

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ  
ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νοήμονες μεθοδολογίες ανάλυσης  
δεδομένων με εφαρμογές στην  
επεξεργασία ερωτηματολογίων

Παναγιώτης Λεονταρίδης

Χανιά 2003

## ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	1
2. Η ΕΝΝΟΙΑ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ Η ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ	3
2.1. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	3
2.2. ΕΞΟΥΣΙΑ ΓΝΩΣΗΣ ΑΠΟ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ.	7
2.2.1 Τι είναι η εξόρυξη γνώσης από βάσεις δεδομένων	7
2.2.2 Πως παράγονται μεγάλοι όγκοι δεδομένων σήμερα	7
2.2.3 Ανάγκη αποθήκευσης δεδομένων	8
2.2.4 Φθινή υπολογιστική ισχύς	8
2.2.5 Ισχυρότατος ανταγωνισμός των επιχειρήσεων	8
2.3. ΧΡΗΣΙΜΟΤΗΤΑ ΤΗΣ ΕΞΟΥΣΙΑΣ ΓΝΩΣΗΣ ΑΠΟ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	9
2.4. ΜΟΝΤΕΛΑ ΕΞΟΥΣΙΑΣ ΓΝΩΣΗΣ ΑΠΟ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	12
2.4.1 Κατηγοριοποίηση	13
2.4.2 Παλινδρόμηση	13
2.4.3 Χρονοσειρές	13
2.4.4 Τμηματοποίηση	13
2.4.5 Ποσοτική ανάλυση	14
2.4.6 Ανάλυση συσχετίσεων	14
2.4.7 Ανακάλυψη ακολουθιών στοιχείων	14
2.5. ΚΑΤΑΣΚΕΥΗ ΜΟΝΤΕΛΩΝ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΜΕ ΤΕΧΝΙΚΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.	15
2.6. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΛΕΙΤΟΥΡΓΙΑ ΜΟΝΤΕΛΩΝ ΕΞΟΥΣΙΑΣ ΔΕΔΟΜΕΝΩΝ.	19
2.7. ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΕΞΟΥΣΙΑΣ ΔΕΔΟΜΕΝΩΝ	21
2.7.1 Ανάλυση καλαθιού αγοράς	21

2.7.2	Συλλογιστική βασισμένη στη μνήμη του συστήματος	22
2.7.3	Εντοπισμός ομάδων	23
2.7.4	Ποσοτική Ανάλυση (Quantitative Analysis)	23
2.7.5	Ανάλυση συνδέσεων	25
2.7.6	Επαγωγικά δένδρα απόφασης	25
2.7.7	Τεχνητά νευρωνικά δίκτυα	26
2.7.8	Γενετικοί αλγόριθμοι	26
2.7.9	Αναλυτική επεξεργασία δεδομένων σε πραγματικό χρόνο.	26
2.8.	ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΕΞΟΥΣΗΣ ΔΕΔΟΜΕΝΩΝ	26
2.8.1	Η κατασκευή προφίλ πελατών (customer profiling)	27
2.8.2	Το στοχευόμενο marketing (targeted marketing)	27
2.8.3	Η ανάλυση του καλαθιού αγοράς (market basket analysis)	27
2.9.	ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ ΤΩΝ ΜΕΘΟΔΩΝ	28
2.10.	ΈΝΑ ΠΑΡΑΔΕΙΓΜΑ ΕΦΑΡΜΟΓΗΣ ΕΞΟΥΣΗΣ ΔΕΔΟΜΕΝΩΝ	28
2.10.1	Η βάση δεδομένων	29
2.10.2	Χρήση του Aira	31
2.10.3	Ερμηνεία της λίστας κανόνων	36
2.11.	ΑΠΟΚΤΗΣΗ ΓΝΩΣΗΣ	37
2.11.1	Η επιλογή ενός μόνο ειδικού παρουσιάζει προβλήματα	38
2.11.2	Αντίθετα η επιλογή ομάδας ειδικών προσφέρει πλεονεκτήματα	38
3.	ΕΞΟΥΣΗ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ΓΙΑ ΜΑΡΚΕΤΙΝΓΚ, ΚΑΙ ΠΩΛΗΣΕΙΣ.	44
3.1.	ΕΙΣΑΓΩΓΗ	44
3.2.	ΈΝΝΟΙΑ ΚΑΙ ΣΥΣΤΑΤΙΚΑ ΕΝΟΣ ΣΥΣΤΗΜΑΤΟΣ ΠΛΗΡΟΦΟΡΙΩΝ ΜΑΡΚΕΤΙΝΓΚ	46

3.3. ΣΥΣΤΗΜΑ ΑΝΙΧΝΕΥΣΗΣ ΠΛΗΡΟΦΟΡΙΩΝ	
ΜΑΡΚΕΤΙΝΓΚ _____	46
3.4. Η ΧΡΗΣΗ ΤΗΣ ΕΡΕΥΝΑΣ ΣΤΟ ΧΩΡΟ ΤΟΥ	
ΜΑΡΚΕΤΙΝΓΚ _____	49
3.5. ΜΙΑ ΜΙΚΡΗ ΑΝΑΦΟΡΑ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ	
ΕΡΕΥΝΑΣ ΜΑΡΚΕΤΙΝΓΚ _____	50
4. ΠΡΟΓΡΑΜΜΑΤΑ ΝΟΗΜΟΝΩΝ ΠΡΟΣΕΓΓΙΣΕΩΝ (PC TOOLS)	52
4.1. ΤΟ ΠΡΟΓΡΑΜΜΑ ΕΥΡΕΣΗΣ ΑΝΩΜΑΛΙΩΝ	
ΔΕΔΟΜΕΝΩΝ GRITBOT _____	52
4.1.1 Προετοιμασία δεδομένων εισόδου στο Gritbot _____	52
4.1.2 Αρχεία εφαρμογών _____	53
4.1.3 Περιβάλλον χρήσης _____	55
4.1.4 Ελέγχοντας τα δεδομένα για ανωμαλίες _____	55
4.1.5 Συμπεράσματα για το πρόγραμμα Gritbot _____	56
4.2. ΤΟ ΠΡΟΓΡΑΜΜΑ MAGNUM OPUS _____	57
4.2.1 Οι δυο τύποι δεδομένων που είναι αποδεκτοί από το Magnum Opus _____	58
4.2.2 Τα αρχεία εισόδου. _____	58
4.2.3 Πως γίνεται η εισαγωγή δεδομένων στο πρόγραμμα _____	61
4.2.4 Το περιβάλλον του προγράμματος _____	62
4.2.5 Επεξεργασία δεδομένων _____	63
4.2.6 Τα αποτελέσματα της αναζήτησης _____	64
4.2.7 Συμπεράσματα για το πρόγραμμα Magnum Opus _____	67
5. ΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝ	
ΠΡΟΓΡΑΜΜΑΤΑ ΝΟΗΜΟΝΩΝ ΠΡΟΣΕΓΓΙΣΕΩΝ (SEE5/C5.0) _____	68
5.1. ΕΙΣΑΓΩΓΗ. _____	68
5.2. ΠΑΡΑΔΕΙΓΜΑΤΑ. _____	68
5.2.1 Κατανέμοντας σπίτια στη Βοστώνη, ανάλογα με την αξία. _____	68

5.2.2	Εντοπίζοντας αγγελίες στο διαδίκτυο	70
5.2.3	Ελέγχοντας μια παραγωγική διαδικασία	72
6. ΠΡΑΓΜΑΤΙΚΗ ΕΦΑΡΜΟΓΗ ΝΟΗΜΟΝΟΣ ΠΡΟΣΕΓΓΙΣΕΩΣ:		
«ΑΝΑΛΥΣΗ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ΓΙΑ ΤΟ ΒΙΟΤΕΧΝΙΚΟ ΠΑΡΚΟ ΧΑΝΙΩΝ»		75
6.1.	ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	75
6.2.	ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ΜΕ ΤΟ ΠΡΟΓΡΑΜΜΑ SEE5	96
6.2.1	Εισαγωγή-Παρουσίαση των μεταβλητών	96
6.2.2	Μεθοδολογία	100
6.2.3	Τα αποτελέσματα του προγράμματος	101
6.2.4	Άλλες δυνατότητες του See5	108
7. ΠΡΑΓΜΑΤΙΚΗ ΕΦΑΡΜΟΓΗ ΠΟΣΟΤΙΚΗΣ ΑΝΑΛΥΣΗΣ:		
«ΑΝΑΛΥΣΗ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ICAP»		109
7.1.	Η ΜΕΛΕΤΗ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΤΗΣ Q-ANALYSIS	109
7.1.1	Η μεθοδολογία της Μελέτης.	109
7.1.2	Μελέτη του πρώτου πίνακα δεδομένων με τη βοήθεια της Q-Analysis.	110
7.1.3	Μελέτη του δεύτερου πίνακα δεδομένων με τη βοήθεια της Q-Analysis.	112
7.1.4	Μελέτη του τρίτου πίνακα δεδομένων με τη βοήθεια της Q-Analysis.	113
7.2.	ΠΙΝΑΚΕΣ ΔΕΔΟΜΕΝΩΝ ΤΗΣ Q-ΑΝΑΛΥΣΗΣ	115
7.2.1	Ερώτηση 3 <sup>η</sup> : Είστε ευχαριστημένος από τα προϊόντα της και αν όχι γιατί;	116
7.2.2	Ερώτηση 4 <sup>η</sup> : Προτάσεις βελτίωσης προϊόντων της	116
7.2.3	Ερώτηση 10 <sup>η</sup> : Για ποιους λόγους προτιμάτε τα ανταγωνιστικά προϊόντα;	118

7.2.4	Ερώτηση 5 <sup>η</sup> : Για ποιους λόγους αγοράζουν τα προϊόντα της	119
8.	Συμπεράσματα	124
9.	Βιβλιογραφία	126

## ΠΡΟΛΟΓΟΣ

Η εργασία αυτή πραγματοποιήθηκε στα πλαίσια ικανοποίησης μέρους των απαιτήσεων για την απόκτηση του διπλώματος μηχανικού παραγωγής και διοίκησης στο Πολυτεχνείο Κρήτης. Η εργασία άρχισε να εκπονείται το 1999 στο εργαστήριο ανάλυσης δεδομένων και μάρκετινγκ, υπό την επίβλεψη των διδασκόντων του τμήματος ΜΠΔ κκ Ι. Δημοτίκαλη και Γ. Δούνια. Επαγγελματικές και στρατιωτικές μου υποχρεώσεις καθυστέρησαν την άμεση ολοκλήρωση της εργασίας. Επιπλέον, κατά την πορεία εκπόνησης οι ανωτέρω επιβλέποντες διδάσκοντες διέκοψαν τη συνεργασία τους με το τμήμα, ωστόσο συνέχισαν το έργο της επίβλεψης της υπάρχουσας μελέτης. Έτσι, με τη σημαντική στήριξη του Αναπληρωτή Καθηγητή ΜΠΔ κ Β. Μουστάκη (αλλά και του Αναπληρωτή Καθηγητή κ. Χ. Σκιαδά παλαιότερα) η εργασία πρόσφατα ολοκληρώθηκε επιτυχώς. Θέλω, λοιπόν, να εκφράσω τις ευχαριστίες μου τόσο στους παραπάνω διδάσκοντες όσο και σε όλους τους γνωστούς και φίλους που με βοήθησαν στην διεκπεραίωση της εργασίας. Ιδιαίτερα δε ευχαριστώ το συνάδελφο και φίλο μου ΜΠΔ, Νίκο Θωμαΐδη.

## 1. ΕΙΣΑΓΩΓΗ

Η παρούσα εργασία επιχειρεί να παρουσιάσει τις δυνατότητες εφαρμογής τόσο συμβατικών όσο και νοημόνων μεθοδολογιών στον χώρο του μάρκετινγκ, με ιδιαίτερη έμφαση στην ανάλυση ερωτηματολογίων. Μερικές ενδεικτικές εφαρμογές του πραγματικού κόσμου οι οποίες σχετίζονται με το χώρο του μάρκετινγκ είναι οι ακόλουθες<sup>1</sup>:

- Η κατασκευή προφίλ πελατών (Customer Profiling).
- Το στοχευόμενο μάρκετινγκ (targeted marketing).
- Η ανάλυση του καλαθιού αγορών (market basket analysis).

Η εργασία ξεκινά με μία θεωρητική παρουσίαση κάποιων βασικών αρχών που διέπουν τις νοήμονες μεθόδους ανάλυσης δεδομένων. Γίνεται συνοπτική αναφορά σε μεθόδους απόκτησης γνώσης, σε τεχνικές εξόρυξης δεδομένων και σε αλγοριθμικές προσεγγίσεις αποκάλυψης γνώσης. Στη συνέχεια, η μελέτη εξειδικεύεται στην παρουσίαση συγκεκριμένων αλγοριθμικών σχημάτων ευφυούς ανάλυσης δεδομένων μέσα από κάποια ενδεικτικά παραδείγματα από το χώρο του μάρκετινγκ. Τέλος, θεωρούμε δύο πραγματικές περιπτώσεις ανάλυσης ερωτηματολογίων σχετικά με:

α) τη διερεύνηση δημιουργίας βιοτεχνικού πάρκου στα Χανιά της Κρήτης και

β) τη μελέτη της ACAP για την έρευνα εταιρειών του κλάδου ελληνικού χαρτονιού.

Στην μεν πρώτη περίπτωση γίνεται αναλυτική παράθεση και σχολιασμός των κανόνων που προκύπτουν από την εφαρμογή αλγορίθμου επαγωγικής μηχανικής

---

<sup>1</sup> Σύμφωνα με έρευνα του Two Corps Corp.



μάθησης, ενώ στη δεύτερη περίπτωση εφαρμόζεται και σχολιάζεται η κλασική μεθοδολογία της Q-ανάλυσης. Έτσι, ο αναγνώστης έχει την ευκαιρία να δει και να συγκρίνει τις δυνατότητες τόσο κλασικών όσο και σύγχρονων μεθοδολογιών ανάλυσης ερωτηματολογίων.

## 2. Η ΕΝΝΟΙΑ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

### 2.1. Τεχνητή νοημοσύνη

Η επιστήμη της τεχνητής νοημοσύνης (Artificial Intelligence) εμφανίστηκε στην δεκαετία του '70 και αναπτύχθηκε παράλληλα με τους σύγχρονους ηλεκτρονικούς υπολογιστές από επιστήμονες ποικίλων ειδικοτήτων όπως μηχανικούς, μαθηματικούς, ψυχολόγους. Ασχολείται με μεθόδους συμβολικής εξαγωγής συμπερασμάτων μέσω του υπολογιστή. Σκοπός είναι η συμβολική αναπαράσταση της γνώσης, ώστε αυτή να χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων [Degewij, 1999]. Η τεχνητή νοημοσύνη διερευνά τις πιθανότητες που έχει ένας υπολογιστής ώστε μέσω της εκπαίδευσης να συμπεριφέρεται με τρόπο τέτοιο που θα αναγνωρίζονταν ως έξυπνη συμπεριφορά μεταξύ ανθρώπων [Edelstein, 1996]. Με άλλα λόγια ο εν λόγω υπολογιστής μοιάζει να συμπεριφέρεται ως νοήμων. Ο ορισμός αυτός ταυτίζεται με το αποκαλούμενο πρόβλημα του Turing, ο οποίος προσπαθώντας να δώσει τον ορισμό της έξυπνης μηχανής, είπε ότι αν βρίσκεσαι μπροστά από 2 οθόνες στις οποίες προβάλλονται τα αποτελέσματα ανθρώπινης λειτουργίας και τα αποτελέσματα λειτουργίας μιας μηχανής αντίστοιχα και δεν μπορείς να εντοπίσεις ποια οθόνη είναι η καθεμία, τότε η μηχανή είναι έξυπνη (Kodratoff, 1999). Έτσι λοιπόν για να ορίσουμε την τεχνητή νοημοσύνη, προσπαθούμε να ταυτίσουμε τα αποτελέσματα μηχανικών διαδικασιών και ανθρώπινης διαδικασίας σκέψης και εξαγωγής συμπερασμάτων. Βέβαια ανθρώπινη και μηχανική γνώση είναι δυο τελείως διαφορετικές έννοιες στην πραγματικότητα.

Οι κυριότερες διαφορές είναι:

#### Ανθρώπινη γνώση

Φθίνει με τον χρόνο  
Δύσκολο να μεταδοθεί  
Ευαίσθητη σε άγχος και συναισθήματα  
Δημιουργική  
Έχει αίσθηση της κοινής λογικής

#### Μηχανική γνώση

Παραμένει άφθαρτη  
Εύκολο να κλωνοποιηθεί  
Σταθερή, ανεπηρέαστη απόδοση  
Περιορισμένης δημιουργικότητας  
Χρειάζεται πλήρη εντολές

Κάθε βασικό γνώρισμα που είναι πλεονέκτημα για τον ένα τύπο γνώσης είναι μειονέκτημα για τον άλλο και αντίστροφα, ανάλογα με την εφαρμογή. Για παράδειγμα στις αυτοματοποιημένες εφαρμογές αποτελεί πλεονέκτημα που η μηχανική γνώση απαιτεί πλήρεις και λεπτομερείς εντολές σε αντίθεση με τον άνθρωπο που επηρεάζεται από την κοινή λογική, αλλά σε απρόβλεπτες ή έκτακτες καταστάσεις που απαιτείται η λήψη έκτακτων μέτρων αποτελεί μειονέκτημα. Η επιστήμη της τεχνητής νοημοσύνης εξελίχθηκε ραγδαία την τελευταία δεκαετία κυρίως λόγω της τεχνολογικής επανάστασης, αφού δημιουργήθηκε η δυνατότητα ανάπτυξης αποτελεσματικών αλγορίθμων που μπορούν να επεξεργαστούν μεγάλο όγκο δεδομένων ώστε να αποκτηθεί χρήσιμη και έγκυρη πληροφορία. Η επέκταση των υπολογιστών σε όλο το εύρος της βιομηχανίας και των επιστημών, σε συνδυασμό με την δυνατότητα αποθήκευσης τεράστιου όγκου πληροφοριών σε εξαιρετικά μικρές αποθηκευτικές μονάδες παρήγαγε πληθώρα δεδομένων. Επίσης η ανάπτυξη του Internet (με πάνω από 20.000.000 χρήστες το 1996) ευνόησε την ανταλλαγή γνώσεων και απόψεων μεταξύ των επιστημόνων. Έτσι δημιουργήθηκε πρόσφορο πεδίο για πειραματισμό και εξαγωγή συμπερασμάτων και επιταχύνθηκε η εξέλιξη του χώρου της τεχνητής νοημοσύνης.

Πρόσφατα ο Dr. Neal Gershenfeld, ερευνητής του Massachusetts Institute of Technology (M.I.T.) ανέπτυξε το λειτουργικό πρότυπο κβαντικού υπολογιστή σε πειραματικό επίπεδο, κάτι που αποτελεί μεγάλο τεχνολογικό άλμα και κάνει τους ταχύτερους ηλεκτρονικούς υπολογιστές του σήμερα να μοιάζουν με άβυσσους. Μέχρι σήμερα οι υπολογιστές λειτουργούσαν με δεδομένα υπό την μορφή μπιτ, δηλαδή δυαδικών ψηφίων, 0 και 1, και ο υπολογισμός έγκειται στην αποθήκευση και τον ανασυνδυασμό των μπιτ με τον ταχύτερο δυνατό ρυθμό. Με τα νέα δεδομένα ο κβαντικός υπολογιστής εκμεταλλεύεται τους παράδοξους νόμους του υποατομικού μικρόκοσμου, που επιτρέπουν τη δημιουργία ενός νέου τύπου μπιτ, του κιούμπιτ που δεν αντιστοιχεί εξ' ολοκλήρου στην κατάσταση 0 ή 1, αλλά αποτελεί ένα συνδυασμό και των δυο. Αυτό σημαίνει ότι ένα κιούμπιτ μπορεί να λειτουργεί ως δυο διαφορετικά μπιτ, διπλασιάζοντας την ταχύτητα

υπολογισμού, και μάλιστα όσο περισσότερα κιούμπιτ χρησιμοποιεί ένας κβαντικός υπολογιστής, τόσο πιο σύνθετα μαθηματικά προβλήματα μπορεί να επιλύει. Πενήντα μόλις κιούμπιτ έχουν την δυνατότητα ενός τετράκις εκατομμυρίου μπιτ, δυνατότητα που ξεπερνά την αντίστοιχη οποιουδήποτε συμβατικού σημερινού υπολογιστή. Τα κιούμπιτ μπορούν να δημιουργηθούν με τη χρησιμοποίηση υποατομικών σωματιδίων, όπως τα πρωτόνια, τα οποία χάρη στο σπίν τους (στροφορμή λόγω ιδιοπεριστροφής), επιτρέπουν τον συνδυασμό καταστάσεων 0 και 1 την ίδια χρονική στιγμή. Το μειονέκτημα όμως είναι πως τα κιούμπιτ αποδεικνύονται ιδιαίτερα ευαίσθητα αφού με την παραμικρή διαταραχή μπορούν να χάσουν το «μαγικό» τους συνδυασμό καταστάσεων και να συμπεριφερθούν σαν κανονικά μπιτ. Σήμερα ο Gershenfeld μαζί με τον συνεργάτη του Isaac Chuang της IBM, στο San Jose, έχουν βρει τον τρόπο να παρακάμψουν το παραπάνω πρόβλημα, με την χρήση όχι μεμονωμένων σωματιδίων ως κιούμπιτ, αλλά ενός τεράστιου αριθμού από αυτά. Τα κιούμπιτ παίρνουν την μορφή αναριθμητών πρωτονίων σε ένα μικρό σωλήνα με νερό, που περιβάλλεται από μαγνήτες οι οποίοι χρησιμοποιούνται για να διαβάσουν τη λύση που δίνουν τα κιούμπιτ. Κατά τις μετρήσεις πολλά κιούμπιτ εξακολουθούν να χάνουν τις «μαγικές» τους ικανότητες, αλλά πάντα παραμένει αδιατάρακτος ένας επαρκής αριθμός που επιτρέπει στη συσκευή να συνεχίσει τους υπολογισμούς της. Ο Gershenfeld θεωρεί πολύ πιθανό ότι ο απαραίτητος επιστημονικός εξοπλισμός θα γίνει πιο προσιτός στο ευρύ κοινό. Το εντυπωσιακό αυτό επίτευγμα δίνει την δυνατότητα στους επιστήμονες που ασχολούνται με την τεχνητή νοημοσύνη να δοκιμάσουν νέες ιδέες και τεχνικές.

Η τεχνητή νοημοσύνη, όπως προαναφέρθηκε, συχνά παίρνει την αρχική ιδέα από την παρατήρηση της ανθρώπινης συμπεριφοράς. Η επιστημονική ανακάλυψη μέσω του ηλεκτρονικού υπολογιστή δεν αποτελεί εξαίρεση, καθότι η αρχική έρευνα είχε επικεντρωθεί στην αναπαραγωγή ανακαλύψεων σε επιστημονικούς τομείς αρκετά διαφορετικούς μεταξύ τους όπως μαθηματικά (Lenat, 1997), φυσική (Langley, 1981), χημεία (Zytkow & Simon, 1986) και βιολογία (Kulkarni & Simon, 1990).

Όπως αποκαλύπτεται από την συλλογή των Shrager και Langley (1990), αυτές οι προσπάθειες είχαν αξιοσημείωτη ευρύτητα εφαρμογής στις επιστημονικές δραστηριότητες που προσπάθησαν να μοντελοποιήσουν, αν και η μεγαλύτερη προσπάθεια επικεντρώνονταν στην αναπαράγωγή των ιστορικά γνωστών καταγραφών σε στοιχειώδες επίπεδο, χωρίς την προσθήκη λεπτομερειών. Παρά τους πολύ συγκεκριμένους στόχους αυτής της αρχικής ερευνητικής προσπάθειας, ορισμένες κριτικές (Gilles 1996), αμφισβήτησαν έντονα την συμβολή της έρευνας στην πρόοδο διότι περιστρέφονταν γύρω από επιστημονικούς νόμους και ήδη γνωστές θεωρίες. Αν και οι προσπάθειες μερικών επιστημόνων συνεχίζουν να επικεντρώνονται σε ιστορικά δεδομένα, άλλοι συνάδελφοί τους έχουν στραφεί στην ανακάλυψη χρήσιμης γνώσης σε βάσεις δεδομένων.

## **2.2. Εξόρυξη γνώσης από βάσεις δεδομένων.**

### *2.2.1 Τι είναι η εξόρυξη γνώσης από βάσεις δεδομένων*

«Η εξόρυξη δεδομένων ή η ανακάλυψη γνώσης σε βάσεις δεδομένων (Knowledge Discovery in Databases), ορίζεται ως η απόκτηση πεπλεγμένης, πρόσφατα άγνωστης και πιθανά χρήσιμης πληροφορίας από δεδομένα. Χρησιμοποιεί τεχνικές μηχανικής μάθησης (machine learning), στατιστική και οπτικοακουστική για να ανακαλύψει και να παρουσιάσει γνώση σε μια μορφή που είναι εύκολα κατανοητή από τον άνθρωπο».

Απλούστερα, ως εξόρυξη γνώσης από βάσεις δεδομένων (Data mining) θα μπορούσαμε να ορίσουμε, την ανακάλυψη και ανάλυση, με περισσότερο ή λιγότερο αυτοματοποιημένες τεχνικές, χρησίμων κανόνων και σχέσεων μέσα από μεγάλους όγκους δεδομένων. Στην περίπτωση του μάρκετινγκ, στόχος και σκοπός της εξόρυξης γνώσης από βάσεις δεδομένων είναι να επιτρέψει στις επιχειρήσεις να βελτιώσουν τις λειτουργίες των πωλήσεων, της εξυπηρέτησης των πελατών και των λοιπών διαδικασιών του marketing μέσω της καλύτερης κατανόησης των πελατών τους. Στην πραγματικότητα, οι τεχνικές που χρησιμοποιεί η εξόρυξη γνώσης από βάσεις δεδομένων δεν είναι κάτι καινούριο, αφού έχουν εφαρμοστεί εδώ και πολλά χρόνια σε διάφορα επιστημονικά πεδία, όπως στην φυσική, την ιατρική, αλλά και τον έλεγχο της βιομηχανικής, παραγωγικής διαδικασίας. Γιατί, όμως, έπρεπε να φθάσουμε στη δεκαετία του '90 για να ανακαλύψουμε τη χρηστική ιδιότητα της εξόρυξης γνώσης από βάσεις δεδομένων στον τομέα του εφαρμοσμένου επιχειρηματικού μάρκετινγκ; Μερικοί από τους παράγοντες που δικαιολογούν αυτή την «αργοπορεία» παρουσιάζονται παρακάτω.

### *2.2.2 Πως παράγονται μεγάλοι όγκοι δεδομένων σήμερα*

Από τη στιγμή που οι μεγάλες επιχειρήσεις κατάλαβαν τη χρησιμότητα που είχαν οι έρευνες για την μελλοντική τους εξέλιξη, και δεν τις έβλεπαν σαν μια εκροή - σπατάλη χρηματικών πόρων, άρχισαν να συλλέγουν μανιωδώς πληροφορίες (δεκαετίες '60 - '70). Με τον τρόπο αυτό δημιουργήθηκε ένας

μεγάλος όγκος δεδομένων που συνεχώς αυξανόταν, αφού στοιχεία συλλέγονταν από όλα σχεδόν τα τμήματα οργάνωσης μιας επιχείρησης (παραδοσιακά από το λογιστήριο μέχρι τη διεύθυνση προσωπικού, παραγωγής, έρευνας, κλπ.). Έτσι, φτάσαμε στη δεκαετία του '90 όπου είχαν ωριμάσει οι πεποιθήσεις για την ανάγκη ύπαρξης συνεχούς και ολοκληρωμένης πληροφόρησης στις επιχειρήσεις.

### *2.2.3 Ανάγκη αποθήκευσης δεδομένων*

Η μεγάλη ροή πληροφοριών δημιούργησε την ανάγκη για την ταξινόμηση και αποθήκευση των δεδομένων που καθημερινά λαμβάνονταν. Έτσι δημιουργήθηκαν αποθήκες δεδομένων και συγκεκριμένα άτομα στην επιχείρηση ανέλαβαν τη διαχείριση τους, αναπτύσσοντας ιδιαίτερες τεχνικές. Αρχικά, αυτές οι αποθήκες δεδομένων ελάχιστα διέφεραν από τις παραδοσιακές αποθήκες προϊόντων, αφού οι πληροφορίες «στοιβάζονταν σε ράφια» με κάποια κριτήρια που σκοπό είχαν την εύκολη ανάκτηση τους. Σταδιακά οι τεχνικές αυτές εξελίχθηκαν, για να φτάσουμε σήμερα να μιλάμε για πλήρως αυτοματοποιημένες τράπεζες δεδομένων, που διοχετεύουν σε κάθε επίπεδο της ιεραρχίας και διαφορετικό είδος πληροφορίας για την υποστήριξη λειτουργιών και αποφάσεων (βλέπε σύγχρονα πληροφοριακά συστήματα διοίκησης).

### *2.2.4 Φθινή υπολογιστική ισχύς*

Οι αλγόριθμοι που χρησιμοποιούν οι τεχνικές της εξόρυξης γνώσης από βάσεις δεδομένων απαιτούν τις περισσότερες φορές μεγάλη υπολογιστική ισχύ. Σήμερα η ανάπτυξη της τεχνολογίας των υπολογιστών μας παρέχει αυτή τη δυνατότητα με αρκετά χαμηλό κόστος. Σχεδόν οι περισσότερες σύγχρονες εφαρμογές αυτού του είδους μπορούν να «λειτουργήσουν» σε απλά PCs της τάξεως των 1000\$ (ποσό μάλλον μικρό για αγορά εξοπλισμού σε μία επιχείρηση).

### *2.2.5 Ισχυρότατος ανταγωνισμός των επιχειρήσεων*

Αυτός ο παράγοντας δεν χρήζει ερμηνείας, αφού μπορούμε εύκολα να αντιληφθούμε την κατάσταση που επικρατεί σήμερα σε ένα διεθνοποιημένο -

παγκοσμιοποιημένο επιχειρηματικό περιβάλλον. Έχουμε φτάσει σε σημείο να μιλάμε για εμπορικό πόλεμο. Και τώρα που τα υπόλοιπα όπλα έχουν ατονήσει η δύναμη των πληροφοριών αποτελεί την αιχμή του δόρατος. Η εποχή των πληροφοριών δεν επιτρέπει σε κανέναν να την αγνοήσει. Στο παιχνίδι αυτό κυρίαρχη θέση έχουν οι εταιρίες παροχής υπηρεσιών. Παράδειγμα οι ασφαλιστικές εταιρίες που διαχειρίζονται καθημερινά τεράστιους όγκους πληροφοριών. Η εξόρυξη γνώσης από βάσεις δεδομένων σε μια τέτοια περίπτωση μπορεί να προσθέσει ένα συγκριτικό πλεονέκτημα σε όποιον το εφαρμόζει. Παράλληλα, δεν είναι λίγες οι εταιρίες που το προϊόν τους είναι η πληροφορία αυτή καθαυτή. Όπως οι συμβουλευτικές, ερευνών, κλπ, και έχουν ανάγκη για σύγχρονες και αξιόπιστες μεθόδους αντήλθεως.

### **2.3. Χρησιμότητα της εξόρυξης γνώσης από βάσεις δεδομένων**

Μια μικρή επιχείρηση θεμελιώνει τις σχέσεις της με τον πελάτη παρατηρώντας τις ανάγκες του, γνωρίζοντας τις προτιμήσεις και τις συνήθειές του. Μαθαίνοντας από τις προηγούμενες συναλλαγές του πετυχαίνει να τον ικανοποιήσει καλύτερα στο μέλλον. Οι συνοικιακές επιχειρήσεις όπως το παντοπωλείο «ο καλός επιχειρηματίας», γνωρίζουν προσωπικά τους πελάτες τους και διαμορφώνουν μια ιδιαίτερη συμπεριφορά για τον καθένα. Αυτή η αμεσότητα στις σχέσεις επιχείρησης και πελάτη που παλαιότερα ήταν αριστά εύκολη, παρουσιάζεται ελλιπής στις μέρες μας. Καμία σύγχρονη επιχείρηση δεν μπορεί να γνωρίζει τις ιδιαιτερότητες του κάθε πελάτη της (σκεφτείτε κάτι τέτοιο σε μια αλυσίδα σουπερ μάρκετ). Έτσι, οι σύγχρονες επιχειρήσεις προσπαθούν με διαφορετικούς τρόπους να πετύχουν ένα καλύτερο επίπεδο εξυπηρέτησης. Το στόχο αυτό καλούνται να ενισχύσουν οι τεχνικές της εξόρυξης γνώσης από βάσεις δεδομένων, ώστε να αποκαλύψουν το προφίλ του δυνητικού πελάτη της κάθε επιχείρησης και άλλες πληροφορίες.



Οι αποθήκες δεδομένων εφοδιάζουν την επιχείρηση με μνήμη, ενώ παράλληλα η εξόρυξη γνώσης από βάσεις δεδομένων οπλίζει την επιχείρηση με έξυπνη γνώση.

Η νοημοσύνη αυτή μας δίνει την δυνατότητα να αναζητήσουμε στις αποθήκες γνώσης συνθήκες και νόμους που θα μας επιτρέψουν να κάνουμε προβλέψεις για νέες ιδέες.

Ιστορικά, η βιομηχανική επανάσταση ξεκίνησε με τη χρήση του νερού ως κινητήρια δύναμη στα υφαντουργεία της Αγγλίας. Αυτά τα πρώτα εργοστάσια - εργαστήρια είχαν τις ρίζες τους στους νερόμυλους των παραποτάμιων περιοχών. Κάθε μορφή οικονομική και παραγωγική επανάσταση έχει την ανάγκη μίας τουλάχιστον καινοτομίας που θα διαδραματίσει το ρόλο του κινητήριου μοχλού. Σήμερα τα δεδομένα και κατ' επέκταση η πληροφόρηση είναι η νέα κινητήρια δύναμη της οικονομίας. Ο κύκλος της εξόρυξης γνώσης από βάσεις δεδομένων έχει ως σκοπό να «τιθαρεύσει» αυτή τη δύναμη, μετατρέποντας την σε πραγματικά επιχειρηματικά αποτελέσματα. Όπως το νερό κάποτε κινούσε τους νερόμυλους που με τη σειρά τους έδιναν κίνηση στις μηχανές, έτσι και τα δεδομένα πρέπει να συγκεντρωθούν και να διοχετευθούν με τέτοιο τρόπο ώστε να παράγουν αξία. Στον χώρο των επιχειρήσεων, η εξόρυξη γνώσης από βάσεις δεδομένων παρέχει μια ριζικά νέα προοπτική, τη δυνατότητα να αριστοποιεί τη διαδικασία λήψης αποφάσεων χρησιμοποιώντας αυτοματοποιημένες μεθόδους και μαθαίνοντας από τις πράξεις του παρελθόντος. Τα τελευταία χρόνια η τεχνολογία κατάφερε να μας χαρίσει αυτή τη δυνατότητα και στο μέλλον ευελπιστούμε ότι περισσότερα εξειδικευμένα εργαλεία (software και hardware) θα μας δώσουν καλύτερα αποτελέσματα. Έτσι, ένα πλήθος επιχειρήσεων χρησιμοποιεί την εξόρυξη γνώσης από βάσεις δεδομένων στις αποφάσεις τους κάνοντας φανερό το βαθμό σημαντικότητας τέτοιων τεχνικών. Όλα τα παραπάνω μας δείχνουν την ένταση που έχει πάρει η ιδέα της εξόρυξης γνώσης από βάσεις δεδομένων στο χώρο της επιχειρηματικής σκέψης, αφού κρίνεται ως μια αρκετά ελπιδοφόρα και συνάμα φιλόδοξη προοπτική για την

αποτελεσματικότερη και ορθολογικότερη διαχείριση των πληροφοριών (που οδηγούν στην τελική λήψη αποφάσεων).

Τα δεδομένα αποτελούν το πυρήνα του επιχειρηματικού σχεδιασμού των περισσότερων οικονομικών μονάδων. Η όλη διαδικασία του σχεδιασμού βασίζεται σε μια συνεχή ροή πληροφοριών που με τον ενδεδειγμένο τρόπο - τεχνική μετασχηματίζονται σε ποιοτικές ή ποσοτικές παράμετροι αποφάσεων. Από τα παραπάνω καταλαβαίνουμε ότι η αποτελεσματικότητα της εξόρυξη γνώσης από βάσεις δεδομένων βασίζεται σε μια κυκλική διαδικασία που αναλαμβάνει να μετατρέψει τα δεδομένα σε πληροφορία, τη πληροφορία σε ενέργεια και την ενέργεια σε αξία. Τα στάδια που περιλαμβάνει ο κύκλος λειτουργίας της εξόρυξη γνώσης από βάσεις δεδομένων είναι:



## **2.4. Μοντέλα εξόρυξης γνώσης από βάσεις δεδομένων**

Τα εργαλεία εξόρυξης γνώσης από βάσεις δεδομένων (data mining tools) κατασκευάζουν, με την χρησιμοποίηση δεδομένων και μέσω τεχνικών μηχανικής μάθησης, ένα μοντέλο πραγματικού κόσμου με αποτέλεσμα μια συλλογή γεγονότων και προτύπων.

Τα μοντέλα που προκύπτουν χρησιμοποιούνται με δυο διαφορετικούς τρόπους. Την περιγραφή γεγονότων και συσχετίσεων σε μια βάση δεδομένων με σκοπό την παροχή γνώσης που βοηθά σε επιχειρηματικές ενέργειες. Και τις συσχετίσεις για την πραγματοποίηση προβλέψεων όπως για παράδειγμα η πρόβλεψη ατόμων που ίσως ανταποκριθούν θετικά σε μια διαφημιστική εκστρατεία μέσω του ταχυδρομείου. Άρα στέλνονται επιστολές μόνο στα άτομα τα οποία προβλέφτηκε πως είναι πιθανόν να ανταποκριθούν, με κυριότερο πλεονέκτημα την αξιοσημείωτη μείωση του κόστους.

Τα μοντέλα που κατασκευάζονται από την εξόρυξη δεδομένων και βοηθούν στην επίλυση επιχειρηματικών προβλημάτων είναι τα εξής:

- Κατηγοριοποίησης (classification)
- Παλινδρόμησης (regression)
- Χρονοσειρών (time series)
- Τμηματοποίησης (clustering)
- Ποσοτικής ανάλυσης (quantitative analysis ή Q-Analysis)
- Ανάλυσης συσχετίσεων (association analysis)
- Ανακάλυψης ακολουθιών στοιχείων (sequence discovery)

Οι δυο πρώτοι τύποι χρησιμοποιούνται κυρίως για πρόβλεψη ενώ οι υπόλοιποι για να αποκαλύψουν συμπεριφορά δεδομένων. Ακολουθώς παρατίθενται σύντομες περιγραφές χρήσης των μεθόδων.

#### *2.4.1 Κατηγοριοποίηση*

Η κατηγοριοποίηση ομαδοποιεί παραδείγματα και περιπτώσεις σε κλάσεις υπολογίζοντας την τιμή μιας μεταβλητής κατηγοριοποίησης. Τα δεδομένα που προκύπτουν χρησιμοποιούνται έτσι ώστε να καθοριστεί επαγωγικά το πρότυπο που θα αποκαλύψει την κατηγορία στην οποία ανήκει η κάθε περίπτωση. Το πρότυπο αυτό είναι χρήσιμο τόσο για κατανόηση δεδομένων όσο και για την κατηγοριοποίηση νέων περιπτώσεων.

#### *2.4.2 Παλινδρόμηση*

Η παλινδρόμηση προβλέπει την τιμή συνεχόμενων μεταβλητών χρησιμοποιώντας μια σειρά υπαρχόντων τιμών και των χαρακτηριστικών τους, ως συνεχόμενες ορίζονται οι μεταβλητές που μπορούν να πάρουν οποιαδήποτε τιμή από ένα σύνολο πραγματικών αριθμών.

#### *2.4.3 Χρονοσειρές*

Η πρόβλεψη με χρονοσειρές προβλέπει μελλοντικές τιμές χρησιμοποιώντας σειρές υπαρχόντων τιμών, όπως και η παλινδρόμηση. Η διαφορά έγκειται στο ότι στην πρόβλεψη με χρονοσειρές οι μεταβλητές εξαρτώνται από τον χρόνο.

#### *2.4.4 Τμηματοποίηση*

Η τμηματοποίηση χωρίζει τη βάση δεδομένων σε διαφορετικές ομάδες. Κύριος στόχος είναι να βρεθούν τόσο ομάδες με ομοιότητες όσο και ομάδες με διαφορές. Αντίθετα με την κατηγοριοποίηση, οι ομάδες δεν είναι γνωστές από την αρχή. Επίσης είναι άγνωστο με βάση ποια χαρακτηριστικά (attributes) θα γίνει η κατηγοριοποίηση.

#### *2.4.5 Ποσοτική ανάλυση*

Η ποσοτική ανάλυση είναι ιδανική για την επεξεργασία ερωτηματολογίων λόγω της δυαδικής φύσης των δεδομένων τους. Για να μπορέσει όμως να δουλέψει πρέπει η κατασκευή του ερωτηματολογίου να είναι τέτοια ώστε να δίνει τη σωστή διάσταση του προβλήματος. Βάζοντας στα στοιχεία του προβλήματος τα κατάλληλα σε είδος αλλά και σε αριθμό χαρακτηριστικά. Τα χαρακτηριστικά πρέπει να είναι διαφορετικά μεταξύ τους και να μην περιέχει το ένα το άλλο γιατί έτσι θα δημιουργηθούν παραπλανητικές ομαδοποιήσεις.

#### *2.4.6 Ανάλυση συσχετίσεων*

Η ανάλυση συσχετίσεων είναι πράξεις ή αντικείμενα που συμβαίνουν αλληλένδετα για ένα συγκεκριμένο γεγονός ή καταγραφή. Τα εργαλεία των συσχετίσεων, όπως και των συνεχειών, ανακαλύπτουν κανόνες της μορφής: «ΑΝ η πράξη Α είναι μέρος ενός γεγονότος, ΤΟΤΕ x% των φόρων (επίπεδο σιγουριάς), η πράξη Β είναι επίσης μέρος του ίδιου γεγονότος». Για παράδειγμα «ΑΝ οι καταναλωτές αγοράζουν τυρί με χαμηλά λιπαρά και άπαχο γιαούρτι, ΤΟΤΕ 85% των φορών αγοράζουν κρέμα γάλακτος».

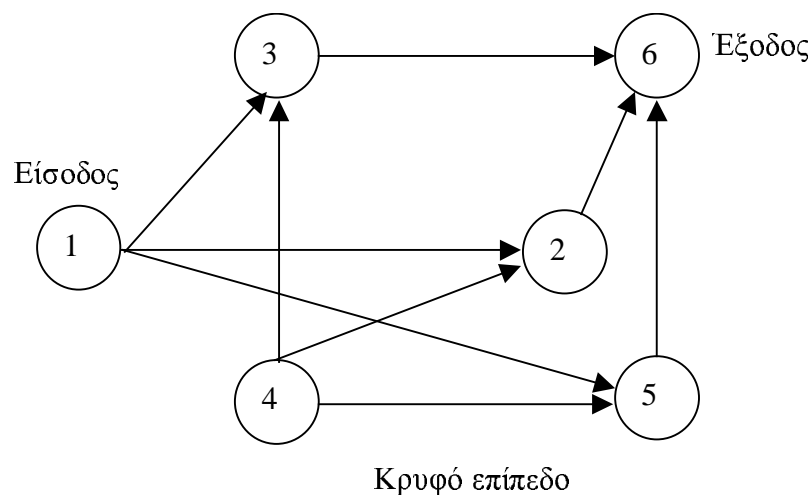
#### *2.4.7 Ανακάλυψη ακολουθιών στοιχείων*

Η ανακάλυψη συνεχειών ή ακολουθιών στοιχείων σχετίζεται στενά με την ανάλυση συσχετίσεων, με βασική διαφορά ότι τα σχετιζόμενα αντικείμενα ή πράξεις είναι διασκορπισμένα στο χρόνο. Για να βρεθούν συνέχειες τέτοιου τύπου, πρέπει να κατανοηθούν οι λεπτομέρειες της κάθε συσχέτισης όπως και η ταυτότητα των συσχετιζόμενων. Για παράδειγμα «Αν μια συγκεκριμένη ημέρα η μετοχή Α έχει άνοδο μεγαλύτερη από 10% και ο γενικός δείκτης του χρηματιστηρίου έχει πτώση, ΤΟΤΕ η μετοχή Β θα παρουσιάσει σε χρονικό διάστημα 2 ημερών 68% των φορών».

## 2.5. Κατασκευή μοντέλων κατηγοριοποίησης με τεχνικές μηχανικής μάθησης.

Μία γενική παρατήρηση είναι ότι δεν υπάρχει εργαλείο που να μπορούμε να εφαρμόσουμε σε όλες τις μορφές προβλημάτων, αλλά ανάλογα με το πρόβλημα και την φύση των δεδομένων επιλέγουμε το κατάλληλο. Είναι λοιπόν λογικό για να βρούμε το καλύτερο δυνατό μοντέλο να εξετάζουμε μια μεγάλη ποικιλία διαφορετικών εργαλείων, με πιο συνηθισμένα τα μοντέλα κατηγοριοποίησης. Δυο είναι οι βασικές στατιστικές τεχνικές κατασκευής μοντέλων. Η λογιστική παλινδρόμηση (logistic regression), που αποτελεί γενική μορφή της γραμμικής παλινδρόμησης, και η διακριτή ανάλυση (discriminant analysis).

Παρόλα αυτά και ενώ η εξόρυξη δεδομένων κερδίζει συνεχώς έδαφος, τα νευρωνικά δίκτυα και τα δέντρα αποφάσεων βρίσκονται στο προσκήνιο, λόγω της μικρότερης στατιστικής πολυπλοκότητας που παρουσιάζουν. Τα νευρωνικά δίκτυα κατασκευάζουν μοντέλα που για ένα σύνολο δεδομένων εισόδου προβλέπουν την τιμή μιας συνεχούς ή κατηγοριοποιημένης παραμέτρου, χρησιμοποιώντας ως παραμέτρους τους κόμβους του κρυφού επιπέδου.



Εικόνα 2-1. Ένα απλό νευρωνικό δίκτυο.

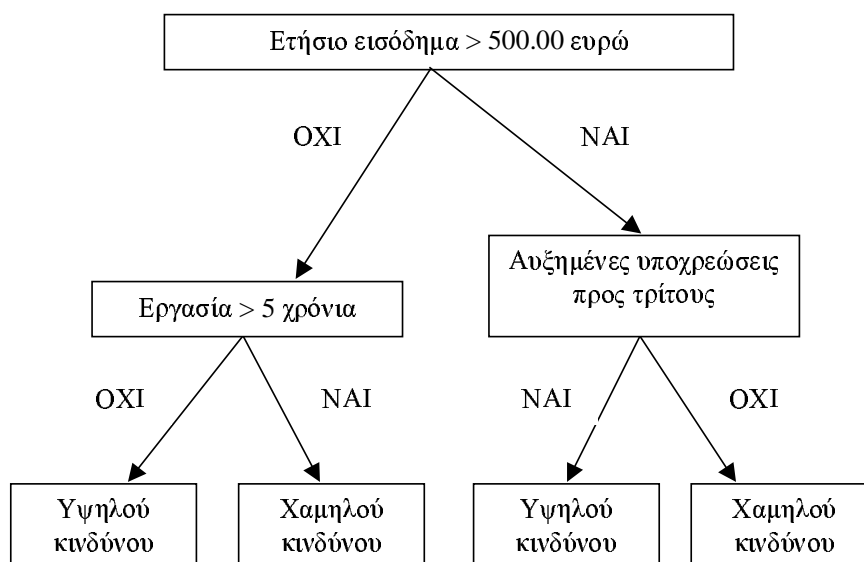
Η τιμή κάθε κρυφού κόμβου είναι συνάρτηση του διαβαθμισμένου αθροίσματος των τιμών όλων των προηγούμενων κόμβων που καταλήγουν στον κρυφό κόμβο. Για να κατασκευαστεί ένα μοντέλο απαιτείται η «εκπαίδευση» (training) του νευρωνικού δικτύου με δεδομένα. Η ανάδραση (back - propagation) είναι η πιο συνηθισμένη μέθοδος εκπαίδευσης στην οποία το αποτέλεσμα που προκύπτει συγκρίνεται με γνωστές ισχύουσες τιμές. Το νευρωνικό δίκτυο αποτελεί ένα πολύ καλό μοντέλο πρόβλεψης, ιδιαίτερα μετά από πολλές επαναλήψεις και χρησιμοποιώντας τα δεδομένα εκπαίδευσης (training data). Όμως τα νευρωνικά δίκτυα έχουν δυο βασικά μειονεκτήματα. Πρώτον όταν οι παράγοντες που οδηγούν στην πρόβλεψη δεν είναι προφανείς τότε η ποιότητα του αποτελέσματος μπορεί να είναι «μειωμένη», και δεύτερον τα νευρωνικά δίκτυα είναι ευάλωτα στην υπερπροσαρμογή (over fitting). Ο τελευταίος όρος έχει το νόημα της παραγωγής πολύ ακριβών μοντέλων πρόβλεψης για συγκεκριμένα δεδομένα δοκιμής (test data), των οποίων η προβλεπτική ικανότητα μειώνεται όταν χρησιμοποιούν νέα δεδομένα.

Τα νευρωνικά δίκτυα προτιμούνται όταν (Mitchell, 1997):

- Τα δεδομένα εισόδου είναι πολυδιάστατα, διακριτά ή πραγματικά. Η έξοδος είναι διακριτή ή πραγματική.
- Η έξοδος είναι διάνυσμα τιμών.
- Τα δεδομένα είναι πιθανό να περιέχουν θόρυβο.
- Ο τρόπος που επιτυγχάνουμε το στόχο είναι άγνωστος.
- Η εξήγηση του αποτελέσματος δεν έχει ιδιαίτερη σημασία.

Τυπικά παραδείγματα εφαρμογής νευρωνικών δικτύων είναι η φωνητική αναγνώριση, η αναγνώριση εικόνας και η χρηματοοικονομική πρόβλεψη (Pίζος, 1996).

Τα δέντρα απόφασης αποτελούνται από μια σειρά κανόνων που οδηγούν σε μια τάξη, κατηγορία ή τιμή. Ένα χαρακτηριστικό παράδειγμα είναι αυτό ενός χρηματοπιστωτικού ιδρύματος το οποίο ενδιαφέρεται να κατατάξει τους αιτούντες για δάνειο σε ομάδες χαμηλού ή υψηλού κινδύνου. Στο ακόλουθο σχήμα φαίνεται ένα δέντρο απόφασης που επιλύει αυτό το πρόβλημα. Είναι σαφές πως για παράδειγμα ένας αιτούμενος με «ετήσιο εισόδημα  $> 500.000$  ευρώ» και «αυξημένες υποχρεώσεις προς τρίτους», θα χαρακτηριζόταν ως υψηλού κινδύνου.



Εικόνα 2-2. Ένα απλό δέντρο απόφασης.

Τα δέντρα απόφασης παράγουν μεγάλης ακρίβειας μοντέλα, και επιτρέπουν την πλήρη κατανόηση του αποτελέσματος πράγμα που τα καθιστά ιδιαίτερα δημοφιλή. Έτσι υπερέχουν έναντι των νευρωνικών δικτύων διότι απαιτούν λιγότερο χρόνο να κατασκευαστούν. Αξίζει να σημειωθεί πως βασικό μειονέκτημα είναι η αδυναμία εύρεσης κανόνων που βασίζονται σε συνδυασμό μεταβλητών και το γεγονός ότι ο διαχωρισμός των κλαδιών του δέντρου δεν είναι ομαλός, δηλαδή η επιλογή του διαχωρισμού δεν λαμβάνει υπόψη τους



διαχωρισμούς των επόμενων επιπέδων. Το πρόβλημα της υπερπροσαρμογής, αν και υπάρχει και σε αυτή την περίπτωση, είναι ασθενέστερο.

Τα δέντρα απόφασης προτιμούνται στις εξής περιπτώσεις (Mitchell, 1997):

- ο Τα παραδείγματα περιγράφονται από ζεύγη (χαρακτηριστικό, τιμή).
- ο Ο στόχος είναι διακριτός.
- ο Απαιτείται διαζευκτική (disjunctive) υπόθεση.
- ο Τα δεδομένα εκπαίδευσης είναι πιθανόν να περιέχουν πολύ θόρυβο

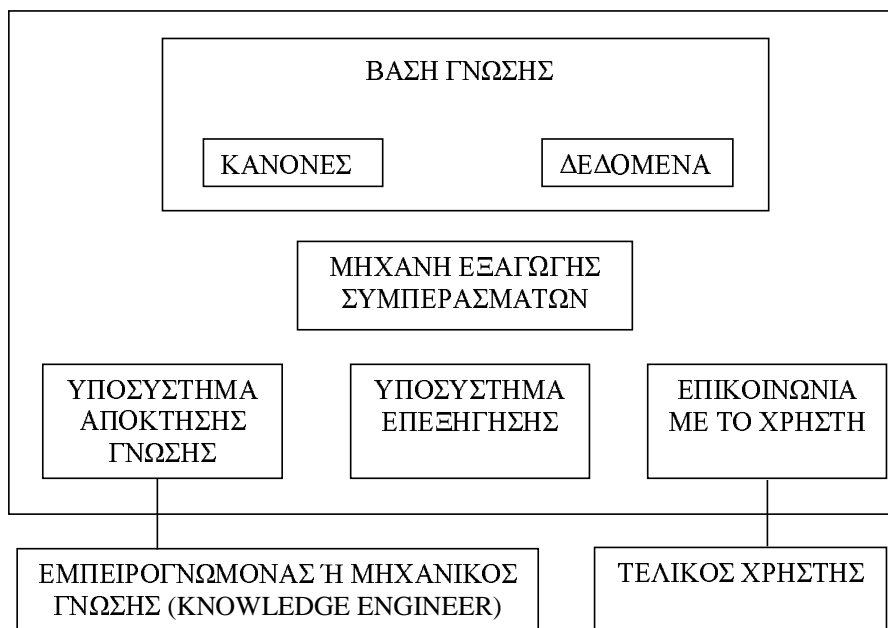
Τυπικά παραδείγματα εφαρμογής δέντρων απόφασης είναι η ιατρική διάγνωση, η ανάλυση χρηματοπιστωτικού κινδύνου, και η μοντελοποίηση προτιμήσεων ημερολογιακού προγραμματισμού. Πρέπει να σημειωθεί πως τα νευρωνικά δίκτυα και τα δέντρα αποφάσεων είναι δυνατόν να χρησιμοποιηθούν για την κατασκευή μοντέλων παλινδρόμησης, και κάποιοι τύποι νευρωνικών δικτύων χρησιμοποιούνται για να κατασκευαστούν μοντέλα ομαδοποίησης.

## 2.6. Αρχιτεκτονική και λειτουργία μοντέλων εξόρυξης δεδομένων.

Δυο είναι τα τμήματα που αποτελούν την βάση των μοντέλων εξόρυξης:

1. Η βάση γνώσης (knowledge base), που αποτελείται από δεδομένα και κανόνες.
2. Η μηχανή εξαγωγής συμπερασμάτων (inference engine) ή επαναληπτικός αλγόριθμος, ο οποίος ψάχνει στη βάση κάνοντας όλους τους πιθανούς συνδυασμούς έτσι ώστε έπειτα από ένα αριθμό επαναλήψεων να φτάσει στη λύση, η οποία ίσως να μην είναι η ιδανική (optimum) αλλά είναι η καλύτερη δυνατή (optimal).

Στο ακόλουθο σχήμα φαίνεται η γενική αρχιτεκτονική μοντέλων εξόρυξης δεδομένων.



Εικόνα 2-1 Γενική αρχιτεκτονική μοντέλων εξόρυξης δεδομένων.

Υπάρχουν πέντε τεχνικές εξερεύνησης μιας βάσης δεδομένων:

1. Πρώτα κατά μήκος (BFS Breadth First Search).
2. Πρώτα κατά βάθος (DFS Depth First Search).
3. Μεγαλύτερου βήματος.
4. Μικρότερου κόστους.
5. Πλήρης εκμετάλλευση των δυνατοτήτων της εκάστοτε γλώσσας προγραμματισμού.

Στο πρωτόκολλο FIFO, κάθε νέο δεδομένο που εισέρχεται γίνεται κομβικό σημείο και διερευνάται πλήρως σε σχέση με τα δεδομένα που αυτό συνδέεται. Η διαδικασία αυτή επαναλαμβάνεται για κάθε επόμενο σημείο. Αντίθετα κατά το πρωτόκολλο LIFO, κάθε νέο δεδομένο που εισάγεται δεν διερευνάται πλήρως, αλλά ο αλγόριθμος προχωράει στο επόμενο σημείο κ.ο.κ. μέχρι να διερευνηθεί το σημείο του τελευταίου επιπέδου, και ο αλγόριθμος να εκτελείται για κάθε ένα από τα προηγούμενα σημεία και τις ανεξερεύνητες συνδέσεις. Η μέθοδος του μεγαλύτερου βήματος επιλέγει το μεγαλύτερο βήμα μεταξύ των δεδομένων, και η μέθοδος του μικρότερου κόστους επιλέγει το μικρότερο κόστος αντίστοιχα. Τέλος, η πέμπτη μέθοδος δεν δίνει κάποια συγκεκριμένη κατεύθυνση στον αλγόριθμο αλλά ο υπολογιστής ψάχνει με τυχαίο τρόπο για την λύση, πράγμα που ίσως οδηγήσει σε πρακτικά άπειρες ανακυκλώσεις. Οι 5 παραπάνω τεχνικές μπορούν να εφαρμοστούν είτε προς τα εμπρός (forward chaining) αν η διερεύνηση εκκινεί από ένα κανόνα που είναι στόχος, είτε προς τα πίσω (backward chaining) αν η διερεύνηση εκκινεί από τα 'ΑΝ' τμήματα των ισχυόντων κανόνων.

## 2.7. Τεχνικές και μέθοδοι εξόρυξης δεδομένων

Στο σημείο αυτό συνοψίζονται όλες οι τεχνικές που χρησιμοποιούνται, με ιδιαίτερη έμφαση στα έργα ή καθήκοντα για τα οποία ενδείκνυνται.

### 2.7.1 *Ανάλυση καλαθιού αγοράς*

Η ανάλυση καλαθιού αγοράς (Market Basket Analysis) είναι ένα είδος ομαδοποίησης (clustering) που χρησιμοποιείται για την εύρεση ομάδων ή συνόλων στοιχείων που εμφανίζονται μαζί σε μία συναλλαγή ή σε ένα καλάθι αγοράς (market basket). Τα μοντέλα που κατασκευάζει δίνουν την πιθανότητα διαφορετικά προϊόντα να αγοράζονται μαζί και μπορούν να εκφραστούν σαν κανόνες. Είναι στενά συνδεδεμένη με αναλύσεις που πραγματοποιούνται στο λιανικό εμπόριο, όπου η πληροφόρηση για προϊόντα που αγοράζονται ταυτόχρονα μπορεί να αποτελεί το μοναδικό διαθέσιμο δεδομένο για την εξαγωγή πελατειακών προτύπων. Ας σημειωθεί ότι άλλες πληροφορίες όπως δημογραφικά και ιστορικά στοιχεία δεν είναι διαθέσιμες δεδομένου του ότι οι συναλλαγές είναι ανώνυμες. Ένα τυπικό παράδειγμα των αποτελεσμάτων από την market basket ανάλυση είναι ότι οι πελάτες που αγοράζουν βαφές αγοράζουν βούρτσες για βάψιμο χωρίς να ισχύει το αντίστροφο.

Σαν μια τεχνική ομαδοποίησης (clustering), η ανάλυση καλαθιού αγοράς (market basket analysis) είναι χρήσιμη σε προβλήματα όπου θέλουμε να ξέρουμε τι είδους στοιχεία εμφανίζονται είτε μαζί είτε σε μια συγκεκριμένη ακολουθία. Τα αποτελέσματα είναι συχνά ιδιαίτερα πρακτικά διότι περιλαμβάνουν τα ειδικά στοιχεία. Η πληροφόρηση που προκύπτει μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς, όπως για τον σχεδιασμό των αποθεμάτων, τον περιορισμό των ειδικών τμημάτων σε ένα (από τα προϊόντα που σε σύνολο τείνουν να εμφανίζονται μαζί), το πακετάρισμα των προϊόντων, την προσφορά κουπονιών για τα άλλα προϊόντα όταν ένα από αυτά πωλείται χωρίς τα υπόλοιπα κ.λ.π. Στην περίπτωση που οι συναλλαγές δεν είναι ανώνυμες, η market basket ανάλυση μπορεί να προσαρμοσθεί για χρήση σε ιστορικά δεδομένα με μία χρονική συνιστώσα.

### 2.7.2 Συλλογιστική βασισμένη στη μνήμη του συστήματος

Η συλλογιστική βασισμένη στη μνήμη του συστήματος (Memory Based Reasoning), είναι μία άμεση τεχνική εξόρυξης δεδομένων (data mining) που χρησιμοποιεί γνωστές καταστάσεις σαν ένα είδος μοντέλου για να κάνει προβλέψεις γύρω από άγνωστες περιστάσεις. Η τεχνική αυτή αναζητά τα πιο γειτονικά στοιχεία στις γνωστές καταστάσεις και συνδυάζει τις αξίες τους για να προσδιορίσει αξίες ταξινόμησης ή πρόβλεψης. Για παράδειγμα, μπορεί να έχουμε μια βάση δεδομένων με διεκδικήσεις, ισχυρισμούς και έστω ότι αυτοί έχουν διευθετηθεί έπειτα από έρευνα. Εάν θέλουμε να προσδιορίσουμε το κατά πόσο μια νέα διεκδίκηση εγγυάται περαιτέρω έρευνα, θα πρέπει να βρούμε παρόμοιες, «γειτονικές» διεκδικήσεις στη βάση δεδομένων και να πάρουμε την απόφαση για περαιτέρω έρευνα ή άμεση εκτέλεση, βασιζόμενοι στο επίπεδο των «γειτόνων». Η απόκλιση από τους «γείτονες» παρέχει ένα μέτρο της ορθότητας των αποτελεσμάτων.

Ένα από τα κυριότερα πλεονεκτήματα της συλλογιστικής βασισμένης στη μνήμη του συστήματος είναι η ικανότητα του να τρέχει κυριολεκτικά κάθε είδους δεδομένα, ακόμη και χωρίς τροποποίηση δεδομένων. Τα δύο συστατικά στοιχεία - κλειδιά στην συλλογιστική βασισμένη στη μνήμη του συστήματος είναι η λειτουργία της απόκλισης που χρησιμοποιείται για την εύρεση των κοντινότερων «γειτόνων» και η συνδυαστική λειτουργία που συνδυάζει αξίες στους πιο κοντινούς «γείτονες» προκειμένου να προβεί σε μια πρόβλεψη. Για ορισμένες περιπτώσεις, οι λειτουργίες αυτές μπορούν να εκφραστούν σαν δηλώσεις γλώσσας SQL σε μια συγγενική βάση δεδομένων, με άμεση δυνατότητα ενεργοποίησης για τους σκοπούς της εξόρυξης δεδομένων. Αν και είναι δυνατό, η χρήση μιας συγγενικής βάσης δεδομένων τείνει να είναι ανεπαρκής. Σε άλλες περιπτώσεις, η συλλογιστική βασισμένη στη μνήμη του συστήματος μπορεί να στηρίζεται σε περισσότερο πολύπλοκους τύπους δεδομένων όπως κείμενα, εικόνες, και ήχους όταν είναι διαθέσιμη η λειτουργία της απόκλισης σε αυτά τα πεδία.

Ένα ακόμη θεμελιώδες πλεονέκτημα της συλλογιστικής βασισμένης στη μνήμη του συστήματος είναι η ικανότητα της να μαθαίνει για νέες ταξινομήσεις, εισάγοντας απλά νέες καταστάσεις στη βάση δεδομένων. Μόλις βρεθούν οι κατάλληλες λειτουργίες απόκλισης και συνδυασμού, τείνουν να παραμείνουν αρκετά σταθερές ακόμη και καθώς οι νέες καταστάσεις για τις νέες κατηγορίες είναι ενσωματωμένες μέσα στα γνωστά δεδομένα. Αυτή η ευκολία ενσωμάτωσης αλλαγών στο πεδίο και τη σειρά, διαφοροποιεί την συλλογιστική βασισμένη στη μνήμη του συστήματος από τις περισσότερες άλλες τεχνικές εξόρυξης δεδομένων, οι οποίες πρέπει να εφαρμοστούν εκ νέου προκειμένου να ενσωματωθεί ουσιαστικά νέα πληροφορία.

### 2.7.3 Εντοπισμός ομάδων

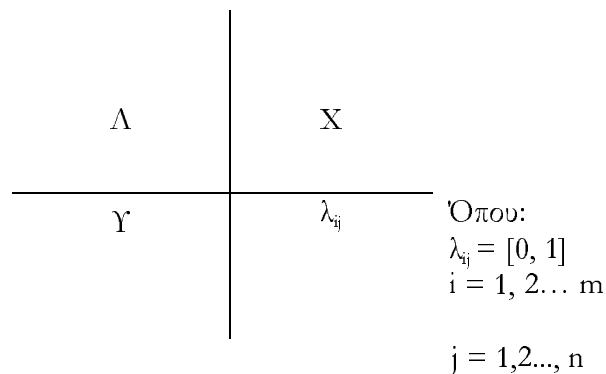
Ο εντοπισμός ομάδων (cluster detection) είναι η κατασκευή μοντέλων που βρίσκουν αρχεία δεδομένων που είναι παρόμοια μεταξύ τους. Αυτοί οι όγκοι ομοιότητας αποκαλούνται ομάδες (clusters). Πρόκειται, επομένως, για έμμεση εξόρυξη δεδομένων, εφόσον ο σκοπός είναι η εκ των προτέρων εύρεση άγνωστων ομοιοτήτων μέσα στα δεδομένα. Υπάρχουν πολυάριθμες τεχνικές για την εύρεση των ομάδων, συμπεριλαμβανομένων γεωμετρικών και στατιστικών μεθόδων καθώς και νευρωνικών δικτύων. Ο εντοπισμός ομάδων αποτελεί έναν πολύ καλό τρόπο για την έναρξη οποιασδήποτε ανάλυσης δεδομένων. Παρόμοιες ομάδες μπορούν να αποτελέσουν το σημείο εκκίνησης για να μάθουμε τι περιλαμβάνεται στα δεδομένα και για να ανακαλύψουμε τον καλύτερο τρόπο αξιοποίησης τους.

### 2.7.4 Ποσοτική Ανάλυση (*Quantitative Analysis*)

Η ποσοτική ανάλυση (Quantitative Analysis ή Q-Analysis) είναι κατασκευασμένη για να επεξεργάζεται 0-1 στοιχεία (0 - ΟΧΙ, 1 - ΝΑΙ). Σαν πίνακα δεδομένων έχει έναν πίνακα που στις γραμμές και στήλες του υπάρχουν διάφορες μεταβλητές που η σχέση τους συμβολίζεται με ένα 0 - όχι ή 1 - ναι. Είναι σαν να έχουμε ένα δίκτυο που οι κορυφές είναι όλες οι παραπάνω μεταβλητές στις γραμμές και στήλες του πίνακα και όπου συσχετίζονται οι

μεταβλητές με 1 υπάρχει μεταξύ των αντιστοιχών κορυφών διαδρομή και όπου δεν υπάρχει συσχέτιση δεν υπάρχει διαδρομή. Παρακάτω θα δούμε μια πιο ουσιαστική γραφική αναπαράσταση της μεθόδου.

Έχουμε λοιπόν δύο ομάδες, την  $X$  που είναι οι στήλες του πίνακα και την  $Y$  που είναι οι γραμμές του πίνακα που συνδέονται με μια σχέση  $\lambda$ . Το όλο πρόβλημα αναπαρίσταται ως ακολούθως.



Όπου  $m$  είναι ο αριθμός των μεταβλητών της ομάδας  $Y$  και  $n$  είναι ο αριθμός των μεταβλητών της ομάδας  $X$ .

Η αντιστοιχία  $\lambda$  της ομάδας  $Y$  στην ομάδα  $X$  μας δίνει με πόσα στοιχεία της ομάδας  $X$  είναι συσχετισμένο κάθε στοιχείο της ομάδας  $Y$ . Τη σχέση αυτή τη συμβολίζουμε  $K_Y(X; \lambda)$ . Η αντίστροφη αντιστοιχία  $\lambda^{-1}$  της ομάδας  $X$  στην ομάδα  $Y$  μας δίνει με πόσα στοιχεία της ομάδας  $Y$  είναι συσχετισμένο κάθε στοιχείο της ομάδας  $X$  και την συμβολίζουμε  $K_Y(Y; \lambda^{-1})$ .

Ο αλγόριθμος λοιπόν της Q-Analysis παίρνει τον παραπάνω πίνακα και κάνει δύο δουλειές. Η πρώτη δουλειά είναι η άθροιση των στοιχείων που είναι 1. Βρίσκει δηλαδή ποια μεταβλητή έχει τα περισσότερα στοιχεία 1 στις στήλες της αν κάνει αντιστοιχία  $\lambda^{-1}$  (μεταβλητές  $X$ ) ή στις γραμμές της αν κάνει αντιστοιχία  $\lambda$  (μεταβλητές  $Y$ ). Κάνει δηλαδή μια ιεράρχηση των μεταβλητών κάθε ομάδας σύμφωνα με το ποια μεταβλητή έχει τα περισσότερα στοιχεία 1.

Η δεύτερη δουλειά που κάνει ο αλγόριθμος της Q-Analysis είναι η ομαδοποίηση, βρίσκει δηλαδή τα κοινά στοιχεία 1 που έχουν ανά δύο οι μεταβλητές μιας ομάδας και τους κατατάσσει στην ίδια ομάδα.

Η όλη διαδικασία γίνεται σε διάφορα επίπεδα που λέγονται q-επίπεδα συσχέτισης και ξεκινούν από τις περισσότερες μονάδες συν μία που έχει μια μεταβλητή και κατεβαίνει επίπεδο προς επίπεδο αφαιρώντας μια μονάδα κάθε φορά μέχρι που να φτάσει στο μηδέν.

### *2.7.5 Ανάλυση συνδέσεων*

Η μέθοδος αυτή βασίζεται στην ακολουθία σχέσεων ανάμεσα στις εγγραφές (πεδία) μιας βάσης δεδομένων. Πάνω σε αυτή τη σχέση, που παρουσιάζουν τα πεδία, δημιουργούνται μοντέλα που ανιχνεύουν περαιτέρω συσχετίσεις. Ουσιαστικά αυτή η μέθοδος αποτελεί εφαρμογή της θεωρίας δημιουργίας γραφημάτων στην εξόρυξη δεδομένων. Οι σχέσεις που εμφανίζονται στους πελάτες αποτελούν ένα εξαιρετικά σημαντικό θέμα τις ομάδες μάρκετινγκ (marketing groups), αφού οι επιχειρήσεις εντείνουν τις προσπάθειες τους να πλησιάσουν ώστε να εξυπηρετήσουν ομάδες πελατών, νοικοκυριά, κλπ. Ως εργαλείο της εξόρυξης δεδομένων η ανάλυση συνδέσεων (link analysis) μάλλον υποστηρίζεται ανεπαρκώς από τις αλληλοσυσχετιζόμενες (relational) βάσεις δεδομένων. Το μεγαλύτερο πεδίο εφαρμογών αυτής της μεθόδου είναι η εγκληματολογία, όπου στοιχεία για διάφορα εγκλήματα συνδέονται μεταξύ τους, ώστε αν δοθεί κάποια λύση.

### *2.7.6 Επαγωγικά δένδρα απόφασης*

Τα δένδρα αποφάσεων όπως προαναφέρθηκε, χρησιμοποιούνται ως απευθείας διαδικασία εξόρυξης δεδομένων και περισσότερο για ταξινομήσεις. Με τη μέθοδο αυτή χωρίζονται τα δεδομένα (κάτω από μια αλληλουχία) σε διαφορετικά υποσύνολα καθένα από τα οποία περιγράφεται από έναν απλό κανόνα. Τα κυριότερα πλεονεκτήματα αυτής της μεθόδου είναι η εύκολη χρήση της και ότι βασίζεται σε απλά μοντέλα και ρητούς κανόνες. Αυτό επιτρέπει



ακόμη και σε μη ειδικούς να ελέγχουν τη διαδικασία καθώς και τα αποτελέσματα.

### *2.7.7 Τεχνητά νευρωνικά δίκτυα*

Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks) τα οποία αναλύθηκαν προηγουμένως, αποτελούν ίσως την πιο διαδεδομένη εφαρμογή εξόρυξης δεδομένων. Ουσιαστικά είναι μια προσπάθεια τεχνικής απομίμησης του ανθρώπινου εγκεφάλου και του τρόπου με τον οποίο αυτός σκέφτεται. Στην πιο συνηθισμένη τους μορφή τα νευρωνικά δίκτυα «εκπαιδεύονται» με τη χρήση σετ εντολών και γενικευμένων συνθηκών, για να κάνουν ταξινομήσεις και προβλέψεις. Νέες εφαρμογές και νέες δομές για νευρωνικά δίκτυα ερευνώνται και παρουσιάζονται σχεδόν καθημερινά. Αυτό είναι και το μεγαλύτερο πλεονέκτημα τους, αφού η ευρεία διάδοση τους καθιστά «εύκολη» την όποια εφαρμογή τους.

### *2.7.8 Γενετικοί αλγόριθμοι*

Οι γενετικοί αλγόριθμοι (genetic algorithms) χρησιμοποιούν τον μηχανισμό της γενετικής και φυσικής επιλογής για την αναζήτηση και εύρεση των άριστων παραμέτρων που περιγράφουν ένα πρόβλημα. Οι γενετικοί αλγόριθμοι κάνουν χρήση της επιλογής, διασταύρωσης και μετάλλαξης δεδομένων, ώστε να δώσουν επιτυχημένες γενεές λύσεων.

### *2.7.9 Αναλυτική επεξεργασία δεδομένων σε πραγματικό χρόνο.*

Οι μέθοδοι αναλυτικής επεξεργασίας σε πραγματικό χρόνο (Online Analytic Processing - OLAP) βασίζονται στις πολυδιάστατες και πολυκριτήριες βάσεις δεδομένων. Αυτού του είδους οι βάσεις είναι στην πραγματικότητα μια αναπαράσταση των δεδομένων που επιτρέπει στον χρήστη να αντλήσει από αυτά ένα σύνολο από κατανοητές και σημαντικές συνοψίσεις.

## **2.8. Εφαρμογές της εξόρυξης δεδομένων**

Σε όλες τις σύγχρονες επιχειρήσεις, η εξόρυξη δεδομένων αποτελεί στρατηγικής σημασίας εφαρμογή, και αυτό ενισχύεται από μια έρευνα της Two

Corps Corp. η οποία δείχνει πως οι τρεις πιο δημοφιλείς χρήσεις της προέρχονται από το χώρο του μάρκετινγκ, και είναι οι ακόλουθοι.

### 2.8.1 Η κατασκευή προφίλ πελατών (*customer profiling*)

Στην περίπτωση αυτή ορίζονται τα χαρακτηριστικά γνωρίσματα του «καλού» πελάτη και στόχος είναι τόσο η πρόβλεψη του ποιοι πελάτες θα εξελιχθούν σε «καλούς», όσο και η διερεύνηση προοπτικών για νέες αγορές. Συγκεκριμένα η εξόρυξη δεδομένων βρίσκει σε μια βάση δεδομένων πρότυπα (patterns) πελατών, τα οποία χρησιμοποιούνται έτσι ώστε η όλη διαδικασία να λειτουργήσει πιο αποτελεσματικά. Χαρακτηριστικό παράδειγμα είναι ο εντοπισμός «καλών» πελατών για κάποιες προσφορές μέσω του ταχυδρομείου, μια διαδικασία η οποία βοηθά τους υπεύθυνους marketing της εταιρίας να μειώσουν τα έξοδα και ταυτόχρονα να βελτιώσουν το αποτέλεσμα.

### 2.8.2 Το στοχευόμενο marketing (*targeted marketing*)

Θετικά αποτελέσματα παρόμοια με την προηγούμενη περίπτωση έχουμε και στην περίπτωση επικέντρωσης συγκεκριμένων διαφημιστικών προσπαθειών σε υπάρχοντες και πιθανούς πελάτες.

### 2.8.3 Η ανάλυση του καλάθιού αγοράς (*market basket analysis*)

Η περίπτωση αυτή βοηθά τα καταστήματα λιανικών πωλήσεων ποια προϊόντα δύνανται να αγοραστούν μαζί ή από έναν καταναλωτή. Με αυτό τον τρόπο καθορίζεται ποια προϊόντα πρέπει να αποθηκευτούν και σε ποια καταστήματα, καθώς και σε ποιο ράφι. Η εξόρυξη δεδομένων μπορεί επίσης να καθορίσει ποιες προσφορές ή κουπόνια είναι περισσότερο αποτελεσματικά.

Άλλη χρήση της εξόρυξης δεδομένων είναι η ανακάλυψη απάτης, που χρησιμοποιείται ευρέως από εταιρίες τηλεπικοινωνιών, ασφαλίσεων, καθώς και κυβερνητικές υπηρεσίες που υφίστανται τεράστιες ζημιές από απάτες. Οι συναλλαγές υψηλού κινδύνου εντοπίζονται έγκαιρα με την χρήση της εξόρυξης δεδομένων. Τέλος η εξόρυξη δεδομένων χρησιμοποιείται για την κατανόηση

των χαρακτηριστικών της αγοράς, για την πρόβλεψη της πορείας εταιρειών και μετοχών, καθώς και σε εφαρμογές στο ιατρικό πεδίο π.χ. στην πρόβλεψη της αποτελεσματικότητας χειρουργικών επεμβάσεων, διαγνωστικών εξετάσεων, λήψης φαρμάκων κ.τ.λ.

## **2.9. Παραδείγματα εφαρμογής των μεθόδων**

Πριν προχωρήσουμε στην εφαρμογή όλων των προαναφερθέντων σε δικές μας βάσεις δεδομένων, μπορούμε να ρίξουμε μια ματιά στις έτοιμες βάσεις που παρέχει το Aira\*, ώστε να εξοικειωθούμε με τις τεχνικές που ακολουθούν. Πατάμε, λοιπόν, Αρχείο - Άνοιγμα (file - open) και ανοίγουμε το φάκελο Aira for Excel, ο οποίος βρίσκεται στην τοποθεσία αποθήκευσης που επιλέξαμε κατά την εγκατάσταση. Ο φάκελος αυτός περιέχει το αρχείο airasamples. Το επιλέγουμε και πατάμε άνοιγμα (open). Θα ανοίξει το βιβλίο εργασίας που περιέχει έξι φύλλα: Το πρώτο φύλλο περιέχει οδηγίες για τον τρόπο εγκατάστασης του Aira, καθώς και λεπτομέρειες για τις βάσεις δεδομένων που περιέχονται στα υπόλοιπα φύλλα. Μπορούμε επίσης να δούμε τα αρχεία βοήθειας του Aira πατώντας βοήθεια (Aira HELP), καθώς και αν παρακολουθήσουμε ένα animation movie σε μορφή Shockwave Flash, σχετικό με το πώς θα πρέπει να χρησιμοποιήσουμε το Aira, πατώντας Βοήθεια-Aira for Excel: The movie.

## **2.10. Ένα παράδειγμα εφαρμογής εξόρυξης δεδομένων**

Υπάρχουν πολλοί και αρκετά καλοί ορισμοί του «Data mining», ωστόσο η βασική έννοια του παραμένει ίδια σε όλους τους ορισμούς. Η IBM ορίζει το Data mining ως «διαδικασία εξαγωγής άγνωστων μέχρι στιγμής, έγκυρων και χρήσιμων πληροφοριών από μεγάλες βάσεις δεδομένων, και μετέπειτα χρήση των πληροφοριών αυτών για τη λήψη σημαντικών επιχειρηματικών αποφάσεων». Το σημαντικό στοιχείο στον ορισμό του Data mining αφορά σε αυτό που εννοούμε όταν κάνουμε λόγο για «άγνωστες πληροφορίες». Οι πληροφορίες

---

\* Το AIRA είναι εφαρμογή που λειτουργεί σαν πρόσθετο του Microsoft Excel

αυτές, που θα μπορούσαν, για παράδειγμα, να είναι το σύνολο των πωλήσεων μιας εταιρίας για το συγκεκριμένο έτος, δεν προέρχονται απευθείας από τα δεδομένα των φύλλων εργασίας, αλλά από τις σχέσεις που δημιουργούνται ανάμεσα στις στήλες των δεδομένων. Οι σχέσεις αυτές είναι έτσι δομημένες ώστε η αξία μίας ή περισσότερων στηλών να προβλέπει το αποτέλεσμα μιας άλλης. Πρόκειται, δηλαδή, για μοντέλα πρόγνωσης, τα οποία όμως πρέπει να είναι έγκυρα. Διαφορετικά δεν υπάρχει κανένας λόγος να τα χρησιμοποιούμε. Επιπλέον, η τελική πληροφορία θα πρέπει να είναι χρήσιμη στο χρήστη, ώστε από τα αποτελέσματα της διαδικασίας του Data mining να συμπεράνει και να ενεργήσει κατάλληλα. Όπως γίνεται φανερό, η διαδικασία του Data mining, της εξαγωγής πληροφοριών από μεγάλες βάσεις δεδομένων, μπορεί να είναι και ακριβή και χρονοβόρα. Το Aira είναι ένα μικρό, ισχυρό και εύχρηστο εργαλείο για data mining, το οποίο λειτουργεί όχι ως ανεξάρτητη εφαρμογή αλλά ως add-in μέσα από το Excel.

### *2.10.1 Η βάση δεδομένων*

Αρχικά, για να το χρησιμοποιήσουμε, θα πρέπει να έχουμε ήδη ένα φύλλο εργασίας που να περιέχει μία βάση δεδομένων. Η δημιουργία βάσης δεδομένων στο Excel είναι σχετικά εύκολη υπόθεση: Το μόνο που έχουμε να κάνουμε είναι να δημιουργήσουμε μερικά πεδία στην πρώτη γραμμή του φύλλου και απλά να εισαγάγουμε τις εγγραφές στις στήλες. Για παράδειγμα θα χρησιμοποιήσουμε την ακόλουθη βάση δεδομένων με πέντε πεδία: ID, Age, Sex, Condition, Magazine.

Η βάση περιέχει ένα δείγμα 100 αναγνωστών με συγκεκριμένα στοιχεία που τους αφορούν. Στην περιοχή ID εισάγεται ο αύξων αριθμός κάθε ατόμου της ομάδας των αναγνωστών. Για κάθε έναν από τους αναγνώστες εισάγουμε την ηλικία, το φύλλο (Male, Female), την οικογενειακή κατάσταση (Single-Married) και το περιοδικό που διαβάζει συχνότερα, ανάμεσα από τα τρία περιοδικά που ρωτάμε.

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	Age	Sex	Condition	Magazine							
2	1	34	M	Single	CGO							
3	2	23	M	Single	Fantastic							
4	3	28	F	Married	CGO							
5	4	38	M	Married	CGO							
6	5	16	M	Single	PC Master							
7	6	26	F	Single	CGO							
8	7	28	M	Married	CGO							
9	8	34	M	Married	CGO							
10	9	23	F	Single	Fantastic							
11	10	15	M	Single	PC Master							
12	11	19	M	Single	PC Master							
13	12	34	M	Married	CGO							
14	13	44	M	Married	Fantastic							
15	14	29	F	Married	CGO							
16	15	39	M	Married	CGO							
17	16	20	F	Single	Fantastic							
18	17	26	F	Single	Fantastic							
19	18	14	M	Single	PC Master							
20	19	19	M	Single	PC Master							
21	20	38	M	Married	CGO							
22	21	33	M	Single	CGO							
23	22	24	M	Single	Fantastic							
24	23	26	F	Married	CGO							
25	24	38	M	Married	CGO							
26	25	15	M	Single	PC Master							
27	26	26	F	Single	CGO							
28	27	28	M	Married	CGO							
29	28	37	M	Married	CGO							

Η βάση αυτή αποτελεί μία μορφή «δείγματος αγοράς». Δηλαδή, μια ομάδα ατόμων με τις προτιμήσεις τους σε συγκεκριμένα προϊόντα. Με τη μορφή που έχει η συγκεκριμένη βάση είναι ελάχιστα τα συμπεράσματα που μπορούμε να εξαγάγουμε σχετικά με τις τάσεις του συνόλου, δηλαδή τις σχέσεις που δημιουργούνται ανάμεσα στις στήλες των δεδομένων. Αν μπορούσαμε να ανακαλύψουμε τους διάφορους συσχετισμούς, θα κατανοήσουμε ποιες ομάδες αναγνωστών διαμορφώνονται, ποια είναι τα χαρακτηριστικά τους και πώς θα μπορούσαμε να τα εκμεταλλευτούμε. Αν π.χ. βρούμε ποιο είναι το προφίλ των αναγνωστών που διαβάζουν το περιοδικό 'CGO' θα είμαστε σε θέση να εκμεταλλευτούμε αυτή την πληροφορία προκειμένου να προσεγγίσουμε διαφημιστικά όλα τα άτομα που ταιριάζουν με το συγκεκριμένο προφίλ, αυξάνοντας έτσι την κυκλοφορία του περιοδικού. Βλέπουμε, λοιπόν, στο παράδειγμα μας ότι η χρήση ενός εργαλείου έχει ένα αμιγώς διαδικαστικό σκέλος δηλαδή την εξαγωγή κανόνων, προφίλ, συμπερασμάτων και ένα άλλο αμιγώς πρακτικό δηλαδή την επιχειρησιακή αξιοποίηση των αποτελεσμάτων.

### 2.10.2 Χρήση του Aira

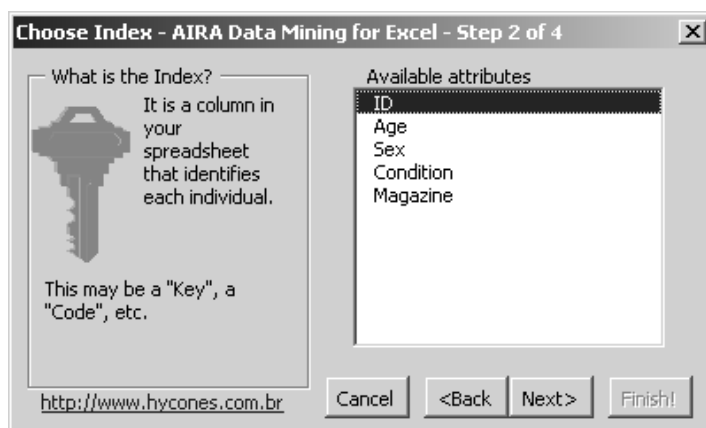
Το πρώτο που θα πρέπει να κάνουμε είναι να επιλέξουμε ένα οποιοδήποτε κελί της βάσης δεδομένων. Κατόπιν, πατάμε το κουμπί Aira Data Mining. Το Aira εκτελείται μέσα από έναν οδηγό τεσσάρων βημάτων.



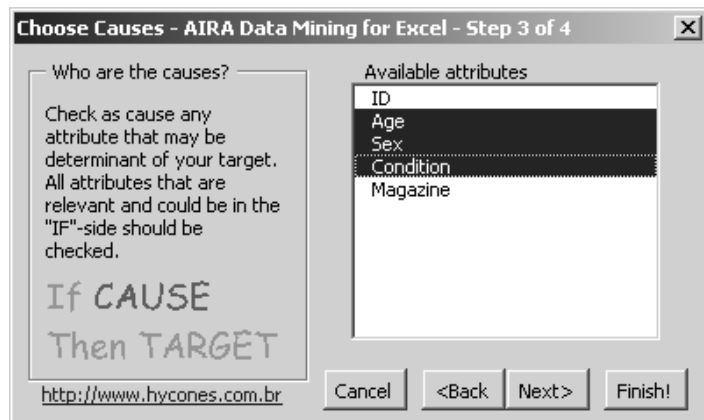
Στη δεξιά πλευρά της οθόνης φαίνονται τα πεδία της βάσης. Για το Aira τα πεδία αυτά ονομάζονται ιδιότητες (attributes). Η αριστερή πλευρά της οθόνης επεξηγεί τι ακριβώς πρέπει αν κάνουμε σε αυτό το βήμα. Ο σκοπός του πρώτου βήματος είναι απλός. Υποτίθεται ότι θέλουμε να μάθουμε τι προκύπτει για το προϊόν μας βάσει δεδομένων συνθηκών. Αλλιώς, θέλουμε να μάθουμε ποιοι και γιατί προτιμούν το Α, Β ή Γ περιοδικό, με βάση τα στοιχεία που έχουμε διαθέσιμα για την ομάδα στην βάση δεδομένων. Άρα έχουμε μια πρόταση IF (αν συμβαίνει αυτό και εκείνο), THEN (τότε προτιμάται αυτό το περιοδικό). Στο πρώτο βήμα του οδηγού, καλούμαστε να επιλέξουμε το THEN. Το αντικείμενο για το οποίο γίνεται η έρευνα είναι φυσικά τα περιοδικά. Συνεπώς, επιλέγουμε Magazine από τη δεξιά πλευρά της οθόνης και πατάμε «Next» για να περάσουμε στην επόμενη οθόνη.

Στη δεύτερη οθόνη του οδηγού θα πρέπει να επιλέξουμε το Index. Το Index είναι ένα κλειδί ταξινόμησης. Επιλέγουμε τη στήλη που περιέχει το μοναδικό αριθμό κάθε ατόμου της βάσης δεδομένων. Κάθε άτομο της βάσης μπορεί να έχει την ίδια ηλικία, το ίδιο φύλλο και τις ίδιες προτιμήσεις με κάποιο άλλο άτομο, ωστόσο θα πρέπει να έχει μοναδικό χαρακτηριστικό που θα το κάνει αν

ξεχωρίσει από κάθε άλλο μέλος της βάσης. Στη συγκεκριμένη βάση δεδομένων, η στήλη που περιέχει το μοναδικό αριθμό κάθε ατόμου είναι η ID. Από τη δεξιά πλευρά, λοιπόν, της οθόνης, επιλέγουμε ID και πατάμε «Next» για να συνεχίσουμε.



Όπως προαναφέρθηκε, ουσιαστικά έχουμε αν εργαστούμε με μια πρόταση IF...THEN... Το σκέλος THEN καλύφθηκε στο πρώτο βήμα του οδηγού. Το σκέλος IF καλύπτεται στο τρίτο βήμα. Σε αυτή την οθόνη θα πρέπει να επιλέξουμε τις αιτίες που θέλουμε να εξετάσουμε. Αν π.χ. επιλέξουμε την ηλικία, η εφαρμογή θα αναζητήσει τη σχέση της ηλικίας κάθε ατόμου με το περιοδικό που επιλέγει. Αν πάλι επιλέξουμε το φύλλο, η εφαρμογή θα ψάξει να βρει τη σχέση ανάμεσα στο φύλλο και την επιλογή περιοδικού. Φυσικά, μπορούμε αν επιλέξουμε περισσότερα από ένα κριτήρια. Έτσι επιλέγουμε τα: Sex, Age και Condition από τη δεξιά πλευρά της οθόνης και πατάμε «Next».

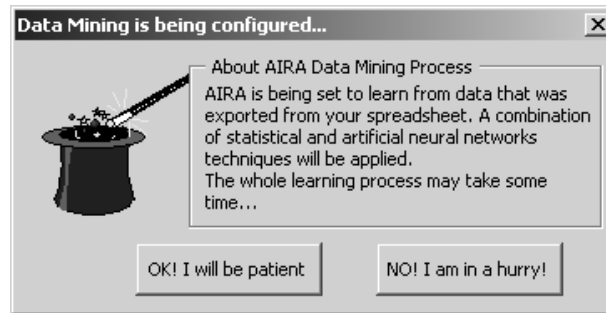


Το τελευταίο βήμα του οδηγού είναι το βήμα του διαχωρισμού (Discrimination). Στο βήμα αυτό εμφανίζονται μόνο αριθμητικά πεδία της βάσης δεδομένων. Αν έχουμε κάποιο πεδίο όπως το Age, θα πρέπει να το επιλέξουμε ώστε να προχωρήσουμε αργότερα σε κατηγοριοποίηση ηλικιών, διαφορετικά το πρόγραμμα θα αντιμετωπίσει κάθε ηλικία ως διαφορετική κατηγορία. Είναι πολύ πιο κατανοητό άλλωστε, να έχουμε να κάνουμε με δυο κατηγορίες ηλικιών (Ενήλικοι - Ανήλικοι) παρά με πολλές και διαφορετικές ηλικίες. Επιλέγουμε, λοιπόν, το πεδίο Age και πατάμε «Finish».

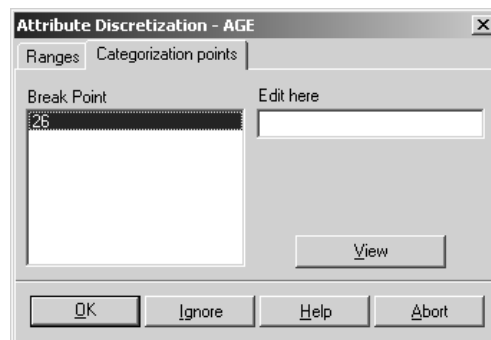
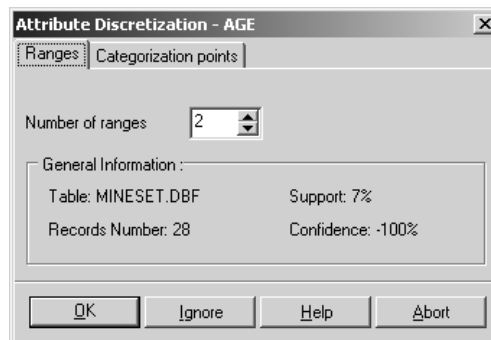


Ακολουθεί η οθόνη που μας ενημερώνει ότι το Aira ετοιμάζεται για την επεξεργασία των δεδομένων. Το μόνο που έχουμε να κάνουμε εδώ είναι να πατήσουμε στο κουμπί «OK, I will Be Patient».

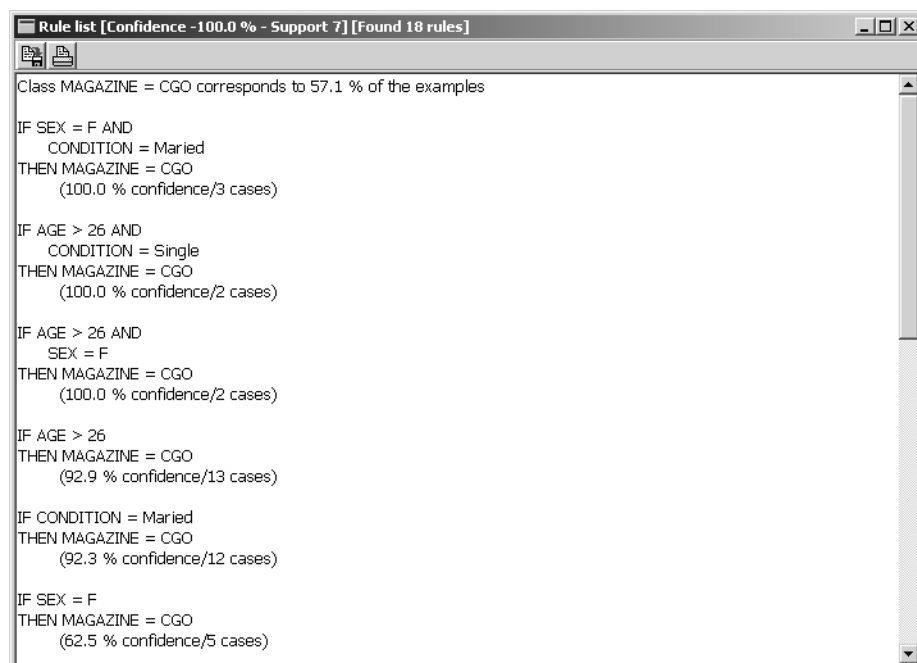




Θα εμφανιστεί η οθόνη των ιδιοτήτων του διαχωρισμού. Αναφέραμε ότι στο τέταρτο βήμα έπρεπε να επιλέξουμε ηλικία, ώστε να δημιουργηθούν ηλικιακές κατηγορίες για την καλύτερη ανάγνωση των αποτελεσμάτων. Εδώ θα πρέπει να καθορίσουμε τον αριθμό των κατηγοριών. Αφήνουμε την εξ' ορισμού επιλογή ενεργό. Κάνουμε όμως κλικ στο tab «Categorization Points» και πατάμε το κουμπί «View». Στην περιοχή «Break Point» θα εμφανιστεί ένας αριθμός. Ο αριθμός αυτός είναι το σημείο διαχωρισμού των δυο ηλικιακών κατηγοριών. Ας επιχειρήσουμε να τον αλλάξουμε. Αφού πρώτα τον επιλέξουμε, πατάμε στο πλαίσιο «edit here» και πληκτρολογούμε 18. Ο αριθμός της περιοχής «Break Point» έχει γίνει τώρα 18. Αυτό πρακτικά σημαίνει ότι έχουμε δύο κατηγορίες ηλικιών. Την άνω των 18 και την κάτω των 18. Πατάμε «OK», για να ολοκληρωθεί η διαδικασία.



Μετά το τέλος της διαδικασίας εμφανίζεται η λίστα των κανόνων. Τα αποτελέσματα είναι μπροστά μας. Το μόνο που μένει είναι να τα ερμηνεύσουμε.



### *2.10.3 Ερμηνεία της λίστας κανόνων*

Με μια πρώτη ματιά, η λίστα των κανόνων φαίνεται σαν μία σειρά από ακαταλαβίστικα σύμβολα και αριθμούς. Με λίγη υπομονή, όμως, θα δούμε ότι δεν είναι καθόλου ακαταλαβίστικα. Οι κανόνες (rules), που εξήγαγε το Aira από τη βάση δεδομένων μας, δεν είναι τίποτε άλλο από τις σχέσεις που βρήκε η εφαρμογή ότι ισχύουν μεταξύ των δεδομένων μας. Ας ριζούμε μια πιο προσεκτική ματιά στην εικόνα. Η πρώτη φράση αναφέρει «Class MAGAZINE=CGO corresponds 49,0% of the examples». Σημαίνει ότι το περιοδικό CGO διαβάζεται από το 49,0% των ατόμων της βάσης δεδομένων. Στη συνέχεια υπάρχουν πέντε προτάσεις που αρχίζουν με τη λέξη IF και καταλήγουν με τη THEN. Είναι: «Όλα τα άτομα γένους θηλυκού που είναι παντρεμένα διαβάζουν το περιοδικό 'CGO'». Ακολουθως, «το 86% των παντρεμένων ατόμων (άνδρες και γυναίκες) διαβάζει το περιοδικό CGO». Αυτομάτως εξάγεται το πρώτο συμπέρασμα. Ότι το περιοδικό 'CGO' έχει μειωμένες πωλήσεις στον ανδρικό πληθυσμό σε σχέση με τον γυναικείο. Συνεχίζουμε στην επόμενη: «Το 63% των ανδρών άνω των 18 ετών διαβάζει το περιοδικό». Οπότε εξάγεται το δεύτερο συμπέρασμα. Το ποσοστό των γυναικών άνω των 18 ετών που διαβάζει το περιοδικό είναι μικρότερο από των ανδρών. Επίσης σύμφωνα με τον πέμπτο κανόνα, το 50% όλων των γυναικών του δείγματος διαβάζει το συγκεκριμένο περιοδικό.

Μετά τους κανόνες αυτούς ακολουθεί με παρόμοιο τρόπο η παράθεση των κανόνων που αφορούν το περιοδικό "Tech Business".

Μελετώντας τους εξαγόμενους κανόνες, μπορούμε να βγάλουμε τα συμπεράσματα που χρειαζόμαστε. Τα μόνα που απαιτούνται είναι υπομονή και προσοχή ώστε να μπορέσουμε να συνδυάσουμε όλους τους κανόνες μεταξύ τους και να κατανοήσουμε τους συσχετισμούς και τα συμπεράσματα που προκύπτουν.

Στο επάνω αριστερό μέρος του παραθύρου των κανόνων υπάρχουν δύο μικρά κουμπιά, με τα οποία αυτά μπορούμε να αποθηκεύσουμε τη λίστα σε μορφή TXT ή να την τυπώσουμε κατευθείαν.

## 2.11. Απόκτηση γνώσης

Όπως είδαμε και προηγουμένως τα σημαντικότερα τμήματα ενός μοντέλου εξόρυξης δεδομένων είναι η μηχανή εξαγωγής συμπερασμάτων και το υποσύστημα απόκτησης γνώσης που δέχεται οδηγίες από τον εμπειρογνώμονα ή τον μηχανικό γνώσης. Σημαντικό τώρα είναι να εξετάσουμε ποια είναι τα κριτήρια επιλογής και ποιος ο ρόλος του εμπειρογνώμονα ή του μηχανικού γνώσης και ποιες οι τεχνικές απόκτησης γνώσης.

Η επιλογή του κατάλληλου ειδικού βασίζεται στα παρακάτω κύρια χαρακτηριστικά.

Ειδίκευση στον τομέα, εμπειρία και φήμη. Οι ειδικοί που θα επιλεγούν πρέπει να έχουν την κατάλληλη εκπαίδευση καθώς και εμπειρία στο θέμα πάνω στο οποίο θα αναπτυχθεί το μοντέλο. Πολλές φορές η αξιοπιστία των αποτελεσμάτων του μοντέλου κρίνεται με βάση την φήμη του ειδικού. Επίσης σημαντικό προσόν αποτελεί η ενασχόληση του εμπειρογνώμονα με την τεχνητή νοημοσύνη και τη μηχανική μάθηση.

Προσωπικά χαρακτηριστικά γνωρίσματα. Η κατοχή γνώσης από τους ειδικούς δεν είναι αρκετή αν δεν συνοδεύεται από την αντίστοιχη ικανότητα μετάδοσης της γνώσης. Σημαντικό είναι επίσης, ο ειδικός να έχει αίσθηση του χιούμορ, κοινωνικότητα αλλά και άνεση στις προσωπικές επαφές.

Διαθεσιμότητα. Η μη διαθεσιμότητα του εμπειρογνώμονα αποτελεί ένα από τα πιο κοινά προβλήματα. Είναι λοιπόν σημαντικό να προκαθοριστούν οι ώρες απασχόλησης του ειδικού, κάτι που αποτελεί την βασικότερη προϋπόθεση για την επιτυχή ανάπτυξη του μοντέλου.

Όταν λοιπόν βρεθούν οι κατάλληλοι ειδικοί με όλα τα παραπάνω θετικά χαρακτηριστικά, πρέπει να γίνει η τελική επιλογή. Είναι αποδεκτό ότι μια ομάδα ειδικών είναι προτιμητέα έναντι ενός ειδικού μεμονωμένα. Αυτό μπορεί να γίνει καλύτερα αντιληπτό από τα παρακάτω χαρακτηριστικά:

### *2.11.1 Η επιλογή ενός μόνο ειδικού παρουσιάζει προβλήματα*

Δυσκολία εύρεσης αρκετού διαθέσιμου χρόνου. Λόγω της φύσης της εργασίας του ειδικού η διαθεσιμότητα του είναι μειωμένη γεγονός που ίσως καθυστερήσει σημαντικά την ανάπτυξη του μοντέλου, ιδίως όταν απαιτείται η συμβολή του σε καίρια σημεία.

Προσωπική αντίληψη. Κάθε άνθρωπος φιλτράρει και αναπαράγει σύμφωνα με τις εμπειρίες και τις αντιλήψεις του τις πληροφορίες. Η επιβολή της προσωπικότητάς του μπορεί να επηρεάσει την αποτελεσματικότητα του μοντέλου.

Αιτιολόγηση περιορισμένης ευρύτητας. Ο κάθε ειδικός ακολουθεί τον δικό του τρόπο σκέψης, χωρίς να μπορεί να προσφέρει διαφορετική αιτιολόγηση των αποτελεσμάτων, κάτι που επιδρά αρνητικά στον πλούτο της αποκτούμενης γνώσης.

Μη πλήρης γνώση του αντικειμένου. Εξ ορισμού οι ειδικοί είναι «πλήρεις γνώσεων μόνο για το τμήμα του αντικειμένου που ασχολούνται».

### *2.11.2 Αντίθετα η επιλογή ομάδας ειδικών προσφέρει πλεονεκτήματα*

Πλήρης βάση δεδομένων. Ο κάθε ειδικός αντιμετωπίζει διαφορετικές περιπτώσεις σχετικά με το πεδίο εφαρμογής του μοντέλου.

Μεγαλύτερη κάλυψη υποπεριπτώσεων. Λόγω της ενασχόλησης ομότεχνων ανθρώπων σε κοινό περιβάλλον με κοινό σκοπό.

Αυξημένη ποιότητα και αξιοπιστία αποτελεσμάτων. Τα αποτελέσματα κρίνονται από διαφορετική σκοπιά ανάλογα με τον εμπειρογνώμονα, μα αποτέλεσμα την μεγαλύτερη ποσοστιαία αποκάλυψη πληροφοριών και συνεπώς μεγαλύτερη αξιοπιστία.

Ενθάρρυνση των προσωπικών επαφών. Οι ειδικοί ανταλλάσσουν απόψεις με αποτέλεσμα την γρηγορότερη και αποτελεσματικότερη αντιμετώπιση τυχόν προβλημάτων κατά την ανάπτυξη του μοντέλου.

Ο ρόλος του μηχανικού γνώσης περιλαμβάνει την καθοδήγηση της ανάπτυξης του μοντέλου, τον ορισμό του πεδίου δράσης, την επιλογή λογισμικού και ηλεκτρονικών συσκευών, την απόκτηση και παρουσίαση της γνώσης, την εφαρμογή του μοντέλου, την αξιολόγηση της απόδοσης του μοντέλου, την εκπαίδευση των χρηστών την υποστήριξη της λειτουργίας του μοντέλου και την παροχή συμβουλών και συστάσεων. (Prerau, 1990). Πρέπει να σημειωθεί πως οι τελικοί χρήστες εμπλέκονται στην φάση ανάπτυξης του μοντέλου, παρέχοντας έτσι μια σαφή άποψη σχετικά με το πώς πρέπει να λειτουργεί το σύστημα για να τους βοηθήσει όσο το δυνατόν περισσότερο.

Όταν οριστικοποιηθεί η επιλογή του ειδικού ή των ειδικών, πρέπει να αντιμετωπιστεί το πρόβλημα απόκτησης της απαραίτητης ποιότητας γνώσης, με την μικρότερη δυνατή κατανάλωση χρόνου, ώστε να αναπτυχθεί αποτελεσματικά το μοντέλο. Οι τρόποι απόκτησης γνώσης είναι οι εξής:

Συνέντευξη. Είναι η πιο συνηθισμένη μέθοδος, και ως βασική μορφή έχει τη δομή ελεύθερων ερωτήσεων και απαντήσεων μεταξύ του μηχανικού γνώσης και του ειδικού. Βασικά μειονεκτήματα είναι πως «οι άνθρωποι βλέπουν αυτό που περιμένουν να δουν» και ο εμπειρογνώμονας αντιμετωπίζει το πρόβλημα υποκειμενικά, χωρίς να μπορεί να περιγράψει αναλυτικά τον τρόπο σκέψης του.

Δομημένη συνέντευξη. Κατά την χρησιμοποίηση αυτής της τεχνικής, ο ειδικός είτε συμπληρώνει ένα ερωτηματολόγιο δομημένο από έναν μηχανικό γνώσης, είτε απαντά σε συγκεκριμένες ερωτήσεις του μηχανικού γνώσης.

Παρατήρηση. Με την χρησιμοποίηση αυτής της μεθόδου, ο ειδικός μπορεί να δουλεύει στο οικείο περιβάλλον του χωρίς παρεμβολές. Βασικό μειονέκτημα είναι ότι τις περισσότερες φορές, η αιτιολόγηση των ενεργειών του ειδικού δεν

γίνεται προφανής μέσω του αποτελέσματός τους, και συνεπώς το μόνο που μπορούμε να δούμε είναι το αποτέλεσμα των πράξεων του ειδικού.

Ανάλυση πρωτοκόλλου. Μια τεχνική που απαιτεί από τον ειδικό να «σκέφτεται φωναχτά». Περιλαμβάνει την ανάπτυξη σεναρίων από τον μηχανικό γνώσης, ο οποίος ζητά από τον ειδικό να αναπτύξει με λόγια τον τρόπο σκέψης του κατά την πορεία εύρεσης της λύσης. Αποτέλεσμα είναι η καταγραφή του τρόπου σκέψης του εμπειρογνώμονα, με αποτέλεσμα την καλύτερη κατανόηση του προβλήματος.

Επαναλαμβανόμενη ανάλυση. Η τεχνική αυτή αποσκοπεί στην κατάταξη των χαρακτηριστικών του προβλήματος ανάλογα με την σημαντικότητά τους. Ο ειδικός καλείται να συγκρίνει τα χαρακτηριστικά αυτά κατά τριάδες, κάνοντας όλους τους πιθανούς συνδυασμούς, και να προτείνει το αντίθετο του πιο σημαντικού από αυτά. Η μέθοδος αυτή κρίνεται ακατάλληλη όταν ο αριθμός των χαρακτηριστικών είναι μεγάλος γιατί πρέπει να πραγματοποιηθούν πολλές συγκρίσεις.

Όπως είδαμε η χρησιμοποίηση ομάδας ειδικών είναι η καταλληλότερη, και γίνεται ακόμη πιο αναγκαία λόγω της απαίτησης για πληρέστερη κάλυψη του προβλήματος και συνεπώς μεγαλύτερης πολυπλοκότητας του μοντέλου. Αν και μερικές από τις προαναφερθείσες μεθόδους όπως η συνέντευξη και η παρατήρηση μπορούν να χρησιμοποιηθούν και στην περίπτωση ομάδας ειδικών και μηχανικών γνώσης, εντούτοις υπάρχουν πολύ πιο κατάλληλες μέθοδοι:

Καταιγισμός: ιδεών (brainstorming). Για την εκκίνηση της μεθόδου, είναι απαραίτητη η παροχή ερεθίσματος στους ειδικούς με την μορφή ερώτησης ή ανάπτυξης σεναρίου. Ο κάθε ειδικός αναπτύσσει τις σκέψεις του, οι οποίες καταχωρούνται σε αρχείο υπολογιστή ή γράφονται σε χαρτί. Επακόλουθα, η πρόταση του ειδικού προσφέρεται στους υπόλοιπους συμμετέχοντες, οι οποίοι είτε συμφωνούν με τον τρόπο σκέψης του ειδικού είτε προσφέρουν νέες ιδέες.

Τελικά, όλες οι προτάσεις αναλύονται και οργανώνονται, παρέχοντας ένα πλήρες σκεπτικό προσέγγισης του προβλήματος.

Τεχνική της ονομαστικής ομάδας (Nominal Group Technique). Η NGT απαιτεί την δημιουργία μικρών ομάδων και την εφαρμογή τεσσάρων βημάτων:

1. Σιωπηλή γραπτή ανάπτυξη ιδεών.
2. Κυκλική καταγραφή των ιδεών.
3. Σειριακή συζήτηση του καταλόγου των ιδεών.
4. Ψηφοφορία.

Με την χρησιμοποίηση της NGT αποφεύγουμε τα αρνητικά επακόλουθα της κατά πρόσωπο αντιπαράθεσης μεταξύ των συμμετεχόντων. Το κάθε μέλος της ομάδας είναι ισότιμο επηρεάζοντας την τελική απόφαση στον ίδιο βαθμό.

Η τεχνική των Δελφών (Delphi Technique). Η τεχνική αυτή είναι ιδανική για την ελαχιστοποίηση της επιρροής πιθανά κυριαρχικών προσωπικοτήτων. Περιλαμβάνει την διανομή ερωτηματολογίων, την αξιολόγηση των απαντήσεων και την δημιουργία νέων ερωτηματολογίων βασισμένων στα αποτελέσματα των πρώτων. Η ανωνυμία διατηρείται σε όλα τα στάδια ανάπτυξης της τεχνικής.

Ψηφοφορία. Στην μέθοδο αυτή κάθε μέλος της ομάδας ενθαρρύνεται να ψηφίσει υπέρ μιας εκ των προτεινόμενων λύσεων. Η ψηφοφορία χρησιμοποιείται για την εύρεση συμβιβαστικών λύσεων μεταξύ των μελών της ομάδας.

Επαναλαμβανόμενη ανάλυση για ομάδες. Αποτελεί την προσαρμοσμένη μορφή της αντίστοιχης απλής τεχνικής.

Συστήματα υποστήριξης αποφάσεων (DSS). Ο όρος αυτός δεν περιγράφει κάποια τεχνική αλλά την υποστήριξη των προαναφερθέντων τεχνικών με ηλεκτρονικούς υπολογιστές. Η χρησιμοποίηση υπολογιστών επιταχύνει τη



διαδικασία ανάπτυξης του μοντέλου, παρέχει βάση δεδομένων σε ηλεκτρονική μορφή και μειώνει τις αντιπαραθέσεις μεταξύ των μελών της ομάδας.

Στην βιβλιογραφία επισημαίνεται πως η έλλειψη μεθοδολογικού σχεδιασμού της διαδικασίας απόκτησης γνώσης μειώνει την ποιότητα των αποτελεσμάτων. Η μεθοδολογία που προτείνεται παρέχεται περιληπτικά στον παρακάτω πίνακα και περιλαμβάνει τέσσερα βασικά στάδια.

Στάδιο	Χαρακτηριστικά
1. Σχεδιασμός	Κατανόηση του πεδίου Επιλογή εμπειρογνομώνων Ορισμός του προβλήματος Αναγνώριση του τύπου της εφαρμογής Σχεδιασμός της διαδικασίας Κατάρτιση χρονοδιαγράμματος
2. Εξαγωγή γνώσης	Εξήγηση της προσέγγισης απόκτησης γνώσης Ορισμός στόχων Διεξαγωγή της συνάντησης Ενημέρωση των ειδικών
3. Ανάλυση	Ανάλυση των συμπερασμάτων Αναπαράγωγή των πληροφοριών
4. Πιστοποίηση	Ανάπτυξη σεναρίων δοκιμής Αξιολόγηση του μοντέλου

Πίνακας 2-1. Μεθοδολογία απόκτησης γνώσης

*Σχεδιασμός της απόκτησης γνώσης.* Στο στάδιο αυτό ο μηχανικός γνώσης συζητά με τους ειδικούς ώστε να κατανοήσει το πεδίο εφαρμογής και να εκτιμήσει το είδος της απαραίτητης γνώσης, επιλέγει τους κατάλληλους εμπειρογνώμονες, αναγνωρίζει τον τύπο της εφαρμογής, καθορίζει τα κατάλληλα μοντέλα υποστήριξης αποφάσεων και καταρτίζει το χρονοδιάγραμμα των απαραίτητων συναντήσεων με τους ειδικούς.

*Εξαγωγή γνώσης.* Ο μηχανικός γνώσης συγκαλεί την ομάδα, αναλύει τον τρόπο προσέγγισης, θέτει τους στόχους της συνάντησης, ηγείται της διαδικασίας και ενημερώνει τους ειδικούς για τα αποτελέσματα της συνάντησης. Η εφαρμογή

διαφορετικών εργαλείων υποστήριξης των χρησιμοποιούμενων τεχνικών απόκτησης γνώσης είναι απαραίτητη για την επιτυχή έκβαση του σταδίου.

*Ανάλυση των αποκτηθέντων πληροφοριών.* Ο κύριος στόχος αυτού του σταδίου είναι η ανάπτυξη ενός πρωτότυπου μοντέλου. Για το σκοπό αυτό οι πληροφορίες που αποκτήθηκαν κατά τα προηγούμενα στάδια αναλύονται και αναπαράγονται υπό την μορφή δομών που υποστηρίζονται από εργαλεία ανάπτυξης μοντέλων εξόρυξης δεδομένων. Το μοντέλο που προκύπτει χρησιμοποιείτε κατά τα επόμενα στάδια.

*Πιστοποίηση γνώσης.* Η γνώση που αποκτήθηκε, παρουσιάζεται στους ειδικούς σε τυποποιημένη μορφή. Αναπτύσσονται σενάρια δοκιμής και γίνεται μια πρώτη επίδειξη στους χρήστες του συστήματος με βάση το πρωτότυπο μοντέλο. Αν εντοπιστούν αδυναμίες στο σύστημα, τότε ίσως θα πρέπει να επέμβουμε είτε με την διεξαγωγή συμπληρωματικών συναντήσεων της ομάδας, είτε με την εφαρμογή διαφορετικού σχεδιασμού της διαδικασίας και την χρησιμοποίηση διαφορετικών εργαλείων. Αν το μοντέλο είναι πλήρες, τότε, ο μηχανικός γνώσης είναι υπεύθυνος για την διατήρηση του συστήματος, την παροχή συμβουλών και την περαιτέρω εκπαίδευση των χρηστών.

### **3. ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ΓΙΑ ΜΑΡΚΕΤΙΝΓΚ, ΚΑΙ ΠΩΛΗΣΕΙΣ.**

#### **3.1. Εισαγωγή**

Γνωρίζουμε πόσο σημαντικό είναι να αρχίσουμε το μάρκετινγκ και τον στρατηγικό σχεδιασμό εξετάζοντας τα και από την εσωτερική και από την εξωτερική τους πλευρά. Το μάνατζμεντ πρέπει να παρακολουθεί τις μεγαλύτερες δυνάμεις στο περιβάλλον του μάρκετινγκ αν θέλει να είναι σύγχρονα τα προϊόντα του και οι μέθοδοι μάρκετινγκ που χρησιμοποιεί. Πώς, όμως, μπορεί το μάνατζμεντ να πληροφορηθεί τις μεταβαλλόμενες επιθυμίες των πελατών, τις νέες πρωτοβουλίες των ανταγωνιστών, τις νέες μεθόδους διανομής και ούτω καθεξής; Η απάντηση είναι σαφής: το μάνατζμεντ πρέπει να δημιουργήσει και να διατηρήσει ένα σύστημα πληροφοριών μάρκετινγκ και να έχει τη δυνατότητα να πραγματοποιήσει μια έγκυρη έρευνα μάρκετινγκ.

Στη μακρόχρονη ιστορία των επιχειρήσεων, το μάνατζμεντ των επιχειρήσεων αφιέρωνε το μεγαλύτερο μέρος της προσοχής του στο μάνατζμεντ του χρήματος, των υλικών, των μηχανημάτων και των εργαζομένων. Το μάνατζμεντ έχει δώσει λιγότερη προσοχή στον πέμπτο κρίσιμο πόρο της επιχείρησης: τις πληροφορίες. Είναι δύσκολο να βρει κανείς μάνατζερ εταιριών, οι οποίοι να είναι πολύ ικανοποιημένοι με τις πληροφορίες μάρκετινγκ που διαθέτουν. Μερικά από τα παράπονα τους είναι: ότι δεν γνωρίζουν σε ποιο τμήμα της εταιρίας βρίσκονται οι πληροφορίες, ότι ή συγκεντρώνουν πάρα πολλές πληροφορίες που δεν μπορούν να χρησιμοποιήσουν ή πολύ πιο λίγες απ' αυτές που πραγματικά χρειάζονται, ότι τις σημαντικές πληροφορίες τις παίρνουν πολύ αργά και τέλος ότι έχουν αμφιβολίες για την ακρίβεια των πληροφοριών. Πολλές εταιρίες δεν έχουν ακόμα προσαρμοσθεί στις σημαντικές απαιτήσεις για πληροφόρηση για την άσκηση αποτελεσματικού μάρκετινγκ στη δεκαετία του '90. Τρεις εξελίξεις δημιουργούν την ανάγκη ύπαρξης πιο εκτεταμένων πληροφοριών του μάρκετινγκ περισσότερο από κάθε άλλη φορά στο παρελθόν:

Απ' το τοπικό στο εθνικό και στο παγκόσμιο μάρκετινγκ. Καθώς οι εταιρίες επεκτείνουν τη γεωγραφική κάλυψη των αγορών, οι μάνατζερ τους χρειάζονται περισσότερες πληροφορίες για την αγορά από ποτέ άλλοτε.

Απ' τις ανάγκες του καταναλωτή στις επιθυμίες του καταναλωτή. Καθώς το εισόδημα των καταναλωτών αυξάνεται, γίνονται πιο εκλεκτικοί όταν επιλέγουν προϊόντα. Οι επιχειρηματίες διαπιστώνουν ότι είναι δύσκολο να προβλέψουν την ανταπόκριση των αγοραστών προς τα διαφορετικά γνωρίσματα, το στυλ και άλλα χαρακτηριστικά, εκτός αν στραφούν προς την έρευνα μάρκετινγκ.

Απ' τον ανταγωνισμό σε επίπεδο τιμής, στον ανταγωνισμό που δεν στηρίζεται στις τιμές. Καθώς οι επιχειρηματίες αυξάνουν τη χρήση της εμπορικής μάρκας, τη διαφοροποίηση του προϊόντος, τη διαφήμιση και την προώθηση των πωλήσεων, χρειάζονται πληροφορίες για την αποτελεσματικότητα αυτών των εργαλείων του μάρκετινγκ.

Οι εκρηκτικές ανάγκες για πληροφόρηση έχουν αντιμετωπιστεί από την πλευρά της προσφοράς με εντυπωσιακές νέες τεχνολογίες πληροφόρησης. Στα τελευταία τριάντα χρόνια έκαναν την εμφάνιση τους οι ηλεκτρονικοί υπολογιστές, τα μικροφίλμ, η καλωδιακή τηλεόραση, τα φωτοαντιγραφικά μηχανήματα, τα φαξ, τα κασετόφωνα, τα βίντεο, τα μηχανήματα βιντεοδίσκων και άλλες συσκευές οι οποίες εισήγαγαν επαναστατικές μεθόδους διαχείρισης των πληροφοριών. Παρόλα αυτά από τις περισσότερες επιχειρήσεις λείπουν σύγχρονες μέθοδοι διαχείρισης πληροφοριών. Πολλές επιχειρήσεις δεν διαθέτουν τμήμα ερευνών μάρκετινγκ. Άλλες διαθέτουν μικρό τμήμα ερευνών μάρκετινγκ και του οποίου οι δραστηριότητες περιορίζονται σε προβλέψεις ρουτίνας, ανάλυση των πωλήσεων και περιστασιακές έρευνες. Μερικές μόνο επιχειρήσεις έχουν αναπτύξει προηγμένα συστήματα πληροφοριών μάρκετινγκ που εξασφαλίζουν στην εταιρεία σύγχρονες πληροφορίες μάρκετινγκ και ανάλυση.

### **3.2. Έννοια και συστατικά ενός συστήματος πληροφοριών μάρκετινγκ**

Κάθε επιχείρηση πρέπει να οργανώνει τη ροή των πληροφοριών μάρκετινγκ προς τους μάρκετινγκ μάνατζερ που διαθέτει. Οι εταιρίες μελετούν τις ανάγκες για πληροφόρηση των μάνατζερ τους και σχεδιάζουν τα συστήματα πληροφοριών μάρκετινγκ για να καλύψουν αυτές τις ανάγκες. Ο ορισμός ενός συστήματος πληροφοριών μάρκετινγκ είναι ο εξής: «Ένα σύστημα πληροφοριών μάρκετινγκ αποτελείται από άτομα, εξοπλισμό και διαδικασίες που συγκεντρώνουν, ταξινομούν, αναλύουν, αξιολογούν και διανέμουν απαραίτητες, έγκαιρες και ακριβείς πληροφορίες στους διαμορφωτές αποφάσεων μάρκετινγκ.»

Η έννοια του συστήματος πληροφοριών μάρκετινγκ εξηγείται με τον ακόλουθο τρόπο. Οι μάρκετινγκ μάνατζερ, προκειμένου να κάνουν την ανάλυση, τον σχεδιασμό, την υλοποίηση και τον έλεγχο που εμπίπτουν στις υποχρεώσεις τους, χρειάζονται πληροφορίες για τις εξελίξεις στο περιβάλλον μάρκετινγκ. Ο ρόλος του συστήματος πληροφοριών μάρκετινγκ είναι να εκτιμήσει τις ανάγκες του μάνατζερ για πληροφόρηση, να εξασφαλίσει τις απαραίτητες πληροφορίες και να τις διανείμει έγκαιρα στους μάρκετινγκ μάνατζερ. Οι απαραίτητες πληροφορίες εξασφαλίζονται μέσω των εσωτερικών αρχείων μιας εταιρίας (αποθήκες δεδομένων), μέσω των μεθόδων πληροφόρησης μάρκετινγκ, μέσω της έρευνας μάρκετινγκ και μέσω της ανάλυσης του συστήματος στήριξης των αποφάσεων μάρκετινγκ (Data mining).

### **3.3. Σύστημα ανίχνευσης πληροφοριών μάρκετινγκ**

Ενώ το σύστημα των εσωτερικών αρχείων παρέχει στους μάνατζερ στοιχεία για τα αποτελέσματα, το σύστημα ανίχνευσης πληροφοριών μάρκετινγκ παρέχει στους μάνατζερ στοιχεία για τα γεγονότα. Ορίζουμε το σύστημα ανίχνευσης πληροφοριών μάρκετινγκ ως εξής:

Το σύστημα ανίχνευσης πληροφοριών μάρκετινγκ είναι ένα σύνολο διαδικασιών και πηγών που χρησιμοποιούν οι μάνατζερ για να εξασφαλίζουν την

καθημερινή τους πληροφόρηση σχετικά με εξέχοντος ενδιαφέροντος εξελίξεις στο περιβάλλον μάρκετινγκ.

Οι μάνατζερ ανιχνεύουν το περιβάλλον με τέσσερις τρόπους:

Μη Κατευθυνόμενη παρατήρηση. Γενική έκθεση σε πληροφορίες, όπου ο μάνατζερ δεν έχει κάποιο συγκεκριμένο σκοπό στο μυαλό του.

Κατευθυνόμενη παρατήρηση. Κατευθυνόμενη έκθεση, που δεν περιλαμβάνει ενεργό διερεύνηση, σε μια περισσότερη ή λιγότερο σαφώς καθορισμένη περιοχή ή είδος πληροφοριών.

Άτυπη διερεύνηση. Μια σχετικά περιορισμένη και μη σχεδιασμένη προσπάθεια για την απόκτηση συγκεκριμένης πληροφορίας ή πληροφοριών για ένα ειδικό σκοπό.

Τυπική διερεύνηση. Μια σκόπιμη προσπάθεια - συνήθως μετά από ένα προκαθορισμένο σχέδιο, διαδικασία ή μεθοδολογία- για την απόκτηση συγκεκριμένης πληροφορίας ή πληροφοριών που αφορούν ένα ειδικό θέμα.

Οι μάρκετινγκ μάνατζερ συλλέγουν και μόνοι τους πληροφορίες με την ανάγνωση βιβλίων, εφημερίδων και εμπορικών δημοσιευμάτων, συζητώντας με πελάτες, προμηθευτές, αντιπροσώπους, τρίτα πρόσωπα ή συζητώντας με άλλους μάνατζερ και στελέχη της εταιρίας τους. Όμως το σύστημα αυτό είναι μάλλον πρόχειρο και είναι πιθανό κάποιες πολύτιμες πληροφορίες να χαθούν ή να φτάσουν πολύ αργά. Οι μάνατζερ μπορεί να πληροφορηθούν μια κίνηση των ανταγωνιστών, μια καινούρια ανάγκη των καταναλωτών ή ένα πρόβλημα με κάποιον προμηθευτή πολύ αργά με αποτέλεσμα η ανταπόκριση τους να μην είναι η καλύτερη δυνατή.

Οι καλά διοικούμενες εταιρίες κάνουν κάποια επιπρόσθετα βήματα για να βελτιώσουν την ποιότητα και την ποσότητα των πληροφοριών μάρκετινγκ. Εκπαιδεύουν και παρωθούν τη δύναμη των πωλήσεων να παρατηρεί και να

αναφέρει τις νέες εξελίξεις. Η εταιρία ωθεί αντιπροσώπους, λιανοπωλητές και άλλους ενδιαμέσους φορείς να μεταδίδουν σημαντικές πληροφορίες. Η εταιρία αγοράζει πληροφορίες από εξωτερικούς προμηθευτές. Μερικές εταιρίες έχουν δημιουργήσει ένα εσωτερικό κέντρο πληροφοριών μάρκετινγκ.

Οι πληροφορίες μάρκετινγκ είναι κρίσιμης σημασίας στοιχείο για το αποτελεσματικό μάρκετινγκ. Σαν αποτέλεσμα της τάσης προς το εθνικό και το διεθνές μάρκετινγκ, τη μετάβαση από τις ανάγκες του αγοραστή, στις επιθυμίες του αγοραστή και τη μετάβαση από τον ανταγωνισμό με βάση τις τιμές, στον ανταγωνισμό που δεν γίνεται με βάση τις τιμές. Όλες οι επιχειρήσεις διαθέτουν ένα σύστημα πληροφοριών μάρκετινγκ που συνδέει το εξωτερικό περιβάλλον με τους μάνατζερ του, αλλά τα συστήματα διαφέρουν σημαντικά από άποψη τεχνολογίας. Σε πάρα πολλές περιπτώσεις ή δεν υπάρχουν πληροφορίες ή δεν μπορεί κανείς να τις εμπιστευθεί. Κάθε μέρα όλο και περισσότερες εταιρίες παίρνουν μέτρα για να βελτιώσουν το σύστημα πληροφοριών μάρκετινγκ που διαθέτουν.

Ένα καλά σχεδιασμένο σύστημα πληροφοριών της αγοράς αποτελείται από τέσσερα υποσυστήματα. Το πρώτο είναι το σύστημα των εσωτερικών αρχείων, το οποίο παρέχει τρέχοντα στοιχεία για τις πωλήσεις, το κόστος, τα αποθέματα, τη ροή ρευστού χρήματος, καθώς και τα γραμμάτια εισπρακτέα και πληρωτέα. Πολλές εταιρίες έχουν αναπτύξει συστήματα εσωτερικών αρχείων που βασίζονται σε προηγμένης τεχνολογίας Η/Υ που παρέχουν πιο γρήγορες και πιο κατανοητές πληροφορίες.

Το δεύτερο είναι το σύστημα πληροφοριών μάρκετινγκ, το οποίο παρέχει καθημερινά στους μάρκετινγκ μάνατζερ πληροφορίες σχετικά με τις εξελίξεις στο εξωτερικό περιβάλλον μάρκετινγκ. Εδώ μια καλά εκπαιδευμένη δύναμη πωλήσεων αγοράζει στοιχεία από ειδικές πηγές και ένα γραφείο συλλογής πληροφοριών μπορεί να βελτιώσει τις πληροφορίες μάρκετινγκ που έχουν στη διάθεση τους οι μάρκετινγκ μάνατζερ της εταιρίας.

Το τρίτο σύστημα είναι η έρευνα μάρκετινγκ, η οποία περιλαμβάνει συλλογή πληροφοριών που έχουν σχέση με ένα συγκεκριμένο πρόβλημα μάρκετινγκ που αντιμετωπίζει η εταιρία. Η διαδικασία της έρευνας μάρκετινγκ αποτελείται από πέντε στάδια: προσδιορισμός του προβλήματος και αντικειμενικοί στόχοι της έρευνας, κατάρτωση σχεδίου της έρευνας, συλλογή των πληροφοριών, ανάλυση των πληροφοριών (Data Mining) και παρουσίαση των ευρημάτων. Χαρακτηριστικά της καλής έρευνας μάρκετινγκ είναι η επιστημονική μέθοδος, η δημιουργικότητα, οι πολλαπλές μεθοδολογίες, η δημιουργία μοντέλων και τα μέτρα κόστους ή οφέλους της αξίας των πληροφοριών.

Το τέταρτο σύστημα είναι το σύστημα υποστήριξης των αποφάσεων μάρκετινγκ, το οποίο αποτελείται από στατιστικές τεχνικές και μοντέλα αποφάσεων που βοηθούν τους μάρκετινγκ μάνατζερ στην πραγματοποίηση καλύτερων αναλύσεων και αποφάσεων.

### **3.4. Η χρήση της έρευνας στο χώρο του Μάρκετινγκ**

Η έρευνα του μάρκετινγκ έχει σκοπό να βοηθήσει τη διοίκηση της επιχείρησης να καταλάβει τη φύση της αγοράς στην οποία προσπαθεί να διαθέσει επικερδώς τα προϊόντα της, καθώς επίσης τους προμηθευτές της, τους χονδρεμπόρους της, τους λιανοπωλητές της, τους ανταγωνιστές και τέλος στους πελάτες της.

Για να επιτύχει μια επιχείρηση τους οικονομικούς της στόχους, είναι απαραίτητο η διοίκηση της να έχει μια συνεχή ροή πληροφοριών, που θα τη βοηθούν να λαμβάνει σωστότερες αποφάσεις. Με την έρευνα μάρκετινγκ οι πληροφορίες αυτές μπορούν χωρίς άλλο να εξασφαλιστούν και να αξιοποιηθούν κατάλληλα. Απαραίτητη, όμως, για αυτό προϋπόθεση είναι η οργάνωση και λειτουργία ενός συστήματος πληροφόρησης μάρκετινγκ, μέσω του οποίου θα εξασφαλίζονται στη διοίκηση της επιχείρησης οι ζητούμενες εκ μέρους της πληροφορίες, από εσωτερικές και εξωτερικές πηγές, με τη βοήθεια πρωτογενούς ή και δευτερογενούς έρευνας.



Για να είναι αποτελεσματική η έρευνα που διεξάγει μια επιχείρηση ή ένας οργανισμός διεξαγωγής ερευνών για λογαριασμό της, πρέπει απαραίτητα να προγραμματιστεί κατάλληλα, καθώς επίσης να επιλεγεί η πιο ενδεδειγμένη μεθοδολογία έρευνας που θα ακολουθηθεί προς το σκοπό αυτό. Συμπερασματικά, μπορούμε να πούμε ότι η έρευνα μάρκετινγκ μπορεί οπωσδήποτε να συμβάλει στην ποιοτική βελτίωση των αποφάσεων που λαμβάνει η διοίκηση της επιχείρησης και κατ' επέκταση στην αύξηση της αποτελεσματικότητας τους, πλην όμως σε καμία περίπτωση δεν μπορεί να την υποκαταστήσει.

### **3.5. Μια μικρή αναφορά στην επιστήμη της έρευνας μάρκετινγκ**

Έρευνα μάρκετινγκ είναι κάθε προγραμματισμένη και οργανωμένη προσπάθεια για την συγκέντρωση νέων στοιχείων και απόκτηση νέας γνώσης που θα βοηθήσουν στην λήψη καλύτερων αποφάσεων μάρκετινγκ. Αν και δεν λέγεται έτσι, οι άνθρωποι των πωλήσεων και του μάρκετινγκ πάντοτε πάνω στην δουλειά τους κάνουν ανεπίσημη έρευνα μάρκετινγκ γιατί όλοι γνωρίζουν πόσο σπουδαίο είναι να ξέρουν όσο το δυνατόν περισσότερα πράγματα για τους πελάτες τους, τις διαθέσεις τους και τις αγοραστικές τους συνήθειες. Η διαδικασία απόκτησης αυτών των γνώσεων αγγίζει τα όρια της πραγματικής έρευνας μάρκετινγκ. Παρ' όλ' αυτά, πολλές από τις επιχειρηματικές γνώσεις μπορούμε τώρα να τις αποκτήσουμε με ένα συστηματικότερο τρόπο.

Οργανώνοντας την έρευνα για την αναζήτηση στοιχείων, θέτοντας ξεκάθαρους και συγκεκριμένους σκοπούς, δίνοντας προσοχή στην συγκέντρωση πλήρων και αμερόληπτων στοιχείων και βάζοντας όλα αυτά τα στοιχεία μαζί, για την επιλογή του πιο επικερδούς τρόπου δράσης, δίνει τουλάχιστον την δυνατότητα για μια ημιεπιστημονική μέθοδο που στο παρελθόν γινόταν με τυχαίο τρόπο. Με το πέρασμα του χρόνου έχει αναπτυχθεί ένας αριθμός επίσημων μεθόδων συλλογής γνώσεων αγοράς, οι οποίες δεν είναι τίποτε περισσότερο από «εργαλεία» που βοηθούν τους επιχειρηματίες αν κάνουν

καλύτερη δουλειά (βλέπε Data mining). Ναι μεν μερικές από αυτές τις τεχνικές μεθόδους (εργαλεία) δεν ανταποκρίνονται λόγω της πολυπλοκότητας των στις δυνατότητες του τυπικού μικρού επιχειρηματία αλλά το κύριο μέρος των είναι απλές και εύχρηστες. Οι επιχειρηματικές δραστηριότητες πλησιάζουν σ' ένα τέτοιο εξελικτικό σημείο όπου μπορούν να αποφευχθούν άσκοπες δαπάνες χάρη στην έρευνα μάρικετινγκ.

## 4. ΠΡΟΓΡΑΜΜΑΤΑ ΝΟΗΜΟΝΩΝ ΠΡΟΣΕΓΓΙΣΕΩΝ (PC TOOLS)

### 4.1. Το πρόγραμμα εύρεσης ανωμαλιών δεδομένων GRITBOT

Το Gritbot είναι ένα εργαλείο εύρεσης ανωμαλιών σε δεδομένα προερχόμενα από σύνολο αρχείων με πολλές παραμέτρους και μπορεί να θεωρηθεί σαν ένας αυτόνομος ελεγκτής ποιότητας δεδομένων που βρίσκει τιμές εκτός κλίμακας της εξεταζόμενης περίπτωσης σε αρχεία διακριτών (discrete) και συνεχών (continuous) μεταβλητών και παραμέτρων. Αν μια τιμή μιας παραμέτρου είναι εκτός κλίμακας τιμών (σε σύγκριση με αντίστοιχες τιμές της ίδιας παραμέτρου των λοιπών αρχείων επεξεργασίας), τότε η τιμή αυτή μπορεί να είναι πιθανή ανωμαλία. Κάθε εφαρμογή του Gritbot αποτελείται από text files που καθορίζουν τις παραμέτρους και περιγράφουν την προέλευσή τους. Τα δεδομένα των αρχείων αρχικά μορφοποιούνται στην μορφή εισόδου και στην συνέχεια επιλέγονται οι παράμετροι που θα ελεγχθούν. Αποφασίζεται το επίπεδο ελέγχου των δεδομένων, και ακολουθεί επεξεργασία και εμφάνιση των ανωμαλιών που ευρέθησαν.

#### 4.1.1 Προετοιμασία δεδομένων εισόδου στο Gritbot

Για να γίνει ευκολότερα κατανοητή η μέθοδος προετοιμασίας και επεξεργασίας των δεδομένων Gritbot θα χρησιμοποιηθεί μια ιατρική εφαρμογή μια βάση δεδομένων εξέτασης θυρεοειδούς από το Gravan Institute of Medical Research, Sydney. Κάθε περίπτωση ασθενή περιέχει πληροφορίες όπως αποτελέσματα εξετάσεων, στοιχεία ασθενή, διαγνωστικά κ.τ.λ.:

Attribute	Case 1	Case 2
Age	41	23
Sex	F	F
On thyroxin	F	F
Query on thyroxin	F	F
On antithyroid medication	F	F
Sick	F	F
Pregnant	F	F
Thyroid surgery	F	F

I131 treatment	F	F
Query hypothyroid	F	F
Query hyperthyroid	F	F
Lithium	F	F
Tumor	F	F
Goiter	F	F
Hypopituitary	F	F
Phych	F	F
TSH	1.3	4.1
T3	2.5	2
TT4	125	102
T4U	1.14	Unknown
FTI	109	Unknown
Referral source	SVHC	Other
Diagnosis	Negative	Negative
ID	3733	1442

#### 4.1.2 Αρχεία εφαρμογών

Κάθε αρχείο εφαρμογής του Gritbot ονομάζεται file stem όπως για παράδειγμα το παραπάνω file stem hypothyroid. Το αρχείο απαιτεί συγκεκριμένο format εισόδου :

##### Αρχεία ονομάτων

Τα αρχεία ονομάτων names file (hypothyroid.names), περιγράφουν τον τύπο των παραμέτρων και ορίζουν ποιες θα εξεταστούν για ανωμαλίες.

Attributes included:

Attributes excluded:

Υπάρχουν δυο βασικοί τύποι παραμέτρων:

- Διακριτές και συνεχείς παράμετροι.
- Τιμές παραμέτρων ορισμένες ρητά και άρρητα.

Το αρχείο hypothyroid.names έχει την εξής μορφή :

Attributes	Diagnosis.Target attribute for See5/C5.0
Age	Continuous
Sex	M, F
On thyroxin	F, t
Query on thyroxin	F, t
On antithyroid medication	F, t
Sick	F, t
Pregnant	F, t
Thyroid surgery	F, t
I131 treatment	F, t
Query hypothyroid	F, t
Query hyperthyroid	F, t
Lithium	F, t
Tumor	F, t
Goiter	F, t
Hypopituitary	F, t
Phych	F, t
TSH	Continuous
T3	Continuous
TT4	Continuous
T4U	Continuous
FTI	TT4 / T4U
Referral source	WEST, STMW, SVHC, SVI, SVHD, other.
Diagnosis	Primary, compensated, secondary, negative.

### Αρχεία δεδομένων

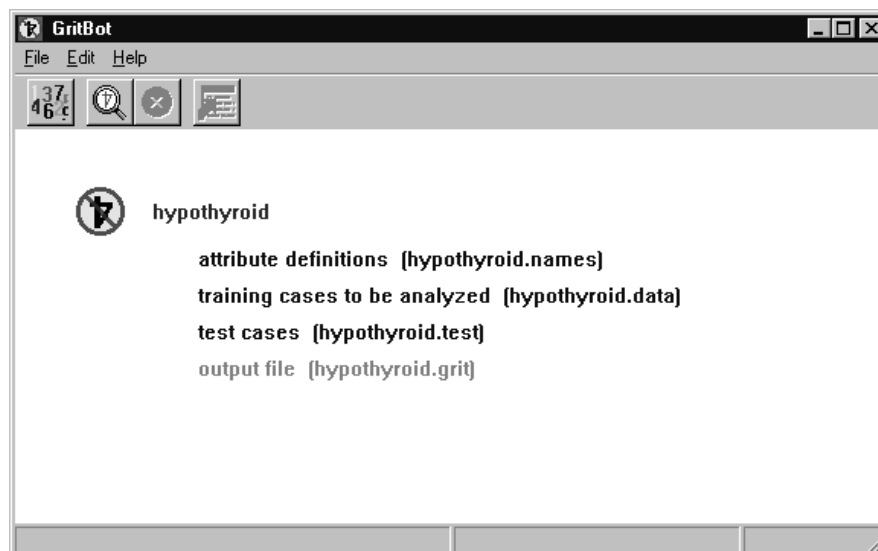
Ο δεύτερος σημαντικός τύπος αρχείων στην εφαρμογή είναι τα data files (hypothyroid.data), τα οποία παρέχουν πληροφορίες για τις περιπτώσεις που θα ελέγξει το Gritbot.

Οι δυο cases από το αρχείο hypothyroid.data είναι:

41, F, f, f, f, f, f, f, f, f, f, f, f, f, f, f, f, 1.3, 2.5, 125, 1.14, SVHC, negative, 3733.  
23, F, f, f, f, f, f, f, f, f, f, f, f, f, f, f, f, 4.1, 2, 1 02, ?, other, negative, 1442.

### 4.1.3 Περιβάλλον χρήσης

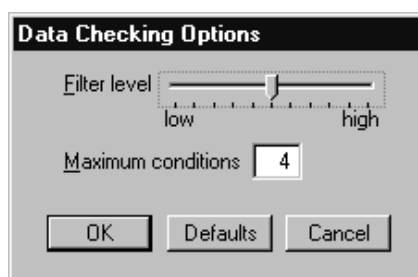
Το κύριο παράθυρο του προγράμματος είναι το εξής:



Παρέχονται οι λειτουργίες Locate Data, Check Data, Stop, Review Output

### 4.1.4 Ελέγχοντας τα δεδομένα για ανωμαλίες

Όταν ετοιμαστούν τα names.data ακολουθεί η επεξεργασία από το Gritbot. Η ανάλυση του Gritbot επηρεάζεται από δυο παραμέτρους. Από το Check Data εμφανίζεται ένα παράθυρο διαλόγου για την επιλογή των εξής τιμών :



Χρησιμοποιώντας τις αρχικές ρυθμίσεις του προγράμματος εξάγονται τα εξής αποτελέσματα:

```

Gritbot [Release 1. 03] Tue Sep 12:25:05 2000
Read 2772 cases (24 attributes) from hypothyroid.data
Read 1 000 cases from hypothyroid.test
while checking age: excluding 1 missing value
while checking sex: excluding 150 missing values
while checking TSH: excluding 369 missing values excluding high
tail (200 cases above 11.4)
while checking T3: excluding 769 missing values
while checking TT4: excluding 231 missing values
while checking T4U: excluding 387 missing values
while checking FTI: excluding 389 missing values
5 possible anomalies identified

data case 1365: (label 861)[0.002]
age = 455 (3771 cases, mean 52, 99.97%<=94)
data case 2618: (label 2079)[0.006]
TSH = 1.1 (45 cases, mean 0.068,98%<=0.25)
data case 373: (label 769)[0.013]
T3 = 7.6 (35 cases, mean 2.06, 97%<=3.7) TSH<=0.16[0.15]
TT4<=155[120] T4U>0.9 and <=1.2[1.04] referral source in {SVI,
SVHD} [SVI]
data case 2224: (label 1562)[0.014]
age = 75 (53 cases, mean 32, 96%<= 42) pregnant = t
data case 1610: (label 3023)[0.016]
age = 73 (53 cases, mean 32, 96%<=42) pregnant = t
Time: 4.1 sec.

```

#### 4.1.5 Συμπεράσματα για το πρόγραμμα Gritbot

Είναι λοιπόν φανερό πως το πρόγραμμα αυτό είναι ένα πολύ χρήσιμο εργαλείο το οποίο βελτιώνει την ποιότητα των δεδομένων πριν αυτά να υποστούν την επεξεργασία για την οποία προορίζονται. Σημαντικό είναι το γεγονός πως σε ελάχιστο χρόνο ελέγχει σε ικανοποιητικό βαθμό τα δεδομένα εξοικονομώντας με αυτόν τον τρόπο το χρόνο που θα απαιτούνταν αν γινόταν ανθρώπινος έλεγχος. Κατά την ανάλυση χρησιμοποιούνται από το πρόγραμμα αρκετά heuristics ή rules με σκοπό να προσδιορίσει τις ανωμαλίες στις τιμές.

Το πρόγραμμα Gritbot αφενός είναι σαφές πως δεν θα ανακαλύψει όλες τις ανωμαλίες σε μια βάση δεδομένων και αφετέρου το αποτέλεσμα είναι πιθανές ανωμαλίες και όχι σίγουρες ανωμαλίες και αυτό με βάση την ευαισθησία και τους κανόνες που έχουν ορισθεί. Αφού εντοπιστεί μια πιθανή ανωμαλία δεδομένων, πρέπει να βρεθεί και η αιτία που την προκάλεσε. Πρέπει δηλαδή να διασαφηνιστεί αν πρόκειται για απλή λανθασμένη εγγραφή της, λανθασμένη μέτρηση, λανθασμένη μεταφορά της στο πρόγραμμα, κ.τ.λ. Με αυτό τον τρόπο καταγράφονται τα προβλήματα στην συλλογή των δεδομένων και εξάγονται

χρήσιμα συμπεράσματα που βοηθούν στην βελτίωση της μελλοντικής τους ποιότητας.

Πρέπει να σημειωθεί πως το Gritbot είναι ένα πρόγραμμα που αναπτύχθηκε από την Rulequest Research © 1999-2000.

#### **4.2. Το πρόγραμμα Magnum Opus**

Το Magnum Opus σχεδιάστηκε με κύρια λειτουργία του το να βρίσκει κανόνες συσχέτισης (association rules) των δεδομένων. Οι κανόνες συσχέτισης έχουν ανταγωνιστικό πλεονέκτημα αφού μπορούν να προσδιορίσουν τις αλληλεπιδράσεις των παραγόντων των δεδομένων. Συχνά χρησιμοποιούνται για τον προσδιορισμό των αιτιών συμπεριφορών όπως για παράδειγμα το πώς η αγορά ενός προϊόντος επηρεάζει την αγορά άλλων προϊόντων.

Το Magnum Opus έχει σχεδιαστεί για να χειρίζεται μεγάλες ομάδες δεδομένων που περιλαμβάνουν χιλιάδες καταγραφές, πράγμα που όμως εξαρτάται και από τη μνήμη του υπολογιστή. Ο χρήστης έχει την δυνατότητα να επιλέξει ανάμεσα σε δυο κριτήρια σημαντικότητας μιας συσχέτισης, το leverage και το lift, δίνοντας έτσι εναλλακτικές αναλύσεις των ιδίων δεδομένων. Έτσι αντίθετα με τα περισσότερα λογισμικά τα οποία είναι σχεδιασμένα για την διαδικασία επεξεργασίας δεδομένων που είναι σποραδικά, το Magnum Opus δεν βασίζεται σε σποραδικά δεδομένα με αποτέλεσμα να αυξάνεται η χρησιμότητά του για την αναζήτηση μιας ποικιλίας δεδομένων και ιδιαίτερα για τον προσδιορισμό απροσδόκητων αλληλεπιδράσεων μέσα στα δεδομένα. Τέλος πρέπει να αναφέρουμε πως το Magnum Opus επιτρέπει στο χρήστη να καθορίσει τον μέγιστο αριθμό των κανόνων συσχέτισης που πρόκειται να μελετηθούν. Μέσα σε αυτούς τους περιορισμούς το πρόγραμμα βρίσκει τις μη ασήμαντες συσχετίσεις (non trivial associations) με την μεγαλύτερη σημασία μέσα σε καθορισμένα όρια σημαντικότητας. Αν βέβαια ο χρήστης τερματίσει την αναζήτηση ή αν δεν υπάρχουν λιγότερες από τον καθορισμένο αριθμό μη ασήμαντες συσχετίσεις το πρόγραμμα θα βρει λιγότερους από τον καθορισμένο αριθμό κανόνες συσχέτισης.



#### 4.2.1 Οι δυο τύποι δεδομένων που είναι αποδεκτοί από το *Magnum Opus*

Έχει ήδη αναφερθεί πως το *Magnum Opus* μπορεί να ανταποκριθεί σε δυο τύπους δεδομένων, στα *basket data* και στα *attribute-value data*. Τα *basket data* αποτελούνται από μια συλλογή δεδομένων, όπως για παράδειγμα τα δεδομένα από την αγορά, όπου το κάθε καλάθι είναι μια συλλογή αντικειμένων που αγοράζονται από ένα πελάτη σε μια συναλλαγή. Τα *attribute-value data* είναι μια συλλογή από περιπτώσεις η καθεμία από τις οποίες περιγράφεται από έναν αριθμό χαρακτηριστικών. Η κάθε περίπτωση αποτελεί μια μονή αξία για κάθε χαρακτηριστικό, και τα χαρακτηριστικά μπορεί να είναι ονομαστικά ή αριθμητικά. Ένα χαρακτηριστικό παράδειγμα τέτοιων δεδομένων είναι η βάση δεδομένων ενός πελάτη, όπου οι περιπτώσεις είναι οι πελάτες και τα χαρακτηριστικά τους μπορεί να περιλαμβάνουν το ποσό που ξοδεύουν σε κάθε κατάστημα, πληροφορίες για την σε κάθε περίπτωση συμπεριφορά κ.τ.λ.

#### 4.2.2 Τα αρχεία εισόδου.

Τα *basket data* μπορούν να εισαχθούν από δυο διαφορετικούς τύπους αρχείων. Τα *identifier item files* και τα *item list files*.

Η πρώτη κατηγορία αποτελεί ουσιαστικά ένα *text file* το οποίο κατηγοριοποιεί τα καλάθια για να αναλυθούν σε μια φόρμα τύπου *identifier-item*. Κάθε καλάθι έχει ένα μοναδικό *identifier* και κάθε γραμμή συσχετίζει τον *identifier* του ενός καλαθιού με ένα *item* σε αυτό το καλάθι. Στο ακόλουθο παράδειγμα βλέπουμε το περιεχόμενο ενός *identifier item file* το οποίο περιλαμβάνει δυο στήλες από τις οποίες η πρώτη περιλαμβάνει τον *identifier* και η δεύτερη το *item*.

```
001, apples
001, oranges
001, bananas
002, apples
002, carrots
002, lettuce
002, tomatoes
```

Το καλάθι 001 περιλαμβάνει μήλα, πορτοκάλια και μπανάνες και το καλάθι 002 περιλαμβάνει μήλα, καρότα, μαρούλια και ντομάτες. Ένα αρχείο item list είναι και αυτό text file που κατηγοριοποιεί όμως τα καλάθια για να αναλύσει ένα σε κάθε γραμμή:

```
apples, oranges, bananas  
apples, carrots, lettuce, tomatoes
```

Το παραπάνω αρχείο είναι ένα item list file το οποίο περιέχει δυο καλάθια.

Οι δυο τύποι αρχείων που επεξεργάζεται η ανάλυση χαρακτηριστικών (attribute-value) είναι τα names file και τα data files. Ένα name file είναι ένα text file που κατηγοριοποιεί τα χαρακτηριστικά και καταγράφει τις περιπτώσεις στο αρχείο δεδομένων (data file) για να αναλυθούν. Κάθε χαρακτηριστικό περιγράφεται σε μια ξεχωριστή γραμμή ενώ κάθε γραμμή αρχίζει με το όνομα του χαρακτηριστικού.

Contents of file: tutorial.nam

```
Profitability99: numeric 3  
Profitability98: numeric 3  
Spend99: numeric 3  
Spend98: numeric 3  
NoVisits99: numeric 3  
NoVisits98: numeric 3  
Dairy: numeric 3  
Deli: numeric 3  
Grocery: numeric 3  
SocioEconomicGroup: A, B, C, D1, D2, E  
Promotion1: t, f  
Promotion2: t, f
```

Τα αριθμητικά χαρακτηριστικά πρέπει να χωριστούν σε υποπεριοχές, οι οποίες πρέπει να καθορίζονται στα αρχεία ονομάτων (name files). Εναλλακτικά τα αρχεία ονομάτων μπορούν να αναγνωρίζουν απλά τον αριθμό των υποπεριοχών και το πρόγραμμα θα επιλέγει αυτόματα τις υποπεριοχές. Για κάθε ένα αριθμητικό χαρακτηριστικό με καθορισμένες υποπεριοχές, το όνομα του χαρακτηριστικού ακολουθείται από μια λίστα υποπεριοχών. Αυτό υποδεικνύει το πώς οι αριθμητικές αξίες για κάθε χαρακτηριστικό είναι χωρισμένες σε

υποπεριοχές. Κάθε σημείο διαχωρισμού φαίνεται με τις σχέσεις  $<$  ή  $\geq$  το οποίο ακολουθείται από μια αξία η οποία κλείνει την υποπεριοχή. Αν η σχέση είναι  $<$ , η υποπεριοχή περιλαμβάνει όλες τις αξίες που είναι μικρότερες από την καθορισμένη αξία, ενώ αν η σχέση είναι  $\geq$ , η υποπεριοχή περιλαμβάνει όλες τις αξίες που είναι μεγαλύτερες ή ίσες με την καθορισμένη αξία. Για παράδειγμα:

```
Spend < 10 <= 100
```

Αυτό καθορίζει ότι το χαρακτηριστικό spend έχει τρεις υποπεριοχές, κάτω από τον πρώτο περιορισμό, ανάμεσα στους δυο περιορισμούς και μετά τον τελευταίο περιορισμό:

```
Spend < 10  
10 <= spend <= 100  
Spend > 100
```

Για την επιλογή υποπεριοχών στο Magnum Opus πρέπει να χρησιμοποιηθεί η λέξη numeric ακολουθούμενη από τον αριθμό των υποπεριοχών που απαιτούνται όπως για παράδειγμα:

```
Spend: numeric 5
```

Ο δεύτερος τύπος αρχείων είναι τα data file που και αυτά είναι τύπου text file και στην συγκεκριμένη περίπτωση κατηγοριοποιούν περιπτώσεις (cases) για να αναλυθούν στην ανάλυση attribute value. Οι περιπτώσεις μπαίνουν μία σε κάθε γραμμή και για κάθε περίπτωση η αξία πρέπει να προσδιοριστεί για κάθε χαρακτηριστικό. Οι attribute values για κάθε περίπτωση πρέπει να κατηγοριοποιηθεί ώστε να εμφανιστεί στο names file.

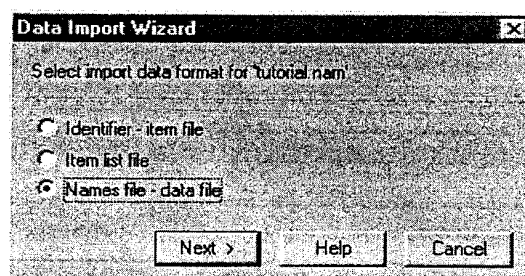
```
Head of file: tutorial
```

```
829, 789, 5250, 6560, 70, 82, 1074, 390, 078, 1995, C, f, f  
141, 118, 722, 928, 19, 16, 15, 155, 139, 404, C, f, f  
1044, 783, 3591, 4026, 63, 61, 81, 218, 232, 2908, 02, f, t  
78, 63, 331, 336, 7, 8, 54, 68, 63, 167, 01, t, f  
511, 419, 2142, 1947, 34, 33, 59, 106, 239, 1477, C, f, f  
987, 1402, 4032, 5376, 56, 64, 891, 681, 995, 1411, C, f, f  
313, 286, 1137, 1008, 22, 18, 153, 63, 146, 762, D1, t, f  
1800, 859, 7350, 3159, 75, 81, 441, 2315, 1433, 1837, D1, f, t  
226, 126, 1034, 612, 11, 6, 351, 377, 259, 196, C, f, f
```

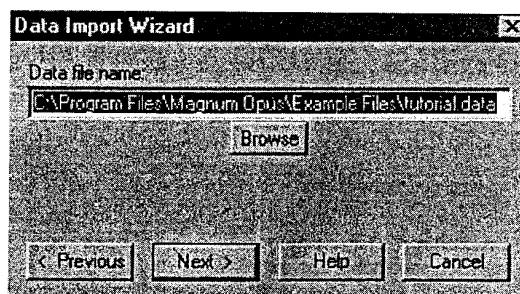
58, 28, 343, 140, 24, 14, 24, 18, 35, 248, A, t, f  
 1136, 597, 4602, 3068, 59, 59, 554, 870, 949, 2623, D1, f, f  
 376, 274, 1980, 1675, 22, 25, 356, 261, 344, 792, C, f, f  
 223, 172, 1656, 1400, 18, 14, 355, 430, 323, 579, C, f, f  
 1808, 976, 7600, 7396, 80, 86, 501, 718, 852, 5928, C, f, f  
 114, 180, 462, 1008, 14, 16, 4, 28, 27, 364, D2, f, f  
 1169, 1125, 4356, 3723, 45, 51, 359, 427, 134, 2187, D1, t, f  
 226, 235, 1230, 1575, 15, 15, 414, 284, 267, 418, D1, f, f  
 493, 189, 2408, 1035, 28, 23, 318, 503, 344, 1083, D1, f, f

#### 4.2.3 Πως γίνεται η εισαγωγή δεδομένων στο πρόγραμμα

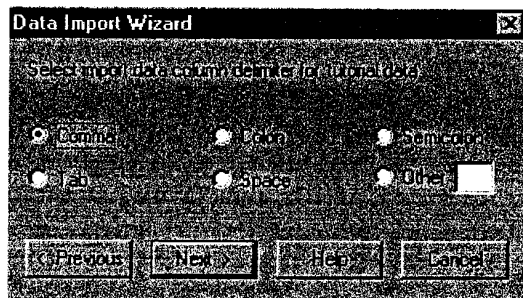
Για να εισάγουμε δεδομένα στο πρόγραμμα έτσι ώστε να τα επεξεργαστούμε και να προσδιορίσουμε τις συσχετίσεις επιλέγουμε από το toolbar την εντολή import data και στην συνέχεια επιλέγουμε το αρχείο που επιθυμούμε. Στο παράδειγμά μας είναι το αρχείο tutorial.nam από το example files. Στην συνέχεια το πρόγραμμα, μας ζητάει να επιλέξουμε τον τρόπο με τον οποίο θέλουμε να γίνει η επεξεργασία. Στο συγκεκριμένο παράδειγμα, από τα τρία αρχεία που υποστηρίζει το Magnum Opus για την εισαγωγή επεξεργασίας δεδομένων, χρησιμοποιούμε την εισαγωγή κατά names file-data format



Στην συνέχεια το Magnum Opus ζητάει να επιλεγεί το αρχείο των δεδομένων.

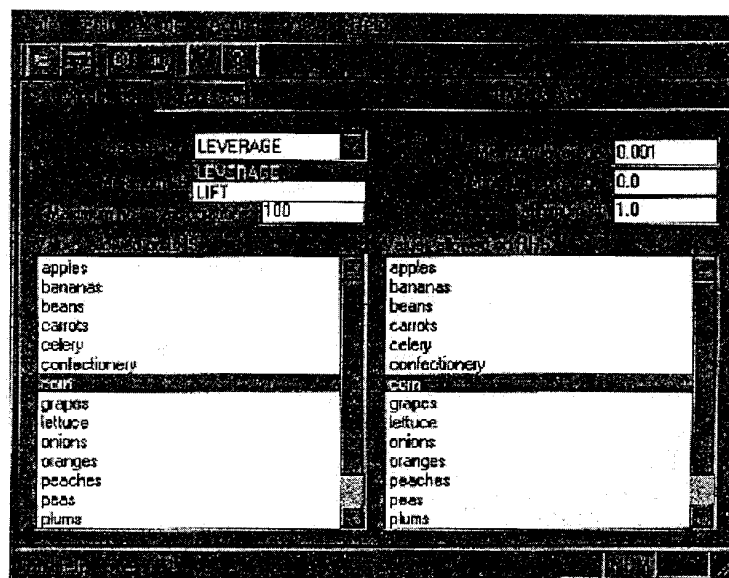


Το επόμενο βήμα είναι ο προσδιορισμός του τρόπου διαχωρισμού των πεδίων στο αρχείο δεδομένων. Στο παράδειγμά μας επιλέγουμε ο διαχωρισμός να γίνει με κόμμα.



#### 4.2.4 Το περιβάλλον του προγράμματος

Αφού έχουμε εισάγει τα δεδομένα με την προαναφερθείσα διαδικασία, μπορούμε να δούμε το κυρίως περιβάλλον όπως αυτό φαίνεται στην παρακάτω εικόνα.



Είναι φανερό πως το πρόγραμμα μας δίνει την δυνατότητα αναζήτησης είτε με το leverage είτε με το lift. Η διαφορά ανάμεσα στους δυο τρόπους έγκειται στο ότι στην αναζήτηση με το leverage το πρόγραμμα βρίσκει τους κανόνες συσχέτισης με το υψηλότερο leverage ενώ στην αναζήτηση με το lift το

πρόγραμμα βρίσκει τους κανόνες συσχέτισης με το υψηλότερο lift. Για να γίνει η αναζήτηση της συσχέτισης των δεδομένων πρέπει να τεθούν κάποιοι περιορισμοί όπως ο μέγιστος αριθμός των συσχετίσεων (maximum no of associations), το μέγιστο leverage, το ελάχιστο coverage και το ελάχιστο lift. Πριν όμως προχωρήσουμε στα αποτελέσματα είναι απαραίτητο να εξηγήσουμε τι περιλαμβάνει το leverage, το lift και το coverage.

Το leverage ενός κανόνα συσχέτισης είναι οι περιπτώσεις που καλύπτονται από την LHS και την RHS δια τις αναμενόμενες αν η LHS και η RHS ήταν ανεξάρτητες μεταξύ τους. Για παράδειγμα αν υπάρχουν 1000 περιπτώσεις, η LHS καλύπτει τις 200 περιπτώσεις, η RHS καλύπτει τις 100 περιπτώσεις και η RHS καλύπτει 50 από τις περιπτώσεις που καλύπτονται και από την LHS. Το ποσοστό των περιπτώσεων που καλύπτονται από την LHS και την RHS είναι  $50/1000=0.05$ , και το ποσοστό των περιπτώσεων που θα αναμενόταν να καλυφθεί και από τις δυο αν ήταν ανεξάρτητες μεταξύ τους είναι  $(200/1000)*(100/1000)=0.02$ . Το leverage είναι  $0.05-0.02=0.03$ . Ο συνολικός αριθμός των περιπτώσεων που αντιπροσωπεύει είναι 30.

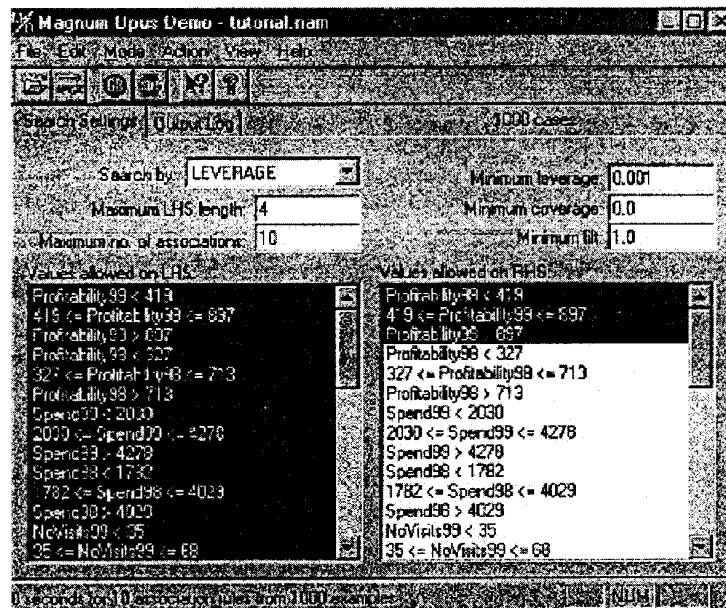
Το lift ενός κανόνα συσχέτισης είναι το strength διαιρεμένο με το ποσοστό όλων των παραδειγμάτων που καλύπτονται από την RHS. Για παράδειγμα, με τα προαναφερθέντα δεδομένα, το strength είναι  $50/200=0.25$  και το ποσοστό των παραδειγμάτων που καλύπτονται από την RHS είναι  $100/1000=0.1$ . Σε αυτή την περίπτωση το lift είναι  $0.25/0.1=2.5$ .

Το coverage είναι το ποσοστό των παραδειγμάτων που καλύπτονται από την LHS δια το σύνολο των περιστάσεων. Άρα με τα ίδια δεδομένα το coverage είναι  $200/1000=0.2$ .

#### *4.2.5 Επεξεργασία δεδομένων*

Έχοντας το παράδειγμα με τα attribute values data θα κάνουμε αναζήτηση συσχέτισης με το leverage. Θα περιορίσουμε το όριο της αναζήτησης στις 10 συσχετίσεις. Αυτό ουσιαστικά γίνεται με την προσθήκη του αριθμού 10 στο

“maximum no. of associations”. Για να κάνουμε το παράδειγμα πιο απλό, θα υποθέσουμε ότι υπάρχει ενδιαφέρον στις συσχετίσεις για την κερδοφορία, ανάμεσα σε άλλα χαρακτηριστικά. Για το σκοπό αυτό πρέπει να περιορίσουμε τις αξίες των χαρακτηριστικών που ίσως εμφανιστούν στην RHS από μια συσχέτιση των τριών αξιών για την κερδοφορία.



Αφού τελειώσει και ο καθορισμός των στοιχείων που θα επεξεργαστεί το πρόγραμμα πατάμε το κουμπί με την ένδειξη GO για να εμφανιστούν τα αποτελέσματα της αναζήτησης. Λόγω του ότι η αναζήτηση έγινε με το leverage, το Magnum Opus ψάχνει να βρει συσχετίσεις με τις υψηλότερες αξίες στο leverage και σύμφωνα και με τους άλλους περιορισμούς που έχουμε θέσει.

#### 4.2.6 Τα αποτελέσματα της αναζήτησης

Όταν ολοκληρωθεί η αναζήτηση, οι συσχετίσεις καταγράφονται στο “Output Log Page”. Οι συσχετίσεις που βρίσκονται καταγράφονται σε μια λίστα που στο παράδειγμά μας φαίνεται παρακάτω. Αρχικά στην σελίδα των αποτελεσμάτων αναφέρεται το όνομα και η έκδοση του προγράμματος και στην συνέχεια περιγράφεται η διαδικασία της αναζήτησης, που στην περίπτωσή μας γίνεται

μέσω της leverage. Τέλος πρέπει να σημειωθεί πως αναφέρονται και οι περιορισμοί που έχουμε θέσει.

```
Magnum Opus - Setting new standards for association rule
discovery.
Copyright © 1999 - 2000 G.I. Webb & associates Pty Ltd.
Thur Mar 30 18:07:40 2000
Data File: tutorial.data
Search by Leverage
Maximum number of attributes on LHS = 4
Maximum number of associations = 10
Minimum Leverage = 0.001
Minimum coverage = 0.0
Minimum Lift = 1.0
All values allowed on LHS
Values allowed on RHS:
Profitability99 < 419 419 <= Profitability 99 <= 897
Profitability99 > 897
Sorted best association rules
Spend99 < 2030 -> Profitability99 <419 [coverage=0.333 (333);
Strength=0.907; Lift=2. 72;
Leverage=0.1911 (191)]
Spend99 > 4278 -> Profitability99 >897 [coverage=0.333 (333);
Strength=0. 862;Lift=2. 60;
Leverage=0.1768 (176)]
Spend99 < 2030 & Grocery <873 -> Profitability99 < 419
[coverage=0.278 (278); Strength=0.953; Lift=2.86; Leverage=0.1724
(172)]
Grocery < 873 -> Profitability99 < 419 [coverage=0.333 (333);
Strength=0.832; Lift=2.50; Leverage=0.1661 (166)]
Spend99 < 2030 & NoVisits99 < 35 -> Profitability99 < 419
[coverage=0.272 (272); Strength=0.938; Lift=2.82; Leverage=0.1644
(164)]
Profitability98 < 327 & Spend99 < 2030 -> Profitability99 < 419
[coverage=0.256 (256); Strength=0.961; Lift=2.89; Leverage=0.1608
(160)]
Spend99 < 2030 & Promotion1=f-> Profitability99 < 419
[coverage=0.270 (270); Strength=0.919; Lift=2. 76;
Leverage=0.1581 (158)]
Grocery > 2113-> Profitability99 > 897 [coverage=0.333 (333);
Strength=0.805; Lift=2.43;
Leverage=0.1578 (157)]
Spend99 < 2030 & Spend99 < 1782 -> Profitability99 < 419
[coverage=0.259 (259); Strength=0.942; Lift=2.83; Leverage=0.1578
(157)]
Spend99 < 2030 & NoVisits98 < 31 -> Profitability99 < 419
[coverage=0.255 (255); Strength=0.945; Lift=2.84; Leverage=0.1561
(156)]
0 seconds for 10 associations rules from 1000 examples
```

Στην τελευταία γραμμή γίνεται γνωστός ο χρόνος που απαιτήθηκε για την ολοκλήρωση της αναζήτησης, ο αριθμός των συσχετίσεων που βρέθηκαν, και ο μέγιστος αριθμός των περιπτώσεων που εξετάστηκαν. Λόγω του περιορισμού



όσον αφορά τον μέγιστο αριθμό των συσχετίσεων οι συσχετίσεις που βρέθηκαν είναι 10.

Για να επεξηγήσουμε τα αποτελέσματα ας δούμε τον πρώτο κανόνα:

```
Spend99 < 2030 -> Profitability99 < 419 [coverage=0.333 (333);  
Strength=0.907; Lift=2.72;  
Leverage=0.1911 (191)]
```

Στο κανόνα αυτό το LHS παρουσιάζεται πριν βέλος -> και το RHS ακολουθεί το βέλος και προηγείται της αγκύλης. Η συσχέτιση αυτή δείχνει πως στις περιπτώσεις που το Spend99 έχει τιμή μικρότερη από 2030 συσχετίζεται με τις περιπτώσεις για τις οποίες η Profitability99 έχει τιμή μικρότερη από 419 πιο συχνά από ότι στις άλλες περιπτώσεις. Μέσα στην αγκύλη φαίνονται οι συσχετίσεις:

Coverage=0.333 (333). Αυτή η τιμή δείχνει το ποσοστό όλων των περιπτώσεων που ικανοποιεί τον κανόνα LHS (spend99 < 2030). Ο αριθμός των περιπτώσεων που ικανοποιούν την LHS είναι 333 για το σύνολο των 1000 περιπτώσεων.

Strength=0.907. Αυτό δείχνει το ποσοστό από τις περιπτώσεις που ικανοποιούν την LHS και την RHS ταυτόχρονα. Από τις 333 περιπτώσεις που ικανοποιούν την LHS οι 302 ικανοποιούν και την RHS. Άρα το Strength υπολογίζεται ως  $302/333=0.907$ .

Lift=2.72. Το Lift μετράει το πόσο πιο ισχυρή από το κανονικό είναι η συσχέτιση ανάμεσα στην LHS και στην RHS. Από το σύνολο των περιπτώσεων οι 333 ικανοποιούν τόσο την LHS όσο και την RHS, γιατί κάθε χαρακτηριστικό χωρίζεται πάνω σε μια τιμή που διαιρείται σε τρία ίσα μέρη. Έτσι αν οι LHS, RHS είναι ανεξάρτητες μεταξύ τους το αναμενόμενο strength της συσχέτισης θα ήταν  $333/1000$ . Διαιρώντας το πραγματικό strength με το αναμενόμενο έχουμε:  $(302/333)/(333/1000)=2.72$ .

Leverage=0.1911. Το leverage είναι ένα μέτρο του μεγέθους της επίδρασης που δημιουργείται από τη συσχέτιση. Αυτό είναι το ποσοστό των περιπτώσεων που εμφανίζουν αυτή την συσχέτιση σε αντιπαράβολές με αυτές που θα αναμενόταν αν οι LHS, RHS ήταν ανεξάρτητες μεταξύ τους. Αν οι τιμές των δυο χαρακτηριστικών ήταν ανεξάρτητες μεταξύ τους και διαφορετικές από το αναμενόμενο ποσοστό των περιπτώσεων που εμφανίζεται στην LHS (0.333) επί το ποσοστό που εμφανίζεται στην RHS (0.333) δηλαδή  $0.333 \times 0.333 = 0.1109$ . Το ποσοστό που εμφανίζεται στη συσχέτιση είναι 0.302 και η διαφορά ανάμεσα σε αυτές τις δυο τιμές είναι το 0.191

#### 4.2.7 Συμπεράσματα για το πρόγραμμα *Magnum Opus*

Όπως έχουμε ήδη αναφέρει το πρόγραμμα *Magnum Opus* έχει ως κύρια λειτουργία την εύρεση κανόνων συσχέτισης (association rules). Τα κυριότερα χαρακτηριστικά του είναι:

Το *Magnum Opus* βρίσκει τους κανόνες συσχέτισης τόσο από τα «καλάθια» δεδομένων (basket data) όσο και από τις ομάδες χαρακτηριστικών (attribute value data).

Έχει σχεδιαστεί για να αναλύσει σημαντικές βάσεις δεδομένων που περιλαμβάνουν μεγάλο πλήθος καταγραφών.

Είναι εξαιρετικά γρήγορο πρόγραμμα. Αν τα δεδομένα διπλασιαστούν, θα διπλασιαστεί και ο χρόνος επεξεργασίας τους από τον υπολογιστή.

Οι συσχετίσεις που γίνονται εκφράζονται σε μια φόρμα εύκολα κατανοητή όπως έχουμε άλλωστε στο παράδειγμα.

Η χρήση του προγράμματος δεν απαιτεί εξειδικευμένη γνώση στατιστικής ή χειρισμού του υπολογιστή.

Πρέπει να σημειωθεί πως το *Magnum Opus* αναπτύχθηκε από την RuleQuest Research ©1999-2000.

## **5. ΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝ ΠΡΟΓΡΑΜΜΑΤΑ ΝΟΗΜΟΝΩΝ ΠΡΟΣΕΓΓΙΣΕΩΝ (SEE5/C5.0)**

### **5.1. Εισαγωγή.**

Στο κεφάλαιο αυτό θα γίνει μια προσπάθεια ώστε να κατανοηθούν τα αποτελέσματα που επιτυγχάνονται με τα C5.0 και See5. Για λόγους συντομίας θα αναφερθούμε μόνο στο See5 (ένα εργαλείο των windows), αφού και το C5.0 δίνει τα ίδια αποτελέσματα.

Τι είναι λοιπόν αυτό που κάνει το See5 τόσο ξεχωριστό; Μια σύντομη απάντηση στο ερώτημα είναι η προσοχή που επιδεικνύει πάνω στο θέμα της αντιληπτικότητας. Τα εργαλεία τεχνητής νοημοσύνης πρέπει να βρίσκουν καλούπια που παρέχουν εύστοχες προβλέψεις και εξασφαλίζουν διορατικότητα. Έτσι λοιπόν το See5 δίνει έμφαση στην ταξινόμηση βάσει κανόνων γιατί έτσι γίνονται πολύ πιο εύκολα κατανοητά -κάθε κανόνας μπορεί να εξεταστεί και να εκτιμηθεί ξεχωριστά, χωρίς να είναι απαραίτητο να πρέπει να εκτιμηθεί το περιεχόμενό τους ως μέρος του συνόλου. Το See5 μπορεί επίσης να παράγει δέντρα αποφάσεων, τα οποία είναι εξαιρετικά χρήσιμα σε καταστάσεις που η ταξινόμηση πρέπει να γίνει γρήγορα.

Για να γίνουν όμως όλα αυτά περισσότερο κατανοητά πρέπει να δούμε ορισμένα παραδείγματα.

### **5.2. Παραδείγματα.**

#### *5.2.1 Κατανέμοντας σπίτια στη Βοστώνη, ανάλογα με την αξία.*

Η πρώτη εφαρμογή χρησιμοποιεί δεδομένα από τιμές σπιτιών στην Βοστώνη. Αντί να προσπαθεί να προβλέψει τις αξίες των σπιτιών διαχωρίζει το 20% των καλύτερων από το 80% των υπολοίπων σπιτιών, μια εφαρμογή παρόμοια με αυτή που χρησιμοποιείται στις έρευνες αγοράς.

Κάθε περίπτωση χρησιμοποιεί χαρακτηριστικά των σπιτιών σε τέτοια έκταση που είναι πιθανόν να επηρεάσει την τιμή τους. Παρακάτω εμφανίζονται ορισμένα παραδείγματα :

<u>Attribute</u>	<u>Tract 1</u>	<u>Tract 2</u>	<u>Tract 3</u>
Crime rate	7.67	2.24	0.08
Proportion large lots	-	-	45
Proportion industrial	18.1	19.6	3.4
Nitric oxides ppm	.69	.61	.44
Av rooms per dwelling	5.7	5.9	7.2
Proportion pre-1940	98.9	91.8	38.9
Distance to employment centers	1.6	2.4	4.6
Accessibility to radial highways	24	5	5
Property tax rate per \$1 0,000	666	403	398
Pupil-teacher ratio	20.2	14.7	15.2
Percentage low income earners	19.9	11.6	5.4
<b>Category</b>	<b>Bottom</b>	<b>Bottom</b>	<b>Top</b>
	<b>80%</b>	<b>80%</b>	<b>20%</b>

Από 506 παρόμοιες περιπτώσεις, το See5 επιλέγει 8 αντιπροσωπευτικούς κανόνες, 4 για τα κορυφαίας αξίας σπίτια και 4 για τα υπόλοιπα :

Rule 1: (57, lift 4.9)  
 av rooms per dwelling > 7.007  
 pupil-teacher ratio <= 19.1  
 -> class top 20% [0.983]

Rule 2: (18, lift 4.8)  
 proportion industrial > 2.68  
 proportion industrial <= 4.39  
 proportion pre-1940 > 17.8  
 percentage low income earners <= 5.39  
 -> class top 20% [0.950]

Rule 3: (5, lift 4.3)  
 crime rate <= 24.3938  
 distance to employment centers <= 1.5888  
 percentage low income earners <= 10.74  
 -> class top 20% [0.857]

Rule 4: (76/13, lift 4.1)  
 percentage low income earners <= 5.39  
 -> class top 20% [0.821]

Rule 5: (407/18, lift 1.2)  
 av rooms per dwelling <= 7.007  
 percentage low income earners > 5.29  
 -> class bottom 80% [0.954]

Rule 6: (236/10, lift 1.2)

```
pupil-teacher ratio > 19.1  
-> class bottom 80% [0.954]
```

```
Rule 7: (405/18, lift 1.2)  
av rooms per dwelling <= 6.794  
percentage low income earners > 4.5  
-> class bottom 80% [0.953]
```

```
Rule 8: (17, lift 1.2)  
av rooms per dwelling <= 6.794  
proportion pre-1940 <= 17.8  
-> class bottom 80% [0.947]
```

Κάθε κανόνας χαρακτηρίζεται από τα στατιστικά δεδομένα (  $N$  /  $E$ , lift  $L$  ) όπου:

- ο  $N$  ο αριθμός των περιπτώσεων που καλύπτονται από τον κανόνα.
- ο  $E$  (εάν υπάρχει) ο αριθμός αυτών που δεν ανήκουν στον κανόνα.
- ο  $L$  ο υπολογισμός της ορθότητας του κανόνα (ο αριθμός βρίσκεται μέσα σε άγκιστρο π.χ. [0.947]).

Οι κανόνες δεν είναι αποκλειστικοί, για παράδειγμα ο πέμπτος κανόνας είναι απλά μια πιο «εξειδικευμένη έκδοση» του τέταρτου. Τα δεδομένα του παραπάνω παραδείγματος δείχνουν πως οι ομάδες των κανόνων παρουσιάζουν στις προβλέψεις ακρίβεια της τάξεως του 92%.

Είναι λοιπόν φανερό ότι το See5 μπορεί με μεγάλες ταχύτητες να επεξεργαστεί μεγάλες βάσεις δεδομένων και να αποδώσει ακριβή μοντέλα.

### 5.2.2 Εντοπίζοντας αγγελίες στο διαδίκτυο

Το επόμενο παράδειγμα χρησιμοποιεί καινοτόμο βάση δεδομένων προερχόμενη από τον Nick Kushmerick. Υπάρχουν 3279 περιπτώσεις, από τις οποίες κάθε μια περιγράφει ένα έγγραφο HTML. Το 15% των αγγελιών αυτών είναι ξεχωριστές αγγελίες, και σκοπός είναι να δημιουργηθούν κανόνες που προβλέπουν αν η εικόνα είναι αγγελία ή όχι. Το σύστημα αυτό του Kushmerick

ονομάζεται AdEater και χρησιμοποιεί την πρόβλεψη έτσι ώστε να επιταχύνει την διαδικασία φόρτωσης σελίδων με διαφημίσεις στον υπολογιστή.

Αυτή η βάση δεδομένων είναι πολυδιάστατη. Υπάρχουν 1558 χαρακτηριστικά, σχεδόν τα μισά σε σχέση με τον αριθμό των περιπτώσεων. Τα χαρακτηριστικά περιλαμβάνουν τρεις αριθμούς: το μήκος και το πλάτος της εικόνας, την λήψη λογικών τιμών 0 ή 1 που παραστούν την παρουσία ή απουσία φράσεων στην επικεφαλίδα της διαφήμισης, και την εικόνα URL.

Το See5 εξάγει 19 κανόνες από αυτή τη βάση δεδομένων, από τις οποίες 19 περιγράφουν αγγελίες και μια όχι αγγελία:

```
Rule 1 : (65, lift 7 .0)
ancurl*http+www = 1
ancurl*click = 0
-> class ad [0.985]

Rule 2: (39, lift 7.0)
ancurl*adclick = 1
-> class ad [0.976]

Rule 3: (103/2, lift 6.9)
url*ads = 0
ancurl*click = 1
-> class ad [0.971]

Rule 4: (24, lift 6.9)
url*ad+gif = 0 ancurl*http+www = 0 ancurl*ad = 1
-> class ad [0.962]

Rule 5: (20, lift 6.8)
ancurl*n+a = 1
-> class ad [0.955]

Rule 6: (152/6, lift 6.8)
url*ads = 1
-> class ad [0.955]

Rule 7: (16, lift 6.7)
ancurl*plx = 1
-> class ad [0.944]

Rule 8: (15, lift 6.7)
url*doubleclick.net = 1
-> class ad [0.941]

Rule 9: (14, lift 6.7)
ancurl*url+http = 1
-> class ad [0.938]

Rule 10: (14, lift 6.7)
```

```

alt*visit+our = 1
-> class ad [0.938]

Rule 11: (13, lift 6.7)
origurl*home.netscape.com = 1
-> class ad [0.933]

Rule 12: (110/8, lift 6.6)
alt*click+here = 1
alt*to = 0
-> class ad [0.920]

Rule 13: (10, lift 6.5)
origurl*zdnet.com = 1
-> class ad [0.917]

Rule 14: (9, lift 6.5)
origurl*messier = 0
ancurl*redirect+cgi = 1
-> class ad [0.909]

Rule 15: (9, lift 6.5)
origurl*jun = 1
-> class ad [0.909]

Rule 16: (8, lift 6.4)
width > 196
url*images+home = 1
-> class ad [0.900]

Rule 17: (4, lift 6.0)
width > 135
url*logo = 0
url*ads = 0
url*marketing = 1
-> class ad [0.833]

Rule 18: (16/3, lift 5.6)
url*logo = 0
url*marketing = 1
-> class ad [0.778]

Rule 19: (3127/313, lift 1.0)
url*ads = 0
-> class nonad [0.900]

```

Σε αυτήν την περίπτωση το See5 ταξινομεί σωστά το 97% των αθέατων εικόνων, πράγμα το οποίο δείχνει πως ακόμα και σε απλές θεωρίες μπορεί να αποδειχθεί πολύ χρήσιμο.

### 5.2.3 Ελέγχοντας μια παραγωγική διαδικασία

Για πολλά χρόνια ο David Stirling ερεύνησε εφαρμογές Data Mining στην παραγωγή ατσαλιού στην BHP, μια από τις μεγαλύτερες εταιρίες της

Αυστραλίας. Μια εφαρμογή αφορά τις κατάλληλες ενέργειες κατά την διάρκεια της επίστρωσης ατσάλινων προϊόντων. Η συγκεκριμένη βάση δεδομένων περιέχει 898 περιπτώσεις με 38 χαρακτηριστικά που καλύπτουν τομείς όπως το εύρος του ατσαλιού, την σκληρότητά του, την σύνθεσή του, την ποιότητά του κ.τ.λ. Υπάρχουν 5 κατηγορίες που αντιστοιχούν σε εναλλακτικές υπό-διαδικασίες επίστρωσης.

Από τα μισά δεδομένα της βάσης το See5 κατασκευάζει ένα ταξινομητή αποτελούμενο από 7 κανόνες.

```
Rule 1: (3, lift 119.7)
steel = S
strength > 350
-> class 1 [0.800]

Rule 2: (4613, lift 7.5)
family = -
temper_rolling = -
condition = S
surface-quality = -
-> class 2 [0.917]

Rule3:(6, lift 7.1)
steel = -
strength > 350
-> class 2 [0.875]

Rule 4: (6, lift 7.1)
steel = M
strength > 350
-> class 2 [0.875]

Rule 5: (428/97, lift 1.0)
hardness <= 70
-> class 3 [0.772]

Rule 6: (39, lift 11.2)
family = TN
-> class 5 [0.976]

Rule 7: (21/1, lift 20.5)
hardness > 70
-> class U [0.913]
```

Αυτοί οι κανόνες ταξινομούν πάνω από το 98% των υπόλοιπων 449 περιπτώσεων .



Η κατασκευή ταξινομητών που βασίζονται σε κανόνες βάσεων δεδομένων είναι ένας μόνο τομέας στον οποίο χρησιμοποιείται το See5 (ίσως ο πιο σημαντικός). Άλλες λειτουργίες περιλαμβάνουν :

Μια τεχνική κατασκευής πολλαπλών ταξινομητών για την βελτίωση της ικανότητας πρόβλεψης η οποία ονομάζεται boosting.

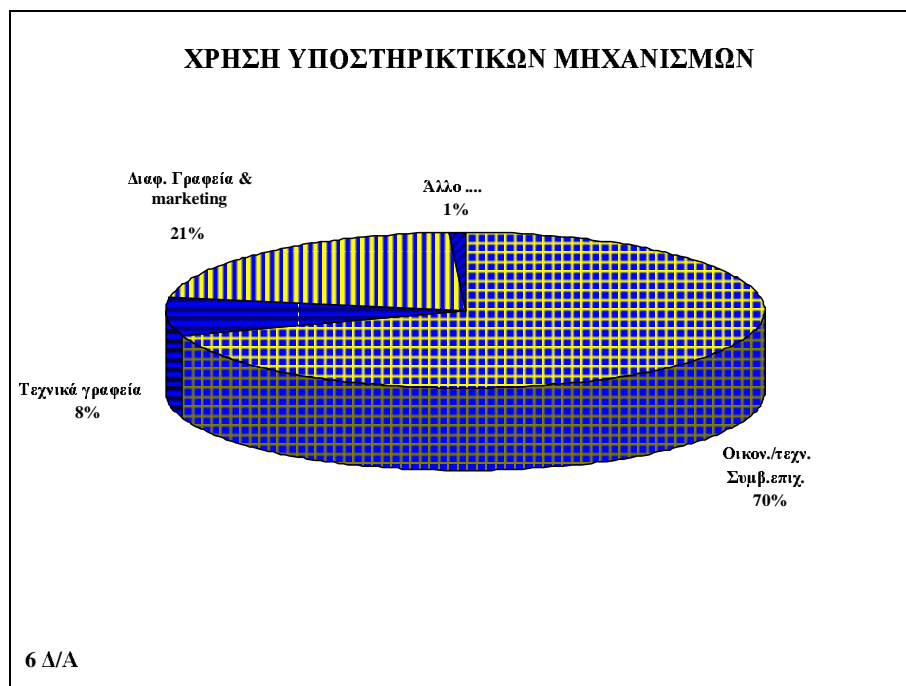
Differential misclassification costs, που βοηθούν στο να αναγνωρίζονται κάποια λάθη ως περισσότερο σημαντικά από άλλα.

Την παράβλεψη των λιγότερο σχετικών χαρακτηριστικών (windowing), και υποστήριξη για δοκιμές εξακρίβωσης και δειγματοληψίας (cross-validation trials and sampling).

## 6. ΠΡΑΓΜΑΤΙΚΗ ΕΦΑΡΜΟΓΗ ΝΟΗΜΟΝΟΣ ΠΡΟΣΕΓΓΙΣΕΩΣ: «ΑΝΑΛΥΣΗ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ΓΙΑ ΤΟ ΒΙΟΤΕΧΝΙΚΟ ΠΑΡΚΟ ΧΑΝΙΩΝ»

### 6.1. Παρουσίαση των δεδομένων

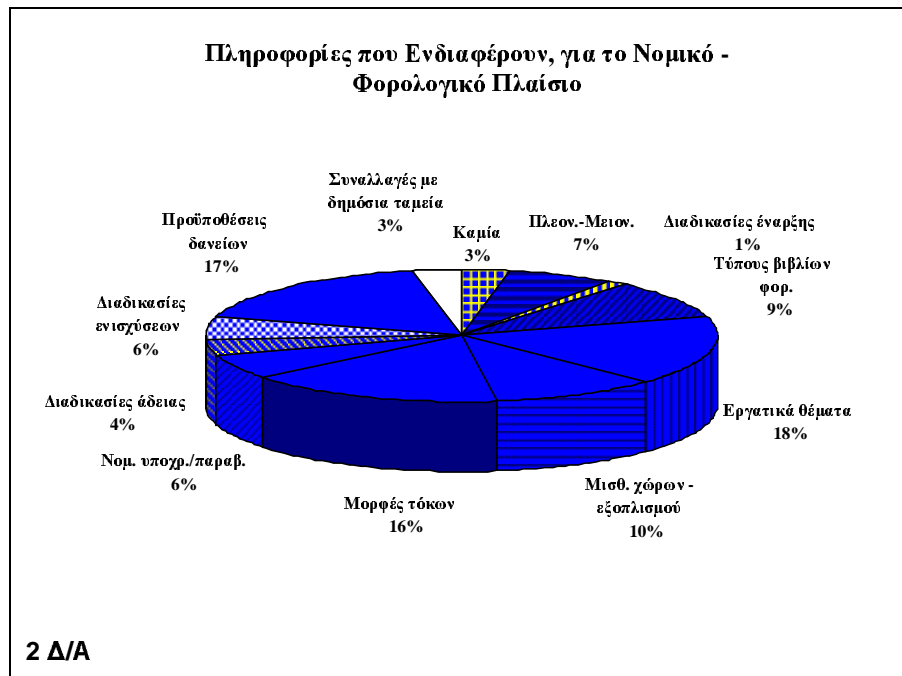
Στο κομμάτι αυτό θα παρουσιάσουμε τα δεδομένα που προέκυψαν έπειτα από έρευνα στην αγορά των Χανίων για την δημιουργία ενός βιοτεχνικού πάρκου, και στα οποία θα στηριχθούμε ώστε να εφαρμόσουμε το πρόγραμμα SEE5.



Εικόνα 6. 1

Στην εικόνα 6.1 βλέπουμε ποιους μηχανισμούς επιλέγουν οι ερωτούμενοι για την υποστήριξη των επιχειρήσεών τους. Το μεγαλύτερο ποσοστό (70%) επιλέγει οικονομικοτεχνικούς σύμβουλους επιχειρήσεων, και ένα μεγάλο ποσοστό προτιμά διαφημιστικά γραφεία ή εταιρίες marketing. Υπάρχει βέβαια και ένα μικρό ποσοστό που επιλέγει τεχνικά γραφεία για την υποστήριξη της

επιχειρησής τους. Είναι φανερό πως όλες οι σύγχρονες επιχειρήσεις επιζητούν κάποια εξειδικευμένη βοήθεια.

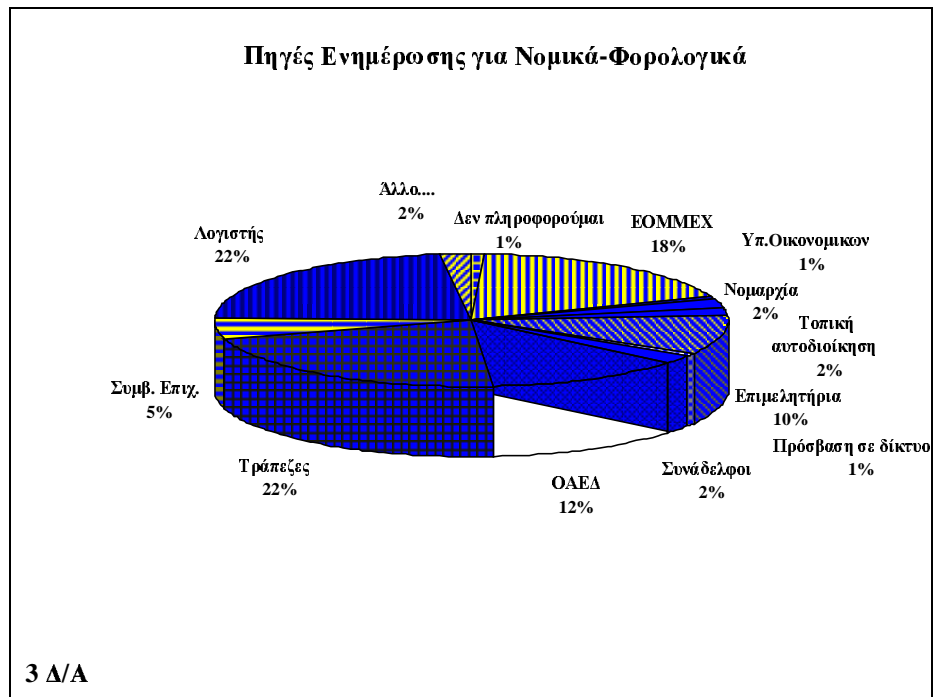


Εικόνα 6.2

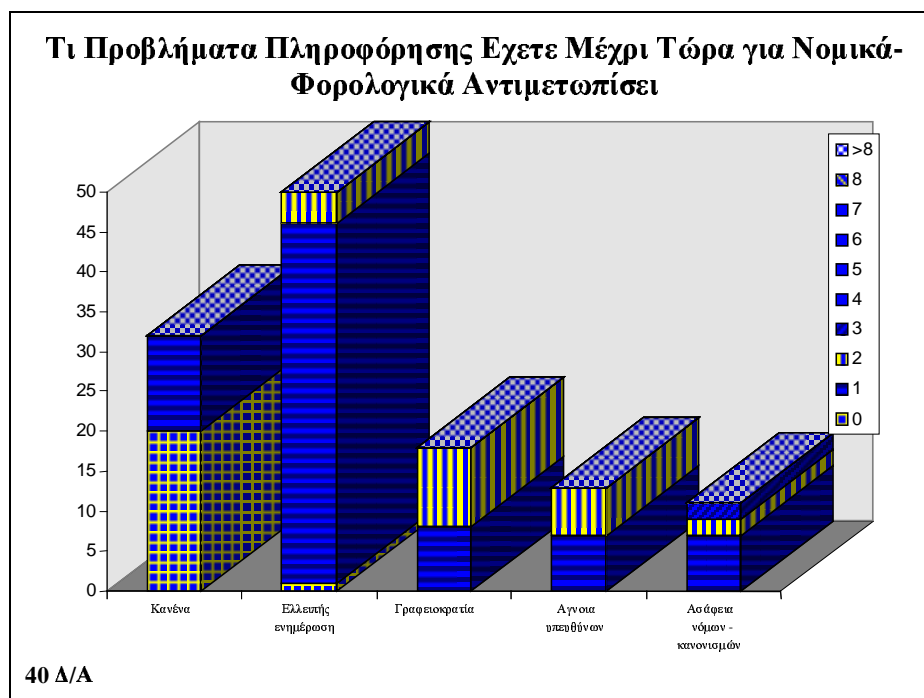
Η εικόνα 6.2 δείχνει ποιες πληροφορίες ενδιαφέρουν τους ερωτούμενους όσον αφορά το νομικό και φορολογικό πλαίσιο που ισχύει. Οι κυριότεροι τομείς για τους οποίους οι ερωτούμενοι επιζητούν πληροφορίες είναι εργατικά θέματα με ποσοστό 18%, προϋποθέσεις για την λήψη δανείων με ποσοστό 17%, θέματα σχετικά με τόκους με ποσοστό 16%, και θέματα σχετικά με μισθώματα χώρων ή εξοπλισμού με ποσοστό 10%. Επίσης, αρκετοί ενδιαφέρονται για τις διαδικασίες ενίσχυσης των επιχειρήσεων τους, για νομικές υποχρεώσεις ή παραβάσεις, για τύπους φορολογικών βιβλίων, καθώς και για τις συναλλαγές με τα δημόσια ταμεία και τις διαδικασίες χορήγησης άδειας.

Στην εικόνα 6.3 βλέπουμε ποιες είναι οι κύριες πηγές ενημέρωσης των επιχειρήσεων πάνω σε νομικά και φορολογικά σχέδια. Η πλειοψηφία απάντησε πως ενημερώνεται από τον λογιστή (22%) ή τις τράπεζες (22%). Επίσης αρκετοί

από τους ερωτούμενους απάντησαν πως ενημερώνονται από τον ΕΟΜΜΕΧ (18%), από τον ΟΑΕΔ (12%) και από τα επιμελητήρια. Μικρά είναι τα ποσοστά αυτών που απάντησαν πως ενημερώνονται από την τοπική αυτοδιοίκηση, το Υπουργείο Οικονομικών, το διαδίκτυο ή από συναδέλφους.

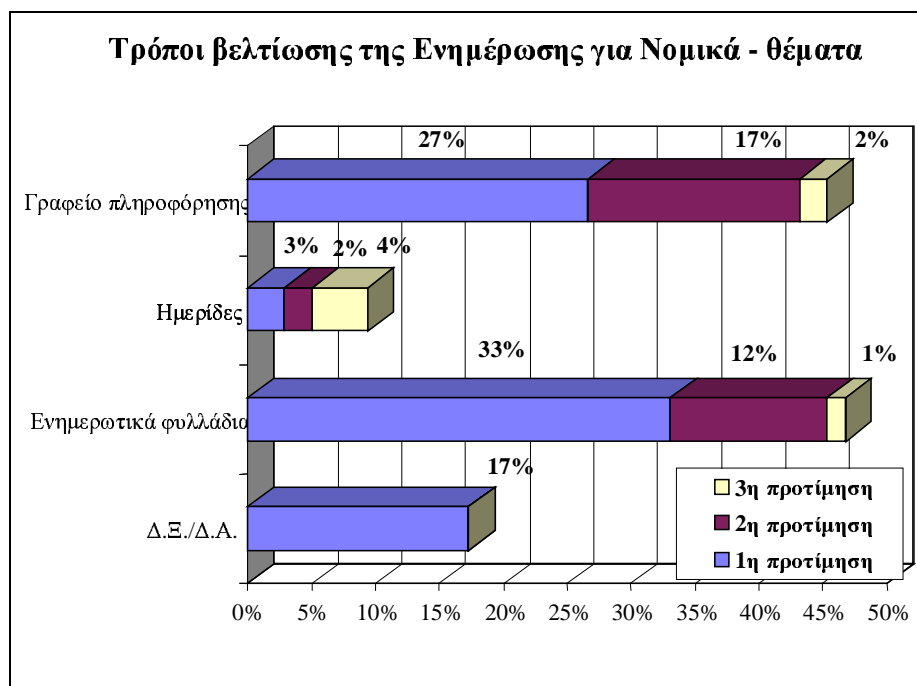


Εικόνα 6.3



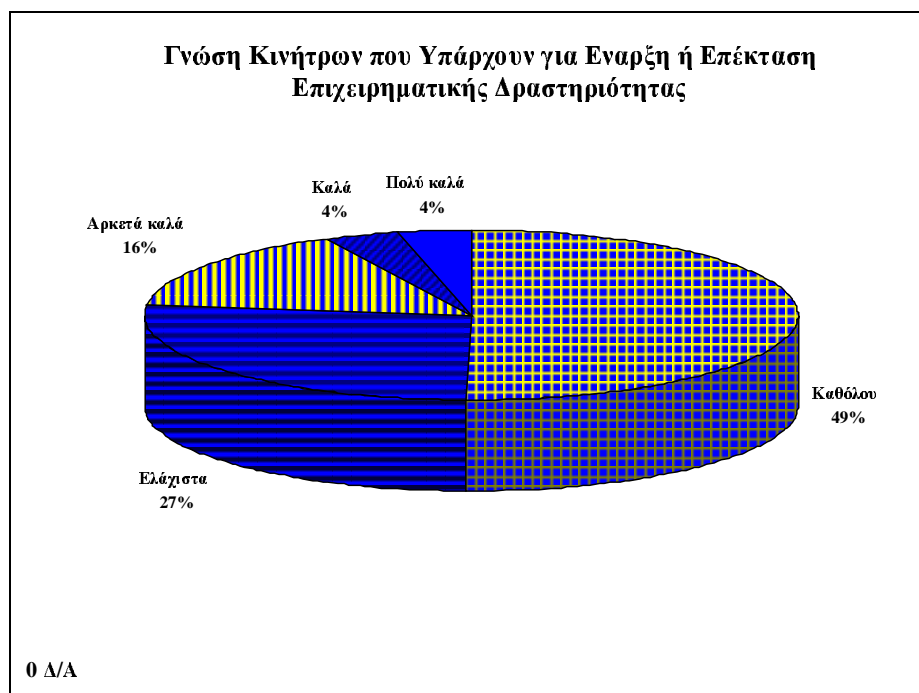
Εικόνα 6.4

Στην εικόνα 6.4 βλέπουμε ποια είναι τα κυριότερα προβλήματα που έχουν αντιμετωπίσει οι ερωτούμενοι για φορολογικά και νομικά θέματα. Αν και το ποσοστό αυτών που δεν αντιμετώπισαν κανένα απολύτως πρόβλημα είναι μεγάλο, αρκετοί ήταν αυτοί που είχαν πρόβλημα ελλιπούς ενημέρωσης, γραφειοκρατίας, άγνοιας από την πλευρά των υπευθύνων και ασάφειας των κανονισμών.



Εικόνα 6.5

Στην ερώτηση 5 δόθηκε η δυνατότητα στους ερωτηθέντες να διαλέξουν ή να προτείνουν τον καταλληλότερο τρόπο βελτίωσης της ενημέρωσής τους σε θέματα φορολογίας και νόμων. Οι περισσότεροι (46%) προτείνουν την έκδοση και διανομή ενημερωτικών φυλλαδίων, τρόπος που είναι ίσως ο πιο κατάλληλος γιατί δίνει τη δυνατότητα σε όλους να ενημερώνονται χωρίς να απομακρύνονται από την επιχείρηση. Περίπου ίδιο ποσοστό (46%) προτείνει τη λειτουργία γραφείου πληροφόρησης, μόνο που σαν πρώτη προτίμηση το ποσοστό είναι 27% έναντι 33% της έκδοσης φυλλαδίων. Για τις ημερίδες υπάρχει εξαιρετικά χαμηλό ποσοστό (9% συνολικά) και μάλιστα η πλειοψηφία του το επιλέγει σαν τελευταία λύση. Τέλος το 17% δεν απάντησε στην ερώτηση. Συμπερασματικά κατά τη γνώμη των επιχειρηματιών η βελτίωση της ενημέρωσης μπορεί να γίνει με την έκδοση ενημερωτικών εντύπων ή την δημιουργία γραφείου πληροφοριών, αν και εντυπωσιακά υψηλό ήταν το ποσοστό που δήλωσε άγνοια ή δεν απάντησε στην συγκεκριμένη ερώτηση (17%).

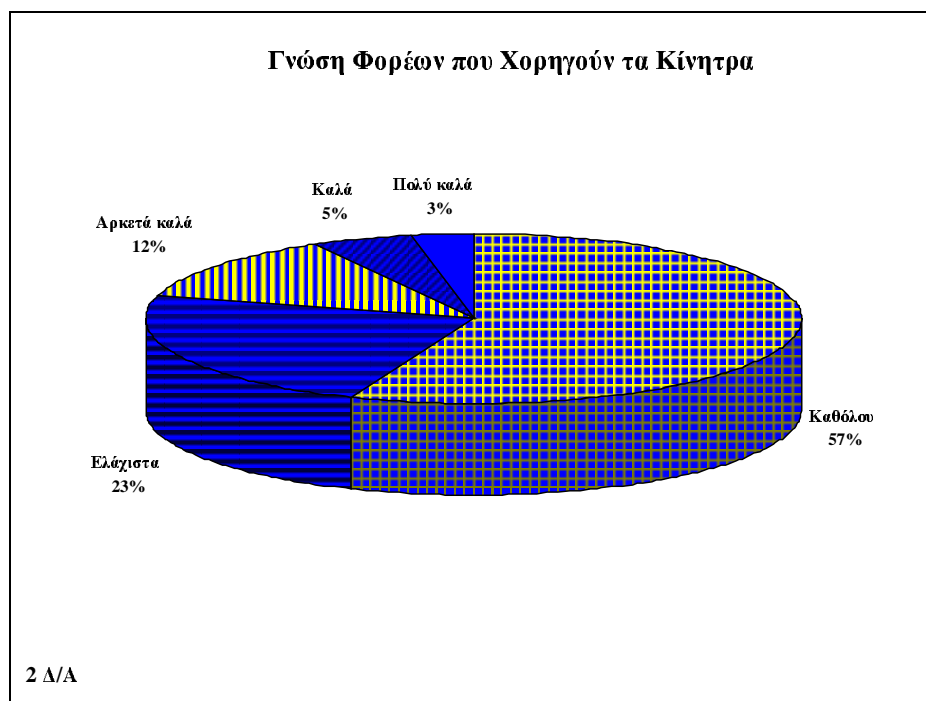


Εικόνα 6.6

Στην ερώτηση για την γνώση των κινήτρων έναρξης ή επέκτασης επιχειρηματικής δραστηριότητας, όπως φαίνεται στην Εικόνα 6 περίπου οι μισοί (49%) δηλώνουν ότι αγνοούν τα υπάρχοντα κίνητρα ενώ είναι επίσης σημαντικό το ποσοστό (27%) αυτών που δηλώνουν ελάχιστα, συνολικά το 76% των επιχειρηματιών έχει ανάγκη αναλυτικών πληροφοριών για το θέμα αυτό, ενώ 16% δηλώνει ότι τα γνωρίζει αρκετά καλά και μόνο 8% (4%+4%) καλά ή πολύ καλά. Δεδομένου ότι σε τέτοιες μορφής ερωτήσεις οποιουδήποτε ερωτηματολογίου είναι δύσκολο να λάβουμε ειλικρινείς απαντήσεις και συνήθως προτιμούνται για ψυχολογικούς λόγους του ερωτώμενου οι «μέσες απαντήσεις» (π.χ. αρκετά), είναι εντυπωσιακό το ποσοστό που απάντησε «καθόλου».

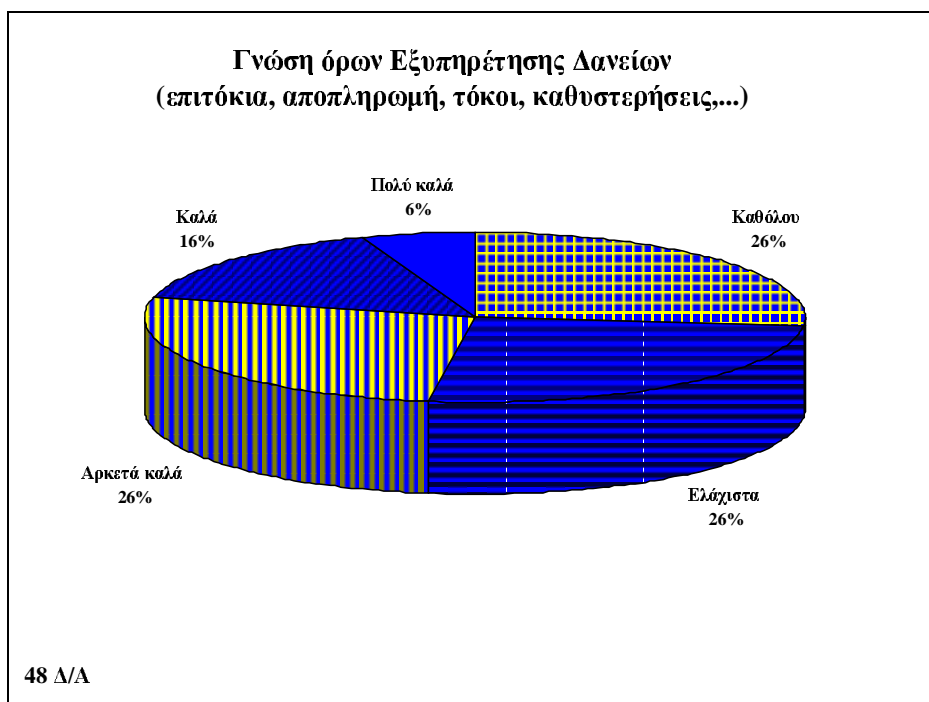
Στην ερώτηση αν γνωρίζουν τους φορείς που παρέχουν τα κίνητρα για έναρξη ή επέκταση της επιχειρηματικής δραστηριότητας (εικόνα 7) το 57% των ερωτούμενων απάντησε πως δεν γνωρίζει καθόλου, το 23% γνωρίζει ελάχιστα, 5% καλά και μόνο 12% και 3% γνωρίζουν αρκετά καλά και πολύ καλά τους εν

λόγω φορείς. Και σε αυτή την περίπτωση το ποσοστό που δήλωσε άγνοια είναι πολύ μεγάλο.



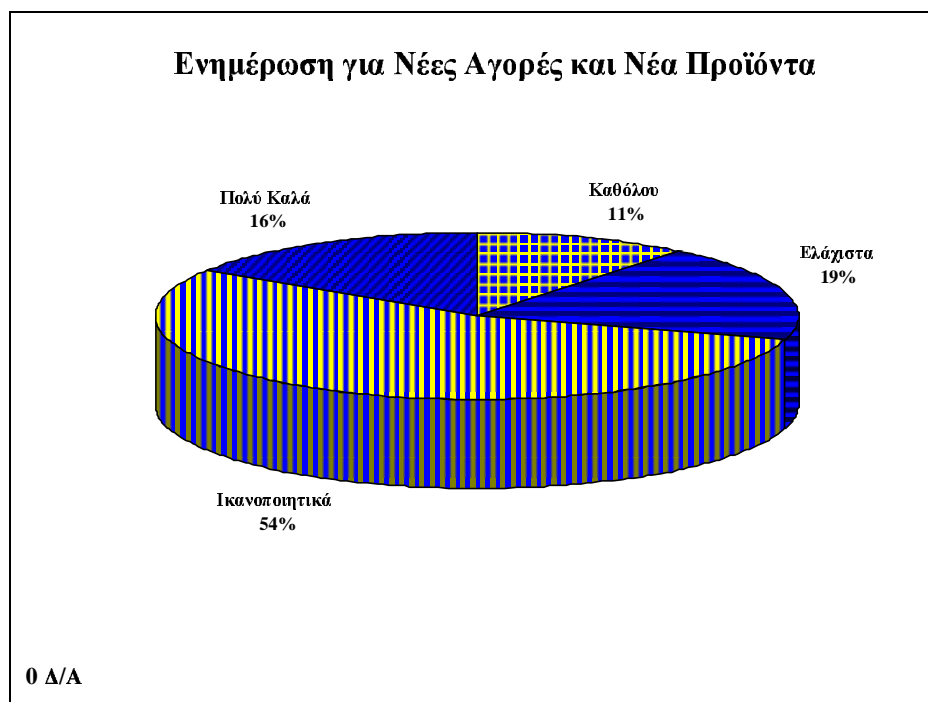
Εικόνα 6.7





Εικόνα 6.8

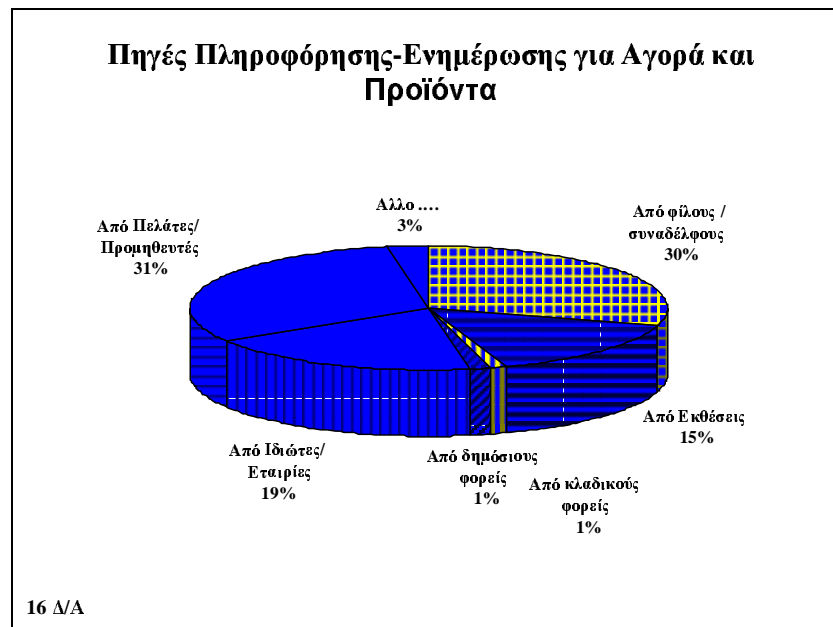
Η εικόνα 6.8 μας δείχνει τις απαντήσεις στην ερώτηση αν γνωρίζουν επαρκώς τους όρους εξυπηρέτησης δανείων. Τα ποσοστά ήταν μοιρασμένα αφού 26% τους γνώριζαν ελάχιστα, το ίδιο ποσοστό αρκετά καλά και καθόλου. Πολύ καλά ή καλά απάντησε το 6% και το 16% αντίστοιχα.



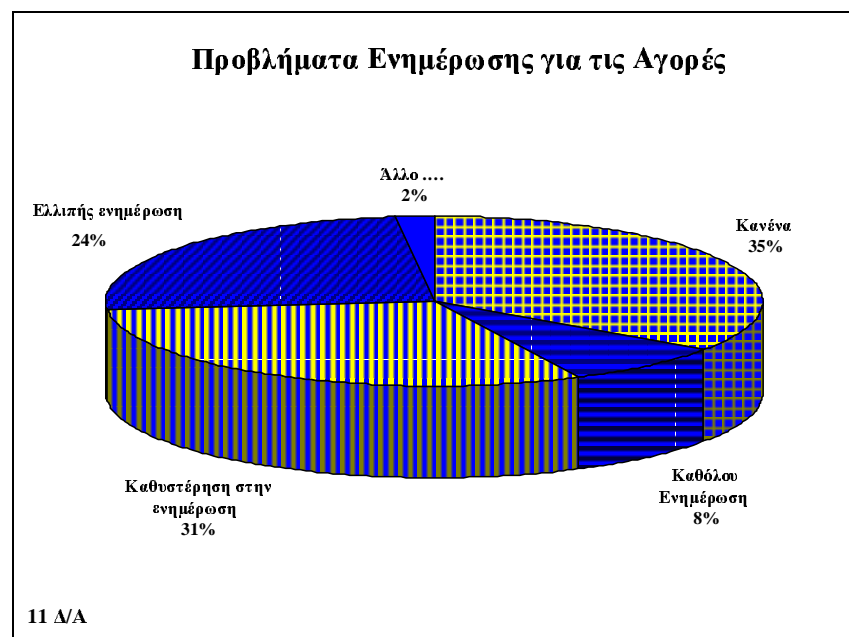
Εικόνα 6.9

Στην ερώτηση για το ποια είναι η ενημέρωση για νέες αγορές ή προϊόντα η πλειοψηφία (56%) απάντησε πως υπάρχει ικανοποιητική ενημέρωση και 16% απάντησαν ότι η ενημέρωση είναι πολύ καλή. Βέβαια υπήρξε και ένα 11% που απάντησε πως δεν υπάρχει καθόλου ενημέρωση, καθώς και ένα 19% που βρίσκει την ενημέρωση ελλιπή. Είναι φανερό πως η πλειοψηφία των ερωτούμενων είναι ικανοποιημένη από την παρεχόμενη ενημέρωση.

Η εικόνα 6.10 αποτελεί ουσιαστικά συνέχεια της 6.9 αφού δείχνει ποιες είναι οι κύριες πηγές πληροφόρησης και ενημέρωσης για τις αγορές και τα προϊόντα. Το 31% ενημερώνεται από πελάτες ή προμηθευτές και το 30% από φίλους ή συναδέλφους. Ακόμη υπάρχει ένα ποσοστό της τάξεως του 19% που ενημερώνεται από ιδιώτες και εταιρίες καθώς και ένα ποσοστό 15% που ενημερώνεται από διάφορες εκθέσεις. Τέλος πολύ μικρό είναι το ποσοστό που ενημερώνεται από δημόσιους ή κλαδικούς φορείς, ένα ποσοστό πολύ μικρό που δείχνει ότι δεν γίνονται σημαντικές προσπάθειες ούτε από συλλογικούς αλλά ούτε και από κλαδικούς φορείς.



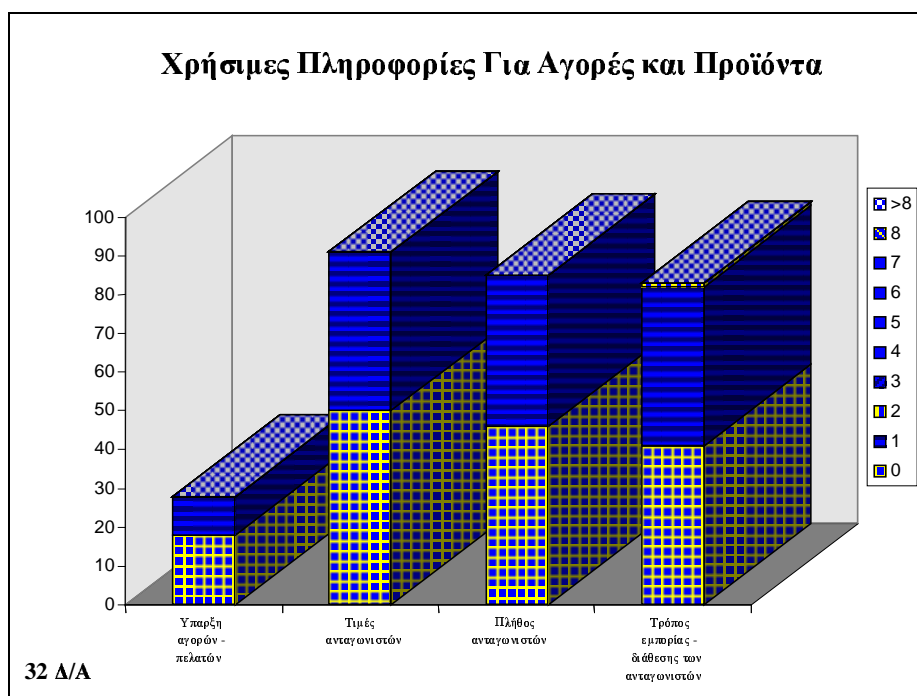
Εικόνα 6.10



Εικόνα 6.11

Προχωρώντας παρακάτω σύμφωνα με όσα μπορούμε να δούμε στην ερώτηση 11 το κυριότερο πρόβλημα που αντιμετώπισαν οι ερωτούμενοι όσον

αφορά την ενημέρωση για τις αγορές είναι ότι υπάρχει καθυστέρηση για ενημέρωση (31%), ενώ το 24% των ερωτούμενων απάντησε πως η ενημέρωση είναι ελλιπής και το 8% πως δεν παρέχεται καθόλου ενημέρωση.

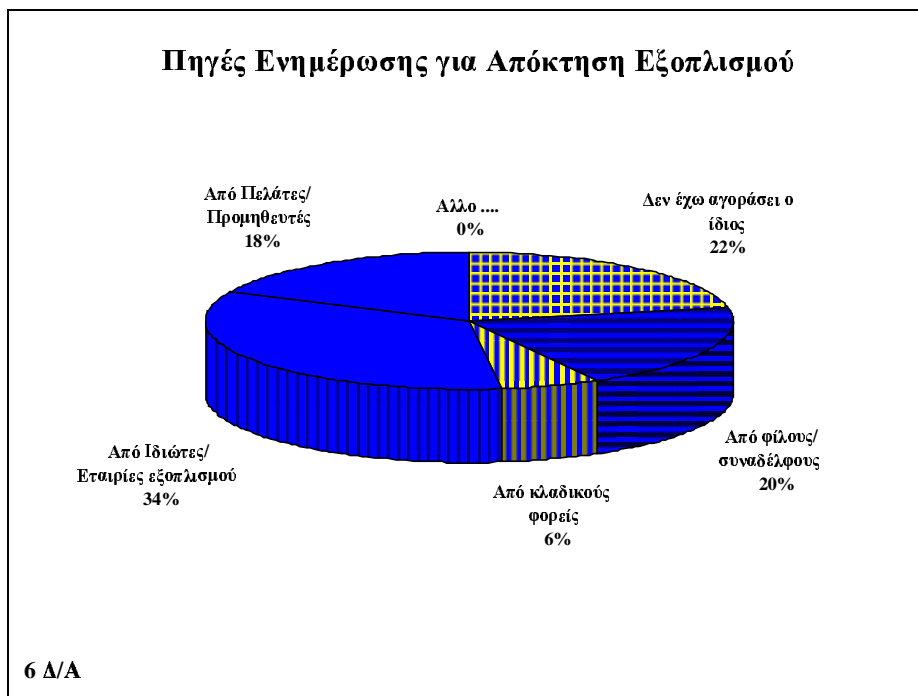


Εικόνα 6.12

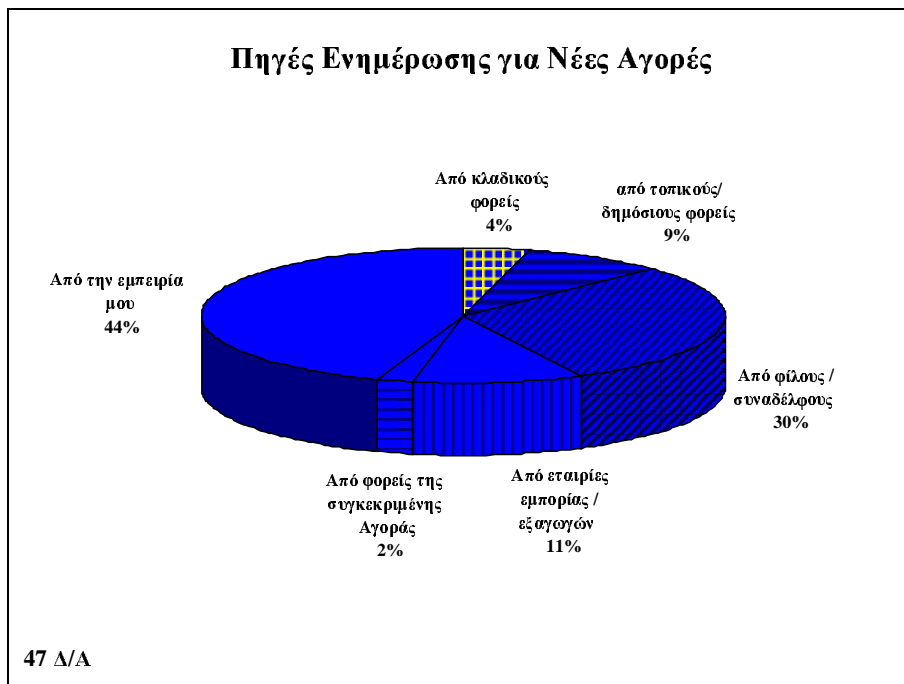
Στην εικόνα 6.12 βλέπουμε ποιες θεωρούν οι ερωτώμενοι τις σπουδαιότερες πληροφορίες για τις αγορές που δραστηριοποιούνται καθώς και για τα προϊόντα τους. Όπως είναι φυσικό όλες οι απαντήσεις έχουν να κάνουν με την αύξηση των κερδών και την θέση της επιχείρησης στην αγορά, είναι όμως εντυπωσιακό ότι όλοι απάντησαν πως τους απασχολούν οι τιμές των ανταγωνιστών τους, ενώ πάρα πολύ μεγάλα είναι τα ποσοστά που απάντησαν πως ενδιαφέρονται για το πλήθος των ανταγωνιστών τους καθώς και για τους τρόπους που χρησιμοποιούν οι ανταγωνιστές τους για την προώθηση και διακίνηση των προϊόντων τους.

Όπως φαίνεται και στην εικόνα 6.13 η κύρια πηγή ενημέρωσης για την απόκτηση εξοπλισμού είναι οι εξειδικευμένες εταιρίες εξοπλισμού σε ποσοστό 34% και ακολουθούν τόσο οι φίλοι ή οι συνάδελφοι με ποσοστό 20% όσο και οι

πελάτες ή οι προμηθευτές (18%). Και σε αυτή την περίπτωση η συνεισφορά των κλαδικών φορέων είναι μικρή.



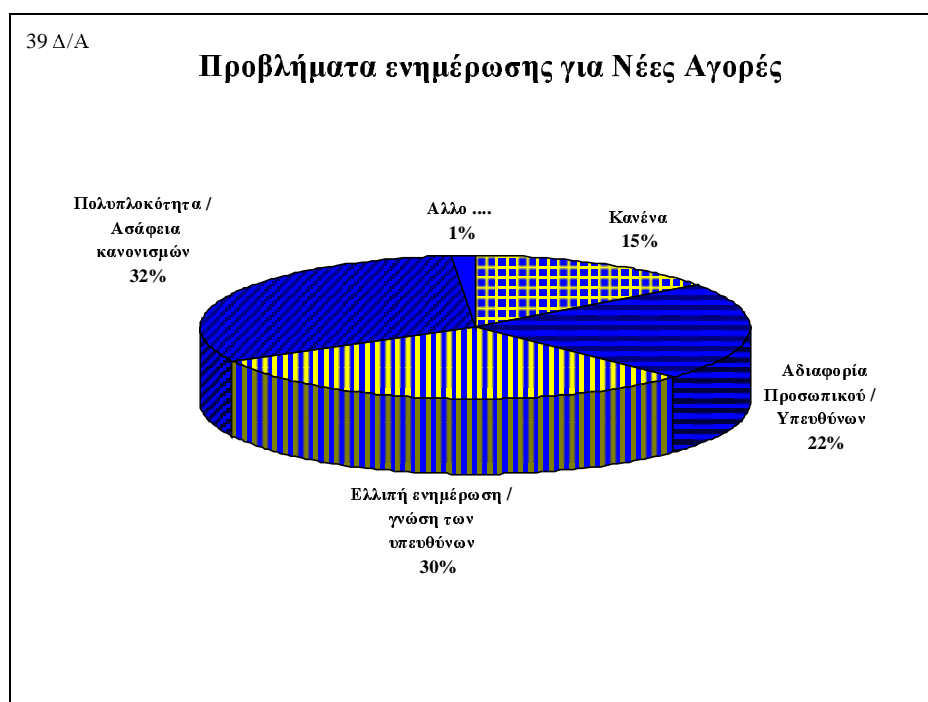
Εικόνα 6.13



Εικόνα 6.14

Η εικόνα 6.14 μας πληροφορεί για το ποια είναι η κύρια πηγή ενημέρωσης για την ύπαρξη νέων αγορών. Αν και οι απαντήσεις ποικίλουν το 34% βασίζεται στην προσωπική του εμπειρία και το 30% ενημερώνεται από φίλους και συναδέλφους. Το 11 % ενημερώνεται από εταιρίες εμπορίας και εξαγωγών και το 9% από δημόσιους φορείς. Εξαιρετικά χαμηλό είναι το ποσοστό που ενημερώνεται από φορείς της συγκεκριμένης αγοράς, μόλις 2%.

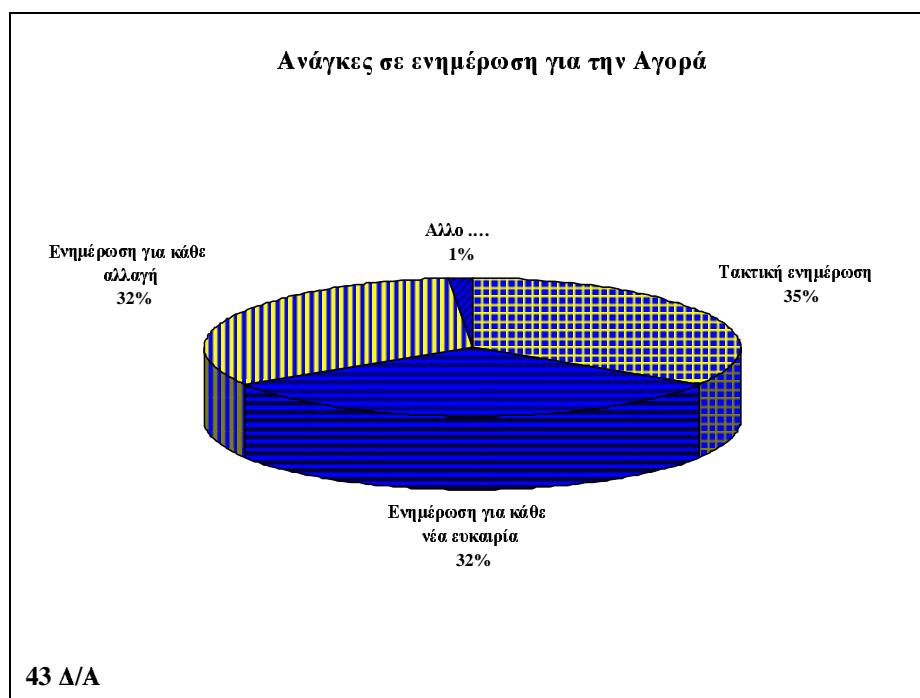
Άμεση σχέση με αυτά έχει και η ερώτηση 15 η οποία αφορά τα κυριότερα προβλήματα που εμφανίζονται στην ενημέρωση για τις νέες αγορές. Το 32% των ερωτηθέντων απάντησε πως υπάρχει μεγάλη πολυπλοκότητα και ασάφεια στους κανονισμούς, το 30% πως τόσο η ενημέρωση όσο και οι γνώσεις των υπευθύνων είναι ελλιπής και το 22% πως υπάρχει αδιαφορία από την πλευρά των υπευθύνων και του προσωπικού. Βέβαια υπήρξε και ένα ποσοστό 15% που απάντησε πως δεν αντιμετωπίσε κανένα πρόβλημα.



Εικόνα 6.15

Στην εικόνα 6.16 παρατηρούμε ότι οι ερωτούμενοι απάντησαν πως η κυριότερη ανάγκη είναι η ανάγκη για τακτική ενημέρωση με ποσοστό

απαντήσεων 35%, και ακολουθούν με το ίδιο ποσοστό (32%) τόσο η ανάγκη για ενημέρωση για κάθε αλλαγή όσο και η ανάγκη για ενημέρωση για κάθε νέα ευκαιρία που παρουσιάζεται. Είναι προφανές πως τα προβλήματα σε αυτή την περίπτωση είναι κοινά μεταξύ των ερωτηθέντων.



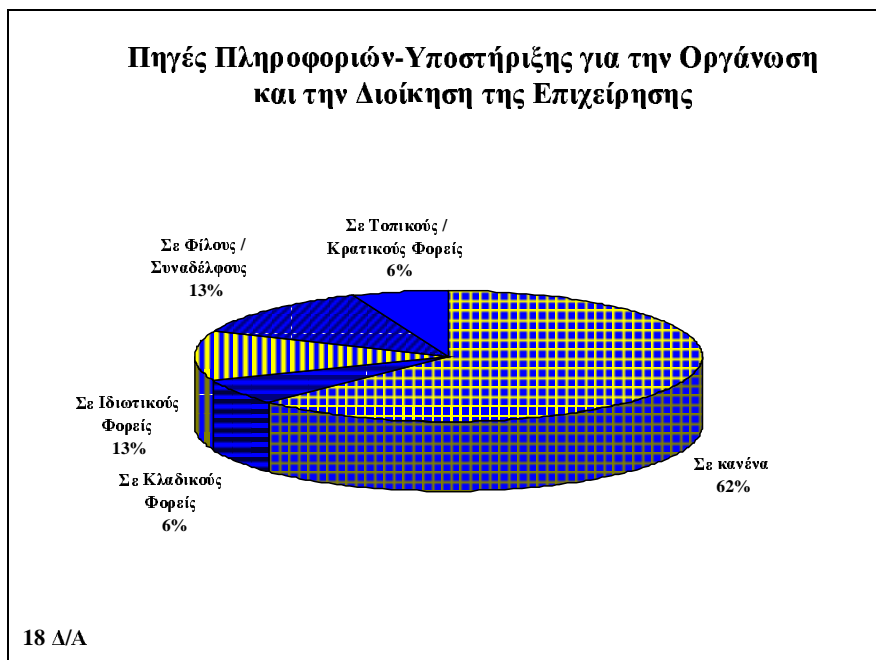
Εικόνα 6.16

Στις ερωτήσεις 17 και 18 οι ερωτούμενοι κλήθηκαν να απαντήσουν για το ποιες είναι οι κύριες πηγές πληροφόρησης και υποστήριξης καθώς και για το ποιες είναι οι σημαντικότερες πληροφορίες που θα τους βοηθήσουν στην οργάνωση και διοίκηση της επιχείρησής τους. Η συντριπτική πλειοψηφία (62%) απάντησε πως δεν απευθύνεται σε κανένα για την υποστήριξη της επιχείρησής του. Το 13% απάντησε πως απευθύνεται σε φίλους ή συναδέλφους και με το ίδιο ποσοστό οι ερωτούμενοι απάντησαν ότι εμπιστεύονται ιδιωτικούς φορείς. Μικρά είναι τα ποσοστά που εμπιστεύονται τοπικούς ή κρατικούς φορείς καθώς και κλαδικούς φορείς (6%).

Στην ερώτηση 18 το 29% θεωρεί πως σημαντικότερη πληροφορία για την διοίκηση της επιχείρησής του είναι η παροχή συμβουλών και ακολουθούν με

αρκετά υψηλά ποσοστά η ενημέρωση για νέες τεχνολογίες (27%) και η πληροφόρηση μέσω ενημερωτικών και εκπαιδευτικών προγραμμάτων. Πολύ υψηλό είναι το ποσοστό που απάντησε πως δεν θεωρεί καμία πληροφορία σημαντική.

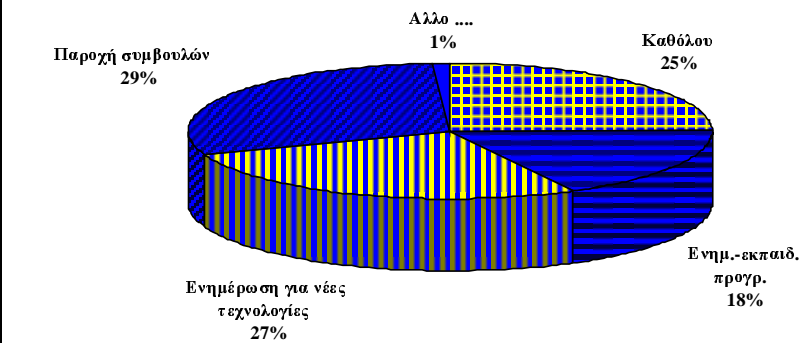
Βέβαια και η ερώτηση 19 είναι σε άμεση συνάρτηση με αυτά αφού το 47% πιστεύει πως τις πληροφορίες αυτές πρέπει να τις παρέχουν δημόσιοι φορείς, το 42% πως πρέπει να παρέχονται από κλαδικούς φορείς και το 9% από ιδιωτικούς φορείς.



Εικόνα 6.17



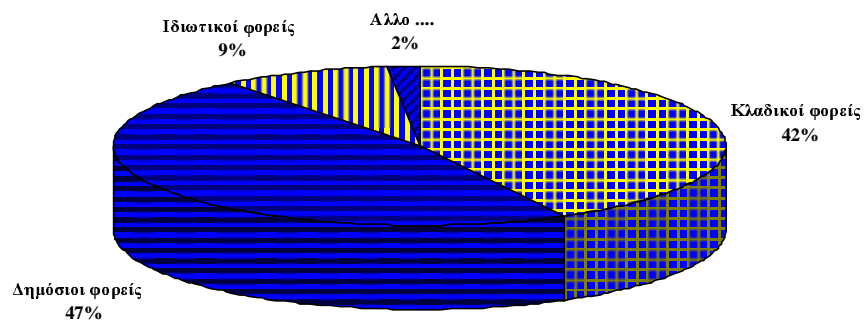
### Χρήσιμες Πληροφορίες για την Οργάνωση και Διοίκηση



10 Δ/Α

Εικόνα 6.18

### Φορέας που θα Επρεπε να Παρέχει τις Πληροφορίες Οργάνωσης και Διοίκησης



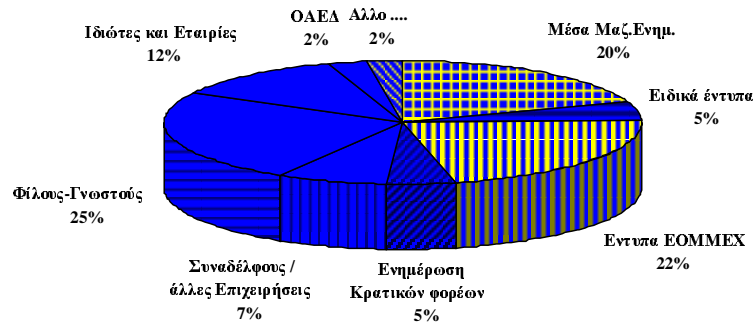
39 Δ/Α

Εικόνα 6.19

Στην ερώτηση 6.20 δόθηκε η δυνατότητα στους ερωτούμενους να εκφράσουν τις απόψεις τους για το ποιοι είναι οι κύριοι φορείς πληροφόρησης για τα προγράμματα εκπαίδευσης. Το 25% απευθύνεται σε φίλους ή γνωστούς και σε αυτή την περίπτωση, και το 22% πληροφορείται από έντυπα του EOMMEX. Μεγάλο ποσοστό των ερωτηθέντων (20%) απάντησε ότι κύριος φορέας ενημέρωσης είναι τα μέσα μαζικής ενημέρωσης, και 12% απάντησε πως είναι ιδιώτες ή εταιρίες. Ακόμη 7% απάντησαν πως κύριος φορέας είναι συνάδελφοι ή άλλες επιχειρήσεις, 5% ενημερώνεται από κρατικούς φορείς, 5% από ειδικά έντυπα και 2% από τον ΟΑΕΔ υπάρχει και ένα μικρό ποσοστό που επέλεξε κάποιον άλλο φορέα ενημέρωσης χωρίς όμως να διευκρινιστεί ποιος.

Στην εικόνα 6.21 που ακολουθεί μπορούμε να δούμε τις απαντήσεις των συμμετεχόντων στην έρευνα στην ερώτηση για το ποια είναι τα κύρια προβλήματα ενημέρωσης σε θέματα κατάρτισης. Το μεγαλύτερο πρόβλημα σύμφωνα με το 40% των ερωτούμενων είναι η έλλειψη ενημέρωσης. Ακολουθεί με επίσης υψηλό ποσοστό (33%) το πρόβλημα «αδυναμία» από την πλευρά των υπευθύνων και η αδιαφορία που επιδεικνύουν οι υπεύθυνοι με ποσοστό 23%. Τέλος υπάρχει και ένα μικρό ποσοστό της τάξεως του 2% που απάντησε πως σημαντικό πρόβλημα είναι οι διάφορες καθυστερήσεις που παρουσιάζονται σε θέματα ενημέρωσης. Και εδώ ένα μικρό ποσοστό καταδεικνύει κάποιο άλλο πρόβλημα.

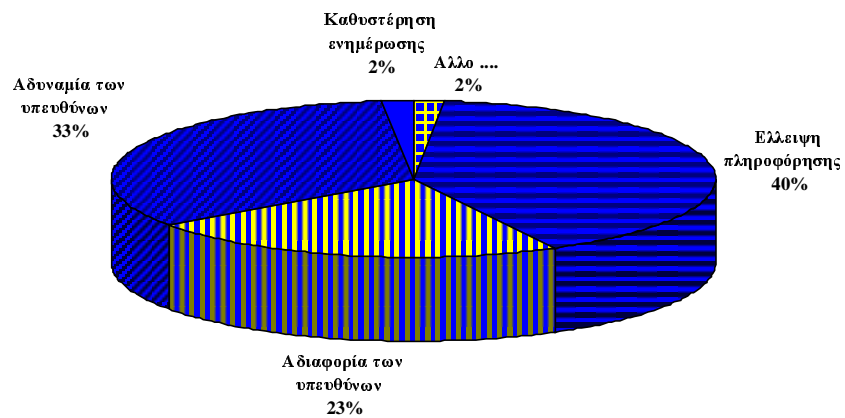
### Φορέας Πληροφόρησης για Προγράμματα εκπαίδευσης



92 Δ/Α

Εικόνα 6.20

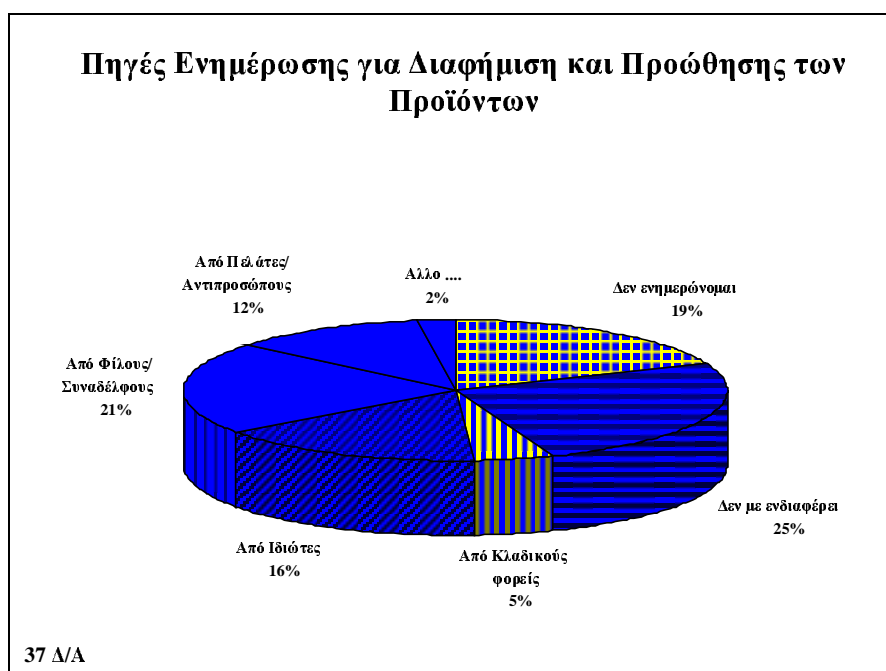
### Προβλήματα Ενημέρωσης για Θέματα Κατάρτισης



78 Δ/Α

Εικόνα 6.21

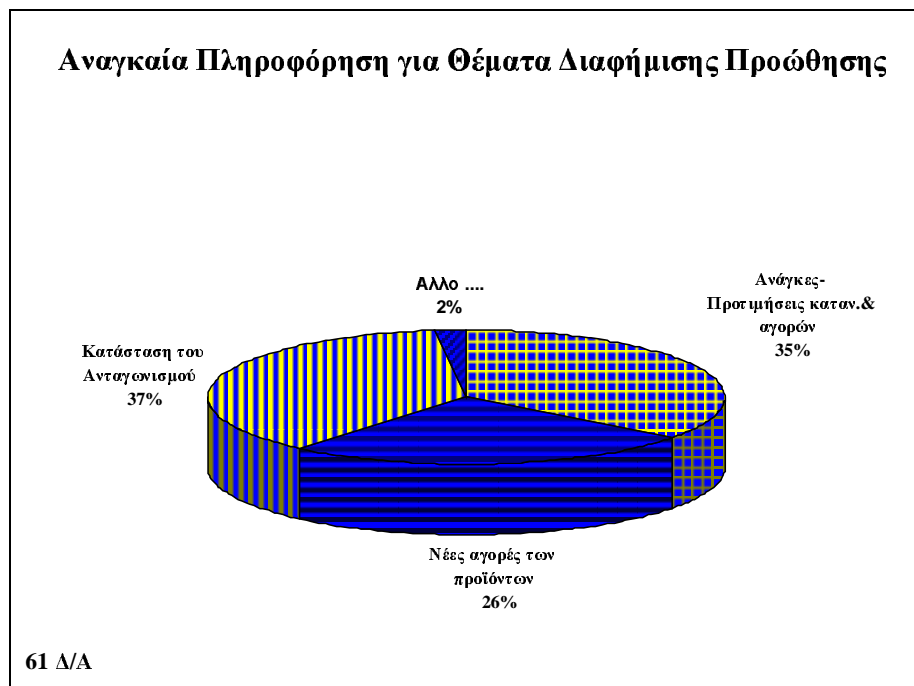
Περνώντας τώρα σε θέματα προώθησης και διαφήμισης των προϊόντων τους, στην ερώτηση 22, το 25% των ερωτούμενων απάντησαν πως κύρια πηγή ενημέρωσης πάνω σε τέτοια θέματα είναι και σε αυτή την περίπτωση φίλοι ή συνάδελφοί τους, 16% απευθύνεται σε ιδιώτες και 12% σε πελάτες ή αντιπροσώπους ενώ μικρό είναι το ποσοστό που ενημερώνεται από κλαδικούς φορείς. Σε αυτή την ερώτηση έκπληξη προκαλούν τα πολύ μεγάλα ποσοστά που είτε δεν ενημερώνονται καθόλου για θέματα διαφήμισης και προώθησης είτε δεν ενδιαφέρονται για τέτοια θέματα. Βέβαια αυτό πιθανόν να προέρχεται από το ότι η έρευνα έγινε στην μικρή σχετικά αγορά των Χανίων όπου σημαντικό ρόλο παίζουν οι διαπροσωπικές σχέσεις και συμπάθειες, καθώς και το γεγονός πως υπάρχουν λίγα και δεδομένα - τοπικά κυρίως - μέσα διαφήμισης.



Εικόνα 6.22

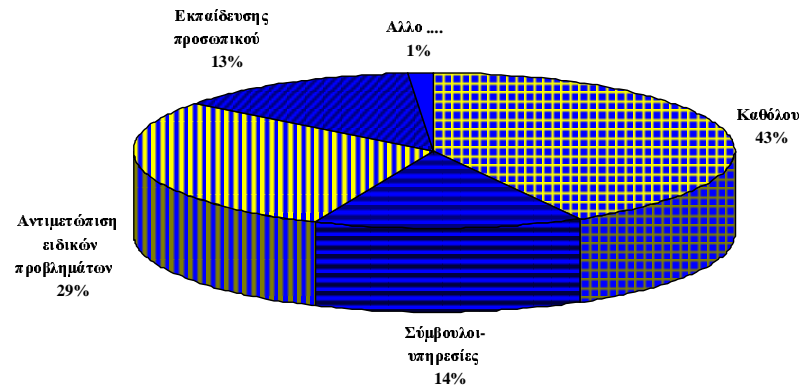
Στην εικόνα 6.23 οι ερωτώμενοι κατέδειξαν 3 βασικούς τομείς στους οποίους θεωρούν αναγκαία την πληροφόρηση για θέματα διαφήμισης και προώθησης. Σημαντικό είναι ότι το 37% ενδιαφέρεται περισσότερο για το τι κάνουν οι ανταγωνιστές του παρά για το πώς θα αυξήσει τις πωλήσεις των δικών του

προϊόντων, αφού πιστεύουν πως σημαντικότερη πληροφορία η κατάσταση των ανταγωνιστικών επιχειρήσεων. Νοοτροπία που σίγουρα δεν βοηθάει, αντίθετα πιθανόν είναι να δημιουργήσει καταστάσεις αθέμιτου ανταγωνισμού. Βέβαια υπάρχουν και διαφορετικές απόψεις από την πλειοψηφία αφού το 35% ενδιαφέρεται για τις ανάγκες και τις προτιμήσεις των καταναλωτών και το 26% για νέες αγορές προϊόντων και υπηρεσιών.



Εικόνα 6.23

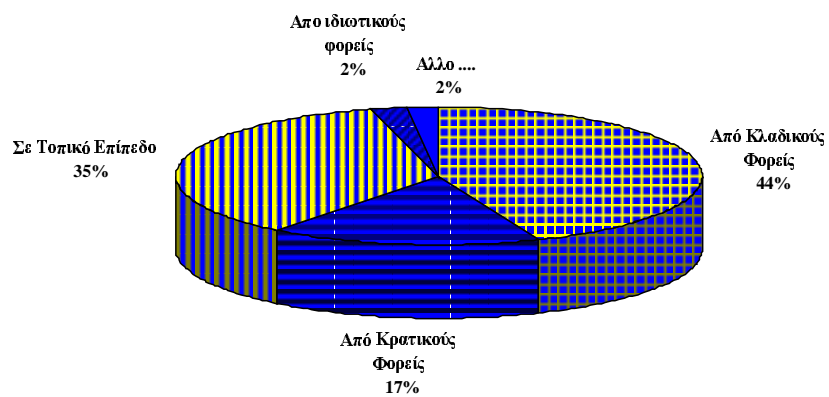
### Απαραίτητες Πληροφορίες για Θέματα Δημοσίων Σχέσεων



53 Δ/Α

Εικόνα 6.24

### Κατάλληλοι Φορείς Παροχής Πληροφοριών Δημοσίων Σχέσεων



84 Δ/Α

Εικόνα 6.25

Όσον αφορά θέματα δημοσίων σχέσεων στην ερώτηση 24 και εδώ όπως στην περίπτωση των διαφημίσεων το 43% δεν ασχολείται καθόλου, το 29% θεωρεί σημαντικότερη πληροφορία την αντιμετώπιση ειδικών προβλημάτων, το 14% απευθύνεται σε συμβούλους ή ειδικές υπηρεσίες και το 13% ενδιαφέρεται για θέματα εκπαίδευσης προσωπικού. Στην εικόνα 6.25 βλέπουμε πως το 44% πιστεύει πως κλαδικοί φορείς είναι οι καταλληλότεροι για θέματα ενημέρωσης πάνω στις δημόσιες σχέσεις το 35% απευθύνεται σε τοπικούς φορείς και το 17% σε κρατικούς φορείς, ενώ μόλις το 2% σε ιδιωτικούς φορείς.

## **6.2. Επεξεργασία των δεδομένων του ερωτηματολογίου με το πρόγραμμα See5**

### *6.2.1 Εισαγωγή-Παρουσίαση των μεταβλητών*

Στο σημείο αυτό θα επεξεργαστούμε με το πρόγραμμα νοημόνων προσεγγίσεων SEE5, τα δεδομένα που προέκυψαν. Η βάση δεδομένων δημιουργήθηκε με επιλογή των κυριότερων ερωτημάτων μέσα από το σύνολο των ερωτήσεων που περιλαμβάνονται στο ερωτηματολόγιό μας. Έτσι επιλέχθηκαν 26 ερωτήματα κάθε ένα από τα οποία αποτελεί και μια μεταβλητή. Έτσι έχουμε τελικά 25 ποιοτικές μεταβλητές και μια συνεχή, για κάθε μια από Τις οποίες δόθηκαν 141 απαντήσεις, και όλα αυτά τα στοιχεία αποθηκεύονται σε ένα νέο αρχείο excel. Οι τέσσερις πρώτες μεταβλητές αφορούν τα γενικά στοιχεία για τον κάθε ερωτούμενο -φύλο, ηλικία, επίπεδο μόρφωσης και γνώση ξένων γλωσσών. Από αυτές η μεταβλητή που αφορά την ηλικία είναι συνεχής. Οι υπόλοιπες μεταβλητές αν και έχουν ήδη αναφερθεί λεπτομερώς στην προηγούμενη παράγραφο, συνοπτικά είναι:

#### Χρήση υποστηρικτικών μηχανισμών:

Η μεταβλητή αυτή αφορά την χρησιμοποίηση υποστηρικτικών μηχανισμών για την κάθε επιχείρηση. Οι απαντήσεις ήταν: γραφεία μάρκετινγκ, οικονομικοί σύμβουλοι επιχειρήσεων, λογιστές και κανέναν υποστηρικτικό μηχανισμό.

Που είναι εγκατεστημένη η επιχείρηση;

Εδώ διερευνάται αν η επιχείρηση βρίσκεται σε ιδιόκτητες εγκαταστάσεις ή νοικιάζει κάποιο χώρο για τις ανάγκες της.

Γνωρίζετε τέτοιους χώρους στα Χανιά;

Η ερώτηση αυτή διερευνά αν οι συμμετέχοντες στην έρευνα γνωρίζουν βιοτεχνικά πάρκα στην πόλη των Χανίων.

Υπάρχουν κίνητρα για επέκταση των επιχειρήσεων;

Εδώ προσπαθούμε να δούμε κατά πόσο ωθούνται οι επιχειρηματίες να επεκτείνουν τις επιχειρήσεις τους. Οι απαντήσεις ήταν κλιμακωτές: καθόλου, λίγο, αρκετά, πολλά.

Έχετε κάνει αίτηση για δάνειο;

Εδώ οι απαντήσεις αφορούσαν το πόσες φορές έχουν πάρει οι ερωτώμενοι δάνειο: καμία, 1 φορά, 2-3 φορές και περισσότερες από 3 φορές.

Ενημερώνεστε για νέες αγορές ή προϊόντα;

Η ερώτηση αυτή ερευνά τον βαθμό που οι επιχειρηματίες ενδιαφέρονται για τις νέες ευκαιρίες που παρουσιάζονται στην αγορά.

Σε ποιες αγορές πουλάτε;

Η μεταβλητή αυτή ερευνά σε ποιες αγορές δραστηριοποιούνται οι επιχειρήσεις που συμμετέχουν στην έρευνα. οι απαντήσεις που δόθηκαν είναι: Χανιά, Κρήτη, Ελλάδα, Αμερική και Ευρωπαϊκή Ένωση.

Γνωρίζετε το θεσμικό πλαίσιο για την προώθηση προϊόντων;



Και εδώ υπάρχει διαβάθμιση στις απαντήσεις (καθόλου, λίγο, αρκετά, πλήρως) οι οποίες διερευνούν κατά ποσό οι ερωτούμενοι γνωρίζουν το θεσμικό πλαίσιο που αφορά την προσπάθεια για αύξηση των πωλήσεων των προϊόντων τους.

Από πού προέρχονται οι προμηθευτές:

Οι απαντήσεις αφορούν το μέρος από το οποίο προέρχονται οι προμηθευτές των επιχειρήσεων.

Από πού προέρχονται οι πελάτες:

Σε συνάρτηση με την προηγούμενη ερώτηση για να διαπιστωθεί σε ποια γεωγραφικά διαμερίσματα γίνονται οι πωλήσεις.

Πως πουλάτε τα προϊόντα σας εκτός από απευθείας στους καταναλωτές:

Εδώ διερευνάται αν υπάρχει συνεργασία όσον αφορά τις πωλήσεις με συνεργάτες, αντιπροσώπους ή χονδρεμπόρους από την επιχείρηση που συμμετείχε στην έρευνα.

Πως διοικείτε την επιχείρησή σας:

Η μεταβλητή αυτή ερευνά αν η διοίκηση των επιχειρήσεων που συμμετείχαν στην έρευνα γινόταν προσωπικά από τον φορέα τους ή σε συνεργασία με ειδικό προσωπικό και εξειδικευμένα γραφεία συμβούλων.

Τι διοικητικά προβλήματα αντιμετωπίζετε:

Εδώ προσπαθήσαμε να διαπιστώσουμε το είδος των προβλημάτων που αντιμετωπίζονται κατά την διοίκηση των επιχειρήσεων. Οι απαντήσεις αφορούσαν: έλλειψη ενημέρωσης από τους αρμόδιους φορείς, έλλειψη προσωπικού, το μικρό μέγεθος της επιχείρησης, όμως υπάρχουν και επιχειρήσεις που δεν αντιμετωπίζουν προβλήματα.

Συμμετέχετε σε προγράμματα κατάρτισης προσωπικού:

Οι επιχειρηματίες κλήθηκαν να απαντήσουν απλά με ναι /όχι χωρίς να ζητούνται περαιτέρω διευκρινίσεις.

Πώς διαφημίζετε / προωθείτε τα προϊόντα σας:

Η μεταβλητή αυτή διερευνά τους τρόπους με τους οποίους οι επιχειρήσεις προσπαθούν να αυξήσουν τις πωλήσεις τους. Οι απαντήσεις ήταν: στα μέσα μαζικής ενημέρωσης, στον ειδικό τύπο, μέσω χορηγιών, ενώ υπάρχουν και πολλές επιχειρήσεις που δεν διαφημίζουν / προωθούν τα προϊόντα τους.

Πώς ενημερώνεστε για θέματα προώθησης ή διαφήμισης:

Σε άμεση συνάρτηση με την προηγούμενη ερώτηση οι επιχειρηματίες απάντησαν για το πώς μαθαίνουν τρόπους ώστε να προωθήσουν ή να διαφημίσουν τα προϊόντα τους.

Τι προβλήματα αντιμετωπίζετε με την προώθηση / διαφήμιση των προϊόντων σας:

Τα προβλήματα που οι επιχειρηματίες απάντησαν πως αντιμετωπίζουν στην προσπάθειά τους να προωθήσουν τα προϊόντα τους είναι το υψηλό κόστος, η αναποτελεσματικότητα και η έλλειψη κατάλληλου προσωπικού.

Τι πληροφόρηση θέλετε για την καλύτερη προώθηση / διαφήμιση των προϊόντων σας:

Εδώ διερευνάται ποιες πληροφορίες βοηθούν τους επιχειρηματίες στην προσπάθειά τους να αυξήσουν τις πωλήσεις των προϊόντων τους.

Με ποιους έχετε προβλήματα επικοινωνίας:

Εδώ οι επιχειρηματίες κλήθηκαν να απαντήσουν αν έχουν προβλήματα στην συνεργασία τους με τους πελάτες, τους προμηθευτές και την τοπική κοινωνία.

Τι προβλήματα επικοινωνίας αντιμετωπίζετε;

Εδώ γίνονται γνωστά τα προβλήματα που αντιμετωπίζουν οι επιχειρήσεις σε θέματα επικοινωνίας.

Θέλετε να εγκατασταθείτε σε οργανωμένο χώρο;

Η τελευταία ερώτηση η οποία θα συνδυαστεί με όλες τις προηγούμενες έτσι ώστε να εξαγάγουμε συμπεράσματα για τα αποτελέσματα της έρευνας μέσω ερωτηματολογίων. Διερευνά αν οι επιχειρηματίες που συμμετέχουν στην έρευνα είναι θετικοί σε ενδεχόμενη συμμετοχή τους στο βιοτεχνικό πάρκο Χανίων.

### *6.2.2 Μεθοδολογία*

Στη συγκεκριμένη εφαρμογή ακολουθήθηκε μια συγκεκριμένη μεθοδολογία που αποσκοπεί στην εξαγωγή χρήσιμων συμπερασμάτων μέσω του προγράμματος see5, και η οποία συνοπτικά περιλαμβάνει τα εξής βήματα:

- Προετοιμασία δυο αρχείων με τις προεκτάσεις «data» και «names».
- Αποθήκευση του αρχείου excel ως «csv».
- Αντικατάσταση των πολλαπλών κομμάτων που εμφανίζονται με μονά κόμματα, και αλλαγή της προέκτασης «csv» σε «data».
- Δημιουργία του αρχείου «names», στο οποίο ανά γραμμή αναγράφεται το όνομα του κάθε χαρακτηριστικού, ακολουθούμενο από άνω και κάτω τελεία και από τις τιμές οι οποίες χωρίζονται με

κόμμα. Τα ερωτηματικά (?) παραλείπονται, αφού αναπαριστούν τις άγνωστες τιμές. Οι συνεχείς τιμές δεν γράφονται και την θέση τους παίρνει η λέξη «continuous»

Πρέπει να σημειωθεί ότι οι τιμές στο αρχείο «names» αναγράφονται με την ίδια σειρά όπως και στην βάση δεδομένων. Για κάθε ανακολουθία το see5 ειδοποιεί τον χρήστη κατά την εκτέλεση του αλγορίθμου. Γενικά κόμματα, κενά και ερωτηματικά δεν πρέπει να χρησιμοποιούνται στην ονομασία των μεταβλητών.

Στην συνέχεια ακολουθεί το τρέξιμο του προγράμματος.

### 6.2.3 Τα αποτελέσματα του προγράμματος

Χρησιμοποιήθηκαν και οι 26 μεταβλητές. Κλάση απόφασης είναι η ερώτηση «θέλετε να εγκατασταθείτε σε οργανωμένο χώρο». Το πρόγραμμα παράγαγε 33 κανόνες.

Τα αποτελέσματα έχουν ως εξής:

```
See5 [Release 1.16] Mon Sep 09 17: 14:45 2002
Options:
Rule-based classifiers
Pruning confidence level 30%
Test requires 2 branches with >= 1 cases
Class specified by attribute
'THELETE_NA_EGATASTATHITE_SE_ORGANOMENO_XORO'
Read 141 cases (26 attributes) from business-park.data
Rules:
Rule 1: (8, lift 1.9)
LANGUAGES = KAMIA
NOMIKI_MORFI = ATOM
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
POU_EINAI_EGATASTIMENI_H_EPIX = ENOIKIO
IPARXOYN_KINITRA_GIA_EPEKTASI_TON_EPIX = KANENA
-> class Y [0.900]

Rule 2: (7, lift 1.9)
LANGUAGES = KAMIA
NOMIKI_MORFI = ATOM
IPARXOYN_KINITRA_GIA_EPEKTASI_TON_EPIX = KANENA
ME_POIOYS_EXETE_PROBLIMATA_EPIKOINONIAS = PELATES_PROMITHEYTES
-> class Y [0.889]

Rule 3: (4, lift 1.8)
ILIKIA > 28
```

```

LANGUAGES = ENG
NOMIKT_MORFI = OE
POS_DIAFIMIZETE_PROOTHITE_TA_PROIONTA_SAS = DEN_DIAFIMIZOMAI
POS_ENIMERONESTE_GIA_THEMATA_PROOTHISIS_DIAFIMISIS =
DEN_ENIMERONOMAI
-> class Y [0.833]

Rule 4: (4, lift 1.8)
XRISI_IPOSTIRIKTIKON_MTXAN = GRAFEIA_MARKETING
GNORIZETE_TETOIOYS_XOROYS_STA_XANTA = Y
-> class Y [0.833]

Rule 5: (4, lift 1.8)
XRISI_IPOSTIRIKTIKON_MTXAN =
GRAFEIA_MARKETING&LOGIS_GNORIZETE_TETOIOYS_XOROYS_STA_XANTA = Y
-> class Y [0.833]

Rule 6: (4, lift 1.8)
XRISI_IPOSTIRIKTIKON_MTXAN = LOGIS_
SIMMETEXETE_SE_PROGRAMMATA_KATARTISIS_PROSOPIKOY = Y
ME_POIOYS_EXETE_PROBLIMATA_EPIKOINONIAS = KANENA
-> class Y [0.833]

Rule 7: (3, lift 1.7)
ILIKTA <= 31
LANGUAGES = KAMTA
XRISI_IPOSTIRIKTIKON_MTXAN = OTK_SIMBOYLOI_EPTX_
-> class Y [0.800]

Rule 8: (3, lift 1.7)
LANGUAGES = ENG-GER
-> class Y [0.800]

Rule 9: (3, lift 1.7)
LANGUAGES = ENG
NOMIKT_MORFI = ATOM
POS_POYLATE_TA_PROIONTA_EKTOS_APO_APEYTHEIAS_STOYS_KATANALOTES =
MONO APEYTHEIAS
TI_PROBLIMATA_ANTIMETOPIZETE_ME_THN_PROOTHISI_DIAFIMISI_PROIONTON
= KANENA
-> class Y [0.800]

Rule 10: (3, lift 1.7)
XRISI_IPOSTIRIKTIKON_MTXAN = LOGIS_
POS_DIAFIMIZETE_PROOTHITE_TA_PROIONTA_SAS = STA MME
-> class Y [0.800]

Rule 11: (8/1, lift 1.7)
GNORIZETE_TETOIOYS_XOROYS_STA_XANTA = Y
POS_DIAFIMIZETE_PROOTHITE_TA_PROIONTA_SAS = STON_EDIKO_TYPO
-> class Y [0.800]

Rule 12: (2, lift 1.6)
LANGUAGES = KAMTA
NOMIKI_MORFI = ATOM
IPARXOYN_KINITRA_GIA_EPEKTASI_TON_EPTX = ARKETA_KALA
TI_DIOIKTIKA_PROBLIMATA_ANTIMETOPIZETE = ELEIPSI_PROSOPIKOU
-> class Y [0.750]

Rule 13: (2, lift 1.6)
LANGUAGES = GER
XRISI_IPOSTIRIKTIKON_MTXAN = OTK_SIMBOYLOI_EPTX

```

```

-> class Y [0.750]

Rule 14: (10/2, lift 1.6)
LANGUAGES = ENG
NOMIKI_MORFI = ATOM
GNORIZETE_TETOIOYS_XOROYS_STA_XANIA = Y
TI_PROBLIMATA_ANTIMETOPIZETE_ME_THN-PROOTHISI_PROIONTON =
YPSILO_KOSTOS
-> class Y [0.750]

Rule 15: (2, lift 1.6)
NOMIKI_MORFI = EE
-> class Y [0.750]

Rule 16: (18/5, lift 1.5)
XRISI_IPOSTIRIKTIKON_MIXAN =
OTK_SIMBOYLOI_EPIX_&GRAFEIA_MARKETING
GNORIZETE_TETOIOYS_XOROYS_STA_XANIA = Y
-> class Y [0.700]

Rule 17: (1, lift 1.4)
LANGUAGES = ENG-FR
GNORIZETE_TETOIOYS_XOROYS_STA_XANIA = Y
-> class Y [0.667]

Rule 18: (1, lift 1.4)
NOMIKI_MORFI = OE
XRISI_IPOSTIRIKTIKON_MIXAN = OTK_SIMBOYLOI_EPIX
POS_ENIMERONESTE_GIA_THEMATA_PROOTHISIS_DIAFIMISIS =
FILOYS_SINADELFOYS
-> class Y [0.667]

Rule 19: (10, lift 1.7)
LANGUAGES = KAMIA
XRISI_IPOSTIRIKTIKON_MIXAN = OTK_SIMBOYLOI_EPIX
POY_ETNAI_EGATASTIMENI_H_EPIX = IDIOKTITES_EGATASTASEIS
ME_POIOYS_EXETE_PROBLIMATA_EPIKOINONIAS = KANENA
-> class N [0.917]

Rule 20: (8, lift 1.7)
ILIKIA > 31
LANGUAGES = KAMIA
NOMIKI_MORFI = ATOM
XRISI_IPOSTIRIKTIKON_MIXAN = OTK_SIMBOYLOI_EPIX
IPARXOYN_KINITRA_GIA_EPEKTASI_TON_EPIX = LIGO
-> class N [0.900]

Rule 21: (5, lift 1.6)
ILIKIA > 27
XRISI_IPOSTIRIKTIKON_MIXAN = KANENAN
-> class N [0.857]

Rule 22: (4, lift 1.6)
XRISI_IPOSTIRIKTIKON_MIXAN = OTK_SIMBOYLOI_EPIX
TI_DIOIKTIKA_PROBLIMATA_ANTIMETOPIZETE = MH_ENHMEROSI_APO_FOREIS
-> class N [0.833]

Rule 23: (4, lift 1.6)
XRISI_IPOSTIRIKTIKON_MIXAN = OTK_SIMBOYLOI_EPIX
POS_POYLATE_TA_PROIONTA_EKTOS_APO_APEYTHEIAS_STOUS_KATANALOTES =
SINERGAZOMENA

```

```

TI_PROBLIMATA_ANTIMETOPIZETE_ME_THN_PROOTHISI_PROIONTON = KANENA
-> class N [0.833]

Rule 24: (16/2, lift 1.6)
XRISI_IPOSTIRIKTIKON_MIXAN = LOGIS_
SIMETEXETE_SE_PROGRAMMATA_KATARTISIS_PROSOPIKOY = N
POS_DIAFIMIZETE_PROOTHITE_TA_PROIONTA_SAS = DEN_DIAFIMIZOMAI
-> class N [0.833]

Rule 25: (9/1, lift 1.6)
GNORIZETE_TETOIOYS_XOROYS_STA_XANIA = N
-> class N [0.818]

Rule 26: (3, lift 1.5)
ILIKTA <= 28
NOMIKT_MORFI = OE
-> class N [0.800]

Rule 27: (3, lift 1.5)
LANGUAGES = KAMTA
IPARXOYN_KINITRA_GIA_EPEKTASI_TON_EPIX = ARKETA_KALA
TI_DIOIKITIKA_PROBLIMATA_ANTIMETOPIZETE = KANENA
-> class N [0.800]

Rule 28: (2, lift 1.4)
NOMIKT_MORFI = OE
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
POS_DIAFIMIZETE_PROOTHITE_TA_PROIONTA_SAS = STON_ETDIKO_TYPO
-> class N [0.750]

Rule 29: (2, lift 1.4)
NOMIKT_MORFI = AE
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
-> class N [0.750]

Rule 30: (2, lift 1.4)
LANGUAGES = ENG_FR_GER
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
-> class N [0.750]

Rule 31: (16/4, lift 1.4)
LANGUAGES = ENG
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
TI_PROBLIMATA_ANTIMETOPIZETE_ME_THN_PROOTHISI_DIAFIMISI_PROIONTON
= ANAPOTELESMATIKOTITA
-> class N [0.722]

Rule 32: (13/4, lift 1.3)
NOMIKT_MORFI = OE
XRISI_IPOSTIRIKTIKON_MIXAN = OIK_SIMBOYLOI_EPIX
-> class N [0.667]

Rule 33: (1, lift 1.3)
LANGUAGES = ENG_FR_GER_IT
-> class N [0.667]

Default class: N
Evaluation on training data (141 cases):
Rules
-----
No          Errors

```

```

33          10(7.1%) <<
(a) (b)      <-classified as
--- ---
62 5   (a): class Y
5 69   (b): class N

Time: 0.1 secs

```

Όπως φαίνεται το μοντέλο κατανοεί ξεκάθαρα τα Yes ή No cases και παράγει αρκετά λογικούς κανόνες. Μπορούμε να διαπιστώσουμε πως 62 επιχειρηματίες είναι θετικοί και 69 είναι αρνητικοί στην συμμετοχή τους στο βιοτεχνικό πάρκο Χανίων. Το ποσοστό λάθους είναι μόλις 7.1%. Ας κοιτάξουμε κάποιους από τους κανόνες, ενδεικτικά.

Στον πρώτο κανόνα:

- Ο επιχειρηματίας δεν γνωρίζει ξένες γλώσσες.
- Δεν υπάρχουν κίνητρα για την επέκταση της επιχείρησής του.
- Η νομική μορφή είναι ατομική επιχείρηση.
- Η επιχείρηση νοικιάζει τους χώρους που είναι εγκατεστημένη.
- Χρησιμοποιούνται οικονομικοί σύμβουλοι.

Σε αυτή την περίπτωση ο επιχειρηματίας είναι θετικός να εγκατασταθεί στο βιοτεχνικό πάρκο, με ποσοστό σωστής πρόβλεψης 90%. Αυτό είναι αναμενόμενο αφού επιβεβαιώνεται και από την κοινή λογική αφού, εκτός των άλλων, ο εν λόγω επιχειρηματίας δεν έχει ιδιόκτητες εγκαταστάσεις, ψάχνει για κίνητρα ώστε να επεκτείνει την επιχείρησή του και χρησιμοποιεί υποστηρικτικούς μηχανισμούς, γεγονότα που καταδεικνύουν ακριβώς το συμπέρασμα του μοντέλου.

Στον δεύτερο κανόνα:



- Ο επιχειρηματίας δεν γνωρίζει ξένες γλώσσες.
- Πρόκειται για ατομική επιχείρηση.
- Δεν υπάρχουν κίνητρα για την επέκταση της επιχείρησης.
- Ο επιχειρηματίας αντιμετωπίζει προβλήματα επικοινωνίας με τους πελάτες ή τους προμηθευτές.

Εδώ ο επιχειρηματίας θέλει να εγκαταστήσει την επιχείρησή του στο βιοτεχνικό πάρκο, με ποσοστό σωστής πρόβλεψης 88,9%. Είναι λογικό αφού αυτή η λύση ίσως τον οδηγήσει στην λύση των προβλημάτων επικοινωνίας και βοηθήσει στο να επεκτείνει την επιχείρησή του.

Στον τρίτο κανόνα:

- Ο επιχειρηματίας είναι 28 χρονών και μιλάει αγγλικά.
- Πρόκειται για ομόρρυθμη επιχείρηση.
- Η επιχείρηση δεν διαφημίζει τα προϊόντα της.
- Ο επιχειρηματίας δεν ενημερώνεται για θέματα προώθησης και διαφήμισης.

Όπως στις προηγούμενες δυο περιπτώσεις ο επιχειρηματίας επιθυμεί την ένταξή του στο βιοτεχνικό πάρκο, με ποσοστό σωστής πρόβλεψης 83,3%.

Στον κανόνα Νο.14:

- Ο επιχειρηματίας μιλάει αγγλικά.
- Ο επιχειρηματίας γνωρίζει παρόμοιους χώρους στα Χανιά.

- Η νομική μορφή της επιχείρησης είναι ατομική επιχείρηση, και υπάρχει πρόβλημα λόγω υψηλού κόστους στην προώθηση / διαφήμιση των προϊόντων της επιχείρησης.

Και εδώ ο επιχειρηματίας είναι θετικός για την ένταξη της επιχείρησής του στο βιοτεχνικό πάρκο με ποσοστό σωστής πρόβλεψης 75%. Είναι και σε αυτή την περίπτωση φανερό πως ο επιχειρηματίας ψάχνει μια νέα ευκαιρία έτσι ώστε να αναπτύξει την επιχείρησή του κρατώντας το κόστος σε χαμηλά επίπεδα.

Στον κανόνα Νο.19:

- Ο επιχειρηματίας δεν γνωρίζει ξένες γλώσσες.
- Η επιχείρηση είναι εγκατεστημένη σε ιδιόκτητο χώρο.
- Χρησιμοποιούνται οικονομικοί σύμβουλοι.
- Δεν αντιμετωπίζονται προβλήματα επικοινωνίας.

Εδώ ο επιχειρηματίας δεν θέλει να εγκατασταθεί στο βιοτεχνικό πάρκο, με ποσοστό σωστής πρόβλεψης 91, 7%. Είναι φανερό πως ο συνδυασμός αυτών των δεδομένων και κυρίως το γεγονός ότι υπάρχουν ιδιόκτητες εγκαταστάσεις και δεν υπάρχουν προβλήματα επικοινωνίας, επηρεάζει αρνητικά την απόφαση.

Επίσης, στον κανόνα Νο.27:

- Ο επιχειρηματίας δεν γνωρίζει ξένες γλώσσες.
- Υπάρχουν αρκετά κίνητρα για την επέκταση της επιχείρησης.
- Δεν αντιμετωπίζονται διοικητικά προβλήματα.

Ούτε σε αυτή την περίπτωση ο επιχειρηματίας θέλει να εγκατασταθεί στο βιοτεχνικό πάρκο, με ποσοστό σωστής πρόβλεψης 80%. Φαίνεται πως η έλλειψη

προβλημάτων και η ύπαρξη ευκαιριών ανάπτυξης αποθαρρύνουν τον εν λόγω επιχειρηματία.

Οι περισσότεροι κανόνες εξάγουν λογικά αποτελέσματα, και τα λάθη που εμφανίζονται οφείλονται περισσότερο στον δικό μας τρόπο μοντελοποίησης.

#### 6.2.4 Άλλες δυνατότητες του *See5*

Το πρόγραμμα είναι ένα πολύ χρήσιμο εργαλείο και αυτό αποδεικνύεται από δυο ακόμα δυνατότητες που παρέχει και οι οποίες είναι πολύ χρήσιμες για το σύγχρονο μάρκετινγκ και την έρευνα αγοράς.

(α) Πατώντας το πλήκτρο με την ένδειξη «Cross-Reference» το πρόγραμμα παρουσιάζει το σύνολο των ερωτήσεων, τονίζοντας τις περιπτώσεις που διέγνωσε κάποιο λάθος. Έτσι μπορούμε οποιαδήποτε στιγμή να ανατρέξουμε στην λίστα και να διαπιστώσουμε τις λάθος περιπτώσεις.

(β) Υπάρχει η δυνατότητα μέσω του πλήκτρου «Use Classifier» να γίνεται για κάποιον επιχειρηματία εισαγωγή ενός μέρους των ερωτήσεων και το πρόγραμμα να ταξινομεί αυτόματα αν ο επιχειρηματίας είναι θετικός ή αρνητικός για την ένταξη της επιχείρησής του στο βιοτεχνικό πάρκο. Αυτό είναι ένα εξαιρετικά χρήσιμο εργαλείο αφού επιτρέπει στον μάρκετινγκ -μάνατζερ να διαπιστώνει εύκολα την βούληση του κάθε επιχειρηματία.

## **7. ΠΡΑΓΜΑΤΙΚΗ ΕΦΑΡΜΟΓΗ ΠΟΣΟΤΙΚΗΣ ΑΝΑΛΥΣΗΣ: «ΑΝΑΛΥΣΗ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ ICAP»**

### **7.1. Η μελέτη ερωτηματολογίου με τη βοήθεια της Q- Analysis**

#### *7.1.1 Η μεθοδολογία της Μελέτης.*

Θα μελετήσουμε τώρα με τη βοήθεια της Q-Analysis τα αποτελέσματα ερωτηματολογίου που έγινε από την ICAP (επιθυμία της ICAP ήταν η αντιστοιχία των επιχειρήσεων με αύξοντες αριθμούς) για λογαριασμό της εταιρείας την οποία και εμείς εξετάζουμε στην εργασία αυτή. Από το ερωτηματολόγιο αυτό επιλέχτηκαν τέσσερις ερωτήσεις τις οποίες απάντησαν 25 εταιρείες που αγοράζουν Ελληνικό χαρτόνι και καλύπτουν περίπου το 90% του χώρου της Ελληνικής αγοράς χαρτονιού. Από τις 25 επιχειρήσεις τέσσερις μόνο δεν αγοράζουν χαρτόνι από την εταιρεία μας και είναι οι 1, 2, 3 και 8. Από τις υπόλοιπες που αγοράζουν έξι είναι ευχαριστημένες από την εταιρεία μας σαν προμηθευτή και είναι οι επιχειρήσεις 4, 10, 14, 15, 18 και 24. Οι υπόλοιπες είναι δυσαρεστημένες. Τα αποτελέσματα των τεσσάρων ερωτήσεων του ερωτηματολογίου της ICAP για να μπορέσουμε να τα επεξεργαστούμε με τον αλγόριθμο της Q-Analysis το κωδικοποιούμε σε τρεις πίνακες δεδομένων.

Ο πρώτος πίνακας περιλαμβάνει τις απαντήσεις για τις ερωτήσεις 3 και 4 του ερωτηματολογίου της ICAP και αναφέρεται στα τεχνικά προβλήματα που παρουσιάζονται στα προϊόντα που παράγει η εταιρεία μας.

Ο δεύτερος πίνακας δεδομένων περιλαμβάνει τις απαντήσεις για την ερώτηση 10 του ερωτηματολογίου της ICAP και αναφέρεται στα προτερήματα των προϊόντων αλλά και της γενικότερης εμπορικής πολιτικής των ανταγωνιστών της εταιρείας μας.

Τέλος ο τρίτος πίνακας δεδομένων περιλαμβάνει τις απαντήσεις για τις ερωτήσεις του ερωτηματολογίου της ICAP και αναφέρεται στα προτερήματα

των προϊόντων αλλά και της εμπορικής πολιτικής γενικότερα της εταιρείας που εξετάζουμε.

Οι απαντήσεις των ερωτήσεων είναι ΝΑΙ ή ΟΧΙ και εισάγονται στο αλγόριθμο της Q-Analysis για επεξεργασία με την μορφή 0-1. Ο πίνακας 4 δεικνύει τη μορφή εισαγωγής των δεδομένων στον αλγόριθμο της Q-Analysis.

### *7.1.2 Μελέτη του πρώτου πίνακα δεδομένων με τη βοήθεια της Q-Analysis.*

Όπως είπαμε και πριν, ο πρώτος πίνακας δεδομένων αναπαριστά κωδικοποιημένες τις απαντήσεις από τις ερωτήσεις 3 και 4 του ερωτηματολογίου της ICAP. Οι ερωτήσεις αυτές δεικνύει ο πίνακας 7 και ο πίνακας 8. Από την συγχώνευση των δύο αυτών ερωτήσεων παίρνουμε τεχνικά προβλήματα που βρίσκουν οι αγοραστές στα προϊόντα της εταιρείας μας. Στους παραπάνω πίνακες δεξιά από τις τεχνικές περιγραφές των προβλημάτων βλέπουμε τον αριθμό των επιχειρήσεων που τα δήλωσαν. Την τελική μορφή της συγχώνευσης των δύο παραπάνω ερωτήσεων δείχνει ο πίνακας 13. Στις γραμμές εδώ έχουμε βάλει τα χαρακτηριστικά, τις  $X_i$  μεταβλητές δηλαδή, λόγω καλύτερης αισθητικά παρουσίασης του πίνακα δεδομένων. Στις στήλες, μεταβλητές  $Y_i$ , βρίσκονται οι 25 εταιρείες που απάντησαν στο ερωτηματολόγιο.

Η μορφή του πίνακα δεδομένων κατά την εισαγωγή του για επεξεργασία δείχνει ο πίνακας 4. Ας δούμε όμως τα αποτελέσματα που προκύπτουν από την επεξεργασία του στον πίνακα 9. Αρχίζουμε από την πρώτη ομάδα την αντίστοιχη  $\lambda^{-1}$  που αναφέρεται στις  $X_i$  μεταβλητές των τεχνικών προβλημάτων των προϊόντων και η ομάδα που μας ενδιαφέρει περισσότερο.

Το βασικό μήνυμα που μας δίνει σ' αυτή την περίπτωση είναι ότι η  $X_1$  μεταβλητή είναι η πρώτη στην ιεράρχηση από τις άλλες και μάλιστα με πολλή μεγάλη διαφορά. Από τις 25 επιχειρήσεις που ερωτήθηκαν οι 16 απάντησαν ότι τα προϊόντα της εταιρείας που μελετάμε δεν είχαν σταθερή ποιότητα.

Δεύτερο στοιχείο που ψάχνουμε να βρούμε από τα αποτελέσματα της  $\lambda^{-1}$  αντιστοιχίας είναι οι σημαντικές ομαδοποιήσεις που σχηματίζονται. Ο αλγόριθμος αρχίζει να ομαδοποιεί σε χαμηλό q-επίπεδο οπότε και εμείς θα πάρουμε ως σημαντική την πρώτη. Βλέπουμε λοιπόν ότι στο 4 q-επίπεδο σχηματίζεται μια ομάδα από τις  $X_1$ ,  $X_2$  και  $X_9$  μεταβλητές. Η  $X_1$  ανατρέχοντας στον πίνακα 13 βλέπουμε ότι είναι το τεχνικό πρόβλημα της μεγάλης φύρας και η  $X_9$  είναι το τεχνικό πρόβλημα του σκεβρώματος.

Η ομαδοποίηση αυτή της Q-Analysis έρχεται να επιβεβαιώσει και σχόλιο στελέχους της εν λόγω εταιρείας. Ότι ο βασικός λόγος της φύρας του χαρτονιού που παράγει η εταιρεία μας, είναι η έλλειψη σταθερής ποιότητας στο προϊόν που παράγει. Επίσης η έλλειψη σταθερής ποιότητας είναι ένας βασικός παράγοντας που προκαλεί και την σκέβρωση. Την σκέβρωση προκαλεί και η  $X_8$  μεταβλητή που είναι η μεγάλη υγρασία και εμφανίζεται στην ίδια ομάδα με τη μεταβλητή  $X_9$  στο 3 q-επίπεδο.

Βλέπουμε λοιπόν ότι ο αλγόριθμος μέσω των δύο ομάδων μας έβγαλε σχέσεις αιτίου και αιτιατού, ανάμεσα στα τεχνικά προβλήματα που παρουσιάζονται στο χαρτόνι που παράγει η εταιρεία μας. Σίγουρα έχοντας σταθερή ποιότητα στην παραγωγή, θα λιγόστευαν όχι μόνο τα παράπονα, αλλά και τα τεχνικά προβλήματα όπως η μεγάλη φύρα και το σκεβρώμα.

Στη δεύτερη ομάδα παρουσιάζεται η  $\lambda$  αντιστοιχία. Εδώ από πλευράς ιεράρχησης βλέπουμε τις επιχειρήσεις που έχουν τα περισσότερα παράπονα όσον αφορά τα τεχνικά προβλήματα του προϊόντος. Τέτοιες είναι η 6 και η 11 πράγμα που σημαίνει ότι θα έπρεπε να το λάβει υπόψη η εταιρεία μας. Οι ομάδες που σχηματίζονται στο 3 q-επίπεδο καθώς και στο 2 q-επίπεδο αποτελούνται καθαρά από επιχειρήσεις που ανήκουν στην ομάδα των εταιρειών που αγοράζουν προϊόντα της εταιρείας μας αλλά είναι δυσαρεστημένες. Άλλη μια φορά έρχεται να επιβεβαιώσει την πραγματικότητα η Q-Analysis με αυτές τις ομαδοποιήσεις.

### 7.1.3 Μελέτη του δεύτερου πίνακα δεδομένων με τη βοήθεια της *Q-Analysis*.

Ο δεύτερος πίνακας δεδομένων παριστάνει κωδικοποιημένες τις απαντήσεις της ερωτήσεως 10 του ερωτηματολογίου της ICAP. Την ερώτηση αυτή στην πρωτογενή της μορφή, όσον αφορά τα αποτελέσματα, μπορούμε να τη δούμε στον πίνακα 10.

Η ερώτηση 10 αναφέρεται στα προτερήματα των ανταγωνιστικών προϊόντων αλλά και γενικότερα της εμπορικής πολιτικής των ανταγωνιστών της επιχείρησης. Από την ερώτηση αυτή ένα προτέρημα των ανταγωνιστών μπορεί να θεωρηθεί μειονέκτημα για την επιχείρηση που μελετάμε. Στο δεξιό μέρος του πίνακα 10 βλέπουμε το πλήθος των επιχειρήσεων που δήλωσαν το αντίστοιχο χαρακτηριστικό ως προτέρημα των ανταγωνιστών. Δίπλα υπάρχει το ποσοστό που αντιστοιχεί στο αριθμό αυτό.

Στον πίνακα 16 βλέπουμε το δεύτερο πίνακα δεδομένων με τα χαρακτηριστικά, μεταβλητές  $X_i$ , που δηλώνουν τα προτερήματα των ανταγωνιστών στις γραμμές του και τις 25 επιχειρήσεις, μεταβλητές  $Y_i$ , στις στήλες του πίνακα. Και πάλι εδώ βάλαμε τις  $X_i$  μεταβλητές στις γραμμές του πίνακα λόγω καλύτερης αισθητικά παρουσίασης του πίνακα δεδομένων.

Όπως και στον προηγούμενο πίνακα δεδομένων η παύλα σημαίνει  $NAI=1$  και το κενό  $OXI=0$ . Τον πίνακα δεδομένων έτοιμο για επεξεργασία από τον αλγόριθμο της *Q-Analysis* τον είδαμε στον πίνακα 4.

Στα αποτελέσματα (βλ. πίνακα 11) τώρα στην αντιστοιχία  $\lambda^{-1}$  που αναφέρεται στις μεταβλητές  $X_i$  εμφανίζονται στα πρώτα έξι επίπεδα οι δύο μεταβλητές  $X_{12}$  και  $X_1$ . Οι μεταβλητές αυτές είναι οι καλύτερης ποιότητας και καλύτερης τιμής. Σύμφωνα με σχόλιο στελέχους της ICAP από όπου πήραμε και τα στοιχεία οι δύο αυτές μεταβλητές όταν βρίσκονται σε ερωτηματολόγιο δεν μπορούν να μας δώσουν σημαντική πληροφορία. Είναι και οι δύο χαρακτηριστικές μεταβλητές που σύμφωνα με παρατηρήσεις σχεδόν πάντα ελέγχονται από τους ερωτώμενους

έστω και αν είναι ευχαριστημένοι, στην συγκεκριμένη περίπτωση, από το προϊόν που αγοράζουν. Ο πελάτης θέλει πάντα το προϊόν που αγοράζει να γίνει καλύτερο και φθηνότερο, έστω και αν είναι ικανοποιημένος από την ποιότητα και την τιμή του.

Η συμπεριφορά αυτή που συνηθίζεται από τους ερωτώμενους φαίνεται στα αποτελέσματα από το να εμφανίζονται το  $X_1$  και  $X_2$  για τρία συνεχή q-επίπεδα χώρια και μετά να συσχετίζονται σε κοινή ομάδα. Παρακάμπτουμε λοιπόν τις δύο αυτές μεταβλητές και συνεχίζουμε τη μελέτη. Το  $X_3$  λοιπόν από πλευράς ιεράρχησης ουσιαστικά είναι το πρώτο προτέρημα των ανταγωνιστών που είναι η σταθερή ποιότητα. Προτέρημα για τους ανταγωνιστές, μειονέκτημα για την εταιρεία μας πράγμα που συμφωνεί και με τον προηγούμενο πίνακα δεδομένων που παρουσίαζε τη μη σταθερή ποιότητα σαν το μεγαλύτερο τεχνικό πρόβλημα που αντιμετωπίζει η εταιρεία μας.

Ομαδοποιήσεις δεν έχουμε σημαντικές σ' αυτή την αντιστοιχία γιατί συμβαίνουν σε πολύ χαμηλό q-επίπεδο.

Στην αντιστοιχία λ βλέπουμε ότι οι επιχειρήσεις 2, 3 και 4 είναι αυτές που βρίσκουν τα περισσότερα προτερήματα στους ανταγωνιστές της εταιρείας μας. Στην ίδια αντιστοιχία στο 2 q-επίπεδο βλέπουμε ότι η ομάδα που σχηματίζεται - η πιο σημαντική αυτής της ομάδας των αποτελεσμάτων- αποτελείται από 2 επιχειρήσεις που δεν αγοράζουν προϊόν από την εταιρεία μας. Εδώ η Q-Analysis όπως και πριν επιβεβαιώνει και πάλι την πραγματικότητα.

#### *7.1.4 Μελέτη του τρίτου πίνακα δεδομένων με τη βοήθεια της Q-Analysis.*

Ο τρίτος πίνακας δεδομένων παριστάνει κωδικοποιημένες τις απαντήσεις της ερώτησης 5 του πίνακα της ICAP (βλ. πίνακα 12).

Η ερώτηση αυτή αναφέρεται στα προτερήματα των προϊόντων αλλά και γενικά στην εμπορική πολιτική της εταιρείας που μελετάμε.



Τον τρίτο πίνακα δεδομένων δείχνει ο πίνακας 15. Στις γραμμές έχουμε τις μεταβλητές  $X_i$  για λόγους αισθητικής, που περιγράφω τα προτερήματα της εταιρείας μας. Στις στήλες υπάρχουν οι 25 επιχειρήσεις που απάντησαν στην ερώτηση. Όπως και στους προηγούμενους πίνακες η παύλα σημαίνει ΝΑΙ και το κενό ΟΧΙ στα προτερήματα που βρίσκονται στις γραμμές του πίνακα. Στον πίνακα 14 βλέπουμε τα αποτελέσματα της επεξεργασίας. Στην αντιστοιχία  $\lambda^{-1}$  Βλέπουμε ότι τα σπουδαιότερα προτερήματα της εταιρείας μας από ιεραρχικής πλευράς είναι το  $X_4$ ,  $X_6$  και  $X_5$ . Ανατρέχοντας στον πίνακα 15 βλέπουμε ότι αυτές είναι ετοιμοπαράδοτα μόλις ζητηθούν, ότι εξυπηρετούν οι άνθρωποι των πωλήσεων και ότι παρέχεται πίστωση αντίστοιχα. Οι μεταβλητές  $X_4$  και  $X_5$  αποκτούν ιδιαίτερη σημασία αν συγκριθούν με τους λογαριασμούς αποτελεσμάτων χρήσης της εταιρείας μας για την περίοδο 1983- 1986.

Αν παρατηρήσουμε την αυξητική τάση των αποθεμάτων για τα παραπάνω και για τέσσερα χρόνια σίγουρα καταλαβαίνουμε γιατί το ισχυρότερο προτέρημα της εταιρείας μας είναι το ότι παραδίδει αμέσως τα προϊόντα της. Το μεγάλο όμως απόθεμα που αναγκάζεται να έχει για να διατηρεί αυτό το προτέρημα βλέπουμε ότι της κοστίζει ακριβά. Ακριβά επίσης της κοστίζει και το  $X_5$  προτέρημα. Η καλή πιστωτική της πολιτική την αναγκάζει κατά την περίοδο των παραπάνω χρόνων να αυξάνει για τον παραπάνω βασικά λόγο συνέχεια τις βραχυπρόθεσμες υποχρεώσεις. Ένας άλλος λόγος είναι και η κάλυψη των ελλειμμάτων της.

Η ομαδοποίηση που συναντάμε στο 4 q-επίπεδο μεταξύ  $X_5$  και  $X_6$  δεν μας δίνει καμιά ένδειξη αφού οι δύο αυτές μεταβλητές στα προηγούμενα q-επίπεδα ήταν χώρια.

Όσον αφορά τώρα την  $\lambda$  αντιστοιχία βλέπουμε ότι η 4 επιχείρηση είναι η πιο ευχαριστημένη. Για άλλη μια φορά οι δυο βασικότερες ομάδες που δημιουργεί ο αλγόριθμος ανήκουν στην ίδια ομάδα εταιρειών που φυσικά αυτή τη φορά είναι επιχειρήσεις που είναι ευχαριστημένες από την επιχείρησή μας. Αυτές είναι η  $Y_4$ ,  $Y_{18}$  και  $Y_{10}$ .

## 7.2. Πίνακες δεδομένων της Q-ανάλυσης

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0
1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Πίνακας 7.1. Ο αρχικός πίνακας δεδομένων

A(25X20)

7.2.1 Ερώτηση 3<sup>η</sup>: Είστε ευχαριστημένος από τα προϊόντα της και  
αν όχι γιατί;

	No	%
ΝΑΙ	4	16
ΟΧΙ – ΛΟΓΟΙ		
- ΜΗ ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ	16	64
- ΜΕΓΑΛΗ ΦΥΡΑ	4	16
- ΕΛΛΕΙΨΗ ΟΜΟΙΟΜΟΡΦΟΥ ΠΑΧΟΥΣ	2	8
- ΥΨΗΛΗ ΥΓΡΑΣΙΑ	2	8
- ΣΤΙΓΜΑΤΑ ΣΤΗΝ ΕΠΙΧΡΙΣΗ	1	4
- ΞΕΝΑ ΣΩΜΑΤΑ ΕΠΙΖΗΜΙΑ	1	4
- ΚΑΘΥΣΤΕΡΗΣΗ ΣΤΗΝ ΠΑΡΑΔΟΣΗ	2	8
- ΣΥΧΝΕΣ ΑΛΛΑΓΕΣ ΠΡΟΣΩΠΙΚΟΥ ΔΥΣΧΕΡΕΙΣ ΣΥΝΑΛΛΑΓΕΣ	1	4
- ΣΚΕΒΡΩΝΕΙ, ΑΚΑΤΑΛΛΗΛΟ ΓΙΑ ΕΚΤΥΠΩΣΗ	3	12
- ΕΤΕΡΟΔΕΙΓΜΑΤΙΚΗ ΠΟΙΟΤΙΚΗ ΠΑΡΑΔΟΣΗ	1	4

Πίνακας 7.2

7.2.2 Ερώτηση 4<sup>η</sup>: Προτάσεις βελτίωσης προϊόντων της

	No	%
ΠΑΡΑΓΩΓΗ ΚΑΙ ΑΛΛΩΝ ΔΙΑΣΤΑΣΕΩΝ ΧΑΡΤΟΝΙΟΥ	1	4
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ ΧΑΡΤΟΝΙΟΥ	10	40
ΑΝΤΑΓΩΝΙΣΤΙΚΕΣ ΤΙΜΕΣ	2	8
ΚΑΛΥΤΕΡΟΥΣ ΟΡΟΥΣ ΠΛΗΡΩΜΗΣ	1	4
ΠΡΟΣΟΧΗ ΣΤΗΝ ΦΟΡΤΩΣΗ ΚΑΙ ΣΥΣΚΕΥΑΣΙΑ	1	4
ΒΕΛΤΙΩΣΗ ΧΡΟΝΟΥ ΠΑΡΑΔΟΣΗΣ	3	12
ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΟΣ	8	32
ΚΑΛΥΤΕΡΟ ΕΠΙΧΡΙΣΜΑ	4	16
ΜΕΙΩΣΗ ΥΓΡΑΣΙΑΣ	3	12
ΞΕΚΟΛΛΑ ΤΟ ΕΠΑΝΩ ΜΕΡΟΣ ΤΟ ΛΕΥΚΟ	1	4
ΣΚΕΒΡΩΝΕΙ	3	12
ΕΧΕΙ ΦΟΥΣΚΕΣ	1	4
ΕΧΟΥΝ ΛΕΚΕΔΕΣ	1	4
ΕΧΟΥΝ ΛΑΚΕΣ	1	4

Πίνακας 7.3

	q	Q
[X1]	15	1
[X1]	14	1
[X1]	13	1
[X1]	12	1
[X1]	11	1
[X1]	10	1
[X1]	9	1
[X1]	8	1
[X1]	7	1
[X1]	6	1
[X1]	5	1
[X1 X2 X9]	4	1
[X1 X2 X4 X7 X9]	3	1
[X1 X2 X4 X7 X9]	2	1
[X1 X2 X4 X7 X9 X12 X14 X16] [X11]	1	2
[ALL]	0	1
Q = (1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1)		

	q	Q
[Y6] [Y11]	6	2
[Y6] [Y11]	5	2
[Y6] [Y11] [Y12]	4	3
[Y5] [Y7] [Y11 Y16] [Y12] [Y13] [Y23]	3	6
[Y5] [Y6 Y11 Y12 Y13 Y16 Y20] [Y7] [Y23]	2	4
[Y5 Y14] [Y6 Y7 Y9 Y11 Y12 Y13 Y15 Y16 Y20 Y23 Y25]	1	2
[Y4 Y5 Y6 Y7 Y9 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y19 Y20 Y22 Y23 Y25]	0	1
Q' = (2 2 3 6 4 2 1) h = .5867188		

Πίνακας 7.4

7.2.3 Ερώτηση 10<sup>η</sup>: Για ποιους λόγους προτιμάτε τα ανταγωνιστικά προϊόντα;

	No	%
ΚΑΛΥΤΕΡΑ – ΤΙΜΗ	8	32
ΣΥΝΕΠΕΙΑ ΣΤΙΣ ΠΑΡΑΔΟΣΕΙΣ	3	12
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ	12	48
ΚΑΛΥΤΕΡΗ ΠΟΙΟΤΗΤΑ	12	48
ΟΜΟΙΟΜΟΡΦΗ ΕΠΙΦΑΝΕΙΑ	1	4
ΕΙΔΙΚΕΣ ΔΙΑΣΤΑΣΕΙΣ	2	8
ΜΙΚΡΕΣ ΠΟΣΟΤΗΤΕΣ	1	4
ΣΥΣΚΕΥΑΣΙΑ ΚΑΛΥΤΕΡΗ	2	8
ΕΥΡΙΣΚΟΥΝ ΤΗΝ ΠΟΙΟΤΗΤΑ ΠΟΥ ΘΕΛΟΥΝ	1	4
ΙΔΙΕΣ ΤΙΜΕΣ	2	8
ΠΙΟ ΓΥΑΛΙΣΤΕΡΑ	1	4
ΕΙΔΗ ΠΟΥ ΔΕΝ ΕΧΕΙ Η ΕΤΑΙΡΕΙΑ	1	4

Πίνακας 7.5

	q	Q
[X4]	9	1
[X4]	8	1
[X4] [X6]	7	2
[X4] [X5] [X6]	6	3
[X1] [X4] [X5] [X6] [X10]	5	5
[X1] [X4] [X5 X6] [X7] [X10]	4	5
[X1 X4] [X5 X6] [X7] [X10]	3	4
[X1 X4 X5 X6 X10] [X7]	2	2
[X1 X3 X4 X5 X6 X7 X10] [X11]	1	2
[ALL]	0	1
Q = (1 1 2 3 5 5 4 2 2 1)		
[Y4]	5	1
[Y4]	4	1
[Y4 Y18]	3	1
[Y4 Y10 Y18] [Y6] [Y7] [Y19]	2	4
[Y4 Y6 Y7 Y10 Y14 Y15 Y18 Y19 Y20 Y24] [Y5 Y12] [Y9] [Y11] [Y13] [Y16] [Y17] [Y22 Y23]	1	8
[Y4 Y5 Y6 Y7 Y9 Y10 Y11 Y12 Y13 Y14 Y15 Y16 Y17 Y18 Y19 Y20 Y22 Y23 Y24 Y25]	0	1
Q' = (1 1 1 4 8 1) H = 0.4485426		

Πίνακας 7.6

7.2.4 Ερώτηση 5<sup>η</sup>: Για ποιους λόγους αγοράζουν τα προϊόντα της

	Νο	%
ΕΙΝΑΙ ΦΘΗΝΑ	6	24
ΚΑΤΑΛΛΗΛΑ ΤΕΧΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	1	4
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ ΣΥΝΕΧΕΙΑ ΣΥΝΕΠΕΙΑ ΕΤΟΙΜΟΠΑΡΑΔΟΤΑ ΜΟΛΙΣ ΖΗΤΗΘΟΥΝ	3	12
ΠΑΡΕΧΕΤΑΙ ΠΙΣΤΩΣΗ	11	44
ΕΞΥΠΗΡΕΤΟΥΝ ΟΙ ΑΝΘΡΩΠΟΙ ΤΩΝ ΠΩΛΗΣΕΩΝ	7	28
ΕΙΝΑΙ ΦΙΛΟΙ	8	32
ΠΑΡΟΧΗ ΕΚΠΤΩΣΕΩΣ ΑΝΑΛΟΓΩΣ ΟΓΚΟΥ ΑΓΟΡΑΣ ΜΕΤΡΗΤΟΙΣ	5	20
ΕΙΝΑΙ ΣΚΛΗΡΑ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΚΥΤΙΟΠΟΠΙΑ	1	4
Η ΕΙΣΑΓΩΓΗ ΕΙΝΑΙ ΠΡΟΒΛΗΜΑΤΙΚΗ - ΔΥΣΚΟΛΗ	1	4
ΩΣ ΕΛΛΗΝΙΚΟ ΠΡΟΪΟΝ	4	16
	2	8

Πίνακας 7.7

<i><b>ΕΠΙΧΕΙΡΗΣΗ:</b></i>	<i><b>1</b></i>	<i><b>2</b></i>	<i><b>3</b></i>	<i><b>4</b></i>	<i><b>5</b></i>	<i><b>6</b></i>	<i><b>7</b></i>	<i><b>8</b></i>	<i><b>9</b></i>	<i><b>10</b></i>	<i><b>11</b></i>	<i><b>12</b></i>	<i><b>13</b></i>	<i><b>14</b></i>	<i><b>15</b></i>	<i><b>16</b></i>	<i><b>17</b></i>	<i><b>18</b></i>	<i><b>19</b></i>	<i><b>20</b></i>	<i><b>21</b></i>	<i><b>22</b></i>	<i><b>23</b></i>	<i><b>24</b></i>	<i><b>25</b></i>
<b>ΑΠΑΝΤΗΣΗ</b>																									
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1
ΜΕΓΑΛΗ ΦΥΡΑ	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1
ΕΛΛΕΙΨΗ ΟΜΟΙΟΜΟΡΦΟΥ ΠΑΧΟΥΣ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΥΨΗΛΗ ΥΓΡΑΣΙΑ	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
ΣΤΙΓΜΑΤΑ ΣΤΗΝ ΕΠΙΧΡΙΣΗ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΞΕΝΑ ΣΩΜΑΤΑ ΕΠΙΖΗΜΙΑ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΚΑΘΥΣΤΕΡΗΣΗ ΣΤΗΝ ΠΑΡΑΔΟΣΗ	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
ΣΥΧΝΕΣ ΑΛΛΑΓΕΣ ΠΡΟΣΩΠΙΚΟΥ	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ΣΚΕΒΡΩΝΕΙ	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
ΕΤΕΡΟΔΕΙΓΜΑΤΙΚΗ ΠΑΡΑΔΟΣΗ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΠΑΡΑΓΩΓΗ ΚΑΙ ΑΛΛΩΝ ΔΙΑΣΤΑΣΕΩΝ	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ΑΝΤΑΓΩΝΙΣΤΙΚΕΣ ΤΙΜΕΣ	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ΚΑΛΥΤΕΡΟΥΣ ΟΡΟΥΣ ΠΛΗΡΩΜΗΣ	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΠΡΟΣΟΧΗ ΣΤΗΝ ΦΟΡΤΩΣΗ ΚΑΙ ΣΥΣΚΕΥΑΣΙΑ	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΟΣ	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΚΑΛΥΤΕΡΟ ΕΠΙΧΡΙΣΜΑ	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΧΟΥΝ ΛΑΚΕΣ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ΞΕΚΟΛΛΑ ΤΟ ΕΠΑΝΩ ΜΕΡΟΣ ΤΟ ΛΕΥΚΟ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΧΕΙ ΦΟΥΣΚΕΣ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΧΟΥΝ ΛΕΚΕΔΕΣ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Πίνακας 7.8

<b>ΕΠΙΧΕΙΡΗΣΗ:</b>		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>
ΑΠΑΝΤΗΣΗ																										
ΕΙΝΑΙ ΦΘΗΝΑ		0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0
ΚΑΤΑΛΛΗΛΑ ΤΕΧΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ		0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ ΣΥΝΕΠΕΙΑ		0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΤΟΙΜΟΠΑΡΑΔΟΤΑ ΜΟΛΙΣ ΖΗΤΗΘΟΥΝ		0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	1	1	1
ΠΑΡΕΧΕΤΑΙ ΠΙΣΤΩΣΗ		0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
ΕΞΥΠΗΡΕΤΟΥΝ ΟΙ ΑΝΘΡΩΠΟΙ ΤΩΝ ΠΩΛΗΣΕΩΝ		0	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0
ΕΙΝΑΙ ΦΙΛΟΙ		0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
ΠΑΡΟΧΗ ΕΚΠΤΩΣΕΩΣ ΑΝΑΛΟΓΩΣ ΟΓΚΟΥ ΑΓΟΡΑΣ ΜΕΤΡΗΤΟΙΣ		0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΙΝΑΙ ΣΚΛΗΡΑ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΚΥΤΙΟΠΟΙΑ		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Η ΕΙΣΑΓΩΓΗ ΕΙΝΑΙ ΠΡΟΒΛΗΜΑΤΙΚΗ – ΔΥΣΚΟΛΗ		0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	0
ΩΣ ΕΛΛΗΝΙΚΟ ΠΡΟΪΟΝ		0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0

ΠΙΝΑΚΑΣ 7.9



<b>ΕΠΙΧΕΙΡΗΣΗ:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>
ΑΠΑΝΤΗΣΗ																									
ΕΙΝΑΙ ΦΘΗΝΑ	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0
ΚΑΤΑΛΛΗΛΑ ΤΕΧΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ ΣΥΝΕΠΕΙΑ	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΤΟΙΜΟΠΑΡΑΔΟΤΑ ΜΟΛΙΣ ΖΗΤΗΘΟΥΝ	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	1	1	1
ΠΑΡΕΧΕΤΑΙ ΠΙΣΤΩΣΗ	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
ΕΞΥΠΗΡΕΤΟΥΝ ΟΙ ΑΝΘΡΩΠΟΙ ΤΩΝ ΠΩΛΗΣΕΩΝ	0	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0
ΕΙΝΑΙ ΦΙΛΟΙ	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
ΠΑΡΟΧΗ ΕΚΠΤΩΣΕΩΣ ΑΝΑΛΟΓΩΣ ΟΓΚΟΥ ΑΓΟΡΑΣ ΜΕΤΡΗΤΟΙΣ	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΕΙΝΑΙ ΣΚΛΗΡΑ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΚΥΤΙΟΠΟΙΑ	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Η ΕΙΣΑΓΩΓΗ ΕΙΝΑΙ ΠΡΟΒΛΗΜΑΤΙΚΗ – ΔΥΣΚΟΛΗ	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	0
ΩΣ ΕΛΛΗΝΙΚΟ ΠΡΟΪΟΝ	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0

Πίνακας 7.10

<b>ΕΠΙΧΕΙΡΗΣΗ:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>
ΑΠΑΝΤΗΣΗ																									
ΚΑΛΥΤΕΡΑ - ΤΙΜΗ	1	1	0	1	1	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	1	0
ΣΥΝΕΠΕΙΑ ΣΤΙΣ ΠΑΡΑΔΟΣΕΙΣ	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ΣΤΑΘΕΡΗ ΠΟΙΟΤΗΤΑ	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
ΚΑΛΥΤΕΡΗ ΠΟΙΟΤΗΤΑ	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΟΜΟΙΟΜΟΡΦΗ ΕΠΙΦΑΝΕΙΑ	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ΕΙΔΙΚΕΣ ΔΙΑΣΤΑΣΕΙΣ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ΜΙΚΡΕΣ ΠΟΣΟΤΗΤΕΣ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΣΥΣΚΕΥΑΣΙΑ ΚΑΛΥΤΕΡΗ	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ΕΥΡΙΣΚΟΥΝ ΤΗΝ ΠΟΙΟΤΗΤΑ ΠΟΥ ΘΕΛΟΥΝ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
ΙΔΙΕΣ ΤΙΜΕΣ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ΠΙΟ ΓΥΑΛΙΣΤΕΡΑ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
ΕΙΔΗ ΠΟΥ ΔΕΝ ΕΧΕΙ Η	0	1	1	0	1	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	1	1

Πίνακας 7.11

## 8. ΣΥΜΠΕΡΑΣΜΑΤΑ

Με την παρούσα εργασία έγινε μια προσπάθεια για την εφαρμογή των προγραμμάτων νοημόνων μεθοδολογιών στον χώρο του μάρκετινγκ. Σε παγκόσμιο επίπεδο υπάρχουν αρκετές εφαρμογές οι οποίες προέρχονται από το χώρο του μάρκετινγκ<sup>2</sup>:

- Η κατασκευή προφίλ πελατών (Customer Profiling).
- Το στοχευόμενο μάρκετινγκ (targeted marketing).
- Η ανάλυση του καλαθιού αγορών (market basket analysis).

Πέρα από την παρουσίαση του θεωρητικού κομματιού, επιχειρήσαμε και μια πρακτική εφαρμογή μέσω του προγράμματος See5. Προσπαθήσαμε να δείξουμε ότι τα προγράμματα αυτά είναι εύκολα εφαρμόσιμα στον χώρο της έρευνας αγοράς και κατ' επέκταση του μάρκετινγκ, και συνεπώς ότι μπορούν πραγματικά να συμβάλουν στο έργο των στελεχών των τμημάτων μάρκετινγκ των επιχειρήσεων .

Στο συγκεκριμένο παράδειγμα καταδείξαμε (μέσω ενός συγκεκριμένου ερωτηματολογίου) ότι με τα προγράμματα νοημόνων προσεγγίσεων μπορούν εύκολα να παραχθούν συμπεράσματα για το εάν είναι εφικτή η δημιουργία ενός βιοτεχνικού πάρκου στην αγορά των Χανίων. Σκοπός ήταν να διαπιστωθεί αν οι επιχειρηματίες της ευρύτερης περιοχής είναι θετικοί σε μια ενδεχόμενη συμμετοχή τους στο συγκεκριμένο χώρο.

Η έρευνα και η εφαρμογή του προγράμματος μας βοηθούν να διαπιστώσουμε τι κίνητρα πρέπει να παρέχουμε στους επιχειρηματίες έτσι ώστε να αυξηθεί το ποσοστό που θέλει να συμμετέχει στο χώρο που επιθυμούμε να δημιουργήσουμε. Για παράδειγμα η παροχή δωρεάν πληροφόρησης σε θέματα προώθησης και διαφήμισης σε όσους συμμετέχουν στον οργανωμένο αυτό χώρο θα οδηγήσει σε αύξηση του ποσοστού των θετικών περιπτώσεων.

Κλείνοντας, πρέπει να επισημάνουμε ότι η συγκεκριμένη εργασία αποτελεί μια πρώτη προσέγγιση στο συγκεκριμένο θέμα. Περαιτέρω έρευνα με διαφορετικά προγράμματα νοημόνων προσεγγίσεων ή με μικρή βελτίωση στην μοντελοποίηση μπορεί να εξάγει ακόμα πιο ποιοτικά

---

<sup>2</sup> Σύμφωνα με έρευνα του Two Corps Corp.

αποτελέσματα, και συνεπώς καλύτερη πληροφόρηση. Αυτό όμως που έχει σημασία είναι ότι υπάρχει πολύ πρόσφορο έδαφος για την εφαρμογή των προγραμμάτων νοσημόνων προσεγγίσεων στο μάρκετινγκ.

## 9. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Berry M.J. & Linoff G., "Data mining techniques", John Wiley & Sons Inc, New York 1997.
- Breen G.& Blankeship A.B., "Έρευνα αγοράς Μάρκετινγκ", Interbooks, Αθήνα, 1984.
- Gillies D., "Artificial intelligence and scientific method", Oxford University Press, Oxford, 1996.
- Internet sites (2002): [www.rulequest.com](http://www.rulequest.com), [www.levkada.physics.auth.gr](http://www.levkada.physics.auth.gr).
- Kotler P., "Marketing Management", Interbooks, Αθήνα, 1991.
- Mitchell T., "Machine Learning", McGraw-Hill, 1997.
- Prerau D.S., "Developing and managing expert systems", AddisonWesley, Massachusetts, 1990.
- Quinlan J.R., "C4.5: Programs for machine learning", Morgan Kaufmann Publishers, California, 1993.
- Δούνιας Γ., «Μηχανική γνώσεων και αποφάσεων» *Σημειώσεις προπτυχιακού μαθήματος*, Πολυτεχνείο Κρήτης, Χανιά, 1999.
- Ρίζος Γ., «Τεχνητά νευρωνικά δίκτυα: θεωρία και εφαρμογές», εκδόσεις Νέων Τεχνολογιών, Αθήνα, 1996