). *Συναισθηματική Νοημοσύνη*.
- Ευσταθίου, Γ. (2005). *Διαταραχές Προσωπικότητας (Σημειώσεις 2ου Έτους)* (Πακέτο σημειώσεων 13 εκδ.). Εκδόσεις του Ινστιτούτου Έρευνας και Θεραπείας της Συμπεριφοράς.
- Κεραυνού, Ε. (2000). *Τεχνητή Νοημοσύνη και Έμπειρα Συστήματα*.
- Λιάκου, Ε. Γ. (2008). *ΗΓΕΣΙΑ ΚΑΙ ΣΥΝΑΙΣΘΗΜΑΤΙΚΗ ΝΟΗΜΟΣΥΝΗ*. Πειραιάς.
- Μ.Βαζιργιάννης, Μ. (2003). *Εξόρυξη Γνώσης από Βάσεις Δεδομένων*.
- Μπαμπαλιάρης, Α. (2011). *ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ*. Καβάλα.
- Παγουρόπουλος, Α. (2006). *Data mining στην Χρηματοοικονομική Ανάλυση*.
- Παντελή, Α. (2012). *Υπολογιστική Νοημοσύνη στην οικονομία για την πρόβλεψη της νομισματικής ισοτιμίας μεταξύ δολαρίου ΗΠΑ και Ευρώ*.
- Παπακωνσταντίνου, Λ. (2009). *Εξόρυξη Κανόνων απο Καρδιαγγειακή βάση με τη χρήση αλγόριθμων συσχέτισης*.
- Πρωτόπαππας, Α. (2008). *Το υποκείμενο στη γνωσιακή επιστήμη*.

## Παράρτημα

### Δεντρα αποφάσεων

```

Decision_Trees_v1 <- function(df.data) {

  set.seed(9850)
  vars <- names(df.data)
  train = sample(1:nrow(df.data), nrow(df.data)*0.8)
  training_data = df.data[train, ]
  testing_data = df.data[-train, ]
#-----Senario 1 -----#
depend.var <- c("birthyear")
ind.vars <- c("emotionality", "self_control", "cityStayMe",
"educationMe", "well_being", "sociability", "sex")
fmla <- as.formula( paste(depend.var, paste(ind.vars, collapse="+"),
sep=" ~ ") )

  ###----- Senario 2 -----###
# depend.var <- c("sex")
# ind.vars <- c("Schizoid", "Antisocial", "Dependent", "Avoissant",
"Schizotypal", "Narcissistic", "Borderline", "self_control",
"Paranoid", "Histrionic", "well_being", "sociability", "birthyear")
# fmla <- as.formula( paste(depend.var, paste(ind.vars,
collapse="+"), sep=" ~ ") )

#-----Senario 3 IPDE---###
#depend.var <- c("birthyear")
#ind.vars <- c("cityStayMe", "educationMe", "sex", "Paranoid",
"Schizoid", "Antisocial", "Avoissant", "Schizotypal", "Narcissistic",
"Borderline", "Histrionic", "Dependent")
#fmla <- as.formula( paste(depend.var, paste(ind.vars,
collapse="+"), sep=" ~ ") )

#----- senario 4 -----#
#depend.var <- c("well_being")
#ind.vars <- c("emotionality", "self_control", "birthyear",
"educationMe", "Borderline", "sex", "Narcissistic")
#fmla <- as.formula( paste(depend.var, paste(ind.vars,
collapse="+"), sep=" ~ ") )
# ptree <- prune(tree_model,
#               cp=
tree_model$cp[which.min(tree_model$cp[,"xerror"]), "CP"])
#mean(pred_ptree != testing_data$well_being)

#----- senario 5 -----#
# depend.var <- c("emotionality")
# ind.vars <- c("Paranoid", "Schizoid", "educationMe", "birthyear",
"Antisocial", "Histrionic", "Borderline")
# fmla <- as.formula( paste(depend.var, paste(ind.vars,
collapse="+"), sep=" ~ ") )

```

```

# ptree <- prune(tree_model, cp = 0.017)
# mean(pred_ptree != testing_data$emotionality)

#----- senario 6 -----#
# depend.var <- c("sociability")
# ind.vars <- c("Paranoid", "educationMe", "birthyear", "Antisocial",
"Hi strionic", "sex", "emotionality")
# fmla <- as.formula( paste(depend.var, paste(ind.vars,
collapse="+"), sep=" ~ ") )
# ptree <- prune(tree_model, cp = 0.025)
# mean(pred_ptree != testing_data$sociability)

tree_model <- rpart(formula = fmla, data=training_data, method =
"class", parms = list(split = 'gini'),
control = rpart.control(minsplit=10, cp=0.01,
xval=100))

prp(tree_model, type=1, under= TRUE, tweak = 1.2, branch.type=0,
gap= 0 , faclen = 0, varlen = 0, leaf.round = 3, box.col =
c("pink", "palegreen3", "slateblue2", "mediumorchi d1")
[tree_model$frame$yval])
#, prefix="fraction\n")

### testing data ##
pred <- predict(tree_model, testing_data[, !(names(testing_data)
%in% depend.var) ], type = "class" )

conf.mat <- table(testing_data[, depend.var], pred)
acc = sum(diag( conf.mat ))*100/sum(sum( conf.mat ))
cl.err <- 100-acc

#----- cross-validation-----#
printcp(tree_model)
plotcp(tree_model)
#----- prune tree-----#
ptree <- prune(tree_model,
cp=
tree_model$scptable[which.min(tree_model$scptable[, "xerror"]), "CP"])
#ptree <- prune(tree_model, cp = 0.0121)

prp(ptree, type=1, under= TRUE, tweak = 1.2, branch.type=0,
gap= 0 , faclen = 0, varlen = 0, leaf.round = 3, box.col =
c("pink", "palegreen3", "slateblue2", "mediumorchi d1")
[tree_model$frame$yval])
#, prefix="fraction\n")
#----- predict prune tree-----#
#pred_ptree <- predict(ptree, testing_data[, !(names(testing_data)
%in% depend.var) ], type = "class" )
#mean(pred_ptree != testing_data$birthyear)

return(list('DT' = tree_model, 'Prediction'=pred, 'CM'=conf.mat,
'Accuracy'=acc, 'Classification.Error'=cl.err) )
}

```

*Apriori*

```

library(scatterplot3d)
library(readxl)
library(arules)
library(arulesViz)
library(CCA)
source("discretize.data.R")

pathname = paste(getwd(), "Emotional intelligence.IPDE.xlsx", sep = "/")
dataset <-
  read_excel(pathname, sheet = 1, col_names = TRUE, na = 'n')
df.data <- discretize.data(dataset)

# ----- Datasets Senariwn-----###
df.data.male <- df.data[df.data$sex=="male", ]
df.data.educationMe <- df.data[df.data$educationMe=="teiwest", ]
df.data.teiIwn <- df.data[df.data$educationMe=="teiIwn", ]
df.data.other <- df.data[df.data$educationMe=="other", ]
df.data.birthyear <- df.data[df.data$birthyear=="[21, 22]", ]
df.data.birthyear1 <- df.data[df.data$birthyear=="[23, 24]", ]
#df.data.birthyear2 <- df.data[df.data$birthyear=="[25, 27]", ]

#----- Sixxetiseis tw n 2 erwtimatologiwn-----
EI <- dataset[, 1:4]
IPDE <- dataset[, 6:15]
a <- matcor(EI, IPDE)
#Creating rules

rules <- apriori(
  df.data.birthyear2,
  parameter = list(supp = 0.35, conf = 0.9)) #,
  # appearance = list(rhs = c("sex=female"), default = "lhs"))

inspect(rules)
#summary(rules)

#Shorting rules by lift
rules <- sort(rules, by = 'lift')

#Finding redundant rules

subset.matrix <- is.subset(rules, rules)
subset.matrix[lower.tri(subset.matrix, diag = TRUE)] <- NA
redundant <- colSums(subset.matrix, na.rm = TRUE) >= 1
which(redundant)

#Clearing redundant rules

rules.pruned <- rules[!redundant]

```

```

rules.final <- rules.pruned

summary(rules.final)      #Descriptive Statistics
inspect(rules.final)     #View rules

#Subsetting rules by lift value

subrules <- subset(rules.final, lift > 1.3)
subrules
inspect(subrules)

#-----Plots-----

#1. Grouped matrix plot
plot (rules.final, method = "grouped")

#2. 3D Matrix
plot (
  rules.final, method = "matrix3D", measure = "lift", control = list
(reorder =
TRUE))

#3. Scatterplot
plot (rules.final)

#4. --Only with a few rules-- Graph
plot (
  subrules, method = "graph", control = list (layout =
igraph::with_graphopt (
  spring.const = 5, mass = 50
)))

```



*Απομονωμένες Τιμές*

```

outlier.detection <- function(dataset)
{

subDir <- "Images"

if (file.exists(subDir)){
  setwd(file.path(root_path, subDir))
} else {
  dir.create(file.path(root_path, subDir))
  setwd(file.path(root_path, subDir))
}

dataset_test <- dataset[, 1:15]

#outlier.scores <- lofactor(dataset_test, k=2)

#new code -- k = 10% of observations (rows)
nr <- nrow(dataset_test)
k = round( nr*0.05 )
outlier.scores <- lof(dataset_test, k, cores=2) # using function
from Rlof package

png('Density.Probability.Funct.png')
c <- density(outlier.scores)
plot(c, main = "Density Probability Function of LOF values")
polygon(c, col = "#4d73ff", border = "#9933ff")
abline(v = 1.4, lty=3, col="#999999")
dev.off()

#-----
# pick top outliers according to Density Function, in this case
thresh = 1.3

sort.outlier.scores.values <- outlier.scores[order(outlier.scores,
decreasing=T)]
sort.outlier.scores.pos <- order(outlier.scores, decreasing=T)

# times kai theseis Outliers (pragmatikes theseis sto sinolo
dedomenwn dataset)
thresh = 1.4
cat("\n Outliers Position: \n")
pos.outliers <- which(outlier.scores > thresh)
cat("\n Outliers Values: \n")
val.outliers <- outlier.scores[outlier.scores > thresh]

# Summary statistics compare normal data vs outliers data
library(pastecs)

```

```

stats.outliers <- stat.desc(dataset_test[pos.outliers, ])
stats.normal <- stat.desc(dataset_test[-pos.outliers, ])

library(psych)
stats.outliers.descr <- describe(dataset_test[pos.outliers, ])
stats.normal.descr <- describe(dataset_test[-pos.outliers, ])

labels <- 1:nr
labels[-pos.outliers] <- "."
png('Biplot.PCA.png')
biplot(prcomp(dataset_test), cex=.7, xlab=labels)
dev.off()

library(ggplot2)

nc <- ncol(dataset_test)
names.ds <- names(dataset_test)

# Create a new column named "label"
dataset_test$label <- ifelse(rownames(dataset_test) %in%
pos.outliers, "Outlier" , "Normal")

for (i in 1:(nc-1)){
  for (j in (i+1):nc){

    fig.name <- paste('Scatterplot', names.ds[i], names.ds[j], "png",
sep=".")
    png(fig.name)
    p <- ggplot(dataset_test, aes_string(x=names.ds[i],
y=names.ds[j], color='label' ) )
    p <- p + geom_point(size=2, aes(shape=factor(label)))
    print(p)
    dev.off()
  }
}

res <- list( "LOF.Values"=sort.outlier.scores.values,
"Pos.LOF"=sort.outlier.scores.pos,
           "Stats.Outliers" = stats.outliers.descr, "Stats.Normal "
= stats.normal.descr)
}

```

*Διακριτοποίηση*

```

discretize.data <- function(data){

  #----- IPDE klimakes

  df7 <- as.data.frame(cbind("Paranoid"=data$Paranoid,
"Schizoid"=data$Schizoid, "Anti social"=data$Anti social))
  for (i in 1:ncol(df7) ){
    x <- round(df7[,i] , digits=2)
    x.disc <- discretize(x, method="fixed", categories = c(-Inf,
0.435, Inf), labels=c("Normal", "Outranged"), digits=2 )
    df7[,i] <- as.data.frame(x.disc)
  }

  df8 <- as.data.frame(cbind("Histrionic"=data$Histrionic,
"Obsessive-compulsive"=data$Obsessive_Compulsive,
"Dependent"=data$Dependent, "Avoidant"=data$Avoidant))

  for (i in 1:ncol(df8) ){
    x <- round(df8[,i] , digits=2)
    x.disc <- discretize( x, method="fixed", categories = c(-Inf,
0.385, Inf), labels=c("Normal", "Outranged"), digits=2 )
    df8[,i] <- as.data.frame(x.disc)
  }

  df9 <- as.data.frame(cbind("Schizotypal"=data$Schizotypal,
"Narcissistic"=data$Narcissistic, "Borderline"=data$Borderline))

  for (i in 1:ncol(df9) ){
    x <- round(df9[,i] , digits=2)
    x.disc <- discretize( x, method="fixed", categories = c(-Inf,
0.34, Inf), labels=c("Normal", "Outranged"), digits=2 )
    df9[,i] <- as.data.frame(x.disc)
  }

  #----- EI klimakes
  x <- round(data$well_being, digits=2)
  step <- diff(range(x))/3
  x.disc <- discretize( x, method="fixed", categories = c(-Inf,
min(x)+step, min(x)+2*step, Inf),
                      labels=c("Low", "Normal", "High"),
digits=2 )
  well_being <- as.data.frame(x.disc)
  names(well_being) = "well_being"

  x <- round(data$self_control, digits=2)
  step <- diff(range(x))/3
  x.disc <- discretize( x, method="fixed", categories = c(-Inf,
min(x)+step, min(x)+2*step, Inf),
                      labels=c("Low", "Normal", "High"), digits=2
)
  self_control <- as.data.frame( x.disc )
}

```

```

names(self_control)='self_control'

x <- round(data$emotinality, digits=2)
step <- diff(range(x))/3
x.disc <- discretize( x, method="fixed", categories = c(-Inf,
min(x)+step, min(x)+2*step, Inf),
                    labels=c("Low", "Normal", "High"), digits=2
)
emotinality <- as.data.frame( x.disc )
names(emotinality)='emotinality'

x <- round(data$sociability, digits=2)
step <- diff(range(x))/3
x.disc <- discretize( x, method="fixed", categories = c(-Inf,
min(x)+step, min(x)+2*step, Inf),
                    labels=c("Low", "Normal", "High"), digits=2
)
sociability <- as.data.frame(x.disc)
names(sociability)='sociability'

# Discretize Other attributes

xmin = min(data$birthyear)
xmax = max(data$birthyear)

x.disc <- discretize( data$birthyear, method="fixed", categories =
c(xmin, 21, 23, 25, 28, xmax),
                    labels=c("[16, 20]", "[21, 22]", "[23, 24]",
"[25, 27]", "[28, 55]"), digits=2 )
birthyear <- as.data.frame( x.disc )
names(birthyear) = 'birthyear'

x.disc <- discretize(data$ageFather, method="interval", categories
= 3, digits=2 )
ageFather <- as.data.frame( x.disc )
names(ageFather) = 'ageFather'

x.disc <- discretize(data$ageMother, method="interval", categories
= 3, digits=2 )
ageMother <- as.data.frame(x.disc)
names(ageMother) = 'ageMother'

sex <- as.data.frame(cbind("sex"=data$sex))
educationMe <-
as.data.frame(cbind("educationMe"=data$educationMe))
educationFather <-
as.data.frame(cbind("educationFather"=data$educationFather))
educationMother <-
as.data.frame(cbind("educationMother"=data$educationMother))
cityBirthMe <-
as.data.frame(cbind("cityBirthMe"=data$cityBirthMe))
cityStayMe <- as.data.frame(cbind("cityStayMe"=data$cityStayMe))
citySameMe <- as.data.frame(cbind("citySameMe"=data$citySameMe))
jobFather <- as.data.frame(cbind("jobFather"=data$jobFather))

```

```

jobMother <- as.data.frame(cbind("jobMother"=data$jobMother))

newdata <- cbind(df7, df8, df9, well_being, self_control, emotionality,
                sociability, birthyear, ageFather, ageMother,
                sex, educationMe, educationFather, educationMother,
                ci tyBi rthMe, ci tyStayMe, ci tySameMe, j obFather, j obMother)

  return(newdata)
}

```

### *Τελικό Script*

```

library(readxl)
library(ggplot2)
library(arules)
library(tree)
library(rpart)
library(rpart.plot)
library(RIof)
library(DMwR)

#root_path <- getwd()
#setwd(root_path)

source("discretize.data.R")

pathname = paste(getwd(), "Emotional intelligence.IPDE.xlsx", sep="/")
dataset <- read_excel(pathname, sheet =1 , col_names = TRUE , na='n')

# Call function to preprocess the dataset for using Apriori
algorithm so as to create Association Rules
df.data <- discretize.data(dataset)
View(df.data)

source("Decision_Trees.R")
DecTree <- Decision_Trees(df.data)

source("Decision_Trees_v1.R")
DecTree <- Decision_Trees_v1(df.data)

new code
source("Outlier_detection.R")
res.outlier.detection <- outlier.detection(dataset)

change directory
setwd(root_path)

```